National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

CHARACTERIZATION OF GRANULINS, A NOVEL FAMILY OF CYSTEINE RICH GROWTH MODULATING PEPTIDES

by

VIJAY BHANDARI

Department of Biochemistry

McGill University, Montreal

February 1994

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

ISBN 0-315-94583-4

Canada

BHANDARI, Vijay
Dept. of Biochemistry

Title of thesis: Characterization of granulins, a novel family of cysteine rich growth modulating peptides.

Short title: Characterization of granulins.

This thesis is dedicated to my parents for their support, encouragement and confidence.

ABSTRACT

Granulins (epithelins) are cysteine rich polypeptides with pleiotropic effects on epithelial cell growth *in vitro*. The human granulin cDNA predicts a 593 amino acid glycoprotein precursor containing seven and one-half granulin-like repeats arranged in tandem. The rat granulin precursor cDNA predicts a 589 residue glycoprotein with an overall identity of 75% with human progranulin. The single copy human granulin gene is located on chromosome 17. The protein coding region of the granulin gene spans 4 kilobases and contains 12 exons with each repeat encoded by two exons. The 5' flanking region of the human granulin gene lacks a TATA box but has several CCAAT boxes, and exhibits heterogeneity in transcription initiation sites. The granulin precursor is processed differently depending on cell type, and secretion of the granulin gene products is both constitutive and regulated. The granulin gene is expressed in a variety of adult and fetal tissues derived from all three embryonic germ layers. *In vitro*, the granulin gene is expressed in cell types of diverse lineages, including epithelial cells, lymphoid and myeloid cells, and fibroblasts, whereas its expression *in situ* is restricted to hematopoietic and some epithelial cells.

RESUME

Les granulines (épithélines) sont des polypeptides riches en cystéine produisant des effets pléiotropiques sur la croissance *in vitro* des cellules épithéliales. Chez l'humain, le cDNA du précurseur de la granuline encode une glycoprotéine de 593 acides animés comportant sept séquences répétées et demi agencées en tandem. Chez le rat, le cDNA du précurseur encode une glycoprotéine de 589 résidus ayant une homologie de 75% avec la pro-granuline humaine. La copie unique du gène humain de la granuline est située sur le chomosome 17. La région codante du gène s'étend sur 4 kilobases comprenant 12 exons dont chaque groupe de deux encode une séquence répétée. La région 5' du gène humain de la granuline est caractérisée par la présence de plusieurs boîtes CCAAT alors qu'aucune boîte TATA n'est observée. De plus, une hétérogénéité au niveau des sites d'initiation de la transcription a été identifiée. Le processus de maturation du précurseur de la granuline est différent selon le type cellulaire, et les variantes formes de granuline sécrétées le sont soit par régulation ou de façon constitutive. L'expression du gène de la granuline est observée dans une variété de tissus foetaux et adultes originant des trois grandes lignes germinales embryoniques. *In vitro*, le gène est exprimé dans différentes lignées cellulaires incluant les cellules épithéliales, lymphoïdes et myéloïdes ainsi que les fibroblastes. Par contre, *in situ*, l'expression de la granuline est restreinte aux cellules hématopoïétiques ainsi qu'à quelques lignées épithéliales.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my research director, Dr. Andrew Bateman, for his close supervision, advice, constructive criticism, encouragement and friendship throughout the course of this project.

I am particularly grateful to Dr. Samuel Solomon for his support and encouragement during the course of these studies.

I would like to thank Drs. H.P.J. Bennett, R.G.E. Palfree and L.F. Congote for their helpful discussions and suggestions.

I would like to thank Dr. Adel Giaid for the opportunity to collaborate on the *in situ* hybridization studies.

The financial support of the Research Institute and Department of Medicine, Royal Victoria Hospital and the Medical Research Council of Canada is gratefully acknowledged.

Finally, I would like to express my deepest gratitude to Lillian for her patience, understanding, support, encouragement and confidence throughout the duration of this PhD.

TABLE OF CONTENTS

CHAPTER 2: METHODS AND MATERIALS

CHAPTER 5: CHROMOSOMAL LOCALIZATION, GENOMIC ORGANIZATION AND STRUCTURAL AND FUNCTIONAL ANALYSIS OF A PROMOTER OF THE HUMAN GRANULIN GENE

ABBREVIATIONS

Measures

g	gram
mg	milligram
µg	microgram
ng	nanogram
Da	dalton
kDa	kilodalton
M	Molar
mM	millimolar
µM	micromolar
nM	nanomolar
L	litre
ml	millilitre
µl	microlitre
cpm	counts per minute
Ci	Curie
mCi	millicurie
°C	degrees Centigrade
bp	base pair
kb	kilobase
nt	nucleotide
min	minutes
sec	seconds
hr	hours
K_d	dissociation constant

Amino Acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Miscellaneous

aa	Amino acid
AML	Acute Myelogenous Leukemia
AP	Activator Protein
cAMP	cyclic Adenosine-3',5'-Monophosphate
CAT	Chloramphenicol Acetyl-transferase
cDNA	complementary Deoxy-ribonucleic Acid
CML	Chronic Myelogenous Leukemia
CNS	Central Nervous System
DEPC	diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetra-acetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
Epi	Epithelin
FBS	Fetal Bovine Serum
FGF	Fibroblast Growth Factor
GF	Growth Factor
GFR	Growth Factor Receptor
GI	Gastrointestinal
Grn	Granulin
HGF	Hepatocyte Growth Factor
HPLC	High Performance Liquid Chromatography
IGF	Insulin-like Growth Factor
MHC	Major Histocompability Complex
mRNA	messenger Ribonucleic Acid
NIH	National Institute of Health
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDGF	Platelet Derived Growth Factor
PEG	Polyethylene Glycol
RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulfate
SV	Simian-Virus
TGF	Transforming Growth Factor
TK	Thymidine Kinase
TLC	Thin Layer Chromatography
tPA	tissue Plasminogen Activator
uPA	urokinase Plasminogen Activator
UV	ultraviolet

LIST OF FIGURES AND TABLES

CHAPTER 4: TISSUE DISTRIBUTION AND CELLULAR LOCALIZATION OF GRANULIN mRNA.

PREFACE

The granulins (grns) are a novel family of cysteine-rich growth modulating proteins. When this PhD project first started, there were only two published papers on the grns; one from researchers in our laboratory, and the other from the research team of Drs. G.J.Todaro and M.Shoyab. During the course of this project, several reports were published describing the isolation and characterization of grn-like peptides in other species. The findings of these papers are briefly discussed.

In 1989, researchers in our laboratory isolated and structurally characterized a novel class of cysteine-rich leukocyte derived peptides, some of which are growth modulating (1). Four peptides were isolated from human peripheral leukocytes. Because these peptides were associated with the granule fraction of leukocytes, they were called granulins. The complete amino acid sequence was determined for one peptide, grnA, and partial amino acid sequences obtained for three others, grns B,C and D. A fifth peptide was purified from rat bone marrow. Partial sequencing of this peptide showed it to be 86% identical with human grnA. The grns are small peptides of approximate molecular weight of 6kDa and are cysteine-rich, with cysteine residues accounting for approximately 20% of the total protein. When the sequences were entered into the National Biomedical Research Foundation PIR data bank, no homologies were found with other proteins, indicating that grns are a novel polypeptide family.

In late 1990, Shoyab et al., while screening conditioned medium from mammalian cells, and tissue and cell extracts for growth-modulatory activities,

identified two proteins with growth regulatory properties on some epithelial cells *in vitro*. Only the amino terminal sequences of the two proteins were reported (2). These proteins, called epithelin (epi) 1 and 2, were isolated from rat kidney and were the rat equivalents of human grns A and B respectively. Epi1 promotes anchorage independent growth of normal rat kidney (NRK) cells in the presence of TGFß and stimulates the growth of keratinocytes *in vitro*, whereas epi2 acts as an antagonist to epi1, inhibiting the epi1-induced growth of keratinocytes. Epi1, and at a lower potency, epi2 inhibit the proliferation of certain epithelial cell lines *in vitro*, including A431 cells, derived from a human epidermal carcinoma of the vulva. Similarly, human grnA, but not grnB, C or D, inhibits proliferation of the epithelial cell line A431 in culture. Human grnA also inhibits the proliferation of A549 lung carcinoma and Chinese Hamster Ovary (CHO) cells in culture (unpublished)

Since the initial characterization of the grns (1) and epis (2), and during the course of this project, several reports have been published describing the isolation and characterization of grn-like peptides in other species.

Cuoto et.al. (1992) reported a 7.2 kDa grn-like antimicrobial peptide called eNAP-1 (equine Neutrophil Antibiotic Peptide-1) from equine neutrophils (3). eNAP-1 was reportedly antimicrobial against several equine uterine pathogens, including *Streptococcus zooepidemicus*, *Escherichia coli*, and *Pseudomonas aeruginosa*. However, micromolar concentrations of eNAP-1 are required to elicit the antimicrobial effect. The peptide's low abundance in the neutrophil suggests this is a non-physiological property, although the possibility that eNAP-1 acts synergistically with other granule components cannot be excluded.

Nakakura et.al. (1992) showed the existance of a grn-like peptide in insect neural tissue (4). Extracts of the neurosecretory *pars intercerebralis* of the insect *Locusta migratoria* yield a peptide, PMP-D1, that has the complete grn/epi cysteine arrangement, with an overall sequence similarity with the vertebrate peptides in the order of 45 to 50%. The biological functions of PMP-D1 in the insect brain have not been established.

Sperm acrosomes in the testis contain a 67 kDa glycoprotein called acrogranin (5) which recently was identified by Baba et.al. (1993) as the product of the guinea pig grn precursor gene (6). As the sperm passes along the epididymis, the 67 kDa acrogranin is broken down into smaller fragments of 62, 51, 39, and 22 kDa respectively. The role of these fragments in sperm maturation or their effects on the oocyte after the acrosome reaction are not yet known.

In 1992, Parnell et.al. reported the partial amino acid sequence of transforming growth factor-e (TGF-e), a TGF-like activity distinct from other known growth factors (7). The reported partial sequence of this factor is indistinguishable from grnA, although the physiochemical properties of TGF-e (e.g. size, amino acid composition etc.) indicate that it is a larger protein than the grns or epis (7). TGF-e activityhas been detected in several tissues of epithelial origin, and is a mitogen and progression factor for epithelial cells and fibroblasts (7).

In 1993, Belcourt et.al. reported the isolation and characterization of 6 kDa grn-like peptides from the spleen and head kidney of a teleost fish (*Cyprinus carpio*) (8). The carp grns, like some of the human and rat grns/epis, can modulate the growth of some cells in culture. The carp grns are mitogenic for certain epithelial cell

lines, including A431 cells (which are growth inhibited by grnA/epi1), and growth inhibitory for a fish embryo cell line (8).

In 1993, Zhou et.al. showed that an 88 kDa autocrine growth factor (PC cell-derived growth factor or PCDGF) for the highly tumorigenic cell line, PC, which is derived from a mouse teratoma, is identical to the murine grn precursor (9). PCDGF can also act as a mitogen for 3T3 fibroblasts (9).

The primary structures of human grnA, rat grnA/epi1 together with the recently characterized grn-like peptides eNAP-1, PMP-D1, carp grn-1, and the amino terminus of TGFe are shown in Figure P.1. The grn family of peptides are characterized by a 12-cysteine motif, represented as $CX_{5-6}CX_5CCX_8CCX_6CCX_5CCX_5CX_{5-6}C$. This cysteine motif represents a novel module not seen in any of the previously characterized growth modulators.

The biological functions of the grns *in vivo* remain poorly understood. However, the strong conservation of the grns from insect to man suggest fundamental biological roles for this family of peptides. Additionally, the reported growth modulating properties of the grns (1,2,8,9) suggest that the grns represent a novel family of proteins capable of regulating cell growth *in vitro*. Researchers in our laboratory have been studying the biology and molecular biology of the grns, to understand the *in vivo* functions of the grn family of peptides. The objective of this PhD. project was to characterize the grn family of peptides at the gene level, with respect to cDNA and genomic structures, gene expression and regulation. The work presented in this thesis represents part of an ongoing study with other workers in our laboratory to elucidate the biological significance of the grn family of peptides.

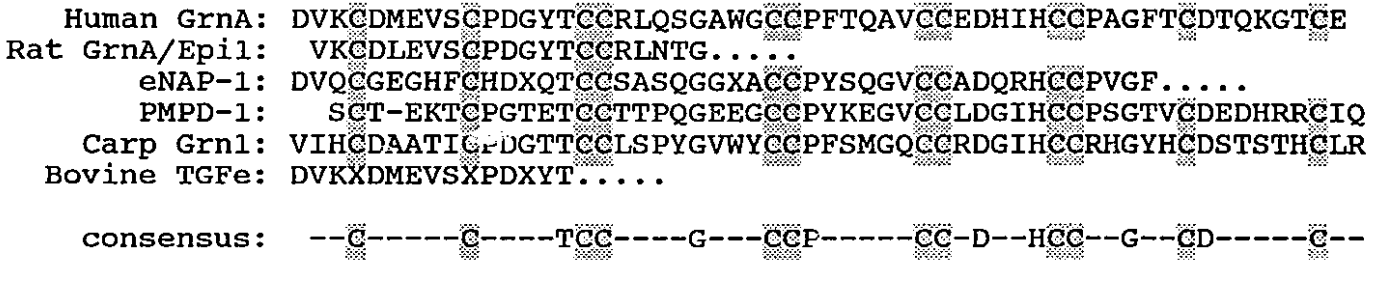


Figure P.1 Alignment of Granulin Peptides

This thesis is organized into five main chapters. Chapter 1 presents an overview of our current understanding of growth factors (GF) with reference to GF structure, expression, and biological properties, and GF receptors, in relation to what is now known about the structure and biology of the grn family of peptides. It should be emphasized that most of the information on the grns presented in Chapter 1 was not known when this project first started. The materials and methods used throughout the course of this project are outlined in Chapter 2. Chapters 3, 4, and 5 constitute the results section of the thesis. Each result section contains a brief introduction followed by a discussion of the significance of the results. Chapter 3 presents the isolation and characterization of the human and rat grn precursor cDNA, and the functional expression of recombinant grn. Chapter 4 presents the tissue distribution and cellular localization of grn mRNA. Chapter 5 presents the genomic organization and structural and functional analysis of a promoter of the human grn gene.

A significant portion of the results presented in this thesis has either been published, or is in preparation for publication as listed below:

1. **Bhandari,V.**, Palfree,R.G.E., Bateman,A. 1992 Isolation and sequence of the granulin precursor cDNA from human bone marrow reveals tandem cysteine-rich granulin domains. *Proc.Natl.Acad.Sci. USA.* 89:1715-1719
2. **Bhandari,V.**, Bateman,A. 1992 Structure and Chromosomal Localization of the Human Granulin Gene. *Biochem.Biophys. Res.Commun.* 188:57-63

3. **Bhandari,V.**, Giaid,A., Bateman,A. 1993 The Complementary Deoxyribonucleic Acid Sequence, Tissue Distribution, and Cellular Localization of the Rat Granulin Precursor. *Endocrinology.* 133:2682-2689

4. **Bhandari,V.**, Bateman,A. 1994 Structural and Functional Analysis of a Promoter of the Human Granulin Gene. In preparation.

Because the granulins represent a recently discovered family of peptides, all results presented in this thesis are contributions to original knowledge.

The main contributions are:

1. Characterization of the human and rat granulin precursor cDNA and first to report that the granulin precursor is a modular protein.
2. First report of a semi-quantitative analysis of the tissue distribution and cellular localization of granulin mRNA.
3. Chromosomal localization and sequence of the human granulin gene and first description of the structural organization of the granulin gene.
4. First analysis of a human granulin gene promoter

CHAPTER 1

GENERAL INTRODUCTION

GRANULINS AND GROWTH FACTORS

The growth and maintenance of tissue is a fundamental process of all living organisms, requiring the coordination of cellular proliferation and differentiation. Many elements contribute to this complex network of intracellular communication, including hormones, neuropeptides, extracellular matrix components, and cell adhesion molecules. A key role, however, is played by a diverse group of regulatory molecules known as polypeptide growth factors. Recently, we and others described a novel family of peptides, the grns, with growth modulating activity (see 1,2, and Preface). To set the stage for an understanding of grn biology, a review of growth factors is presented.

The past decade has seen an explosion in the field of growth factor research. The recent advances in tissue culture, protein biochemistry and molecular biology, has led to the identification, isolation, characterization and cloning of a great number of growth factors. Growth factors mediate intracellular communication by interacting with specific transmembrane cellular receptors, evoking a cascade of events such as tyrosine phosphorylation, phosphatidylinositol turnover, and the expression of cell-cycle dependent genes, resulting in pleiotropic biological responses including the modulation of proliferation, differentiation, adhesion and locomotion. One impetus for studying growth factors is their presumed involvement in a number of clinically

relevant processes including tissue repair, artherosclerosis and cancer.

Much progress has been made in our understanding of the structure and activities of these regulators of cellular homeostasis, and it is clear that many of these growth factors are multifunctional. The exponential increase in the number of publications on growth factors make a comprehensive review on this subject beyond the scope of this thesis. Instead, the intent of this chapter is to summarize our current understanding of growth factors with reference to growth factor structure, expression, biological properties, and growth factor receptors and, where possible relating this to what is presently known about the structure and biology of the grn family of peptides. It should be emphasized that most of the information on the grns was not known when this project first started.

Because of their complex involvement in the immune response, the hematopoietic and lymphopoietic growth factors (e.g. the interleukins, tumor necrosis factor, colony stimulating factors, etc.) are not discussed. Instead, the reader is referred to the following references for recent reviews of the hematopoietic and lymphopoietic growth factors: interleukins (10,11,12), tumor necrosis factor (12,13), intercrines (14), and colony stimulating factors (15,16,17). Here we will discuss the major somatic growth factors, emphasizing the similarities with the grns wherever possible. In this way it may be possible to place the presently fragmented knowledge of the grns in a broader biological context.

1.1 Growth Factor Families.

A detailed comparison of the primary sequences of growth factors has led to

the identification of common structural motifs and residues shared by some growth factors. This has allowed the classification of growth factors into distinct families. For example, members of the epidermal growth factor/transforming growth factor-α family, the transforming growth factor-ß family, the insulin family, and the heparin-binding fibroblast growth factor family are defined by readily identifiable patterns of cysteine residues. Examples of some growth factor families are presented in Table 1.1. Members of each family may have arisen through the duplication and evolution of a common ancestral gene. It is conceivable that the grns, characterised by a unique arrangement of cysteine residues (Figure P.1), may serve as the prototype for a new growth factor family.

1.1.2 The Epidermal Growth Factor Family.

Epidermal growth factor (EGF) was one of the first growth factors to be discovered (18), and serves as the prototype for other members of this family, which includes transforming growth factor-α (19), amphiregulin (20), cripto (21), heparin binding EGF-like factor (22), betacellulin (23), schwannoma-derived growth factor (24), *neu* differentiation factor / heregulins (25,26), and vaccinia virus growth factor (27,28).

The members of the EGF family are characterized by the presence of one or more cysteine-rich EGF modules, represented as $CX_7CX_{3-5}CX_{10-12}CXCX_5GXRC$ (where C is cysteine; G is glycine; R is arginine; X is any other amino acid) (29). The disulfide bridging has been confirmed in EGF (30) and TGF-α (31) [C1-C3, C2-C4 and C5-C6], and presumably other members of the EGF family may fold in a way

Table 1.1: Examples of Superfamilies of Growth Factors

Epidermal Growth Factor
Epidermal Growth Factor
Transforming Growth Factor-α
Amphiregulin
Cripto
Heparin-binding EGF
Betacellulin
Heregulin/*Neu* Differentiating Factor
Vaccinia Virus Growth Factor
Spitz
lin-3

Insulin
Insulin
Insulin-like Growth Factor-I
Insulin-like Growth Factor-II
Relaxin
Bombyxin
Locust Insulin-related Peptide
Molluscan Insulin-related Peptide

Intercrine-ß
LD78
ACT-2
I-309
RANTES
MCAF

Transforming Growth Factor-ß
Transforming Growth Factor-ß1
Transforming Growth Factor-ß2
Transforming Growth Factor-ß3
Transforming Growth Factor-ß4
Transforming Growth Factor-ß5
Mullerian Inhibitory Substance
Decapentaplegic Gene Product
Bone Morphogenetic Proteins
Mesoderm-Inducing Factor (Vg1)
Activins
Inhibins
Dorsalin-1
GDF-3/GDF-9

Interleukin-6
Interleukin-6
G-CSF
cMGF

Platelet-Derived Growth Factor
PDGF-AA
PDGF-AB
PDGF-BB
v-*sis* viral oncoprotein

Fibroblast Growth Factor
FGF-1 (aFGF)
FGF-2 (bFGF)
FGF-3 (*int*-2)
FGF-4 (*hst*-1/*ks*3)
FGF-5
FGF-6
FGF-7 (KGF)
FGF-8 (AIGF)
FGF-9 (GAF)

Neurotrophins
Nerve Growh Factor
Brain-Derived Growth Factor
NT-3
NT-4/NT-5

Intercrine-α
Interleukin-8
Platelet Factor-4
ß-Thromboglobulin
IP-10
Melanoma Growth Stimulating-Factor

Hepatocyte Growth Factor
HGF/Scatter Factor
HGF-like Protein/Macrophage-Stimulating Protein

similar to that of EGF and TGF-α. In most cases, the EGF family members are produced from transmembrane precursors, which are cleaved at the extracellular domain to generate the soluble and active growth factor (29). The precursors are biologically active in many instances, sometimes even equalling or surpassing their soluble products in biological activity. The precursors can activate cognate receptors on adjacent cells i.e. juxtacrine stimulation.

The EGF module is not limited to growth factor precursors, but can also be found in other proteins (32), such as, extracellular matrix proteins, various blood coagulant factors, the low-density lipoprotein receptor and the latent TGF-ß binding protein (33). EGF-like repeats in blood coagulation factors bind calcium (34,35,36), and it is proposed that the calcium binding may be crucial for numerous protein-protein interactions involving EGF-like repeats in coagulant factors, plasma proteins, and membrane proteins (37). Interestingly, the cysteine motif of the grn family of peptides is found in a cold induced tomato thiol protease and germination specific rice seed thiol protease (38), and it is conceivable that the grn-like repeats may also be found in other proteins. The grn motif is found in C-terminal region of the plant thiol proteases. The significance of the grn motif in these plant thiol proteases are not known. The catalytic domain of the plant thiol proteases lies in the N-terminal region, and, while the purpose of the C-terminal region is not clear, it has been proposed that this domain may serve to regulate the protease activity by binding to metal ions (38). However it is not known if the grn repeats can bind to metal ions.

Also included in the EGF superfamily are the products of the *Notch, Delta* and *Serrate* neurogenic genes from *Drosophila*, the *Notch* homologues from human,

mouse and *Xenopus*, the *Caenorhabditis elegans glp*-1, *lin*-3 and *lin*-12 gene products, and the *Drosophila crumbs* and *spitz* gene product. These proteins contain multiple EGF-like repeats, and are involved in a spectrum of activities, including cell-cell interaction, cell signalling, and cell differentiation and development (29,39,40,41 and references therein). Significantly, the grn repeat is also found in insects. Extracts of the brain of the locust yield a peptide, PMP-D1, that has the complete grn cysteine motif (4). However, unlike the gene products containing the EGF-like repeats, the function of PMP-D1 is not known.

1.1.2.1 Epidermal Growth Factor.

Several of the known biological properties of the grns resemble those of epidermal growth factor (EGF). For example, both peptides are growth inhibitory for adenocarcinoma A431 cells, and can stimulate the growth of keratinocytes in culture (2). EGF and grnA/epi1 can induce anchorage independent growth of normal rat kidney (NRK) cells in the presence of TGF-ß (2). Additionally, during the course of this project more similarities were noted between the grns and EGF. The iterative nature of the grn precursor, first revealed through the work described later (see Chapter 3), draws unavoidable comparisons with the EGF precursor (42,43,44), while Northern blot analyses of grn gene expression reveal that both the progrn mRNA (see Chapter 4) and EGF precursor mRNA are abundant in the kidney (45). Additionally, both the grn precursor and EGF precursor are heparin binding proteins, and are eluted from a heparin column with 0.5M NaCl (46, and see Chapter 3). Whether these similarities are physiologically significant is unclear.

EGF was initially isolated from male mouse submaxillary glands based on its ability to cause premature eyelid opening and eruption of incisors in newborn mice (18); its ability to stimulate the growth of cultured cells was recognised later (47,48). An independent study on controlling peptic ulceration by the inhibition of gastric acid secretion led to the isolation of human EGF (urogastrone) from urine (49). EGF is a single polypeptide chain of approximately 6 kDa consisting of 53 amino acids displaying 3 internal disulfide bonds (reviewed in 39,50).

The mouse and human EGFs are synthesized as large precursors of approximately 1200 residues consisting of an extracellular domain, a transmembrane region, and a cytoplasmic tail (42,43,44). In addition to the mature EGF, the EGF precursor contains 8 other EGF-like repeats in the extracellular region, with the mature EGF lying closest to the transmembrane region. In some respects the grn precursor resembles the EGF precursor. Cloning of the grn precursor cDNA (see Chapter 3) reveals that the grn precursor is also a modular protein, consisting of seven and one half grn repeats arranged in tandem.

How EGF is cleaved from the precursor molecule is not known. Not all tissues that produce preproEGF are capable of processing the precursor. For example, in the mouse submaxillary gland the EGF precursor is rapidly processed to the 53 amino-acid form of EGF, while in certain cells of the kidney the precursor accumulates and does not appear to be processed intracellularly to mature EGF (45). This observation has led to the suggestion of a different role for the intact precursor, such as acting as a receptor for membrane transport events in the kidney (45). Interestingly, cell culture transfection studies with the human grn precursor

cDNA (see Chapter 3) suggest that the processing of the grn precursor may also be tissue or cell specific, and this will be discussed in detail in Chapter 3. The EGF precursor can exist as a secreted form and retains EGF-like biological activity (46). Similarly, the grn precursor is also biologically active, acting as a mitogen for 3T3 fibroblasts, and as an autocrine growth factor for PC cells, a highly tumorigenic teratoma derived cell line, *in vitro* (9). Both the membrane-associated and secreted forms of the EGF precursor can bind the glycosaminoglycan heparin (46). Similarly, the grn precursor is also a heparin-binding protein (see Chapter 3).

EGF mRNA is expressed in many tissues, although its abundance varies widely. For example, in the mouse, EGF expression has been observed at the mRNA and/or protein level in the submaxillary gland, kidney, pancreas, small intestine, prostate and brain (45), while in humans, the major sites of production are the gastrointestinal tract, pancreas, kidney, and mammary gland (50). In the kidney, EGF synthesis is localized to the distal tubules of the renal cortex and outer medulla (45). EGF can also be found, as the mature 6 kDa form and as a biologically active secreted precursor form, in all body fluids (46,51). Neither the EGF protein or its mRNA are detectable during fetal development or in early neonates (50), suggesting that EGF is not involved in fetal development. It has been suggested that TGF-α, a member of the EGF family, serves as the alternative ligand for the EGF receptor during fetal development (TGF-α will be discussed separately under Transforming Growth Factors).

EGF displays broad biological effects *in vitro* and *in vivo* (reviewed in 39,50). It is a mitogen for a variety of ectodermal and mesodermal cell and tissue types,

including skin keratinocytes, conjunctival and pharyngeal tissues, corneal endothelial cells, vascular smooth muscle cells, chondrocytes, fibroblasts, liver cells, thyroid follicular cells, granulosa cells, mammary gland epithelium, and glial cells. EGF also has effects on cell differentiation, morphological changes, and the regulation of nutrient and electrolyte transport, and glycolysis. In view of the apparent similarities between EGF and the grns, it will clearly be relevant to compare the activities of the grns in these assays when they become sufficiently available to make it a practical proposition. Despite its well characterised *in vitro* activities, the actual *in vivo* functions of EGF remain rather speculative. However, EGF has been implicated in a number of physiological and/or pathological processes including wound healing and certain cancers (reviewed in 39,50).

1.1.2.2 Other Members of the EGF Family.

In addition to TGF-α, which will be discussed under transforming growth factors, there are other growth modulating proteins that are structurally related to EGF (Table 1.1). For brevity, the other members of the EGF family are not discussed in detail, instead their properties are summarized in Table 1.2.

1.1.3 Transforming Growth Factors.

One of the reported properties of the grns is that they have transforming growth factor-like activity. Transforming growth factors (TGFs) are operationally defined as polypeptides that stimulate the anchorage-independent growth of non-transformed anchorage-dependent cells through high affinity binding to membrane

Table 1.2. Members of the Epidermal Growth Factor Family

Growth Factor	Description	Localization/Sources	Known Target	Known Effects	Comments	Ref.
Amphiregulin (AR)	Two mature soluble forms of 79 and 84 aa released from a 252-residue transmembrane precursor. 38% identical to EGF.	Normal human lung, cardiac muscle,breast, kidney, spleen, ovary, placenta,colon,pancreas and several carcinoma cell lines.	Epithelial cells, Mesenchymal cells.	Mitogen for leukocytes and fibroblasts. Inhibits growth of several tumor cells (e.g. A431, breast carcinoma).	Detected in nucleus of normal and malignant ovarian and colonic epithelial cells. Binds to and is inhibited by heparin.	20,52, 53,54, 55
Schwannoma-derived growth factor (SDGF)	A 31-35 kDa (148 aa) protein derived from a 243 aa precursor. Mature SDGF is 76% homologous to AR.	Newborn rat lung, fetal brain, fetal and adult lung, sciatic nerve, pituitary.	Astrocytes, Schwann cells, Fibroblasts.	Mitogen	Structurally similar to AR but show differences in biological activity e.g. SDFG has no effect on A431 cells.	24
Heparin-binding EGF-like growth factor (HB-EGF).	Multiple forms (19-23 kDa) due to differential processing and post-translational modification of 208 a.a. transmembrane precursor. 41% homologous to EGF.	Human macrophages, Smooth muscle cells. Wide variety of tissue including skeletal muscle, lung, brain, heart,liver.	Fibroblasts, Smooth muscle cells, Keratinocytes.	Mitogen	Heparin binding property postulated to mediate biological effects. Cell-surface associated transmembrane form serves as a receptor mediating the binding and uptake of diphtheria toxin into cells. Implicated in wound healing and atherosclerosis.	22,56, 57

Growth Factor	Description	Localization/Sources	Known Target	Known Effects	Comments	Ref.
Cripto	Predicted product of a cDNA clone (18 aa) isolated from undifferentiated human teratocarcinoma cells.	Undifferentiated teratocarcinoma cells.	Fibroblasts	Transforming for fibroblasts.	Proposed to be involved in maintaining stem cell integrity.	21
Vaccinia Virus Growth Factor (VVGF)	A 25 kDa (77 aa) glycoprotein derived from a 140 residue transmembrane precursor.	Vaccinia virus infected cells.	Epithelial cells, Fibroblasts.	Mitogenic for keratinocytes, fibroblasts. Transforming in the presence of TGF-ß.	Acts through the EGF receptor. Shorter form VVGF containing only the EGF repeat acts as antagonist of other EGF receptor ligands.	27,28, 58
Betacellulin	A 32kDa (80 aa) glycoprotein derived from a 177 residue transmembrane precursor. 50% homology to TGF-α.	Wide variety of normal mouse tissue including thymus, lung, heart, liver, spleen, small intestine, pancreas, kidney, muscle, testis, uterus. Also in tumor cell lines such as sarcomas, fibrosarcomas, and adenocarcinomas.	Retinal pigment epithelial cells, Vascular smooth muscle cells.	Mitogen	Production by proliferating pancreatic ß-cells and mitogenic effects suggest role in vascular complications associated with diabetes.	23

Growth Factor	Description	Localization/Sources	Known Target	Known Effects	Comments	Ref.
Heregulin(HRG)/*Neu* differentiation factor (NDF)	HRG: 45kDa protein derived from a 640 aa precursor. Several isoforms predicted from cDNA cloning. 45% similar to HB-EGF. NDF: 44kDa soluble glycoprotein derived from a 422 residue transmembrane precursor. HRG and NDF are homologs.	HRG: multiple mRNA transcripts detected in wide variety of human tissue including breast, ovary, testis, prostate, heart, lung, liver, kidney, salivary gland, small intestine, brain, spleen. Also detected in breast carcinoma cell lines. NDF: multiple mRNA transcripts detected in several rat tissues including spinal cord, lung, brain, ovary, stomach, skin, kidney, heart. Also various human tumor cell lines and *ras*-transformed fibroblasts.	Epithelial cells	Mitogen. Growth inhibitory and differentiation effects on mammary tumor cells.	Ligands of the *her2/erbB2/neu* protooncogene receptor tyrosine kinase. Overexpression of this receptor is correlated with several human malignancies, including breast, ovarian, gastric and endometrial cancers, and non-small cell lung adenocarcinoma. Contains nuclear targeting motif.	25,26, 59

receptors. Probably the most widely used *in vitro* bioassay for TGFs is the soft-agar colony forming assay. Other indicators of cell transformation include morphological changes, and the ability of the transformed cell to cause tumor formation when injected into nude mice. That the grns are transforming growth factors is shown by the ability of grnA/epi1 to induce the anchorage independent growth of normal rat kidney cells in the presence of TGF-ß (2). The recent report that the partial amino acid sequence of transforming growth factor-e (TGF-e) is indistinguishable from grnA (7), further implicates the grns as transforming growth factor-like proteins. Transforming growth factor-e (TGF-e) is a TGF-like activity present in a wide variety of epithelial tissue and cells that can stimulate soft agar growth of some epithelial cell lines, including small cell carcinomas, squamous cell carcinomas, and embryo-derived cells. (7). However, it is not known if the grns act as transforming growth factors *in vivo*.

TGF-α and TGF-ß are the two major representatives of transforming growth factors, and will be discussed briefly. Also discussed briefly is TGF-e.

1.1.3.1 Transforming Growth Factor-α.

The discovery of transforming growth factor-α (TGF-α) was associated with the observation that culture supernatants of sarcoma-virus transformed cells secreted a mitogen capable of binding to the EGF receptor (60,61). TGF-α is a 50 amino-acid polypeptide that shows significant homology to both human (44%) and mouse (33%) EGF respectively, including the conservation of all six cysteines and the formation of three similar disulfide bonds (reviewed in 50,62), and belongs to the EGF-family

of growth factors. This structural relationship may explain why TGF-α and EGF share a common cellular receptor (63).

Secreted TGF-α exists as multiple forms, ranging in apparent size from 5-20 kDa. The size heterogeneity is due to differential processing of the TGF-α precursor, and N- and O-glycosylation of the excised peptides (62). ProTGF-α is synthesized as a 160 amino acid transmembrane protein in humans (64,65). The mature 50 amino acid TGF-α sequence lies in the extracellular domain, 10 residues from the transmembrane domain, and shows 92% sequence identity between the human and rat (64,65). The cytoplasmic domain (38 amino-acids) is palmitoylated at one or more cysteines and shows only one amino acid difference (Val136 to Ile) between the human and rat (64,65). This structural feature and extreme sequence conservation suggest a conserved function for the cytoplasmic domain of proTGF-α.

TGF-α has many biological actions both *in vitro* and *in vivo*. Because TGF-α acts through the EGF-receptor, its actions are similar to those of EGF, including the stimulation of keratinocyte and fibroblast proliferation, the suppression of gastric acid secretion, and the acceleration of healing of epidermal injuries (50,62). The similar effects of EGF and TGF-α on cells make it essential to establish which of these two molecules is present in the tissue of interest. Despite the similarities, there are some differences between the actions of EGF and TGF-α. For example, TGF-α is a more potent inducer of neurovascularization than EGF, TGF-α induces greater production of prostaclyclins by endothelial cells, and TGF-α is more potent promoter of calcium release from fetal rat long bones than EGF (50,62).

TGF-α has been implicated in many pathophysiological processes. For

example, overexpression of TGF-α in keratinocytes is suggested to be responsible for the initiation and maintenance of psoriasis (66). TGF-α expression is prevalent in tumors, tumor-derived cell lines, and cells transformed by cellular oncogenes, retroviruses, and tumor promoters (62), suggesting an involvement of TGF-α in tumorigenesis. This is further implied by transgenic studies, where overexpression of TGF-α in transgenic mice induces epithelial hyperplasia, pancreatic metaplasia, and mammary and liver neoplasias (67). TGF-α is also produced by a wide variety of non-neoplastic cells including activated macrophages, eosinophils, hepatocytes, keratinocytes, gastrointestinal cells, brain cells, and placental cells (62), suggesting that TGF-α is also involved in normal physiological processes. In adult human tissues, TGF-α is detectable, as mRNA and/or protein in a wide variety of tissue including the digestive tract, liver, pancreas, kidney, thyroid, adrenal, skin keratinocytes, mammary gland, genital organs, salivary gland, pituitary, brain, and macrophages (68), with regenerative and/or hyperplastic epithelial tissue (e.g. skin and gastrointestinal tract) showing elevated levels of expression. In fetal tissue, TGF-α mRNA and/or protein is present in nerve tissue, liver, adrenal, kidney, the nasopharyngeal pouch, and the maternal decidua (68). Interestingly, the distribution of TGF-α is similar to the grns. The grn mRNA is also widely distributed in a variety of neoplastic and non-neoplastic cell lines as well as in adult and fetal tissue (see Chapter 4).

Again the apparent similarities between TGF-α and the grns suggest that, once the grns can be generated in sufficiently large amounts, it may be profitable to compare their activities with TGF-α.

1.1.3.2 Transforming Growth Factor-ß Family.

The transforming growth factor-ß (TGF-ß) family includes a group of closely homologous dimeric proteins which regulate cell growth, cell differentiation, and cell function (reviewed in 69,70,71). The TGF-ßs are structurally unrelated to TGF-α. The TGF-ß family in turn, belongs to a larger, extended superfamily, whose members affect a wide range of differentiation processes during embryonic development. Members of this superfamily include, for example, Mullerian Inhibiting Substance (MIS), a protein that induces the regression of the Mullerian duct system in the developing male genitourinary system (72); the *Drosophila* decapentaplagic gene product, which is required for dorsal-ventral axis formation and morphogenesis of the imaginal discs (73); the bone morphogenetic proteins (BMPs, osteogenin, osteogenic protein (OP-1), which may regulate cartilage and bone formation (74); the *Xenopus* Vg-1 gene product, which localizes the vegetal pole of eggs (75); the activins, proteins involved in the induction of mesoderm and anterior structure formation in *Xenopus* embryos (76); dorsalin-1, a gene selectively expressed in the dorsal neural tube, that appears to regulate cell differentiation within the neural tube (77); and GDF-3 and GDF-9, two novel mammalian members of unknown function identified at the cDNA level (78). Each member of the superfamily is synthesized as part of a larger secretory precursor. The homologies of these proteins with each other reside in the C-terminal domains of the precursor, which are generally cleaved to form the mature bioactive peptides, which share with TGF-ß, 25-40% sequence identity and conservation of at least seven of the nine residues (69,78).

1.1.3.3 Transforming Growth Factor-ß.

Transforming growth factor-ß (TGF-ß) was originally identified by its ability to induce anchorage independent growth of normal rat kidney fibroblasts in soft agar (61). To date, five distinct molecular forms of TGF-ß have been identified; three in mammals (TGF-ß1, TGF-ß2, and TGF-ß3), and one each in chicken (TGF-ß4) and *Xenopus* (TGF-ß5) respectively (79). The TGF-ßs show 68-86% homology to each other, including the conservation and perfect alignment of nine cysteine residues, of which eight are involved in intrachain disulfide bridging (69). The TGFßs are synthesized as larger precursor polypeptides (preproTGF-ß) containing a hydrophobic signal sequence, that are then processed to yield 12.5 kDa monomers and an N-terminal latency associated peptide. Two identical disulfide-linked monomers constitute the biologically active, 25 Kda mature TGF-ß. The TGF-ß isoforms are highly conserved across species. For example, the mature TGF-ß1 is completely conserved across human, bovine, simian and porcine species, and differs by only 1 residue in murine TGF-ß1 (69,70,71).

TGF-ß is secreted as a latent complex that must be converted to an active form before interacting with its ubiquitous high affinity receptors. Three forms of latent TGF-ß have been identified (reviewed in 80): small latent complex, consisting of non-covalently associated dimers of mature TGF-ß and the remainder of the TGF-ß precursor (or latency associated peptide, LAP); large latent complex, comprising the small latent TGF-ß complex disulfide linked to a 125-160 kDaA protein called latent TGF-ß binding protein (LTBP); and a TGFß-$\alpha_2$macroglobulin complex, where mature TGF-ß dimers found in plasma are covalently bound to $\alpha_2$macroglobulin. The

mechanisms involving the conversion of the latent forms to active TGF-ß *in vivo* have not been fully characterized but appear to be cell specific and might involve processes such as acidification or proteolysis. The activation of the latent molecule has been implicated in the biological control of TGF-ß activity. Inappropriate activation of latent TGF-ß has been implicated in a number of pathological conditions, including acute mesangial proliferative glomerulonephritis, diabetic nephropathy, immunosuppression, re-stenosis of vessels after angioplasty, as well as fibrosis of the skin, central nervous system, lungs, liver, and heart (reviewed in 81). Colouscou et al. have postulated the existence of a grn binding protein (82), but at present there is little direct evidence to suggest that the grns are associated with binding proteins.

The TGF-ß gene is expressed in most transformed and normal cells in culture (69,70). In this respect, it is very similar to the grn gene which is also widely expressed in transformed and normal cells in culture (see Chapter 4). In tissues, expression of TGF-ß is detected throughout embryonic development and into adulthood, with the expression pattern of the TGF-ß isoforms mRNA and protein spatially and temporally restricted (69,70). For example, in mouse embryo, TGF-ß1 mRNA is detectable in lung, intestine, and kidney mesenchymes, epithelial structures, megakaryocytes, osteocytes, and centers of hematopoiesis, while TGF-ß2 mRNA is detectable in gastrointestinal and tracheal submucosae, blood vessels, skin, cartilage, and bone. In the adult mice, TGF-ß immunoreactivity is present in cells of the adrenal cortex, bone marrow, cardiac myocytes, chondrocytes, renal distal tubules, ovarian glandular cells, and chorionic cells of the placenta, and in the

extracellular matrices of the cartilage, skin, heart, pancreas, placenta, and uterus (69,70). The widespread expression of TGF-ß mRNA in fetal and adult tissue again draws comparisons to the pattern of grn mRNA expression in fetal and adult tissue (see Chapter 4).

TGF-ß elicits a variety of cellular responses and acts on a wide range of cells depending on the type and state of differentiation of the cell, and cell culture conditions (69,70,71). *In vitro*, TGF-ß is bifunctional, it strongly inhibits the proliferation of many normal and tumor-derived cell types of mesenchymal and myeloid origin as well as nearly all epithelial, lymphoid and endothelial cells, and stimulates the proliferation of a few mesenchymal cell types (fibroblasts and osteoblasts). It is worth noting that the grns also exhibit bifunctional growth modulating properties *in vitro*. The grns gene products inhibit the growth of several epithelial cells in culture, and stimulates the proliferation of keratinocytes and fibroblasts (2,8,9). Several studies suggest that TGF-ß also inhibits cell growth *in vivo*. For example, the *in vivo* administration of TGF-ß inhibits mouse mammary gland ductal growth (83), and inhibits the early phase of hepatocyte proliferation that occurs in response to partial hepatectomy in the rat (84). Apart from modulating cell growth, TGF-ß can also effect cellular differentiation (69,70). Examples include the promotion of chondrogenesis and certain normal and tumor epithelial cell differentiation *in vitro*, as well as the inhibition of adipocyte and myoblast differentiation (69,70). TGF-ß can also stimulate extracellular matrix (ECM) formation through the stimulation of ECM protein synthesis, the inhibition of proteinase synthesis and stimulation of proteinase inhibitor synthesis, and an increase in cell

surface membrane proteins that bind ECM proteins (85). TGF-ß also exhibits diverse immunoregulatory activities including the suppression of T-cells, the inhibition of B-cell Ig production, cytolysis of natural killer cells, the inhibition of cytokine production from lymphocyte activated killer cells, and the inhibition of macrophage respiratory bursts (86 and references therein). In addition, TGF-ß is a potent chemoattractant for macrophages and fibroblasts (87).

Although well studied, the *in vivo* functions of TGF-ß are not completely understood at present. However, the multifunctional properties of TGF-ß strongly implicates TGF-ß as a key modulator of a wide variety of important physiological and pathophysiological processes including embryogenesis, bone formation and remodelling, tissue repair, inflammation, immunosuppression, angiogenesis, fibrosis, and tumor development (69,70,71). That TGF-ß serves a crucial function *in vivo* is further supported by transgenic studies. Animals homozygous for the mutated TGF-ß1 allele show no gross developmental abnormalities, but soon after birth they succumb to a wasting syndrome accompanied by a multifocal, mixed inflammatory cell response and tissue necrosis, leading to organ failure and death (88).

1.1.3.4 Transforming Growth Factor-e.

In 1992, Parnell et.al. reported the partial amino acid sequence of transforming growth factor-e (TGF-e), a TGF-like activity distinct from other known growth factors (7). The reported partial sequence of this factor is indistinguishable from grnA (see Figure P.1), although the physiochemical properties of the reported TGF-e (see below) indicate that it is a larger protein than the 6 kDa grns or epis. In view of the

possibility that TGF-e probably represents a product of the grn gene, TGF-e will be reviewed briefly.

TGF-e was initially characterised as an autostimulatory activity present in the conditioned medium of SW-13 adrenal adenocarcinoma cells capable of inducing the irreversible anchorage-independent growth of the indicator cells and was subsequently shown to be present in a variety of cell lines and in both neoplastic and nonneoplastic tissues of mainly epithelial origin (89,90). For example, extracts of several human solid cancers such as colonic, gastric, renal, lung, ovarian, and squamous cell carcinomas contain TGF-e activity, while no TGF-e activity is detected in extracts of most non-carcinoma malignant neoplasms such as sarcomas, melanomas, lymphomas, and chordomas (89,90). Several human neoplastic cell lines such as SW-13 cells, A431 cells (epidermoid carcinoma of vulva), A549 cells (bronchioalveolar carcinoma of the lung) and D562 (epidermoid carcinoma of nasopharynx) and several colon carcinoma cell lines also secrete TGF-e activity (89,90). TGF-e activity is also detected in extracts of several nonneoplastic tissues, including kidney, lung and mouse embryo (89,90). In addition, TGF-e activity is present in human milk, plasma, platelets, and amniotic fluids (7). Similarly, the grn gene is also expressed in a wide variety of cell lines and tissues, and is abundant in tissues that are rich in epithelia (see Chapter 4).

TGF-e exhibits growth modulating properties *in vitro*. In addition to stimulating the anchorage independent growth of certain carcinoma cells of epithelial origin, and of AKR-2B fibroblast-like cells (90,91), TGF-e acts as a mitogen and progression factor for both epithelial and fibroblastic cells (91). Additionally, TGF-e has affinity

for heparin, and can partially overcome the heparin-induced growth inhibition of SW-13 carcinoma cells in monolayer and soft agar respectively (92). TGF-e is eluted from a heparin-Sepharose column with 0.5M NaCl, as is the grn precursor (see chapter 3). TGF-e differs from other known growth factors such as EGF, TGF-α, TGF-ß, aFGF, and insulin, with regard to its ability to stimulate soft agar growth of SW-13 cells, with the exception of bFGF. However, TGF-e does not bind to FGF receptors indicating that TGF-e is distinct from bFGF.

The TGF-e activity exists as several different molecular weight forms. For example, the TGF-e activity isolated from bovine kidney whose reported partial amino acid sequence is indistinguishable from grnA, exists as a 20-25 kDa protein (7), while human TGF-e has a reported size of 59 kDa (93). Several lower molecular weight forms of TGF-e activity of 13-15 kDa and 6.5 kDa have been reported (94). Whether these different forms are due to glycosylation, proteolytic processing of the precursor or represent different proteins is not known. Interestingly, the grn precursor is also processed in a tissue and cell specific fashion to generate several different sized products (1,2,3,4,6,8,9 and see Chapter 3).

It is likely that TGF-e is a product(s) of the grn precursor gene. However, until the complete amino-acid sequence of the various forms of TGF-e is determined and the TGF-e gene(s) cloned, the possibility that the grns and TGF-e are identical remain speculative, although plausible at present.

1.1.4 PC-derived Growth Factor.

More recently strong evidence has been presented that the intact grn

precursor is secreted as an autocrine growth factor for PC cells, a highly tumorigenic cell line derived from a mouse teratoma (9). PC cells have lost cell surface binding and response to growth factors such as insulin, EGF, and TGF-ß (9). The amino acid sequence of the N-terminal and several proteolytic peptide fragments of the PC-derived growth factor (PCDGF) were identical to sequences found within the mouse grn precursor (9). The PC-derived growth factor (PCDGF) is a heparin binding glycoprotein of 88 kDa (9). Both PCDGF and the grn precursor are eluted from a heparin column with 0.5M NaCl (9, and see Chapter 3). PCDGF can function as a mitogen for 3T3 fibroblasts as well as an autocrine growth factor for PC cells (9).

1.1.5 The Fibroblast Growth Factor Family.

The fibroblast growth factors represent another family of widely distributed growth factors. The fibroblast growth factor (FGF) family presently consists of nine structurally related polypeptides, of which acidic FGF (aFGF or FGF-1) and basic FGF (bFGF or FGF-2) are the best characterized and serve to establish the basis of a larger family of polypeptide growth factors. Other members of this family include *int-2* (FGF-3), *hst-1*/ks3 (FGF-4), FGF-5, FGF-6, KGF (FGF-7), AIGF (FGF-8) and GAF (FGF-9). The genes for each member of the FGF family have been cloned and sequenced, and probably derive from a common ancestral gene. The FGF family members are approximately 30% to 80% identical to each other. The multifunctional properties of the FGFs include mesodermal cell mitogenesis, angiogenesis, neurite extension, neuronal cell survival and inhibition of myoblast differentiation, suggesting that members of the FGF family have important physiological roles in the

development of vascular, nervous and skeletal systems, in promoting the maintenance and survival of certain tissues and in stimulating wound healing and tissue repair (reviewed in 95,96).

1.1.5.1 Fibroblast Growth Factor-1 and Fibroblast Growth Factor-2.

FGF-1 and FGF-2 were initially isolated from extracts of bovine brain (97) and pituitary (98) respectively and subsequently characterized as potent heparin-binding polypeptide mitogens for endothelial cells (99,100). FGF-1 and FGF-2 show 55% sequence identity and act through separate but closely related receptors (101). However, unlike most polypeptide growth factors, both FGF-1 and FGF-2 lack classical consensus signal peptide sequences. Biosynthesis studies suggest that FGF-1 and FGF-2 are matrix- and cell-associated and are not released into the conditioned medium (102,103). The mechanism of secretion of FGF-1 and FGF-2 is not known. Novel secretion pathways activated by heat (104) or secretion by exocytosis independent of the endoplasmic-Golgi apparatus (105) and cellular disruption (106) have been postulated.

The genes for FGF-1 and FGF-2 have been cloned (107,108). The single copy human FGF-2 gene encodes four coexpressed FGF-2 isoforms (24, 23, 22 and 18-kDa)) as a result of cotranslation from a single transcript, from upstream CUG codons (22, 23 and 24 kDa) and from a classical AUG codon (18 kDa) (109). *De novo* transfection studies show that the different FGF-2 isoforms localize to different subcellular compartments. All three higher molecular weight isoforms localize exclusively to the nucleus, while the 18-kDa isoform is exported both to the cell

surface and to the nucleus (109). This suggests that the different isoforms might have different functions dependent upon localization. Interestingly, the grn gene products also exist in several different size forms (1,2,3,4,6,8,9 and see Chapter 3) and it is conceivable that these different forms have overlapping and distinct functions *in vivo*.

FGF-1 and FGF-2 proteins have been isolated from a variety of tissues, including brain, pituitary, corpus luteum, prostate, kidney, adrenal gland, macrophage, retina, liver and testis (95,96). FGF mRNA and protein expression is also detected in many normal and transformed cell types *in vitro*. For example, FGF-2 mRNA and protein is present in bovine retinal pigment epithelial cells, bovine pituitary follicular cells, bovine adrenal cortex cells, and bovine endothelial cells from the brain, aorta, adrenal cortex and capillary bed, human umbilical vein endothelial cells, human retinoblastoma cells, human rhabdomyosarcoma cells, cervical carcinoma cells, embryo lung fibroblasts, melanoma, retinoblastoma, human hepatoma, Ewing's sarcoma, bovine and chicken embryo fibroblasts, and cells derived from murine peritoneal exudate or bladder tumors (96 and references therein). FGF-1 mRNA is detected, *in vitro*, in human glioma, medullablastoma, rhabdomyosarcoma, foreskin fibroblasts and vascular smooth muscle cells (96). Similarly, the grn gene is also widely expressed. Grn mRNA is detected in a variety of transformed and nontransformed cells and in tissues derived from the endoderm, mesoderm and ectoderm (see Chapter 4).

FGF-1 and FGF-2 are multifunctional and can act as mitogens, stimulate chemotaxis and promote cellular differentiation (95,96). FGF-1 and FGF-2 are

mitogenic to a variety of normal cell types from mesoderm and neuroectoderm lineages, including endothelial cells, smooth muscle cells, adrenal cortex cells, prostatic and retinal epithelial cells, oligodendrocytes, astrocytes, chondrocytes, myoblasts, and osteoblasts (95,96). In addition, they are also potent mitogens for a large number of established cell lines of fibroblast, epithelial and mesenchymal origin. Similarly, the grn gene products are also mitogenic for fibroblasts and epithelial cells (2,7,9). FGF-1 and FGF-2 can act as competence factors for a variety of growth factors and FGF-2 can induce division of fibroblasts and epithelial cells in soft agar (95,96). Likewise, grnA also has transforming growth factor activity (2). FGF-1 is chemotactic for endothelial cells, fibroblasts and astroglial cells (95,96), suggesting a role for FGFs in wound repair, angiogenesis, neurite outgrowth and tumor metastasis. In addition, FGF-1 and FGF-2 can promote cellular differentiation *in vitro* (95,96). For example, FGF-2 induces adipocyte differentiation *in vitro*, and both FGF-1 and FGF-2 promote the survival and differentiation of a variety of cells that originate from the neural crest, including hippocampal and cortex derived cells *in vitro* (95,96). FGF-1 and FGF-2 are also thought to play a role in embryogenesis, possibly during the development of the nervous, skeletal and vascular systems (95,96).

1.1.5.2 Other Members of the FGF Family.

Unlike FGF-1 and FGF-2, the restricted distribution of the other members of the FGF family suggests that they act in a more specialized than general fashion. The other FGF family members will not be discussed in detail. Instead the properties

Table 1.3 Members of the Fibroblast Growth Factor Family

Growth Factor	Description	Signal Peptide	Localization/Source	Known Target	Known Effects	Comments	Ref.
Int-2 (FGF-3)	27-32 kDa translation product of *int*-2 protooncogene, originally discovered through transcriptional activation by MMTV. Precursor protein of 231 aa.	Yes	mRNA detected in developing embryo and traces in adult brain and testis. Also in mouse mammary tumor and teratocarcinoma cells.	Fibroblasts	Transforming, but less effective than other members of FGF family.	The growth promoting activities of *Int*-2 are not known. mRNA expression suggests function in specialized cell type rather than in a general capacity.	110 111 112
Hst-1/ *ks3* (FGF-4)	18 kDa glycoprotein product of *hst*(human) or *ks3* (mouse) protooncogenes. Precursor protein of 206 aa. 40% homology to FGF1 and FGF2	Yes	mRNA in germ cell tumors and mouse embryos. Present as a transforming gene in a variety of cancerous and non-cancerous tissue including chronic myelogenous leukocytes, gastric and colon tumors, hepatoma, melanoma, osteosarcoma, and colon mucosa.	Fibroblasts, Endothelial cells.	Mitogen for fibroblasts and vascular endothelial cells. Transforming for fibroblasts. Potent angiogenic activity *in vivo*		113 114 115 116 117 118

Growth Factor	Description	Signal Peptide	Localization/Source	Known Target	Known Effects	Comments	Ref.
FGF-5	Product of a human oncogene able to transform 3T3 fibroblasts. Synthesized as a 268 aa precursor. 40% homologous to FGF-1.	Yes	mRNA expressed through most phases of embryogenesis and weakly but exclusively in adult central nervous system. mRNA also present in certain tumor cell lines including hepatoma, bladder, and endometrial carcinoma	Fibroblasts, Endothelial cells.	Mitogen		119 120
FGF-6	Isolated by homology cloning using a FGF-4 specific probe. Putative protein product of 198 aa showing 79% identity to FGF-4.	Yes	mRNA: low levels in mouse embryo during middle and late gestation, restricted to skeletal muscle lineage. In adult, mRNA detected in testes and cardiac and skeletal muscle tissue.	Fibroblasts	Mitogen	FGF-6 transformed cells are tumorigenic when injected into nude mice.	121 122 123 124
Keratinocyte growth factor (KGF) or FGF-7	Single polypeptide chain of 25-30 kDa derived from a precursor protein of 194 aa 37 and 39% identical to FGF-1 and FGF-2 respectively.	Yes	mRNA detected in stromal fibroblast cells derived from embryonic, neonatal and adult sources. In adult tissues, mRNA detected in kidney, GI tract.	Epithelial cells	Mitogen	Epithelial specific mitogenic activity, and expression by stromal cells suggests KGF may function as a specific paracrine mediator of normal epithelial cell proliferation.	125 126 127

Growth Factor	Description	Signal Peptide	Localization/Source	Known Target	Known Effects	Comments	Ref.
Androgen-induced growth factor (AIGF) or FGF-8	28-32 kDa heparin-binding factor secreted by an androgen dependent mouse mammary carcinoma cell line following testosterone stimulation. Predicted precursor protein of 215 aa. 30-40% homology with other FGFs.	Yes	Androgen stimulated mouse mammary carcinoma cells	Mammary carcinoma cells	Mitogen	Growth stimulating effects and induction by androgen suggests possible role in mediating hormone-induced growth of certain cancer cells.	128
Glia Activating Factor (GAF) or FGF-9	Three mature forms of 25-30kDa derived from differential processing of a 208 aa precursor. 30% similar to other FGFs.	No, but transfection studies show that FGF-9 is a secreted protein, suggesting the presence of a unique secretory pathway.	mRNA restricted to brain and kidney of adult rat, and human glioma cells .	Glial cells	Mitogen	Unlike other members of the FGF family, FGF-9 is not mitogenic for endothelial cells.	129 130

of these proteins are summarized in Table 1.3.

There are, as discussed, similarities between the grns and EGF, TGF-α, TGF-ß, and the FGFs. The similarities with other growth factors are less clear, but to place the grns in a broader biological context, the other well known somatic growth factors will be reviewed briefly.

1.1.6 Hepatocyte Growth Factor / Scatter Factor.

Hepatocyte growth factor (HGF) was initially identified as a factor present in sera of partially hepatectomized rats able to stimulate hepatocyte proliferation *in vitro* (131) and subsequently purified and characterized from rat platelets (132), human plasma (133) and rabbit serum (134). Scatter factor (SF) is a fibroblast-derived protein that causes separation of contiguous epithelial cells and increased mobility of unanchored cells (135,136). HGF and SF have recently been found to be synonymous (137).

HGF contains four kringle domains and a serine-protease-like domain in the mature peptide and is structurally related to plasminogen and other coagulation-related proteases, but lacks proteolytic activity. Purified HGF is a heterodimeric glycoprotein composed of a large α-subunit of 69 kDa and a small ß-subunit of 34 kDa. The HGF cDNA predicts a precursor protein of 728 amino acids containing both a putative signal peptide, and the α- and ß-chains (138,139). The α- and ß-chains are likely processed from the precursor by a trypsin-like protease (138).

HGF mRNA and protein have been detected in a wide variety of cells and

tissue including stromal fibroblasts from human adult skin, lung, gastrointestinal tract and prostate and embryonic lung fibroblasts (138), in adult rat liver, kidney, brain and thymus (139). In adult rabbit, HGF immunoreactivity has been detected in acinar cells of the pancreas, neurons of the brain, C cells of the thyroid, ductal cells of the salivary glands and Brunners glands of the duodenum (140).

HGF exhibits multiple biological effects, depending on target cell types (reviewed in 141). In addition to its enhancement of cell motility resulting in dissociation and scattering of epithelial cells (135,136), HGF is strongly mitogenic for mammary, bronchial and renal-tubular epithelial cells, keratinocytes, endothelial cells, melanocytes and hematopoietic progenitor cells, but not for fibroblasts (141). Conversely, HGF inhibits the growth of tumor cells, including hepatoma, melanoma, and squamous cell carcinoma (141). HGF also acts as an epithelial morphogen, inducing tubule formation of kidney epithelial cells grown in collagen gel matrix (142).

The broad expression of HGF mRNA and protein, and the multiple biological effects of HGF suggest that HGF is a pleiotropic factor involved in many physiological and pathophysiological processes (141). The observation that HGF mRNA and activity markedly increase in the liver and kidney following hepatic insults and acute renal failure respectively (143), and that HGF is a potent mitogen for hepatocytes (131), and possesses mitogenic and morphogenic activities for renal epithelial cells (141) suggest that HGF functions as a hepatotropic and renotropic factor. HGF can prime the neutrophil oxidative response (144) and may be involved in combating infection. Other possible functions of HGF include a role in early hematopoiesis and hematopoietic disorders (145).

1.1.7 Insulin-like Growth Factors.

The insulin-like growth factors (IGFs) belong to a family that currently comprises two proteins, IGF-I and IGF-II, which are similar to proinsulin (reviewed in 146,147,148) and other insulin-like proteins (149). IGF-I and II are single chain polypeptides consisting of an A, B, C, and D domain. The sequence homology between IGF-I and IGF-II is 62%. The IGFs circulate bound to carrier proteins (binding proteins). It is hypothesized that the binding proteins stabilize the IGFs in the circulation, regulate the distribution of the IGFs to their target cells and modulate the actions of the IGFs through receptor presentation (reviewed in 150). As previously mentioned, Culouscou et al. have postulated the existence of a grn binding protein (82), but at present there is little direct evidence to suggest that the grns are associated with binding proteins.

1.1.7.1 Insulin-like Growth Factor-I.

Human IGF-I was first isolated from human serum as a single chain basic polypeptide of 70 amino-acids with three internal disulfide bonds (151). Somatomedin-C, a growth hormone-dependent serum mitogen, has been shown to be identical to IGF-I (152).

Multiple IGF-I mRNA transcripts are detected in a wide variety of adult and fetal tissue (148,153), localised mainly to connective tissue or cells of mesenchymal origin (154). IGF-I is detected in body fluids with highest levels in the serum (148). In addition, IGF-I is also present in various tissue with the highest concentrations of IGF-I detected in kidney, liver, lung, and testis. The liver is the principal source of

circulating endocrine IGF-I. IGF-I in other tissues is believed to support autocrine and paracrine functions (148).

IGF-I exerts a wide range of biological effects *in vivo* and *in vitro* (146,147,148). IGF-I plays an important role in postnatal development, appearing to be the most important mediator of growth in the human. *In vitro*, IGF-I (i) exerts insulin-like effects in insulin target tissues (e.g. adipose tissue, striated and heart muscle) such as stimulation of glucose metabolism to CO_2, lipids, and glycogen, and inhibition of epinephrine-stimulated lipolysis and of glycogen breakdown. (ii) promotes sulfate incorporation into cartilage tissue, which reflects growth *in vivo*, and (iii) promotes cell differentiation and cell proliferation. IGF-I is mitogenic for a wide variety of cells in culture, including fibroblasts, Sertoli cells, fetal brain cells, myoblasts, lens epithelium, pancreatic-β cells, and lectin-activated lymphocytes. Additionally IGF-I stimulates the differentiation of muscle cells (myoblasts to myotubules) and osteoblasts, and stimulates the late stages of erythroid differentiation independently of erythropoietin (146,147,148).

1.1.7.2 Insulin-like Growth Factor-II.

Human IGF-II was first isolated from serum as a single-chain slightly acidic polypeptide with 67 amino-acid residues cross-linked by three disulfide bridges (155). The IGF-II gene is expressed in multiple tissues with IGF-II mRNA levels highest during fetal development and declining in the postnatal period (153). In human fetal tissue, multiple IGF-II mRNA transcripts are detected in a wide variety of tissue with the liver, adrenal and skeletal muscle showing the highest levels of

expression. Expression is also detected in the kidney, skin, pancreas, placenta, intestine, heart, stomach, lung, spleen, thymus and brain stem. The IGF-II mRNA in human fetal tissue is localized to regions that are predominantly populated by fibroblasts and mesenchymal cells (154). In addition, a number of tumors appear to synthesize IGF-II mRNA constitutively (147). The IGF-II protein is detected in a wide variety of fetal and adult tissue and in body fluids (148).

IGF-II is believed to play a role in fetal growth or development (146,147,148). *In vitro*, IGF-II stimulates sulfate incorporation into cartilage, displays insulin-like effects on fat cells and is mitogenic to a variety of cells, including fibroblasts, smooth muscle cells, chondrocytes and bone cells. IGF-II effects the differentiation of some mesodermal cells and stimulates myoblast differentiation and myotubule formation in primary cultures of chick embryonic cells (146,147,148).

1.1.8 The Platelet Derived Growth Factor.

The platelet-derived growth factor (PDGF) family comprises of two related polypeptides, the PDGF A- and B- chains, which form homo (AA, BB)- and hetero (AB)-dimers of equal activity. The v-*sis* viral oncoprotein makes up the third member of this family (reviewed in 156,157,158).

PDGF was first observed as a factor in platelets that supported the growth of chick embryo fibroblasts in culture (159,160) and subsequently identified as the principal mitogenic protein in serum for cells of mesenchymal origin (158). PDGF from human platelets is a heterodimer of two chains, termed A and B. In contrast, porcine PDGF appears to consist of B-B homodimers, while PDGF produced by

several human tumor cell lines, such as osteosarcomas, gliomas and melanomas consist of A-A homodimers (156,157,158).

The cDNA of the PDGF A-chain predicts a 23 kDa precursor comprising a 18-residue signal sequence, a propeptide of 66 amino-acids and a 125 amino acid mature A-chain protein (161). The PDGF B-chain precursor is encoded by the c-*sis* gene, the cellular counterpart to the transforming gene v-*sis* of the simian sarcoma virus (SSV) (162). The human c-*sis* gene produces a 4.2 kb transcript that encodes a 27 kDa polypeptide (162). PDGF-A and B chains show 60% identity.

Normal human plasma contains undetectable levels of PDGF (156), suggesting that the ability of connective tissue cells throughout the body to respond to PDGF probably depends not on circulating levels of PDGF but on local synthesis and secretion or on infiltrates of blood cells. A wide variety of cells are known to secrete PDGF, including circulating cells such as platelets and mononuclear phagocytes; resident cells such as endothelial cells, vascular smooth muscle cells, embryonic cells, and megakaryocytes; and transformed cells such as SSV- and SV40-transformed cells, and osteosarcomal cell lines (158). These cells also produce other mitogens that may complement or modulate the effects of PDGF.

PDGF has a number of biological effects (157,158,163). It acts as a potent vasoconstrictor, induces phospholipase activation and prostaglandin metabolism, modifies the connective tissue matrix, acts as a chemoattractant and stimulates mitogenesis. PDGF is chemotactic for fibroblasts, smooth muscle cells, monocytes, and neutrophils. It also stimulates the release of granular contents by monocytes and neutrophils. In addition, PDGF MRNA and protein have been detected in neurons

of embryonic and adult mice (164,165) and, together with the known mitogenic effects of PDGF on glial cells *in vitro*, suggests that PDGF may also act as a neuronal regulatory agent.

The pleiotropic biological effects have implicated PDGF in a number of physiological and/or pathological processes, including wound repair, embryogenesis, atherosclerosis, neoplasia, desmoplasia, bone marrow fibrosis and myeloproliferative disease (158).

1.2 Other Growth Factors.

There are many other growth factors that do not belong to any of the growth factor families discussed above. Some of these growth factors, like the grns, may serve as the prototype for a larger family of yet undiscovered growth factors. Several examples can be presented.

Two closely related developmentally regulated cytokines, **pleiotrophin** (PTN) and **midkine** (MK), which display both neurotrophic and mitogenic properties. PTN was originally identified as a heparin-binding mitogen for murine fibroblasts and as a neurite growth promoting activity in neonatal brain cultures (166), and subsequently shown to be mitogenic for endothelial cells and normal rat kidney (NRK) cells, and to induce tumor formation in nude mice (167). MK is the product of a retinoic acid responsive gene, and is mitogenic for fibroblasts and PC12 cells, and can promote neurite extension from PC12 cells (168,169).

Platelet-derived endothelial cell growth factor (PD-ECGF) is an endothelial cell mitogen purified from human platelets that has chemotactic activity for

endothelial cells *in vitro* and angiogenic activity *in vivo* (170). Unlike other endothelial cell mitogens (e.g. FGFs) PD-ECGF does not bind heparin, and is not mitogenic for fibroblasts.

Vascular permeability factor / vascular endothelial growth factor (VEGF) is a recently discovered heparin binding mitogen related to PDGF that can induce endothelial proliferation *in vitro* and angiogenesis *in vivo* (171, 172). In addition, VEGF induces vascular permeability, promotes monocyte migration *in vitro*, induces intracellular Ca^{+} shifts and plasminogen activator and plasminogen activator inhibitor 1 synthesis in endothelial cells, and stimulates glucose transport into endothelial cells (173,174).

Autocrine motility factor (AMF) is a tumor cell-derived factor that was initially characterized by its ability to stimulate the random and directed migration of the producer cells (175), and recently shown to function as a paracrine mitogen for 3T3-fibroblasts (176).

1.3 Uncharacterised Growth Modulators.

The increased sophistication of protein purification methods coupled with improved tissue culture techniques has led to the identification of numerous growth factor activities in tissue extracts and in conditioned media of cell lines. The purification, sequencing, and cloning of the factors responsible for the growth modulating activities may well reveal that many of these biological activities are mediated by known growth factors. However, there are many reported factors whose physio-chemical properties (e.g. size, pI, acid/base/heat stability, Ab recognition) are

different from other well studied mitogens, suggesting that they represent novel growth factors. Because the grn gene products exists in several forms ranging in size from 6-90 kDa, have known growth modulating activity for epithelial cells and fibroblasts, has TGF-like activity, and is widely expressed, and has only recently been described, it is conceivable that some of these yet to be characterized activities may be products of the grn gene. Several examples of such factors can be cited: nephroblast growth factor (NB-GF), a growth factor activity secreted by a Wilms' tumor (a neoplasm arising in nephroblasts) cell line that is mitogenic for, and maintains the undifferentiated state of renal mesenchymal progenitor cells (nephroblasts) (177); A human myelomonocytic cell line THP-1 secretes a 50-70 kDa factor that is mitogenic for a wide spectrum of target cells in culture, including myeloid, T-, B-, mastocytoma, melanoma, and fibroblast cells (178); A 67-69 kDa factor secreted by a preadipocyte cell line derived from the 3T3 fibroblast cell line that stimulates the growth of normal breast epithelial cells, and inhibits the growth of breast and colon cancer cell lines (179); A human mesothelioma (tumors of the mesodermally derived serosal surfaces) cell line produces an autocrine growth factor that is also mitogenic for several lung-derived tumor cell lines (180); a 12 kDa muscle derived mitogen from rat muscle that is stimulatory for fibroblasts, and induces soft agar colony formation of 3T3 cells (181); human prostatic cancer cells, PC13, secrete a 26-30 kDa basic mitogen that selectively stimulates human bone cells (182); human malignant gliomas secrete growth factors with TGF-like activity (183);

It would clearly be naive to suggest that all these activities are due to the grns,

but by analogy with the FGFs for which several different activities were identified, it is possible that the same may be true for the grns.

1.4 Granulins and Wound Healing.

It is well established that somatic growth factors are frequently associated with cells of the immune response, and it is generally assumed that this implies important functions for these peptides in tissue repair processes. Many of the growth factors discussed in the preceding pages are implicated in this process. Examples include platelet-derived growth factor, transforming growth factor-ß, fibroblast growth factor, epidermal growth factor, and heparin-binding EGF. Some of these peptides may be involved in initiating or sustaining an inflammatory response (87,184), in tissue remodelling at the site of a wound by regulating matrix formation (85,185), and in the recruitment of neighbouring fibroblasts by chemotaxis (186). The similarity between the molecular biology of wound repair and tumor development has been discussed (187), and it is likely that many of the same regulatory polypeptides are involved in both processes.

The isolation of grns from inflammatory leukocytes (1), and the widespread distribution of grn mRNA in hematopoietic cells (see Chapter 4) implicate the grns in the immuno-inflammatory response. Although it has not been possible to determine their role in these processes directly, one working hypothesis by researchers in our laboratory is the possible involvement of the grn gene products in tissue repair. The role of growth factors in wound healing will be discussed briefly.

1.4.1 Growth Factors and Wound Healing.

Wound healing is a complex biological process that involves the integration of inflammation, mitosis, angiogenesis, and synthesis and remodelling of the extracellular matrix. Peptide growth factors have been shown to regulate many of these processes *in vitro*, strongly suggesting that peptide growth factors are also important regulators of the key processes of wound healing *in vivo*. The association of human and rat grns with inflammatory leukocytes, and the mitogenic properties of the grn gene products suggests roles for the grn gene products in wound healing.

Wound healing is initiated by the process of blood clotting and platelet degranulation, which results in the release of several growth factors which are potent chemotactic factors for inflammatory cells. The inflammatory cells (neutrophils and macrophages) are rapidly drawn into the wound and begin to synthesize and secrete additional growth factors. The secreted growth factors stimulate the migration of fibroblasts and epithelial cells that proliferate and repopulate the wound area, and vascular endothelial cells that extend new capillaries into the wound site, establishing a new blood supply. As the number of inflammatory cells in the wound begins to decrease, other cells in the wound, such as fibroblasts, endothelial cells, and keratinocytes, continue to synthesize and secrete growth factors that continue to stimulate cell proliferation, synthesis of extracellular matrix proteins to form scar tissue, and new capillary formation. As proliferation and neovascularization ceases, and the extracellular matrix is remodelled through a balance of synthesis of new components of the scar matrix and their degradation by proteases.

Peptide growth factors regulate many of these processes *in vitro*. These include the chemotactic migration of cells. For example, TGF-ß is a potent chemotactic factor for monocytes (87) and PDGF is chemotactic for fibroblasts and smooth muscle cells (188,189). bFGF and IGF-I are chemotactic for vascular endothelial cells (190), and EGF for epithelial cells (191). Additionally, growth factors are mitogenic to the cells involved in wound healing i.e. fibroblasts, keratinocytes, and vascular endothelial cells, with some degree of selectivity observed. For example EGF/TGF-α and KGF are effective mitogens for epithelial cells (50,126), bFGF for fibroblasts and vascular endothelial cells (95), IGF-I and PDGF for mesodermal cells such as fibroblasts and smooth muscle cells (146,156). Several growth factors can induce neovascularization, including bFGF, EGF, TGF-α and TGF-ß (192). The synthesis and degradation of extracellular matrix components is also important in wound healing. For example, EGF stimulates the synthesis of fibronectin (193) and TGF-ß stimulates the synthesis of collagen, elastin, and tissue inhibitors of metalloproteinases, while decreasing the production of collagenase (85).

The direct involvement of peptide growth factors in wound healing is further supported by experiments using animal models of wound healing. Several examples can be cited. Topical treatment with EGF, TGF-α, FGF, or the combination of PDGF plus IGF-I accelerates the rate of epidermal regeneration in skin wounds (194,195,196,197). PDGF increases the formation of granulation tissue in diabetic rats (198), and enhances healing of chronic pressure ulcers (199).

1.5. Granulin Receptors.

In 1993, the preliminary biochemical analysis of a grn/epi receptor on a human breast carcinoma cell line was reported (82). This high affinity 140 kDa receptor (K_d $2X10^{-10}M$) binds both grnA/epi1 and grnB/epi2 with the same affinity. However, a discrepancy exists between the relatively high affinity (0.2 nM) of the receptor for its ligands, and the high levels of ligand required to elicit a growth response (half maximal biological activity at approximate 60 and 300 nM for grnA/epi1 and grnB/epi2 respectively). This can be compared with the K_d of other growth factor receptors such as the EGF receptor (K_d = 0.1-1 nM) (48), TGF-ß receptor-I and II (K_d = 5-50 pM) (200), FGF receptors (K_d = 20-600 pM) (101), IGF-I receptor (K_d = 1.5 nM) (201), IGF-II receptor (K_d = 2 nM) (201), and PDGF receptors (K_d = 0.1-10 nM) (156). However, unlike the grns, the half-maximal concentrations of ligand required to elicit a growth response correlates closely with the K_d of the above mentioned receptors respectively.

It is not known if the recently characterized grn receptor possesses an intrinsic protein kinase activity. Additionally the presence of other receptors for the grn family of peptides cannot be excluded. In view of the recent identification of a grn/epi receptor on an epithelial cell line, our current understanding of growth factor receptors will be discussed briefly.

1.5.1 Growth Factor Receptors.

Growth factors are believed to exert effects on target cells through integral membrane proteins that are coupled to cytoplasmic signal transducers (i.e. growth

factor receptors). Binding of growth factors to their cognate receptors evokes pleiotropic cellular responses, including the stimulation of Na^+/H^+ exchange and glucose and amino acid transport, Ca^{2+} influx, activation of membrane kinases and a number of cytoplasmic pathways, pinocytosis, membrane ruffling, and other cytoskeletal and morphological changes; events which in many cell types culminate in cell cycle progression, DNA synthesis, and cellular replication.

As with the growth factors, it has become clear that growth factor receptors can also be divided into several types based on common structural features. Members of the EGF, IGF, PDGF, FGF, NGF, and HGF receptor families all share in common an intrinsic, ligand-sensitive, protein-tyrosine-kinase activity. This class of receptor tyrosine kinases (RTK) comprise an extracellular ligand binding domain connected by a hydrophobic membrane spanning segment to a cytoplasmic catalytic domain involved in signal transduction. They can be classified into 6 subclasses based on primary sequence similarity and distinct structural characteristics (202,203).

Subclass I

Representatives of subclass I receptors are the EGF receptor, and the HER2/*neu* (c-*erb*2) and HER3 (c-*erb*3) receptor-like gene products (204). Characteristic structural features of this subclass are two cysteine rich repeat regions in the extracellular ligand binding domain. Following the intracellular catalytic domain is a C-terminal extension that contains the major sites for autophosphorylation, and is proposed to modulate the activity of the tyrosine kinase domain.

SubclassII

Class II RTKs include the receptors for insulin and IGF-I (201). These receptors comprise two α and ß subunits disulfide -linked to form a $\alpha_2\beta_2$ tetrameric structure. Each α subunit possesses a cysteine-rich repeat and together contribute to form the ligand binding domain; the transmembrane ß subunits contain the catalytic kinase domains. The α and ß subunits are cleavage products of a receptor precursor that is the product of a single gene.

Subclass III

This receptor subclass is defined by the presence of five immunoglobulin-like repeats in the extracellular domain, and, unlike the class I and II RTKs, the intracellular catalytic domain is interrupted by a long, structurally unique, hydrophilic insertion sequence. This kinase domain insertion sequence may be involved in defining substrate specificity and other receptor-specific functions. Members of the subclass III RTK include the α and ß PDGF receptors (163).

Subclass IV

The receptors for the FGFs (e.g. *flg*, *bek*, *cek*) constitute the class IV RTKs (101). Class IV RTKs have similar structural characteristics found in subclass III RTKs, with the difference of one to three instead of five immunoglobin-like repeats in the extracellular domain. The second and third immunoglobulin-like loops which are implicated in ligand binding, and the tyrosine kinase domains are highly conserved between members of this family. Several forms of these receptors,

generated by differential splicing, with different ligand-binding specificities and affinities have been identified.

Subclass V

The product of the c-*met* protooncogene (the HGF receptor) serves as the prototype of this subclass of RTKs (205). The c-*sea* protooncogene product was recently identified as the second member of this family (206). This subclass of receptors is characterized by a heterodimeric ($\alpha\beta$) subunit structure, in contrast to the monomeric structure of class I, III, and IV RTKs, and tetrameric subunit structure of class II RTKs respectively. The α and ß- subunits are cleavage products of a single precursor protein. Both the α and ß-chains are glycosylated and exposed at the cell surface; the ß-chain spans the membrane and contains the cytoplasmic tyrosine kinase domain.

Subclass VI

The neurotrophin receptors constitute the class VI RTK family, which currently numbers three members, *trk*A, *trk*B, and *trk*C (207,208,209). The *trk* proteins have a tripartite structure consisting of two immunoglobulin-like modules in the extracellular domain involved in ligand binding, a single transmembrane domain, and a cytoplasmic catalytic domain containing the tyrosine kinase. These receptors are responsible for mediating the trophic properties of the neurotrophins. The neurotrophins also interact with a second type of cell surface receptor. This low affinity receptor, known as the low-affinity neurotrophin receptor (LANR), contains a

highly glycosylated cysteine-rich extracellular region and a short cytoplasmic domain, and binds the neurotrophins with equal affinity but is not able to mediate signal transduction processes. The physiological role of the LANR, and its relationship with the *trks* is unclear.

Although most peptide growth factors act via tyrosine kinases, it should not be assumed that the grns will as well. Additionally, not all growth factor receptors have intrinsic tyrosine kinases activity, some, like the TGF-ß type II receptor (210) is a serine/threonine kinase, while others, like the TGF-ß type III receptor (betaglycan) (200) and the IGF type II receptor (211) are not coupled to any cytoplasmic structures involved in signal transduction.

TGF-ß binds to various membrane proteins, two of which, the TGF-ß type I and II receptors, form a signalling receptor complex (210). Type I receptor is a monomeric transmembrane protein, while the type II receptor consists of a cysteine-rich extracellular domain and a cytoplasmic serine/threonine kinase domain. Both type I and II receptors are required for cell signalling (210). The TGF-ß type III receptor (betaglycan), a non-signalling receptor, serves as a direct regulator of TGF-ß access to the type I and II signalling receptors (212).

The IGF type II receptor is a single transmembrane protein with a large extracellular domain consisting of fifteen cysteine-rich repeats, and a small region homologous to the collagen-binding domain of fibronectin (211). The type II receptor lacks intrinsic kinase activity (211). The IGF type II receptor is identical to the cation-independent mannose 6-phosphate receptor, a protein implicated in the targeting of

lysosomal enzymes to the lysosomes (213). Both IGF-II, and with a reduced affinity, IGF-I, bind to the IGF type II receptor. The significance of the IGFs binding to this nonsignalling receptor is not completely understood. It is proposed that the IGF type II receptor serves as a scavenger receptor (211).

At present there is no published data on the mechanism of action of the grn receptors. The availability of recombinant grns (see Chapter 3) will make it possible to undertake a thorough study of grn receptors and their mode of action.

CHAPTER 2

MATERIALS AND METHODS

2.1 Restriction Enzyme Digests and Modifying Enzyme Reactions.

All restriction and modifying enzymes were purchased from Pharmacia (Piscatoway, NJ), Gibco-BRL (Gaithersburg, MD), Stratagene (La Jolla, CA), New England Biolabs (Beverly, MA), Bio-Can (Mississauga, Ont.), and Boehringer Mannheim (Indianapolis, IN). Restriction enzyme digestions were performed at 37°C for 1 to 16 hrs in the buffers supplied by the manufacturer. DNA modifying reactions (e.g. blunt-ending, end-labeling, ligations, etc.) were carried out according to conditions suggested by the manufacturer or as outlined in Molecular Cloning, a Laboratory Manual (214). Enzyme reaction buffers were either supplied by the manufacturer or made according to protocols provided in Molecular Cloning, a Laboratory Manual (214) or Current Protocols in Molecular Biology (215). DNA nick-translation reactions were performed with enzyme and reagents supplied in kit form (Boehringer Mannheim) according to conditions outlined by the manufacturer.

2.2 Isolation of DNA and RNA.

2.2.1 Extraction of DNA from Tissue.

Freshly obtained or frozen tissue (approximately 2g) was finely chopped, washed in Phosphate Buffered Saline, PBS (137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 1.8 mM KH_2PO_4, pH 7.4) and resuspended in 2 ml of a 10 mM Tris-HCl pH 7.6, 50 mM NaCl solution. 12 ml of 10 mM Tris-HCl pH 7.6, 50 mM NaCl, 0.2% SDS

containing freshly added Proteinase K (200 μg/ml) was added and the reaction mixed gently. Following incubation at 42°C for 16 hrs with gentle agitation to digest the proteins, the DNA was extracted with water-saturated phenol-chloroform-isoamyl alcohol (24:24:1). Extractions were performed by gently rocking the DNA sample in the aqueous phase with an equal volume of the organic solvent for 10 min to obtain a complete emulsion, centrifuging at 600xg (using a Beckman JS-4.2 rotor in a Beckman J6-B centrifuge; Beckman Instruments Inc. Fullerton, CA) for 3 min, and recovering the aqueous phase. The extraction was repeated once with phenol-chloroform-isoamyl alcohol (24:24:1) and twice with chloroform-isoamyl alcohol (24:1). The DNA was precipitated from the aqueous phase by adding 0.02V of 3M NaCl and 1V of isopropanol. The DNA was then removed with a glass rod, washed 3X with 70% ethanol and resuspended in 1mM Tris-HCl pH 7.6, 0.1mM EDTA. The resuspended DNA was then dialysed extensively in 1mM Tris-HCl pH 7.6, 0.1mM EDTA, at 4°C until the DNA was fully dissolved. The concentration of the DNA was determined by measuring the absorbance at 260 nm (Hitachi U-2000 Spectrophotometer, Hitachi Instuments, Danbury, CT).

2.2.2 Plasmid DNA Isolation.

Plasmid DNA was extracted according to a modified procedure of Birnboim and Doly (216). Cells containing the plasmids were grown in Luria Bertani (LB) medium (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, pH 7.5) containing either 100μg/ml ampicillin or 50 μg/ml kanamycin. 5 ml of LB medium was used for a small scale preparation to obtain 1-20 μg of plasmid DNA. 50 ml of LB medium

was used for a large scale preparation to obtain 0.5-2 mg of plasmid DNA. Following 16-24 hrs incubation in a 37°C shaker-incubator, the culture was centrifuged either for 30 sec. at maximum speed in a microcentrifuge for the small scale preparation or at 600xg (JS-4.2 rotor; Beckman J6-B centrifuge) for 15 min. for the large scale preparation to collect the cells. The cells were resuspended in an ice cold solution of 25 mM Tris-HCl pH 8, 10 mM EDTA, 50 mM glucose using 100 µl for the small scale preparation or 3 ml for the large scale preparation, and incubated at room temperature for 5 min. Then 200 µl for the small scale preparation, or 6 ml for the large scale preparation of a freshly prepared solution containing 1% SDS and 0.2N NaOH was added, the sample gently mixed and incubated on ice for 5 min following which an ice cold KOAc solution pH 4.8 (60% 5M KOAc, 11.5% glacial acetic acid) was added, either 150 µl for the small scale preparation or 4.5 ml for the large scale preparation, and the sample mixed briefly by vortexing. Following centrifugation in a microcentrifuge for 5 min at maximum speed for the small scale preparation or at 1750xg (JS-4.2 rotor;Beckman J6-B centrifuge) for 10 min for the large scale preparation, the supernatant was transferred to a fresh tube and RNase A added to a final concentration of 50 µg/ml and incubated at 37°C for 30 min. An equal volume of phenol-chloroform-isoamyl alcohol (24:24:1) was added and the sample vortexed. Following centrifugation in a microcentrifuge at maximum speed for 1 min for the small scale preparation or at 1750xg (JS-4.2 rotor;Beckman J6-B centrifuge) for 3 min for the large scale preparation the aqueous phase was transferred to a fresh tube and the DNA precipitated with 2.5V of 100% ethanol at -70°C. The DNA was collected by centrifugation in a microcentrifuge at maximum speed for the small

scale preparation or at 1750xg (JS-4.2 rotor;Beckman J6-B centrifuge) for the large scale preparation for 20 min. The pellet was rinsed with 70% ethanol, air-dried, and resuspended in H_2O. NaCl was added to a final concentration of 0.8M followed by the addition of 1V 13% polyethylene glycol and the sample incubated on ice for a minimum of 2 hrs. The DNA was recovered by centrifugation in a microcentrifuge at maximum speed for 20 min., the pellet rinsed with 70% ethanol, air-dried and resuspended in H_20.

2.2.3 Bacteriophage Lambda DNA Isolation.

DNA was isolated from bacteriophage lambda either by using a commercially available bacteriophage lambda DNA isolation kit (Qiagen Inc., Chatsworth, CA) as described by the manufacturer, or according to a modified procedure of Sambrook et.al. (214) as described below. Plate lysates of the bacteriophage lambda clones were prepared essentially as described in Sambrook et.al. (214, Preparation of Plate Lysates Stocks : Protocol I, pp 2.65), except that agarose was used in place of agar. Chloroform was added to the plate lysates to a final concentration of 5%, vortexed and the debris removed by centrifugation at 5000xg (JA-20 rotor;Beckman J2-21 centrifuge) for 10 min. at 4°C. RNaseA (200µg/ml) and DNAse1 (10 µg/ml) were then added and the reaction incubated at 37°C for 2 hrs. Polyethelene glycol (PEG) 8000 and NaCl were then added, in solid form, to a final concentration of 10% and 1.25M respectively, and the sample incubated on ice for 12-20 hrs. The bacteriophage lambda particles were recovered from the solution by centrifugation at 12000xg (JA-20 rotor;Beckman J2-21 centrifuge) for 15 min., and resuspended in lambda diluent

(100mM NaCl, 8mM $MgCl_2$, 50mM Tris-Cl pH 7.5, 0.01% gelatin). EDTA, SDS and proteinase K were then added to give final concentrations 12.5mM, 0.5% and 200μg/ml respectively. Following incubation at 65°C for 30 min. the sample was extracted three times with phenol, then twice with chloroform and the DNA precipitated with 2.5V 100% ethanol. The DNA was removed with a glass rod, washed with 70% ethanol, resuspended in H_20 and the concentration determined by measuring the absorbance at 260 nm (Hitachi U-2000 Spectrophotometer).

2.2.4 Total RNA Extraction.

Total cellular RNA was isolated from cell lines and tissues by a modification of the acid guanidinium thiocyanate-phenol-chloroform extraction method as initially described by Chomczynski and Sacchi (217). Briefly, fresh or frozen cells or tissue was homogenized in a solution of 4M guanidinium thiocyanate, 25 mM sodium citrate pH 7, 0.5% sarkosyl, 0.1M mercaptoethanol. Sodium acetate, pH 4, was then added to a final concentration of 0.2M followed by the addition of 1V of water-saturated phenol and 0.2V of chloroform. The homogenate was mixed after each solution was added. The mixture was shaken vigorously for 1 min, placed on ice for 15 min, and then centrifuged at 18,000xg (JA-20 rotor, Beckman J2-21 centrifuge) for 20 min. The upper aqueous phase was then transferred to a fresh tube and the RNA precipitated by the addition of 1V of isopropanol for a minimum of 1 hr at -20°C. The RNA was recovered by centrifugation at 18,000xg (JA-20 rotor, Beckman J2-21 centrifuge) for 30 min. The resulting pellet was then resuspended in 1 ml of 4M LiCl to solubilize the polysaccharides, and transferred to a microcentrifuge tube.

Following centrifugation in a microcentrifuge at maximum speed for 20 min., the resulting pellet was resuspended in 10 mM Tris-HCl pH 7.5, 1mM EDTA, 0.5% SDS. An equal volume of chloroform was added and the mixture vortexed. The aqueous phase was recovered by centrifugation in a microcentrifuge at maximum speed for 15 min and the RNA precipitated at -20°C for a minimum of 1 hr by the addition of 0.1V of 2M NaAcetate and 1V of isopropanol. The RNA was recovered by centrifugation in a microcentrifuge at maximum speed for 30 min, rinsed in 70% ethanol and resuspended in DEPC-treated H_20. The concentration of the RNA was determined by measuring the absorbance at 260 nm (Hitachi U-2000 Spectrophotometer).

2.3 Polymerase Chain Reactions.

2.3.1 Polymerase Chain Reaction to Generate a Granulin Specific Probe.

Degenerate oligonucleotide primers (GrnAfp, 5'-CGATGTGAAGTG(T/C)GA(T/C)AT GGA-3'; GrnArp, 5'-CTGCATGTGGTT(T/C)TC(A/G)CA(G/A)CA-3') were designed based on the amino acid sequence of grnA. Genomic DNA (2μg) from human placenta was amplified with 1μM primers in a 100 μl reaction containing 50mM KCl, 10mM Tris.Cl pH 8.4, 1.5mM $MgCl_2$, 0.01% gelatin with 200 μM each of dATP, dCTP, dGTP, TTP and 2.5 units of *Thermus aquaticus* (*Taq*) DNA polymerase (218). The reaction was subjected to 30 cycles of amplification in a thermal cycler (Perkin Elmer-Cetus Instruments, Norwalk, CT). Each cycle consisted of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, and elongation at 72°C for 1 min. Amplified DNA

fragments were purified from an agarose gel with reagents supplied in kit form (Mermaid, BIO 101 Inc., La Jolla, CA), blunt ended with T4 DNA polymerase and subcloned into the plasmid vector Bluescript KSII+ (Stratagene, La Jolla, CA) by blunt-end ligation. Nucleotide sequence was obtained by double stranded dideoxynucleotide sequencing (section 2.7).

2.3.2 Asymmetric Polymerase Chain Reaction to Generate Single-Strand Templates for Dideoxy Nucleotide Sequencing.

DNA from HBM3 and HBM4 was amplified with 50 pmoles each of a lambda gt11 specific forward primer, lgt11f (5' GACTCCTGGAGCCCG 3') and a grn specific reverse primer, C12S3IPR (5' GGGCACCGCGTTCCAGC 3') corresponding to nucleotides 46 to 62 of the human grn precursor cDNA, in a 100 µl reaction mixture with 0.5U *Taq* DNA polymerase (section 2.3.1). 25 amplification cycles were performed, consisting of denaturation at 94°C for 30 sec., annealing at 55°C for 30 sec., and elongation at 72°C for 1 min. Following the PCR, the reactions were concentrated and excess primers removed using Centricon-30™ microconcentrator filters (Amicon Divison, Beverly, MA) as described by the manufacturer. 6 µl of the concentrated reaction was then used as the template for a second PCR with 50 pmoles of either lgt11f or C12S3IPR as the primer in a 100 µl reaction mixture (section 2.3.1). The PCR consisted of 20 cycles of denaturation at 94°C for 30 sec., annealing at 58°C for 30 sec., and elongation at 72°C for 1 min. The completed reactions were then passed through Centricon-30™ microconcentrator filters to remove the excess primers. 7 µl of the concentrated reaction product (approximately

25 µl) was then used as template for dideoxy sequencing (section 2.7) with lgt11f or C12S3IPR as the sequencing primers.

2.3.3 Polymerase Chain Reaction to Obtain 5'-Untranslated Region of the Human Granulin Precursor cDNA.

An aliquot (1µl) of a human bone marrow cDNA library in lambda gt11 (Clontech, Palo Alto, CA) was diluted to 10 µl with H_2O, freeze/thawed 4X using dry ice and a 37°C water bath, and subjected to a PCR with 0.5 µM of a lambda gt11 specific primer, lgt11r (5'GGTAGCGACCGGCGC 3') in combination with 0.5 µM of a grn specific reverse primer, C9S3IPR (5'GTGATGGCCATCCCCGC 3') corresponding to nucleotides 275 to 291 of the human grn cDNA. 30 cycles of amplification were performed consisting of 1 min at 94°C, 1 min at 55°C, 1 min 30 sec at 72°C. 1µl of this reaction was reamplified using 0.5 µM of primer lgt11r and 0.5 µM of a second grn specific reverse primer, C12S3IPR (5'GGGCACCGCGTTCCAGC 3') corresponding to nucleotides 46 to 62 of the human grn cDNA. 30 cycles of amplification were performed (section 2.3.1). Each cycle consisted of 1 min at 94°C, 1 min at 58°C, and 1 min at 72°C. Reaction products were ligated to the pCR-1000 vector (TA cloning kit, Invitrogen, San Diego, CA) according to conditions suggested by the manufacturer. Bacterial colonies containing recombinant plasmids were then identified by colony lifts (section 2.6.4), with a ^{32}P end-labeled 35-mer oligonucleotide probe, 5'CCCTGCTGTTAAGGCCACCCAGCTCACCAGGGTCC 3' corresponding to nucleotides 5 to 39 of the human grn precursor cDNA. Plasmid DNA was obtained

from positively hybridizing colonies (section 2.2.2) and the insert sequenced by double stranded dideoxy sequencing (section 2.7).

2.4 Library Screening by Hybridization.

2.4.1 Human Bone Marrow cDNA Library.

A human bone marrow library in lambda gt11 (Clontech, Palo Alto, CA) consisting of 1.51×10^6 independent clones was screened with a 200 base-pair (bp) PCR product (see section 2.3.1), corresponding to part of the grnA gene, labeled with ^{32}P by nick-translation. Duplicate nitrocellulose filters (Schleicher & Schuell, Keene, NH) were prehybridized in 5X SSC (1X SSC is 0.15M sodium chloride, 0.015M sodium citrate,pH 7.0), 5X Denhardt's reagent (1X Denhardt's reagent is 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% bovine serum albumin), 0.2% SDS at 37°C for 5 hrs. Hybridization was in 5X SSC, 2.5X Denhardt's, 0.2% SDS, 50% formamide, 10% PEG-8000 at 37°C for 12 hours. Filters were then washed twice, 45 min each, in 2X SSC, 0.1% SDS at 58°C, and exposed at -70°C with Kodak X-OMAT film (Eastman Kodak Co., Rochester, NY) with an intensifying screen. The DNA from positively hybridizing clones were digested with the restriction enzymes *Kpn*I and *Sac*I. The resulting restriction fragments were then subcloned into the *Kpn*I and *Sac*I cloning sites of Bluescript KSII by cohesive-end ligation. Nucleotide sequence of the cDNA inserts were obtained by double stranded dideoxy sequencing (section 2.7).

2.4.2 Rat Kidney cDNA Library.

A rat kidney cDNA library in lambda gt11 (Clontech) was probed with

*Eco*R1/*Sac*I 1890-bp and *Eco*RI/*Pst*I 310 bp restriction fragments corresponding to nucleotides 6 to 1896 and 6 to 316 of the human progrn cDNA respectively. Nitrocellulose filters (Schleicher & Schuell) were prehybridized in 5X SSC, 5X Denhardt's solution, 0.2% SDS at 57°C for 6 hr. Hybridization was in 5X SSC, 5X Denhardt's solution, 0.2% SDS at 57°C for 16 hr with probe labelled with ^{32}P by nick translation. Filters were washed twice for 60 min in 2X SSC, 0.1% SDS at 60°C, and exposed at -70°C with Kodak X-Omat film with an intensifying screen. Several clones were identified and purified by two additional rounds of screening at progressively lower plaque densities. The inserts from positively hybridizing clones were removed by digestion with the restriction enzyme *Eco*RI, and the resulting fragments subcloned into the *Eco*RI cloning site of Bluescript KSII- by cohesive-end ligation. The cDNA fragments were sequenced by double stranded dideoxy sequencing (section 2.7).

2.4.3 Human Genomic Library.

A human genomic library in EMBL-3 SP6/T7 (Clonetech) was probed with: i) a 200-bp PCR-derived genomic fragment (section 2.3.1) corresponding to part of the grnA gene (see Chapter 3, Figure 3.3), ii) a 394 bp *Kpn*I/*Eco*RI restriction fragment corresponding to nucleotides 1706 to 2099 of the 3'-end of the human grn precursor cDNA, and iii) a 108 bp fragment corresponding to the 5' untranslated region of the human grn precursor cDNA obtained by the PCR as described in section 2.3.3. Conditions of hybridization and washes were as described for screening of the human bone marrow cDNA library (section 2.4.1). DNA inserts from positively

hybridizing clones were removed by digestion with the restriction enzymes *Sac*I and *Xho*I, and the resulting restriction fragments were subcloned into the *Sac*I and *Xho*I cloning sites of Bluescript KSII by cohesive-end ligation. Subclones containing the grn gene were identified by Southern hybridization with the same grn specific cDNA fragments used to screen the genomic library. In some cases, controlled digestions with ExonucleaseIII were performed (as described in Current Protocols in Molecular Biology, Section 7.2.1 - Basic Protocol:Using Exonuclease III to Construct Unidirectional Deletions) (215) on the subclones to generate overlapping templates for sequencing. Nucleotide sequence was obtained by double stranded dideoxy sequencing as described in section 2.7., and in some cases by automated nucleotide sequencing (Nucleic Acids Facility, Columbia University, NY).

2.5 Gel Electrophoresis.

2.5.1 Agarose Gel Electrophoresis of DNA.

Electrophoretic separations were carried out in a horizontal electrophoresis apparatus (DNA Sub Cell ™, DNA Wide-Mini Sub Cell ™ and DNA Mini Sub™ Cell; Bio-Rad, Richmond, CA) in 0.7%, 1.0% or 1.5% (w/v) agarose containing ethidium bromide (0.5. μg/ml). Electrophoretic separations were performed at 5-10 V/cm for 1-4 h at ambient temperature. Electrophoresis buffers were either TAE (50 mM Tris base, 30 mM sodium acetate, 10 mM EDTA, pH 8.0) or TBE (90mM Tris base, 90mM boric acid, 2.5mM EDTA, pH 8.0). DNA was visualized in the gel by short wavelength (254 nm) ultraviolet transillumination and photographed using Polaroid(R) film, type 667 (Polaroid Corp. Cambridge, MA).

2.5.2 Polyacrylamide Gel Electrophoresis of Proteins.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) was performed following the Laemmli method (219) with a Mini-Protean Slab-gel™ apparatus (Bio-Rad). The protein samples were resuspended in 62.5 mM Tris-HCl, pH 6.8, 1% SDS, 10% glycerol .00125% bromophenol blue, with 2.5% ß-mercaptoethanol (reducing), or without ß-mercaptoethanol (non-reducing), and heated for 5 min at 100°C to denature the samples. The samples were then applied to 0.75 mm or 1 mm thick polyacrylamide gels comprising a resolving phase of 8-20% acrylamide and a stacking phase of 3 or 4.5% acrylamide. The samples were electrophoresed at 10-20 V/cm with constant volts for 2-4 hrs, fixed for 5-10 min in 20% methanol, 10% acetic acid and then rehydrated in a solution of 20% methanol, 3% glycerol for 30 min to 6 hrs. The gels were then dried under vacuum in a gel dryer at 55°C for 2 to 12 hrs and autoradiographed using X-OMAT X-ray film (Kodak) at room temperature. Prestained protein markers (Amersham Corp., Arlington Heights,IL; Sigma Chemical Co., St.Louis, MO) were run as size standards.

2.6 DNA and RNA Hybridizations.

2.6.1 Southern Hybridization.

Following electrophoresis on an agarose gel (section 2.5.1), the DNA was transferred to nylon membranes (ZetaProbe, BioRad; Biotrans, ICN) either by capillary blotting with 0.4M NaOH or by semi-dry electrophoretic transfer (Trans-blot SD Semi-Dry Transfer Cell, Bio-Rad) with 0.5X TBE or 1X TAE as the solvent. If transfer was by electroblotting, the membranes were placed, following DNA transfer,

on Whattman paper saturated with 0.4N NaOH for 10 min and then briefly rinsed in 2X SSC. The membranes were then baked at 80°C in a vacuum oven for 30 min. The membranes were prehybridized in 0.5M NaH_2PO_4 pH 7.5, 7% SDS, 1mM EDTA for 10 to 30 min, and hybridized at 65°C in 0.5M NaH_2PO_4 pH 7.5, 7% SDS, 1mM EDTA for 12-20 hrs with a ^{32}P labeled cDNA probe. The cDNA fragment was labeled by nick translation with reagents supplied in kit form (Boehringer Mannheim, Indianapolis, IN). The membranes were washed at 65°C for 60 min in 40 mM NaH_2PO_4 pH7.5, 5% SDS, 1mMEDTA, and 60 min in 40mM NaH_2PO_4 pH 7.5, 1% SDS, 1mM EDTA. For oligonucleotide probes (17-25 mer) (end-labeled with ^{32}P using T4 polynucleotide kinase), the membranes were hybridized at 50°C in 5X SSC, 20 mM NaH_2PO_4, 7% SDS, 10X Denhardt's reagent for 12-20 hrs. The membranes were washed twice in 3X SSC, 7.5X Denhardt's reagent, 5% SDS, 25 mM NaH_2PO_4, pH 7 at 50°C for 30 min., and once in 1X SSC, 1% SDS for 30 min at 50°C. Autoradiograms were obtained by exposing the blots to Kodak X-OMAT film with intensifying screens at -70°C.

2.6.2 Northern Blot Analysis.

Total cellular RNA (8 to 30µg) was denatured with deionised glyoxal (1.45M glyoxal, 72.8% DMSO, 14.5mM NaH_2PO_4 pH 6.8, 0.145% SDS) at 50°C for 1 hr, and fractionated by electrophoresis on 1.1% agarose gel in 10 mM NaH_2PO_4, pH 6.8, transferred to nylon membranes (ZetaProbe, Bio-Rad; Biotrans,ICN) either by semi-dry electrophoretic transfer (Trans-Blot SD Semi-Dry Transfer Cell,Bio-Rad) with 0.5X TBE or by capillary blotting with 10 mM NaOH, and the membrane baked at 80°C in

a vacuum oven for 30 min. The membranes were hybridized at 65°C in 0.5M NaH_2PO_4 pH 6.8, 7% SDS, 1mM EDTA for 16-20 hrs with grn specific cDNA fragments labelled with α-^{32}P dCTP or dATP (3000Ci/mmole, ICN) by nick translation. The membranes were washed at 65°C for 60 min each in 40mM NaH_2PO_4 pH 6.8, 5% SDS, 1mM EDTA and in 40mM NaH_2PO_4 pH 6.8, 1% SDS, 1mM EDTA respectively. Autoradiograms were obtained by exposing the blots to Kodak X-OMAT film with intensifying screens at -70°C. In some cases, following autoradiography, the blots were reprobed with a 25-mer oligonucleotide (5'-TGCTTAAATTCAGCGGGTCGCCACG-3') specific for the 18S ribosomal RNA to correct for loading discrepancy. Densitometric scanning was performed using a Beckman Densitometer (Beckman Instruments Inc. Fullerton, CA).

2.6.3 In Situ Hybridization.

Paraformaldehyde-fixed tissue was sectioned (10 μm) on RNase-free poly-L-lysine coated glass slides and dried for 2 hours at 37°C. Following rehydration in PBS, sections were incubated with proteinase K (10 min) and the reaction was stopped by immersing sections in 4% paraformaldehyde for 5 min. The sections were washed three times in PBS, immersed in a 0.1M solution of triethanolamine and 0.25M of acetic anhydride for 10 min, washed in DEPC-treated water, dehydrated in ethanol and air dried. Sections were hybridized with a ^{35}S-labelled cRNA probe ($1x10^6$ cpm/section) for 16 hr at 42°C. Unbound probe was removed by RNase treatment (20 μg/ml) in 2X SSC for 30 min at 42°C. Washes were carried out in graded concentrations of SSC (from 4X to 0.1X) at temperatures from 20 to 55°C.

Following dehydration of sections with increasing concentrations of ethanol containing 0.3M ammonium acetate, sections were dipped in liquid emulsion (Amersham Corp., Arlington Heights, IL). Exposure was for 5-10 days at 4°C. Autoradiograms were developed (D-19 developer,Kodak), fixed, counterstained with haematoxylin, dehydrated and mounted. For negative controls, extra set of sections were either treated with RNaseA prior to hybridization with the grn cRNA probe, or hybridized with a probe having an identical sequence to the coding strand (sense probe).

2.6.4 Bacterial Colony Lifts and Hybridization.

Bacterial colonies were subcultured in grid-form on a LB agar petri-dish containing the appropriate antibiotic. Nitrocellulose discs (Schleicher and Schuell, 0.45 μm, 82mm) were prewetted in LB medium and placed on the bacterial colonies for 1 min. The nitrocellulose filters were peeled from the plates and placed in succession, colony side up, on Whatman 3MM papers saturated with i) 10% SDS for 3 min. ii) 0.5M NaOH, 1.5M NaCl for 5 min. iii) 1.5M NaCl, 0.5M Tris-HCl pH 7.4 for 5 min. and iv) 2X SSC for 5 min. The filters were then allowed to dry at room temperature before being baked for 2 hrs at 80°C in a vacuum oven. Prior to the hybridization, the filters were placed in a solution of 5X SSC, 0.5% SDS, 1mM EDTA pH 8 at 50°C for 30 min. and the bacterial debris gently scrapped from the surface of the filters. The filters were prehybridized in 5X SSC, 5X Denhardt's solution, 0.2% SDS at 65°C for 3-8 hrs. The hybridization conditions were either in 5X SSC, 5X Denhardt's solution, 0.2% SDS with ^{32}P labelled probe at 65°C for 12-20 hrs, or in

50% formamide, 10% PEG-8000, 5X SSC, 5X Denhardt's solution, 0.2% SDS with ^{32}P labelled probe at 55°C for 12-20 hrs. The filters were washed at 68°C for 60 min each in 2X SSC, 0.1% SDS and in 0.5X SSC, 1% SDS respectively and exposed at -70°C with Kodak X-OMAT film with an intensifying screen.

2.7 DNA Sequencing Reactions.

Nucleotide sequences were obtained using the double-stranded dideoxy chain termination DNA sequencing method (220) as modified for using the Sequenase Version 2.0. enzyme and reagents supplied by the manufacturer in kit form (United States Biochemicals, Cleveland, OH). The sequencing reactions were performed as described by the manufacturer. Sequencing primers were either M13, T7, T3, KS, SK or custom-synthesized 17 to 23-mer primers (Sheldon Biotechnology Center, Montreal, Quebec) with either α-^{32}P dATP (3000 Ci/mmole, ICN Biomedicals, Costa Mesa, CA) or α-^{35}S dATP (1326 Ci/mmole, NEN-DuPont Canada Inc., Mississauga, Ont). DNA template for sequencing was purified by PEG precipitation (section 2.2.2). In some cases nucleotide sequence was obtained by dideoxy sequencing of single-stranded templates generated by asymmetric PCR (section 2.3.2). Sequencing reactions were analysed on 5-8% denaturing polyacrylamide gels (31.0 cm X 38.5 cm X 0.4 mm; S2 Sequencing apparatus, GIBCO-BRL, Gaithersburg, MD) at constant power (30-55 W) for 2-12 hrs in 1X TBE buffer. Following electrophoresis, gels were fixed in 10% acetic acid, 10% methanol for 10 min, dried in a gel dryer at 80°C for 2-6 hrs, and autoradiographed at room temperature for ^{35}S sequencing reactions.

For ^{32}P sequencing reactions, the gels were autoradiographed directly without fixing, at -70°C using Kodak X-Omat film with an intensifying screen.

2.8 Chromosomal Mapping.

Grn-specific primers [forward primer,5'GGAAGTATGGCTGCTGC 3',corresponding to nucleotides 681 to 697 of the human grn precursor cDNA; reverse primer 5'GGATCAGGTCACACACA 3' nucleotides 753 to 769 (primer pair A); and foward primer 5'CTGTGTGTGACCTGATC 3' nucleotides 752 to 768; reverse primer 5'CTGGATGTGGTT(T/C)TC(A/G) CA(G/A)CA 3' nucleotides 940 to 959 (primer pair B)] were used in the PCR with DNA from human-hamster somatic cell hybrids (BIOSMAP™, BIOS Corp.,CT) as the template. The PCR consisted of 30 cycles of denaturation at 94°C for 1 min, annealing at 60°C for 1 min, and elongation at 72°C for 1 min. The amplified products were electrophoresed on a 1.5% agarose gel, transferred to nylon membranes by capillary blotting and analysed by Southern hybridization (section 2.6.1) using grn-specific probes (5'TGCTCCGATCACCTCCACTGC 3', nucleotides 715-735, and a 1890-bp *Eco*RI/*Sac*I cDNA restriction fragment, nucleotides 6 to 1896, for PCR products from primer pairs A and B respectively).

2.9 Primer Extension Analysis.

The primers used were a 35-mer (5' CCCTGCTGTTAAGGCCACCCAGCTCACCAGGGTCC 3') and a 17-mer (5' AGCAGCAACCGGGTAGC 3') corresponding to nucleotides 5 to 39, and -35 to -51,

of the human grn precursor respectively. 0.05-0.2 pmole primer, end-labeled with ^{32}P-ATP (7000 Ci/mmole, ICN), was hybridized to 20 µg total RNA in a solution of 40mM PIPES pH 6.4, 1mM EDTA, 0.4M NaCl, 80% formamide. The reaction was heated to 85°C for 10 min, slowly cooled to 50°C over 60 min. and incubated at 37°C for 12-16 hrs. The nucleic acids were precipitated with ethanol, and resuspended in a solution of 50mM Tris.Cl pH 7.6, 60mM KCl, 10 mM $MgCl_2$, 1mM of each dATP, dCTP, dGTP and TTP, 1mM dithiothreitol, 1U/µl placental RNAase inhibitor. The primers were extended with 2U/µl AMV reverse transcriptase at 42°C for 1 hr. The extended products were ethanol precipitated, recovered by centrifugation, resuspended in 10mM Tris.Cl, 1mM EDTA pH 7.6, and analysed on a 8% acrylamide/8M urea sequencing gel (section 2.7).

2.10 Expression Vector Construction.

2.10.1 Human Granulin Precursor Expression Vector.

The cDNA of the human grn precursor was removed from bacteriophage lambda clone HBM4 by restriction digestion with *Eco*RI and cloned into the *Eco*RI cloning site of the mammalian expression vector pcEXV-3 (221). This vector contains the SV40 (simian virus 40) enhancer, replication origin and early promoter, a synthetic polylinker region, and a polyadenylation signal. Positive clones were selected by colony hybridization (section 2.6.4) with a grn specific probe corresponding to nucleotides 6 to 1896 of the human grn precursor cDNA. The orientation of the grn precursor cDNA in pcEXV-3 was verified by restriction mapping

and partial nucleotide sequencing (section 2.7).

2.10.2 Human Granulin-A Expression Vector.

To construct an expression vector for grnA, the cDNA encoding grnA was ligated to the Ly-6C leader sequence and the construct cloned into the eukaryotic expression vector pcEXV-3 (221) as described. Briefly, Ly6C cDNA in Bluescript KSII (a generous gift of Dr. R.G.E.Palfree, R.V.H) was digested with *Sac*I to remove the cDNA coding for the Ly6C precursor leaving behind the first 24 residues of the 26 residue signal peptide (222). Two oligonucleotides were synthesized, GrnAExPf 5'GGAGCTCAGGGAGATGTGAAATGTGACATG 3'; GrnAExPr 5'CGAGCTCGAATTCTCACTGTTCACAGGTACCCTT 3' and used in the PCR with human grn precursor cDNA as template to generate the cDNA encoding grnA flanked by the last 6 nucleotides of the Ly6C leader sequence and a *Sac*I restriction site at the 5' end, and a stop codon followed by *Sac*I and *Eco*RI restriction enzyme sites at the 3' end. This PCR-generated fragment was cloned into the pCR-1000 vector (TA cloning kit, Invitrogen) as described by the manufacturer, removed by digestion with *Sac*I and ligated to the Ly6C leader sequence to generate a Ly6C leader-grnA construct in Bluescript KSII. The fusion gene was then removed from Bluescript KSII by digestion with *Eco*RI and cloned into the *Eco*RI cloning site of pcEXV-3. Positive clones were identified by plaque hybridization (section 2.6.4) with a grn specific cDNA probe and the construct sequenced (section 2.7) to verify orientation and reading frame.

2.10.3 Human Granulin Promoter-CAT Expression Plasmids.

The chimeric grn promoter/CAT plasmids, pGRN-1662, pGRN-638, pGRN-492, and pGRN-342 were constructed by ligation of various grn gene 5'-flanking region restriction fragments into the promoterless CAT-gene containing plasmid pBLCAT3. pBLCAT2, which contains the thymidine kinase (TK) promoter upstream of the CAT-gene, was used as a positive control. The CAT-vectors, pBLCAT2 and pBLCAT3 (223), were generously supplied by Dr.G.Hendy, R.V.H. The 5'-flanking region restriction fragments of the human grn gene were generated by digesting a 1 kb *Sac*I/*Xho*I subclone of HGL1 (see figure X) with *Sst*II and (i) *Pst*I (pGRN-342); (ii) *Sma*I (pGRN-492); (iii) *Xho*I (pGRN-638); and a 8 kb *Sac*I subclone of HGLP26 with *Sst*II and *Xho*I (pGRN-1662). All restriction fragments were purified from low melting point agarose gel by phenol extractions followed by ethanol precipitation. The fragments were cloned into pBLCAT3 by blunt-end ligation into the *Bam*HI cloning site of pBLCAT3, which was filled in with dNTPs using T4 DNA polymerase. Correct orientation of inserts with relation to the CAT gene coding sequence was verified by restriction enzyme analysis using *Pst*I, *Sac*I, and *Xho*I, and DNA sequencing (section 2.7). 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 (215). The 5' end of these grn promoter/CAT constructs were verified by nucleotide sequencing (section 2.7).

2.11 Cell Culture and Transfections.

All cell lines were obtained from American Type Culture Collection (Rockville,

MD) except for Swei cells which were obtained from Dr J.R.Newcomb at Yale University, New Haven, CT. The cells were maintained in the culture medium suggested by the supplier. The type and origin of the cell lines are given in table 4.1.

2.11.1 Human Granulin Precursor and Granulin A-Expression Vectors.

COS-7 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Gibco-BRL) in an atmosphere of 5% CO_2 at 37°C. Prior to transfections, cells were trypsinised and replated in 60mm or 100mm tissue culture dishes and allowed to proliferate until 50-70% confluency (2-4 days). Cells were transfected with 4 μg plasmid by liposome mediated transfection with 10 μl (1mg/ml) Lipofectin™ reagent (Gibco-BRL) for six hours as described by the manufacturer, or with 250 μg/ml DEAE-dextran for 3 hours as described in Molecular Cloning, a Laboratory Manual, section 16.41 (214). 48 hrs after transfection, the cells were washed twice with PBS, and once with serum-free DMEM. The cells were then labeled for 20 hrs in serum-free DMEM (4ml/100-mm plate) supplemented with 13.7 μl/ml ^{35}S-Cysteine (1104 Ci/mmole, ICN). Labeled supernatants were then dialyzed against 0.1N acetic acid (grn precursor) or passed through preconditioned Waters C_{18} SepPak cartridges (grnA) and eluted in 80% acetonitrile in 0.1% trifluoroacetic acid.

2.11.2 Granulin Promoter-CAT Constructs.

Cells were maintained in RPMI-1640 medium (for A431, A549, and Swei cells)

or DMEM (for COS-7 cells) supplemented with 10% FBS and 10 μg/ml gentamycin in an atmosphere of 5% $C0_2$ at 37°C. All DNA used in the transfections was purified by banding on a CsCl gradient essentially as described in Molecular Cloning, a Laboratory Manual, section 1.42 (214). CsCl centrifugations (1 g/ml CsCl and 0.8mg/ml ethidium bromide) were performed for 48 hrs at 48 000 rpm (Beckman L265B and L275B ultracentrifuges, Beckman Instruments Ins, Fullerton,CA) using a Sorval T880 rotor (Sorval Instruments, DuPont, Mississauga, Ont.). The ethidium bromide was removed from the plasmid DNA preparation by repeated extractions with iso-amyl alcohol. The DNA was precipitated with 100% ethanol, recovered by centrifugation, and the DNA pellet resuspended in H_2O. Prior to transfection, adherent cells were trypsinised and replated in six-well (35mm) tissue culture plates and incubated until 50-70% confluency (2-4 days). For cells growing in suspension (Swei cells), 2-3x10^6 cells were used for each transfection. Cells were transfected with 2 μg of chimeric grn promoter/CAT plasmid by the liposome mediated transfection with 10 μl (1mg/ml) Lipofectin™ reagent (Gibco-BRL) for six hrs as described by the manufacturer. Transfection efficiency using the liposome mediated method is consistant (224,225,226) thus eliminating the need to normalize for transfection efficiency. Cell extracts were assayed for CAT activity 48-52 hrs post transfection (section 2.12).

2.12 Chloramphenicol Acetylacyltransferase Assays.

Cells were harvested and resuspended in 100μl 0.25M Tris-Cl, pH 7.8, disrupted by three cycles of freezing and thawing using dry ice and a 37°C water

bath, followed by centrifugation in a microcentrifuge at maximum speed for 30 sec. to remove cell debris. Equal volumes of cell extract were then assayed for CAT activity according to a modified procedure of Gorman et. al. (227). Briefly, the cell extracts were heated for 10 min at 65°C, centrifuged for 1 min at maximum speed in a microcentrifuge, and 70µl of the supernatant added to a solution of 66.5µl 0.25M Tris-Cl pH 7.8, 2µl 14C-Chloramphenicol (60 mCi/mmole, ICN), 10µl 40mM acetyl co-enzyme A, and 1.5µl 0.5M EDTA, pH 8. The samples were incubated for 2 hr at 37°C, and then extracted with 1ml ethyl acetate. 900µl of the organic phase was removed and placed in a fresh tube, and dried under vacuum using a speedvac™ (Savant Instruments Inc., Farmingdale, NY). The resulting pellet was then resuspended in 25 µl ethyl acetate. The unreacted and acetylated forms of the ^{14}C-chloramphenicol were separated by thin layer chromatography (Silica Gel 1B, J.T.Baker Inc., Phillipsburg, NJ) using chloroform:methanol (19:1) as solvent, and, following autoradiography, were cut out from the TLC sheets and the ß-particle emission 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 protein used in each assay. The protein content of each extract was determined by the Bradford Protein Analysis with reagents supplied in kit form (BioRad) as described by the manufacturer.

2.13 High Performance Liquid Chromatography.

The dialysed supernatant from grn precursor cDNA transfected cells (see section 2.11.1.) was lyophilized and fractionated by reversed-phase HPLC using a

Waters C-4 μBondapak column eluted over a 60 min period using a gradient of 0-80% acetonitrile in 0.1% trifluoroacetic acid throughout at an elution rate of 1.5 ml min^{-1}. The eluted proteins were monitored by UV absorbance at 215 and 280 nm. The ß-emission of the eluted fractions were then determined by scintillation counting. For heparin-sepharose chromatography, the sample (fraction 33, C4 column) was applied to a Shodex heparin-affinity HPLC column pre-equilibrated with 0.05M Tris-Cl pH 7.1. The heparin-bound protein(s) were eluted with a gradient of NaCl from 0.01 to 1.0M in 0.05M Tris-Cl pH 7.1, over a 30 min period at a flow rate of 0.6 ml min^{-1}. Fractions containing the labelled grn precursor were monitored by scintillation counting.

CHAPTER 3

ISOLATION AND CHARACTERIZATION OF THE HUMAN AND RAT GRANULIN PRECURSOR cDNA

INTRODUCTION

Granulins are novel candidate growth factors isolated from inflammatory leukocytes and the kidney (1,2). Four members (grns A,B,C and D) of this cysteine-rich family of peptides were isolated and structurally characterized from human neutrophil extracts (1, and see preface). The discovery of a novel family of biologically active peptides raises many questions, the answers to which could lead to an understanding of its biological role(s) in physiological and/or pathological processes. For example, is the protein the product of one or several genes under coordinate or independent control ? What is the tissue distribution and cellular localization of gene expression ? How is the expression of the gene regulated ? Is the biological response(s) mediated through receptor interactions ? What is its mechanism of action? etc. The availability of the cDNA(s) encoding the protein(s) of interest would be an initial step towards addressing these and many other questions. We therefore set out to isolate and sequence the human grn cDNA.

At the outset of these experiments it seemed reasonable that each grn peptide could be encoded for by a distinct gene. We therefore chose to obtain the cDNA for

grnA first. GrnA was chosen because (i) it was the most abundant grn in human leukocyte extracts (1), (ii) it was the only grn for which the complete peptide sequence was available (1), and (iii) it was the only member of the grn family to have a known biological activity (1,2).

There are several approaches towards obtaining the cDNA for a particular protein, of which two are most commonly used. The first method is to screen proteins produced from expression cDNA libraries with antibodies directed against the protein of interest. The second method uses nucleic acid probes to isolate the desired clone(s) from cDNA libraries. Immunological detection was not possible because antibodies against grnA were not available at the start of the project. However, the complete amino acid sequence of grnA was known (1), therefore DNA probes could be designed and used for nucleic acid screening of a cDNA library. To maximize the specificity of our probes we decided to use faithfully copied fragments of the grnA gene generated by PCR amplification rather than use small, highly redundant synthetic oligonucleotide probes of relatively low specificity.

Also presented in this chapter is the characterization of the rat grn precursor cDNA. The rat grn precursor cDNA was isolated and sequenced for two reasons. First, regions in a protein that have been conserved in evolution often provide an indication of residues and/or domains that may be involved in the structure and/or function of the protein. Second, the human is not a good experimental model in which to understand some of the *in vivo* roles played by the grns. Because both we (1), and others (2) have demonstrated the presence of the 6 kDa biologically active peptide grnA/epi1 in rat hematopoietic and epithelial tissue respectively, we chose

to characterize the rat grn precursor cDNA to provide an animal system in which to further study the biology of the grns.

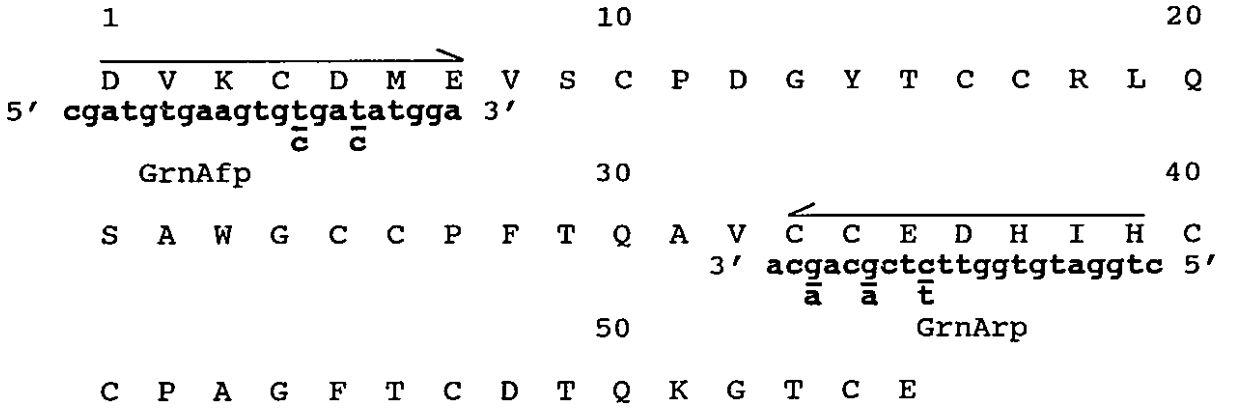
RESULTS

3.1 Isolation and Characterization of the Human Granulin Precursor cDNA.

3.1.1 Generation of a Granulin Specific Probe.

Two degenerate oligonucleotides, GrnAfp (4-fold degeneracy) and GrnArp (8-fold degeneracy), were synthesized based on the probable codon usage (228) and lowest redundancy, and corresponded to the amino terminal and midportion regions of grnA respectively (Figure 3.1). The degenerate oligonucleotides were used as primers in a polymerase chain reaction (PCR) amplification using human genomic DNA isolated from the placenta as the template. In the first PCR performed, the template was denatured at 94°C for 1 min., the primers annealed at 42°C for 1 min., and extended at 72°C for 2 min. This PCR resulted in a series of extension products with no well defined amplification product(s) as observed by agarose gel electrophoretic analysis of the reaction. This result suggested that the primer annealing temperature was too low, resulting in spurious annealing of primer to the template. To increase the specificity of the reaction, the primer annealing temperature was increased to 55°C, while keeping the template denaturation and primer extension temperatures and times unchanged. Agarose gel electrophoretic analysis of the completed reaction revealed five amplification products of

Figure 3.1 Location and sequence of degenerate oligonucieotides corresponding to the amino-terminal (GrnAf, forward primer) and mid-portion (GrnArp, reverse primer) of grnA. The amino-acid sequence of grnA was determined by gas-phase microsequencing (1).

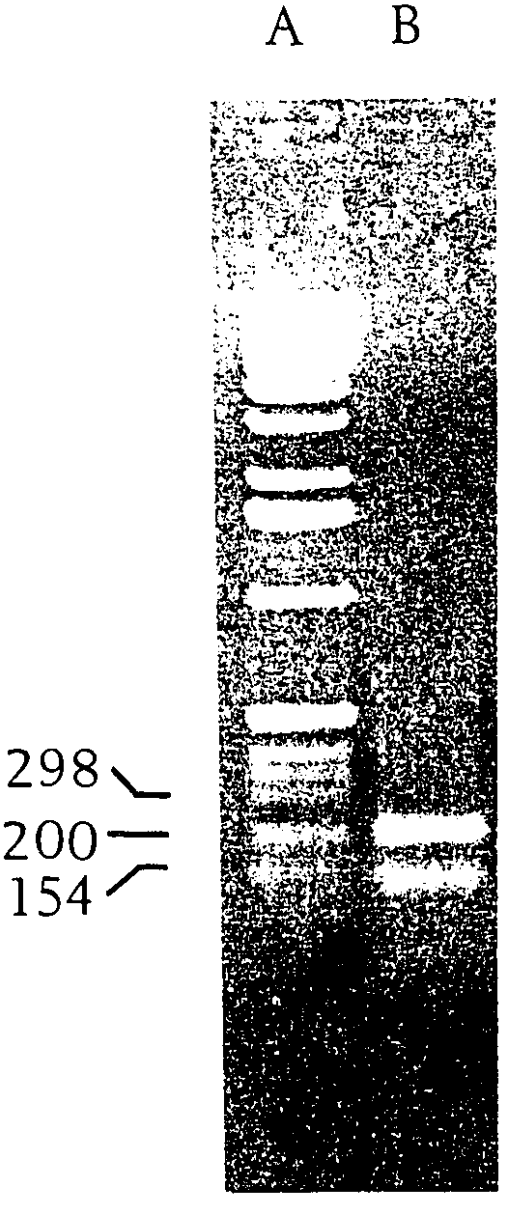

1
10
20
D V K C D M E V S C P D G Y T C C R L Q
5′ cgatgtgaagtgtgatatgga 3′
GrnAfp
30
40
S A W G C C P F T Q A V C C E D H I H C
3′ acgacgctcttggtgtaggtc 5′
50
GrnArp
C P A G F T C D T Q K G T C E

approximately 120, 200, 350, 480 and 600 bp respectively. The annealing temperature was then raised to 60°C to further increase the reaction specificity. This resulted in the elimination of the 480 and 600 bp amplified products respectively. A further increase of the primer annealing temperature to 65°C, resulted in only two amplified products of approximately 120 and 200 bp as determined by agarose gel electrophoresis (Figure 3.2).

Because the primers used in the PCR were based on two regions of grnA that are 40 amino acids apart (Figure 3.1), and, because genomic DNA was used as the template thereby allowing for the possibility of amplifying across an intron(s), any PCR reaction product of 120 bp or greater could represent part of the amplified grnA gene. Additionally, the PCR resulting in only two amplification products of approximately 120 and 200 bp respectively employed stringent reaction conditions (primer annealing at 65°C), strongly suggesting accurate annealing and amplification of the degenerate oligonucleotides. Consequently, both the 120 and 200 bp PCR amplified DNA fragments were purified from an agarose gel and ligated into Bluescript KSII to obtain their nucleotide sequence.

The protein sequence deduced from the DNA sequence of the 120 bp amplified fragment showed that it did not correspond to grnA. A search of nucleic acid and protein databases using the BLAST E-mail server offered by the National Library of Medicine, NIH (229), showed that the 120 bp fragment was unrelated to any other sequence stored in the data banks. This fragment probably results from spurious annealing of the degenerate primers to the genomic DNA. The protein sequence deduced from the DNA sequence of the larger amplified fragment showed

Figure 3.2 Polymerase chain reaction amplification products. Human genomic DNA was subjected to the PCR, under conditions outlined in Materials and Methods, Section 2.3.1, with the degenerate oligonucleotides GrnAfp and GrnArp (Figure 3.1) and the reaction products analyzed on a 1.5% agarose gel. Numbers on left indicate DNA size markers in bp.


A
B
298
200
154

that it represented part of the grnA gene (Figure 3.3). Analysis of this fragment revealed that the first 73 bases following the forward primer correspond to amino acids 8 to 31 of grnA (Figure 3.1) while the last 6 nucleotides preceding the reverse primer corresponds to amino acids 32 and 33 of grnA (Figure 3.1). Between codons 31 (Gln) and 32 (Ala) of grnA is a stretch of 79 nucleotides that, when translated, does not correspond to grnA. Sequence analysis of this 79 nucleotide stretch indicates it probably represents an intron because the sequence at the 5' and 3' ends (5' CAGgta....tagGCT 3', where CAG (Gln) is codon 31 and GCT (Ala) codon 32 respectively) conform to splice site 5'donor/3'acceptor consensus sequences (i.e. (C,A)AGgt(a,g) .....(c,t)agG) (230). The successful amplification and cloning of a DNA fragment representing part of the grnA gene provided us with a suitable probe to screen for the grn cDNA.

3.1.2 Isolation of Human Granulin Precursor cDNA Clones.

The human and rat grns were isolated from circulating leukocytes and bone marrow respectively (1), suggesting that the bone marrow was a possible source of grn mRNA. For this reason, and also because a commercial human bone marrow cDNA library was available, we decided to screen a bone marrow cDNA library for the grn cDNA. Approximately $3x10^5$ clones were screened with the PCR-derived fragment corresponding to part of the grnA gene (Figure 3.3). Sixteen positively hybridizing clones were identified. The human bone marrow library contains $1.51x10^6$ independent clones. This would suggest that the grn mRNA represents approximately 0.05% of the total mRNA in the bone marrow. Of the 16 positively

Figure 3.3 Sequence of the PCR-derived fragment corresponding to part of the grnA gene. Underlined sequences correspond to forward and reverse primers, and the nucleotides in lowercase letters represent an intron in the coding region of grnA.

GATGTGAAGTGTGATATGGAGGTGAGCTGCCCAGATGGCTATACCTGCTGCCGTCTACAGTCGGGG
D V K C D M E V S C P D G Y T C C R L Q S G

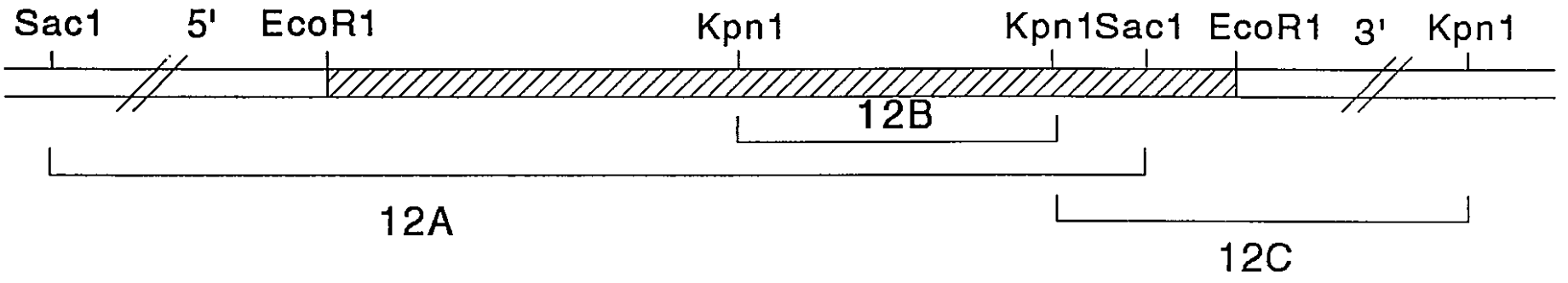
GCCTGGGGCTGCTGCCCTTTTACCCAGgtaccaggtgcggcggtggctgagcacagtgtgcagcag
A W G C C P F T Q

ccggccccagtgcccacctgcccttcttcatctgccctagGCTGTGTGTTGTGAAGACCACATCCA
A V C C E D H I H

hybridizing clones, six, HBM5, HBM9, HBM10, HBM11, HBM12 and HBM16, were plaque purified following three additional rounds of screening at progressively lower plaque densities. The DNA was then isolated from these clones (see Materials and Methods, section 2.2.3 for experimental details). Surprisingly, we were unable to release the cDNA inserts from these clones by digestion with *Eco*RI (the restriction enzyme used in the construction of the human bone marrow cDNA library), even with several different DNA preparations. It is possible the *Eco*RI restrictions sites were destroyed during the construction of the cDNA library, or that specific restriction enzyme inhibitors were present in the DNA preparations. Attempts to amplify the inserts by the PCR with bacteriophage lambda gt11 specific primers flanking the cloning site also proved unsuccessful, which could be due to secondary structures in the cDNA sequence. As such, we could not unequivocally determine which clone contained the longest cDNA insert. One clone, HBM12, was therefore chosen for further analysis.

The cDNA insert from HBM12 was removed by digestion with *Sac*I and *Kpn*I, two restriction enzymes that cleave the lambda gt11 bacteriophage DNA on either side of the cloning site used to construct the lambda gt11 bone marrow cDNA library. The reason why the DNA from clone HBM12 was susceptible to cleavage by the restriction enzymes *Sac*I and *Kpn*I but not *Eco*RI is not clear. Three overlapping restriction fragments containing the cDNA insert of HBM12 were obtained (Figure 3.4); two *Kpn*I fragments of approximately 800 bp (cDNA insert; fragment 12B) and 1.4 kb (400 bp cDNA insert plus 1 kb flanking lambda sequence; fragment 12C) respectively, and one *Sac*I fragment of approximately 3 kb (2 kb cDNA insert plus 1

Figure 3.4 Partial restriction map of clone HBM12. Hatched area represents the cDNA insert of HBM12, and flanking region represents the lambda gt11 DNA. 12 A, B, and C indicate *Sac*I and *Kpn*I restriction fragments that were subcloned into Bluescript KSII to determine the complete nucleotide sequence of the cDNA insert of HBM12.



CLONE HBM12

kb flanking lambda sequence; fragment 12A). The three overlapping restriction fragments were subcloned into Bluescript KSII and restriction digests performed with *Sac*I and *Kpn*I, either separately, together or in combination with *Eco*RI, to provide a partial restriction map of clone HBM12 (Figure 3.4). Now *Eco*RI was able to digest the *Sac*I and *Kpn*I subclones containing both the cDNA insert and lambda flanking sequence (subclones 12C and 12A respectively). This proves that the *Eco*RI cloning site in HMB12 was intact and the inability to release the cDNA insert with *Eco*RI was probably due to enzyme inhibitors in the bacteriophage lambda DNA preparation (231). The three overlapping subclones were then sequenced to obtain the complete nucleotide sequence of the cDNA insert of clone HBM12 (see Materials and Methods, section 2.4.1 and Figure 3.5A for sequencing strategy).

Nucleotide analysis of the 2093 bp cDNA insert of HBM12 showed that it contained the sequence coding for grnA. Using the grnA coding sequence to set the reading frame, the first in-frame ATG codon is located at nucleotides 418 to 420 of the cDNA insert of HBM12. Several things suggested that this was not the ATG start codon. The first 417 nucleotides of HBM12 upstream of this ATG codon maintains an open reading frame with no stop codons. Stop codons are frequently found in all three reading frames upstream of the initiator codon. Additionally, the sequence surrounding this ATG codon, GTTATGG does not show good agreement with the Kozak consensus sequence for eukaryotic initiation, given by ACCATGG (232). Taken together, this would suggest that the ATG codon at nucleotides 418 to 420 does not code for the initiator methionine. Similarly, three other in-frame ATG codons present 3' to nucleotide 420 and upstream of the grnA coding region are not

likely candidates for the initiator methionine. This suggested that clone HBM12 does not contain a full length cDNA insert.

The remaining 15 positive clones were then analyzed to determine if they contained longer cDNA inserts than clone HBM12. This was achieved using a PCR based strategy. The bacteriophage particles were eluted from the positively hybridizing plaques and the bacteriophage DNA released by several cycles of freezing and thawing. The DNA was then subjected to a PCR with a reverse primer complementary to nucleotides 269 to 286 at the 5' end of the cDNA insert of clone HBM12, in combination with either a bacteriophage lambda gt11 specific forward or reverse primer flanking the bone marrow cDNA library cloning site. The amplified products were then analyzed on a 1.5% agarose gel to determine their sizes. While we were unable to amplify the complete cDNA inserts of the grnA positively hybridizing clones using two flanking bacteriophage lambda gt11 specific primers, amplification (under similar conditions) with an internal grn cDNA specific primer in combination with either one of the flanking bacteriophage lambda gt11 specific primers was possible. This may be due to secondary structure(s) within the cDNA insert blocking the *Taq* polymerase extension from the primers, or it may be an idiosyncrasy of the PCR when using the two bacteriophage lambda gt11 specific primers together. Analysis of the products of the PCR revealed that only two clones, HBM3 and HBM4, contained cDNA inserts longer than the insert of HBM12. The 5' sequence of the cDNA inserts of HMB3 and HBM4 were then determined by sequencing single stranded templates generated by asymmetric PCR (233) using a grn specific reverse primer (complementary to nucleotides 46 to 62, Figure 3.5B) and

flanking bacteriophage lambda gt11 specific primers with HBM3 and HBM4 DNA as the template (see Materials and Methods, section 2.3.2 for experimental details).

The 5' sequences obtained from the cDNA inserts of HBM3 and HBM4 overlapped exactly with that of HBM12 and extended the previously determined sequence of the cDNA insert of HBM12 by 104 and 30 nucleotides respectively. Because only the 5'- end of the cDNA inserts of HBM3 and HBM4 were sequenced, the possibility exists that the cDNA inserts of HBM3 and HBM4 may represent alternatively spliced, or different grn transcripts from that of HBM12. This was investigated using a PCR mapping strategy, the premise being that if HBM3, HBM4 and HBM12 are copies from identical transcripts, then amplification of their DNA with the same primer pairs should result in amplification products of equal length for all three clones. Six primer pair combinations using nine different primers complementary to various regions of the cDNA from HBM12 were employed in the PCR with HBM3, HBM4 and HBM12 DNA as the template. The primer pairs used were b (nucleotides 275 to 291 in Figure 3.5) and d (nucleotides 753 to 769) giving an amplification product of 494 bp; e (nucleotides 680 to 696) and k (nucleotides 1643 to 1660) with a 964 bp amplification product; i (nucleotides 1144 to 1160) and k, 516 bp; h (nucleotides 839 to 860) and grnAr (nucleotides 940 to 960), 122 bp; grnBf (nucleotides 752 to 768) and grnAr, 193 bp; and b and grnFr (nucleotides 475 to 491), 217 bp; where grnAr, grnBf, grnFr are primers specific for grnA, grnB and grnF respectively (see section 3.1.4), and primers b,d,e,h,i and k are custom synthesized primers used for sequencing the cDNA insert of clone HBM12 (see legend of Figure 3.5). An identical pattern of amplification products of the expected

size were obtained from clones HBM3, HBM4 and HBM12 respectively when these primer pairs were used in a PCR with the lambda DNA as the template. This result strongly suggests that HBM3, HBM4 and HBM12 are copies from identical grn transcripts.

The complete nucleotide sequence of the human grnA precursor cDNA was therefore determined from the overlapping cDNA inserts of clones HBM12, HBM3 and HBM4. A schematic of the three overlapping clones together with the sequencing strategy used to obtain the complete nucleotide sequence of the human grnA cDNA is shown in Figure 3.5.

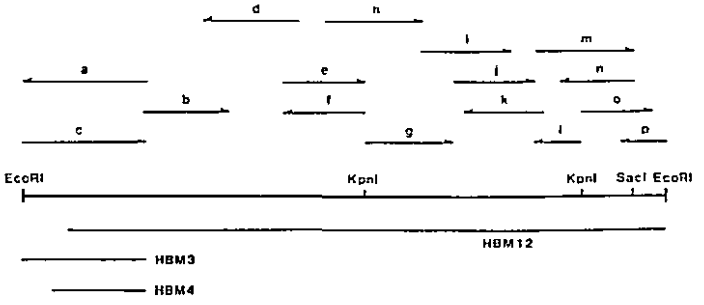
3.1.3 Nucleotide Sequence of the Human Granulin Precursor cDNA.

The complete nucleotide sequence of the human grnA precursor cDNA is shown in Figure 3.5. There is a stop codon at nucleotides 1780 to 1782 predicting an open reading frame of 1779 bp from the initiator methionine, followed by 311 bp of 3' untranslated region. The 5' most ATG is probably the start site because i) the surrounding nucleotide sequence, ACCATGT, shows good agreement with the Kozak consensus sequence for eukaryotic initiation, ACCATGG (232); ii) we have isolated and sequenced a peptide, TRCPDGQFCPVACCLDPGGASYSCCRP, from human granulocytes whose corresponding nucleotide sequence begins 48 bp 3' to the putative initiator methionine, eliminating any AUG codon 3' to this sequence from consideration, and iii) there are in-frame stop codons upstream of the putative initiator methionine.

At the 3' end of the grn precursor cDNA there is a polyadenylation signal

Figure 3.5 (A) Schematic of three overlapping clones and sequencing strategy used to obtain the complete nucleotide sequence of the human grn precursor. Arrows indicate the direction of sequencing. **(B)** Complete nucleotide sequence and deduced polypeptide sequence of the human grn precursor. The nucleotides are numbered from the initiator codon (ATG), and the amino acids are numbered from the probable signal peptide cleavage site. Underlined sequences correspond to sequences previously determined by gas phase microsequencing of purified grns (1). Possible N-glycosylation sites are indicated by an asterix (*), and the stop codon is shown by a #. The boxed sequence correspond to a poly(A) signal. Clone HMB12 starts at nucleotide 6. The sequences of clones HMB3 and HBM4 differ in the 5' untranslated region, as indicated by the bifurcation in the nucleotide sequence.

A



B

```
                5'---GCGTGGATCCTGAGAACTTCAGGCTCCTGGGCAACGTGCTG
GTCTTGTGTGCTGGCCCATCACTTTGCAAAGAATTCACCCCACCAG┐
                                              ├CGCAGGCAGACC
                5'---GCGCGGAGTCGGA┘
ATGTGGACCCTGGTGAGCTGGGTGGCCTTAACAGCAGGGCTGGTGGCTGGAACGCGGTGC    60
 M  W  T  L  V  S  W  V  A  L  T  A  G  L  V  A  G  T  R  C
                                                    1
CCAGATGGTCAGTTCTGCCCTGTGGCCTGCTGCCTGGACCCCGGAGGAGCCAGCTACAGC   120
 P  D  G  Q  F  C  P  V  A  C  C  L  D  P  G  G  A  S  Y  S
          10                  20
TGCTGCCGTCCCCTTCTGGACAAATGGCCCACAACACTGAGCAGGCATCTGGGTGGCCCC   180
 C  C  R  P  L  L  D  K  W  P  T  T  L  S  R  H  L  G  G  P
          30                  40
TGCCAGGTTGATGCCCACTGCTCTGCCGGCCACTCCTGCATCTTTACCGTCTCAGGGACT   240
 C  Q  V  D  A  H  C  S  A  G  H  S  C  I  F  T  V  S  G  T
          50                  60
TCCAGTTGCTGCCCCTTCCCAGAGGCCGTGGCATGCGGGGATGGCCATCACTGCTGCCCA   300
 S  S  C  C  P  F  P  E  A  V  A  C  G  D  G  H  H  C  C  P
          70                  80
CGGGGCTTCCACTGCAGTGCAGACGGGCGATCCTGCTTCCAAAGATCAGGTAACAACTCC   360
 R  G  F  H  C  S  A  D  G  R  S  C  F  Q  R  S  G  N  N  S
          90                 100 *
GTGGGTGCCATCCAGTGCCCTGATAGTCAGTTCGAATGCCCGGACTTCTCCACGTGCTGT   420
 V  G  A  I  Q  C  P  D  S  Q  F  E  C  P  D  F  S  T  C  C
         110                 120
GTTATGGTCGATGGCTCCTGGGGGTGCTGCCCCATGCCCCAGGCTTCCTGCTGTGAAGAC   480
 V  M  V  D  G  S  W  G  C  C  P  M  P  Q  A  S  C  C  E  D
         130                 140
AGGGTGCACTGCTGTCCGCACGGTGCCTTCTGCGACCTGGTTCACACCCGCTGCATCACA   540
 R  V  H  C  C  P  H  G  A  F  C  D  L  V  H  T  R  C  I  T
         150                 160
CCCACGGGCACCCACCCCCTGGCAAAGAAGCTCCCTGCCCAGAGGACTAACAGGGCAGTG   600
 P  T  G  T  H  P  L  A  K  K  L  P  A  Q  R  T  N  R  A  V
         170                 180
GCCTTGTCCAGCTCGGTCATGTGTCCGGACGCACGGTCCCGGTGCCCTGATGGTTCTACC   660
 A  L  S  S  S  V  M  C  P  D  A  R  S  R  C  P  D  G  S  T
                190                 200
```

```
TGCTGTGAGCTGCCCAGTGGGAAGTATGGCTGCTGCCCAATGCCCAACGCCACCTGCTGC   720
 C  C  E  L  P  S  G  K  Y  G  C  C  P  M  P  N  A  T  C  C
                   210                 * 220
TCCGATCACCTGCACTGCTGCCCCCAAGACACTGTGTGTGACCTGATCCAGAGTAAGTGC   780
 S  D  H  L  H  C  C  P  Q  D  T  V  C  D  L  I  Q  S  K  C
                   230                 240
CTCTCCAAGGAGAACGCTACCACGGACCTCCTCACTAAGCTGCCTGCGCACACAGTGGGC   840
 L  S  K  E  N  A  T  T  D  L  L  T  K  L  P  A  H  T  V  G
                *  250                 260
GATGTGAAATGTGACATGGAGGTGAGCTGCCCAGATGGCTATACCTGCTGCCGTCTACAG   900
 D  V  K  C  D  M  E  V  S  C  P  D  G  Y  T  C  C  R  L  Q
                   270                 280
TCGGGGGCCTGGGGCTGCTGCCCTTTTACCCAGGCTGTGTGCTGTGAGGACCACATACAC   960
 S  G  A  W  G  C  C  P  F  T  Q  A  V  C  C  E  D  H  I  H
                   290                 300
TGCTGTCCCGCGGGGTTTACGTGTGACACGCAGAAGGGTACCTGTGAACAGGGGCCCCAC  1020
 C  C  P  A  G  F  T  C  D  T  Q  K  G  T  C  E  Q  G  P  H
                   310                 320
CAGGTGCCCTGGATGGAGAAGGCCCCAGCTCACCTCAGCCTGCCAGACCCACAAGCCTTG  1080
 Q  V  P  W  M  E  K  A  P  A  H  L  S  L  P  D  P  Q  A  L
                   330                 340
AAGAGAGATGTCCCCTGTGATAATGTCAGCAGCTGTCCCTCCTCCGATACCTGCTGCCAA  1140
 K  R  D  V  P  C  D  N  V  S  S  C  P  S  S  D  T  C  C  Q
                   350 *               360
CTCACGTCTGGGGAGTGGGGCTGCTGTCCAATCCCAGAGGCTGTCTGCTGCTCGGACCAC  1200
 L  T  S  G  E  W  G  C  C  P  I  P  E  A  V  C  C  S  D  H
                   370                 380
CAGCACTGCTGCCCCCAGCGATACACGTGTGTAGCTGAGGGGCAGTGTCAGCGAGGAAGC  1260
 Q  H  C  C  P  Q  R  Y  T  C  V  A  E  G  Q  C  Q  R  G  S
                   390                 400
GAGATCGTGGCTGGACTGGAGAAGATGCCTGCCCGCCGCGGTTCCTTATCCCACCCCAGA  1320
 E  I  V  A  G  L  E  K  M  P  A  R  R  G  S  L  S  H  P  R
                   410                 420
GACATCGGCTGTGACCAGCACACCAGCTGCCCGGTGGGCGGAACCTGCTGCCCGAGCCAG  1380
 D  I  G  C  D  Q  H  T  S  C  P  V  G  G  T  C  C  P  S  Q
                   430                 440
GGTGGGAGCTGGGCCTGCTGCCAGTTGCCCCATGCTGTGTGCTGCGAGGATCGCCAGCAC  1440
 G  G  S  W  A  C  C  Q  L  P  H  A  V  C  C  E  D  R  Q  H
                   450                 460
TGCTGCCCGGCTGGCTACACCTGCAACGTGAAGGCTCGATCCTGCGAGAAGGAAGTGGTC  1500
 C  C  P  A  G  Y  T  C  N  V  K  A  R  S  C  E  K  E  V  V
                   470                 480
TCTGCCCAGCCTGCCACCTTCCTGGCCCGTAGCCCTCACGTGGGTGTGAAGGACGTGGAG  1560
 S  A  Q  P  A  T  F  L  A  R  S  P  H  V  G  V  K  D  V  E
                   490                 500
TGTGGGGAAGGACACTTCTGCCATGATAACCAGACCTGCTGCCGAGACAACCGACAGGGC  1620
 C  G  E  G  H  F  C  H  D  N  Q  T  C  C  R  D  N  R  Q  G
                   510                 520
TGGGCCTGCTGTCCCTACGCCCAGGGCGTCTGTTGTGCTGATCGGCGCCACTGCTGTCCT  1680
 W  A  C  C  P  Y  A  Q  G  V  C  C  A  D  R  R  H  C  C  P
                   530                 540
GCTGGCTTCCGCTGCGCACGCAGGGGTACCAAGTGTTTGCGCAGGGAGGCCCCGCGCTGG  1740
 A  G  F  R  C  A  R  R  G  T  K  C  L  R  R  E  A  P  R  W
                   550                 560
GACGCCCCTTTGAGGGACCCAGCCTTGAGACAGCTGCTGTGAGGGACAGTACTGAAGACT  1800
 D  A  P  L  R  D  P  A  L  R  Q  L  L  #
                   570
CTGCAGCCCTCGGGACCCCACTCGGAGGGTGCCCTCTGCTCAGGCCTCCCTAGCACCTCC  1860
CCCTAACCAAATTCTCCCTGGACCCCATTCTGAGCTCCCCATCACCATGGCAGGTGGGGC  1920
CTCAATCTAAGGCCCTTCCCTGTCAGAAGGGGGTTGAGGCAAAAGCCCATTACAAGCTGC  1980
CATCCCCTCCCCGTTTCAGTGGACCCTGTGGCCAGGTGCTTTTCCCTATCCACAGGGGTG  2040
TTTGTGTGTTGGGTGTGCTTTCAATAAAGTTTGTCACTTTCTTAAAAAAAAAAAAAAAA  2099
```

sequence (AAUAAA) together with part of the poly(A) tail. The predicted message, not including a poly(A) tail, is 2062 bp, which is in close agreement with the message length of approximately 2.3 kb including a poly(A) tail, determined by Northern blot analysis (see Chapter 4). A 26 nucleotide polypurine-polypyrimidine tract is found 3 nucleotides upstream of the polyadenylation signal.

The 5' end of the grn cDNA sequence showed heterogeneity in the 5' untranslated region. The cDNA inserts of clones HBM3 and HBM4, whose sequence extended upstream of the translation start site, had identical sequence 12 bp upstream of the initiator AUG, following which the sequence diverged (Figure 3.5). Several possible reasons could account for the anomaly at the 5'-untranslated region of the grn cDNA. The two different transcripts may have arisen from alternate splicing of the grn gene since (i) in later studies we found an intron 7 bp upstream of the initiator methionine and 5 bp downstream from the point of sequence divergence (Figure 3.5 and see Chapter 5, Figure 5.3), and ii) the 3 bp of sequence common to the cDNA inserts of both HBM3 and HBM4, following the intron fit the consensus sequence for a 3' donor exon (230). Alternatively, the heterogeneity could be a cloning artifact resulting from the ligation of a spurious cDNA fragment to the 5'-end of the grn cDNA insert of clone HBM3 by the manufacturer. This possibility is supported by the presence of an *Eco*RI restriction site (5'GAATTC 3', cloning site enzyme) in the 5'-untranslated region of the cDNA insert of clone HBM3 (Figure 3.5).

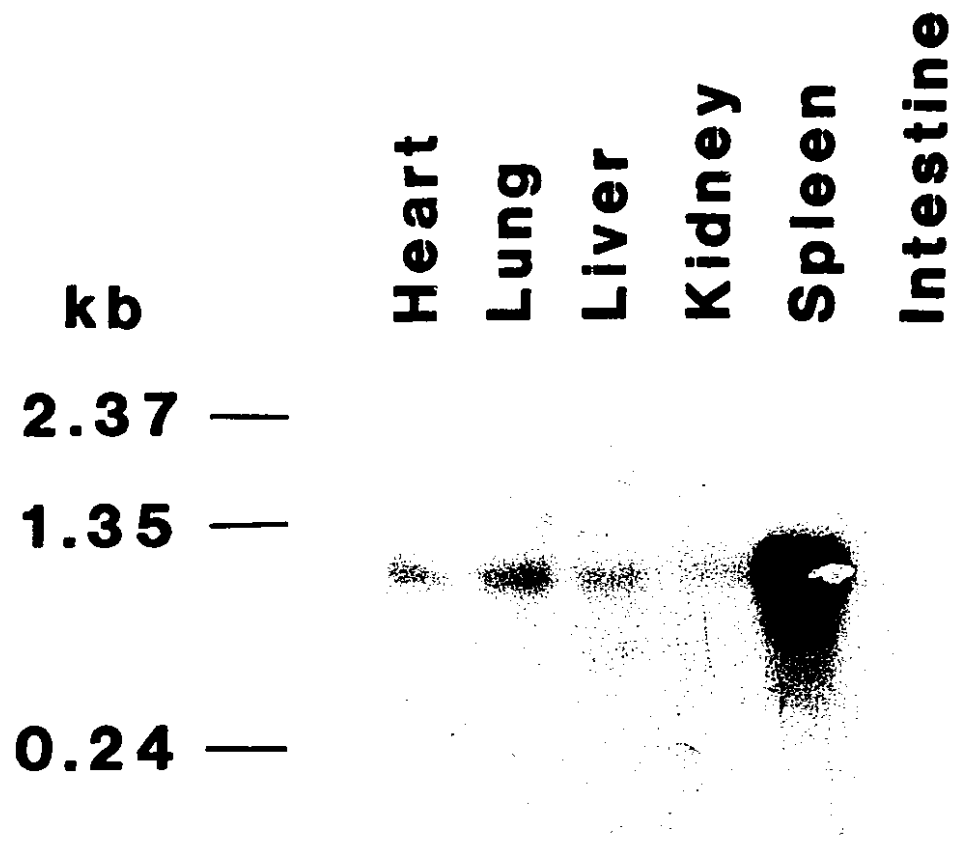
The origin of the heterogeneity in the 5' untranslated region of the grn cDNA was investigated in two ways. First, we cloned several cDNA fragments containing

the 5'-untranslated and coding region of the amino terminal portion of the grn precursor. The cDNA fragments were cloned using the PCR with grn specific reverse primers in combination with a flanking lambda gt11 primer. Human bone marrow cDNA was used as the template (see Materials and Methods, section 2.3.3 for experimental details). This PCR-based approach resulted in our obtaining 46 clones containing the 5'-end of the grn cDNA, which, after their size had been determined by agarose gel electrophoresis and nucleotide sequencing, could be divided into 3 distinct groups based on the length of the 5' untranslated region of the grn cDNA. Clones from all 3 groups were similar to the previously determined nucleotide sequence of the cDNA insert of clone HBM4 and contained 18, 28 and 142 bp of 5'-untranslated region respectively (HBM4 contains 24 bp of 5'-untranslated sequence). None of the cDNA fragments were similar to the 5'-end nucleotide sequence of the cDNA insert of clone HBM3.

In the second approach, an *EcoRI* restriction fragment corresponding to the first 72 nucleotides of the cDNA insert of clone HBM3 (Figure 3.5) was used as the probe in a Northern blot analysis of various human tissues. An mRNA of approximately 1 kb was detected in some tissues (Figure 3.6). This is smaller than the grn mRNA of 2.3 kb, indicating that the 1 kb transcript does not represent the grn mRNA. A subsequent search of nucleic acid and protein databases using the BLAST E-mail server (229) offered by the National Library of Medicine, NIH, showed that the first 45 nucleotides of the 72 bp cDNA fragment are identical to a part of the human and rabbit ß-globin genes.

The results obtained from the PCR cloning and Northern blot analysis together

Figure 3.6 Northern blot analysis of human tissues probed with the 5'-end *Eco*RI fragment of clone HBM3. Numbers on left indicate RNA size markers in kb.


kb
2.37
1.35
0.24
Heart
Lung
Liver
Kidney
Spleen
Intestine

clearly suggest that the sequence at the 5'-end of the cDNA insert of clone HBM3 is an extraneous fragment attached to the 5'-end of the grn precursor cDNA. However, because the 5'-end *Eco*RI restriction fragment of HBM3 does not include the first 10 nucleotides following the divergence in sequence between the cDNA inserts of HBM3 and HBM4 respectively (Figure 3.5), it is possible that these 10 nucleotides are part of a grn transcript arising from alternate splicing of an exon in the 5' untranslated region. In human bone marrow, this transcript must be at least four times less abundant than the grn transcript/cDNA sequence of the HBM4 insert because this sequence was not detected in the four sets of PCR amplification products of the 5'untranslated region.

3.1.4 Deduced Amino Acid Sequence of the Human Granulin Precursor.

The deduced amino acid sequence of human preprogrn predicts a protein of 593 residues with a calculated Mr 63.6 kDa (Figure 3.5). A computer search of protein and DNA databases showed that this sequence was unrelated to any other sequences stored in the databases. The 16 residues following the initiator methionine, WTLVSWVALTAGLVAG, have the characteristics of a signal peptide; i.e. a short, positively charged amino-terminal region immediately following the initiator methionine, a 7- to 15-residues-long central hydrophobic core, followed by a more polar 3- to 7-residues-long carboxy-terminal region (234). Based on the probability weight matrix described by von Heijne (235), the potential site of cleavage between the signal peptide and the mature grn precursor is assigned between Gly_{17} and Thr_{18}. The predicted 593 residue protein contains the 56-amino acid grnA sequence, as

well as six other cysteine rich grn-like repeats, including grnB, grnC, and grnD, previously known from only N-terminal sequences (1). The precursor contains two additional twelve cysteine repeats, E and F, and a degenerate grn repeat, G, with 10 cysteines. The positions of the cysteines are highly conserved in each grn repeat, as are certain other residues, such as Asp_{37}, His_{40}, and Pro_{43}, numbered according to the grnA sequence (Figure 3.7). An eighth repeat at the amino terminus (paragrn, Figures 3.5 and 3.7) contains only six cysteines that align well with the six amino terminal cysteines of a grn repeat.

The sequences between the repeats show little homology but contain a consensus ΣX(K/R)XPA motif (where Σ is a hydrophobic amino acid, and X is any amino-acid) midway between each repeat except D and E, and F and G (Figure 3.8A). Mono- or dibasic sites, which are frequent sites of proteolysis in peptide processing (236) flank grn domains C, D and E, but not A or B, both of which have been isolated as excised peptides (1). Thus nothing can be reliably inferred about the post translational cleavage mechanisms of the human grn precursor. A Kyte-Doolittle hydrophobicity plot made with a 19-residue window, which gives a better discrimination of membrane-spanning regions than the more commonly used 7-residue window "default" (237), revealed no obvious transmembrane sequences (Figure 3.9A). A search for the N-glycosylation consensus sequence, NX(S/T), (where X is any amino acid except proline) (238), revealed 5 potential N-glycosylation sites (Figure 3.5), including one at Asn5 of grnC. GrnC hydrolysates contain amino sugars, and microsequencing (1) gave a blank at residue 5 (n=2), consistent with glycosylation. This confirms that the grn precursor is a glycoprotein.

Figure 3.7 Comparison of the human grn like repeats. Residues occurring four or more times are boxed, and dashes have been introduced to align the cysteines. The repeats are located at amino acids 264-319 (grnA), 169-244 (grnB), 347-400 (grnC), 425-479 (grnD), 501-556 (grnE), 106-162 (grnF), 41-96 (grnG), and 1-27 (paragranulin). The order of the repeats is paragranulin, G,F,B,A,C,D,E. They are not alphabetically aligned because several repeats were isolated and named as discrete peptides before their common origin was known.

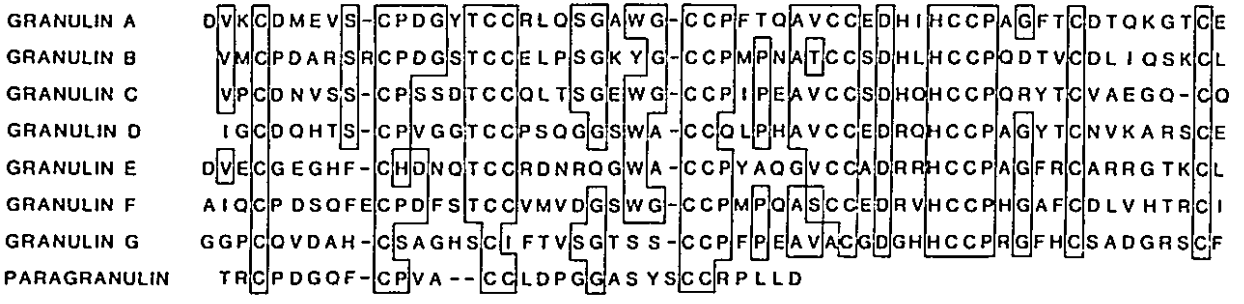
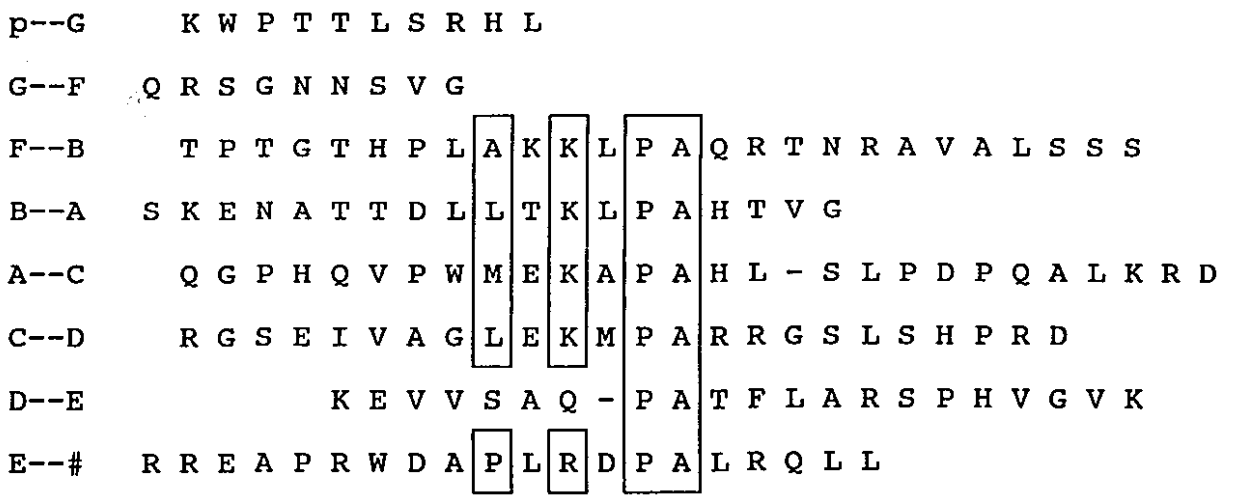

GRANULIN A DVKCDMEVS-CPDGYTCCRLQSGAWG-CCPFTQAVCCEDHIHCCPAGFTCDTQKGTCE
GRANULIN B VMCPDARSRCPDGSTCCELPSGKYG-CCPMPNATCCSDHLHCCPQDTVCDLIQSKCL
GRANULIN C VPCDNVSS-CPSSDTCCQLTSGEWG-CCPIPEAVCCSDHQHCCPQRYTCVAEGQ-CQ
GRANULIN D IGCDQHTS-CPVGGTCCPSQGGSWA-CCQLPHAVCCEDRQHCCPAGYTCNVKARSCE
GRANULIN E DVECGEGHF-CHDNQTCCRDNRQGWA-CCPYAQGVCCADRRHCCPAGFRCARRGTKCL
GRANULIN F AIQCPDSQFECPDFSTCCVMVDGSWG-CCPMPQASCCEDRVHCCPHGAFCDLVHTRCI
GRANULIN G GGPCQVDAH-CSAGHSCIFTVSGTSS-CCPFPEAVACGDGHHCCPRGFHCSADGRSCF
PARAGRANULIN TRCPDGQF-CPVA--CCLDPGGASYSCCRPLLD

Figure 3.8 Comparison of the inter-repeat amino acid sequences of the **(A)** human, and **(B)** rat grn precursor. p,G,F,B,A,C, D,E, and # indicate paragranulin, grnG, grnF, grnB, grnA, grnC, grnD, grnE, and the stop codon respectively. The consensus sequence, ΣX(K/R)XPΘ (where Σ and Θ are hydrophobic and aliphatic amino acids respectively, and X is any amino acid) occurring within some of the inter-repeat sequences is boxed.

A


p--G K W P T T L S R H L
G--F Q R S G N N S V G
F--B T P T G T H P L A K K L P A Q R T N R A V A L S S S
B--A S K E N A T T D L L T K L P A H T V G
A--C Q G P H Q V P W M E K A P A H L - S L P D P Q A L K R D
C--D R G S E I V A G L E K M P A R R G S L S H P R D
D--E K E V V S A Q - P A T F L A R S P H V G V K
E--# R R E A P R W D A P L R D P A L R Q L L

B

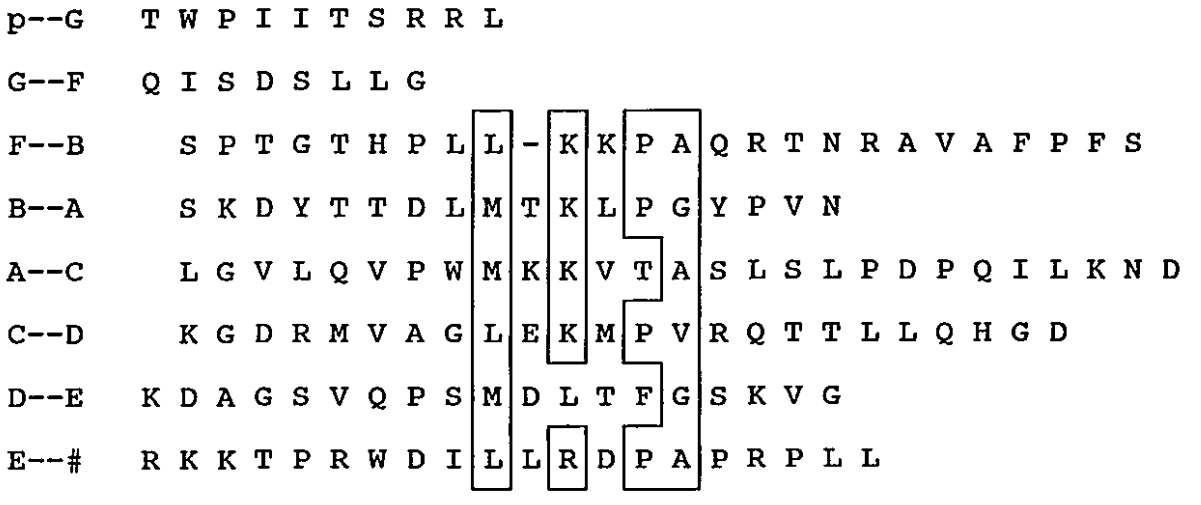
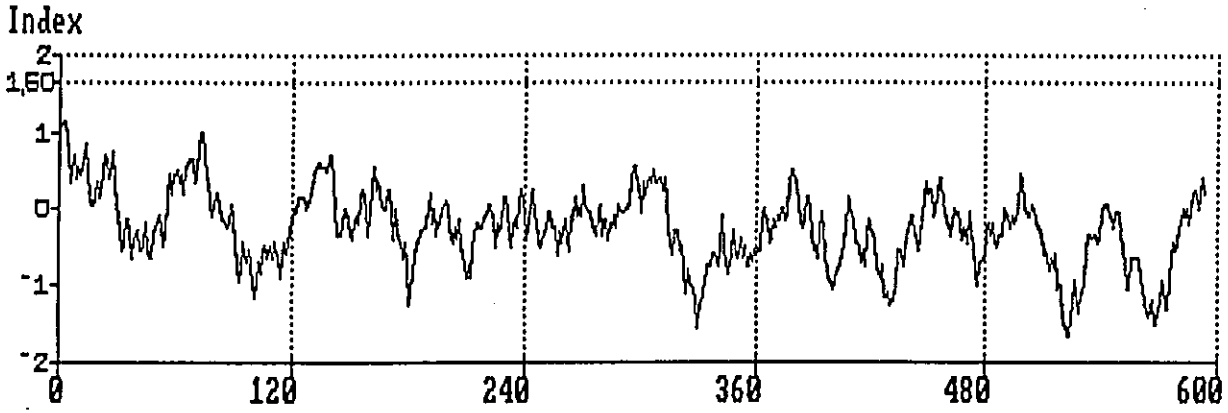
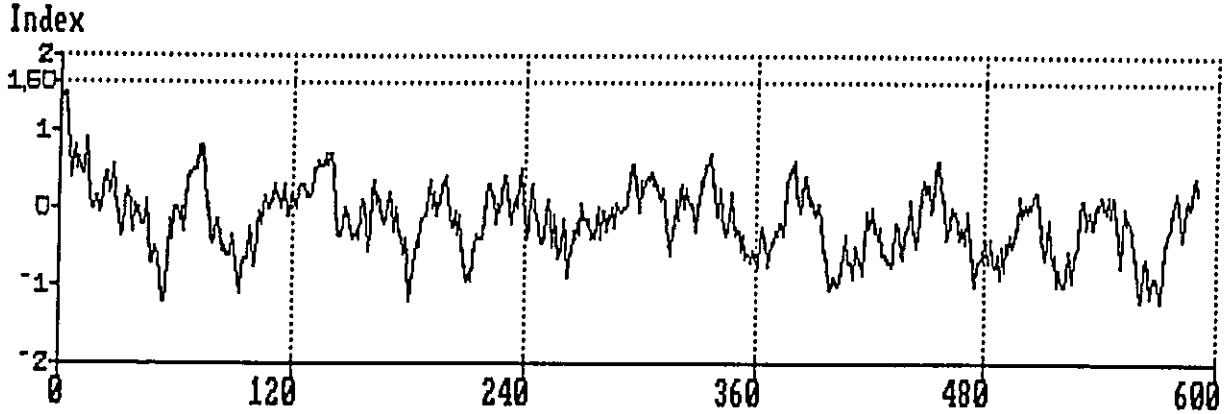

p--G T W P I I T S R R L
G--F Q I S D S L L G
F--B S P T G T H P L L - K K P A Q R T N R A V A F P F S
B--A S K D Y T T D L M T K L P G Y P V N
A--C L G V L Q V P W M K K V T A S L S L P D P Q I L K N D
C--D K G D R M V A G L E K M P V R Q T T L L Q H G D
D--E K D A G S V Q P S M D L T F G S K V G
E--# R K K T P R W D I L L R D P A P R P L L

Figure 3.9 Hydrophobicity analysis of the **(A)** human, and **(B)** rat grn precursor. The Kyte-Doolittle scale (237) with a window size of 19 residues was used for both the human and rat grn precursors. A mean hydrophobicity higher than +1.6 predicts a membrane-spanning segment (237).

A


Index
2
1,60
1
0
-1
-2
0
120
240
360
480
600

B

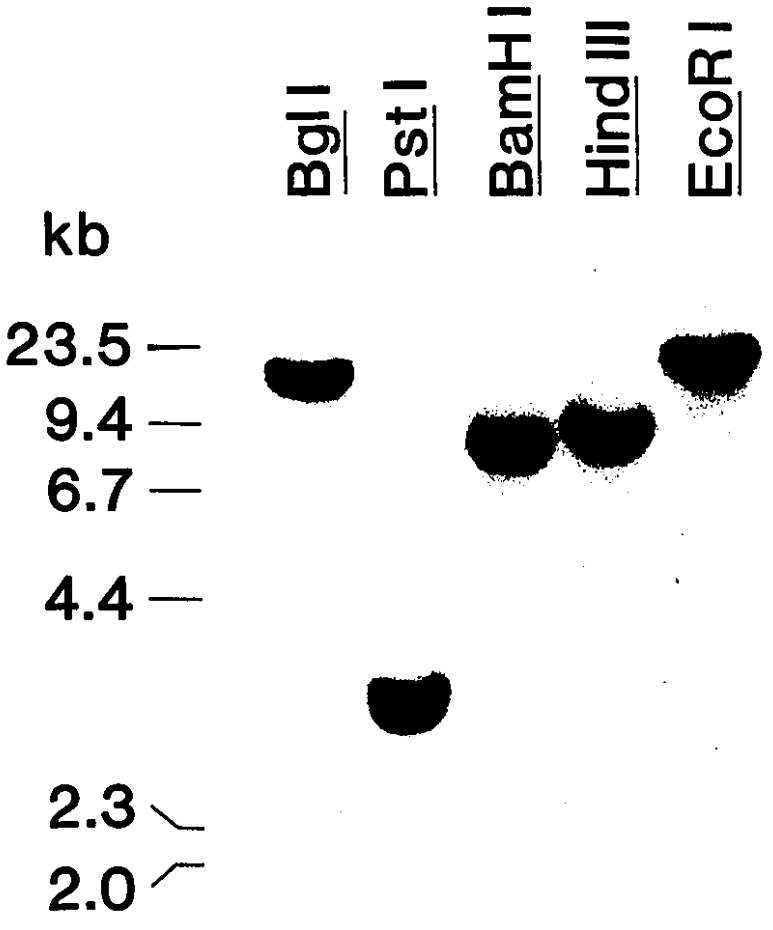

Index
2
1,60
1
0
-1
-2
0
120
240
360
480
600

3.1.5 Southern Hybridizations of the Human Granulin Gene.

The copy number of the human grn gene was determined by Southern blot analysis. A nylon membrane purchased from Clonetech, Palo Alto Ca., containing human DNA digested with *Bgl*II, *Pst*I, *Bam*HI, *Hind*III and *Eco*RI was probed with a 1890 bp cDNA fragment corresponding to nucleotides 6 to 1896 of the human grn precursor cDNA as described in Materials and Methods, section 2.6.1. The result is shown in Figure 3.10. A simple hybridization pattern was obtained implying that the grn gene exists as a single copy in the human genome. Because the Southern hybridization was done under conditions that allow for a 25-35% mismatch between the probe and the target sequence, the simple hybridization pattern obtained does not exclude the possibility of distantly related grn-like genes in the human genome.

To determine if the grn gene is conserved between species, a blot, containing DNA from human, monkey, rat, mouse, dog, bovine, rabbit, chicken and yeast (Clonetech, Palo Alto, Ca.) was probed with a cDNA fragment corresponding to the coding region of the grn precursor (see Materials and Methods, section 2.6.1 for experimental details). A hybridization signal was detected in DNA from human, monkey, rat, mouse, dog, bovine and rabbit, but not chicken or yeast (not shown), demonstrating significant interspecies sequence conservation. This result probably underestimates the species distribution since grn-like peptides have been isolated from teleost hematopoietic tissue (8) and from the brain of the locust (4).

Figure 3.10 Southern blot analysis of human DNA, digested with *Bgl*II, *Pst*I, *Bam*HI, *Hind*III and *Eco*RI, 5 μg per lane, and probed with a ^{32}P nick-translated 1890-bp *Eco*RI/*Sac*I cDNA fragment from clone HBM12 as described in Materials and Methods, section 2.6.1.


Bgl I
Pst I
BamH I
Hind III
EcoR I
kb
23.5
9.4
6.7
4.4
2.3
2.0

3.2 Isolation and Characterization of the Rat Granulin Precursor cDNA.

3.2.1 Isolation of Rat Granulin Precursor cDNA Clones.

Gas phase microsequencing studies showed that human and rat grnA are 86% identical at the protein level (1). This high degree of similarity between the human and rat grnA allowed for homology cloning of the rat grn cDNA. We screened a rat kidney cDNA library for the rat grn cDNA because grn/epi peptides have been isolated from the rat kidney (2). A restriction fragment from the cDNA insert of clone HBM12, corresponding to nucleotides 6 to 1896 of the human grn precursor cDNA was used to screen a rat kidney cDNA library with reduced stringency to allow for cross species hybridization. Hybridization was performed at 57°C in 0.75M Na^+(ionic strength), conditions which allow a 35-45% mismatch between the probe and target sequence (214). Several positively hybridizing clones were obtained and 25 were selected for further analysis. Of the 25 clones, only one clone, RKL6, hybridized following two additional rounds of screening at progressively lower plaque densities. The surprisingly high number of "false-positive" signals obtained in the primary screen is most likely due to insufficient blocking of the membrane during the prehyridization step which may have resulted in high background. The cDNA insert from RKL6 was removed by digestion with *Eco*RI and subcloned into BluescriptKSII to obtain its nucleotide sequence. Unlike the human grn cDNA clones which were not susceptible to restriction cleavage with *Eco*RI, the cDNA insert of RKL6 could be removed with *Eco*RI. This is probably due to the different methods used to isolate the bacteriophage DNA containing the human and rat grn cDNAs respectively. The DNA from clones carrying the human grn cDNA

were isolated from plate lysates prepared with agar, according to a modified procedure of Sambrook et.al.(214), as described in Chapter 2 section 2.2.3, and a commercially available bacteriophage lambda DNA isolation kit (Qiagen) was used to isolate RKL6 DNA from plate lysates that were prepared with agarose. DNA isolated from bacteriophage plate lysates prepared with agar contain enzyme inhibitors (231) which could account for the inability of *Eco*RI to cleave the bacteriophage DNA containing the human grn cDNA. All subsequent bacteriophage lambda DNA were then prepared from plate lysates made with agarose, and were not resistant to restriction enzyme digestion.

The 1645 bp cDNA sequence of RKL6 showed extensive similarity to the human progrn cDNA sequence. The deduced protein sequence from RKL6 when compared with the deduced human grn precursor protein sequence suggested that the cDNA insert of clone RKL6 did not contain the initiator methionine and thus was not a complete cDNA. To obtain the 5'-end of the rat grn cDNA, the rat kidney library was rescreened with an *Eco*RI/*Pst*I restriction fragment from the cDNA insert of clone HBM12 corresponding to nucleotides 6 to 316 of the 5' end of the human grn precursor cDNA under similar hybridization conditions used to obtain RKL6. This screening yielded four positive clones, RKL27, RKL28, RKL42 and RKL46. The size of the cDNA inserts of these clones were then determined by agarose gel electrophoretic analysis of products obtained from a PCR using flanking lambda gt11 specific primers with RKL27, RKL28, RKL42 and RKL46 DNA as the template to obtain approximate sizes of 1700 bp (RKL27), 500 bp (RKL28), 2000 bp (RKL42) and 1900 bp (RKL46) (not shown). The cDNA insert of RKL42, which was the longest of

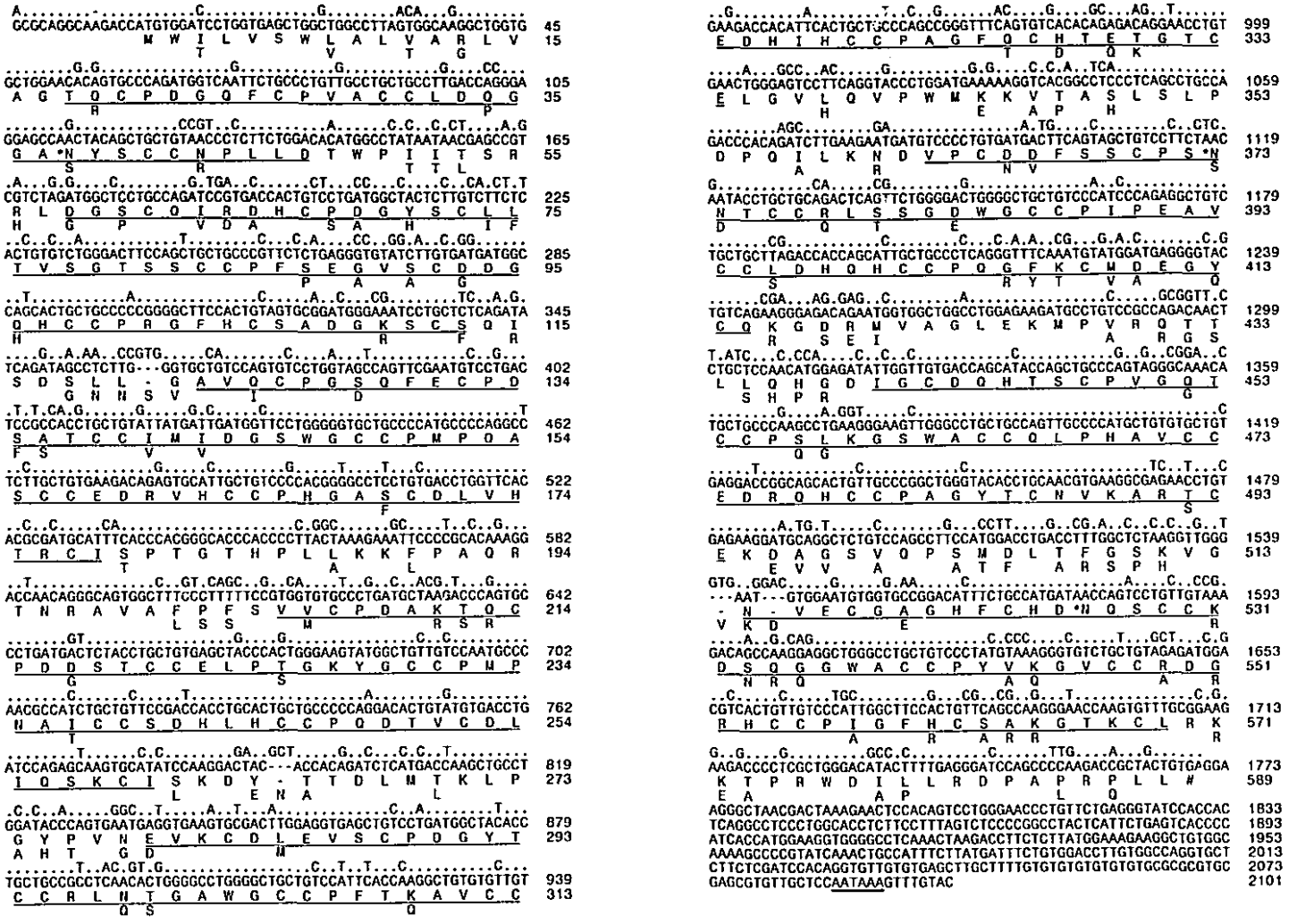
the four cDNA inserts, was removed from RKL42 with *Eco*RI and subcloned into BluescriptKSII to facilitate nucleotide sequencing. Analysis of the sequence showed that it extended the rat grn cDNA sequence derived from clone RKL6 by an additional 342 bases but was still incomplete as determined by comparison with the human progrn precursor cDNA. The remaining 5'-end sequence of the rat progrn cDNA was then obtained from clone RKL28, whose 469 bp cDNA insert contained the probable initiator methionine as determined by comparison with the human progrn cDNA, and overlapped with the first 192 bases of the cDNA insert of clone RKL42. A schematic of the three overlapping clones is shown in Figure 3.11B together with the sequencing strategy used to obtain the complete nucleotide sequence of the rat grn precursor cDNA.

3.2.2 Nucleotide Sequence of the Rat Granulin Precursor cDNA.

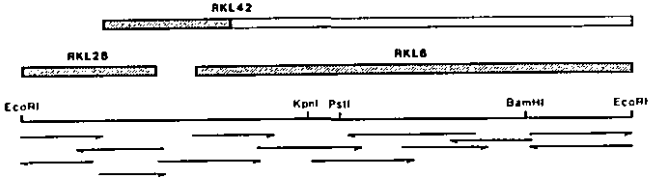
The composite cDNA sequence of the rat grn precursor contains an open reading frame of 1767 bp with 14 bp and 334 bp of 5'- and 3'- untranslated region respectively (Figure 3.11A). The 5' most AUG is most likely to be the true translation start site because i) the sequence around this AUG (ACCATGT) agrees closely to the Kozak consensus sequence for eukaryotic initiation, ACCATGG (232), ii) by comparison to the human preprogrn sequence, and (iii) nucleotide sequence upstream of this predicted initiator methionine in the human sequence contain in-frame stop codons 5' to the putative initiator methionine. Since we obtained only one clone that extended upstream of the translation start site, it is not known if heterogeneity exists at the 5'-end of the rat grn cDNA. The 14 bp of 5' untranslated

Figure 3.11 **(A)** Complete nucleotide sequence and deduced polypeptide sequence of the rat grn precursor. The nucleotides and amino acids are numbered from the initiator codon (ATG) and methionine (M) respectively. Underlined sequences correspond to the grn repeats. Possible N-glycosylation sites are indicated by an asterix (*) and the stop codon is shown by a #. The poly(A) signal is double underlined. Clone RKL6 starts from nucleotide 456. Clone RKL42 extends the sequence of RKL6 by an additional 342 nucleotides, and RKL28, whose 469 bp insert contains the probable initiator methionine and extends the first 192 bases of the sequence of clone RKL42. Homology to the human grn sequence is shown. Differences in sequence is presented above (nucleotide) and below (amino acid) the rat sequence. **(B)** Schematic of overlapping cDNA clones and sequencing strategy used to obtain the complete rat grn sequence. Arrows indicate direction of sequencing. Hatched areas indicate region sequenced. **(C)** Organization of the grn precursor. Closed boxes represent the grn repeats, open boxes represent the intervening spacer regions and the hatched box represents the signal sequence. Potential N-glycosylation sites in the rat and human are shown above and below the schematic respectively. Epi1 and epi2 correspond to grnA and grnB respectively.

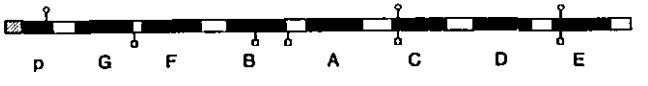
A



B



C



region of the rat cDNA, however, shows a 92% identity to the 5' untranslated region of the human precursor cDNA of clone HBM4 (Figure 3.11A). A polyadenylation signal, AATAAA, is present at the 3'-end but the sequence lacks a poly(A) tail. A polyadenylation signal sequence but no poly(A) tail was also found in the 3' terminal of two other clones. A 26 nucleotide polypyrimidine-polypurine tract is found 13 nucleotides upstream of the polyadenylation signal.

3.2.3 Deduced Amino Acid Sequence of the Rat Granulin Precursor.

The deduced amino acid sequence of rat preprogrn predicts a protein of 589 residues with a calculated size of 63.5 kDa (Figure 3.11A). The rat progrn contains 4 fewer amino acids than the human sequence. The positions of the extra amino acids in the human sequence is shown in Figure 3.11A. Following the initiator methionine are 16 residues, WILVSWLALVARLVAG, predicted to be a leader sequence based on the empirical rules for leader cleavage (234). The structural units of the rat grn precursor are arranged in similar fashion to the human grn precursor (Figure 3.11C). Rat progrn consists of six 12-cysteine grn repeats, including epi 1 and 2, which correspond to grns A and B respectively; a degenerate 10-cysteine domain in the same position as the human grn precursor; and an eighth domain at the extreme amino terminus containing 6 cysteines (Figure 3.12). The grn repeats within the precursor are 36% to 60% identical to each other at the amino-acid level (Figure 3.13). The position of the cysteine residues are highly conserved between each repeat as are several other residues and can be represented as $CX_{4,6}$-C-X_5-(T,S)-C-C-X_4-G-X_3-C-C-X_6-C-C-X-D-X_2-H-C-C-P-X_4-C-X_6-C (Figure 3.12). There

Figure 3.12 Comparison of the rat grn-like repeats. Residues occurring four or more times are boxed. Dashes have been introduced to align the cysteines. Location of the grn repeats within the precursor are indicated by the numbers in brackets.

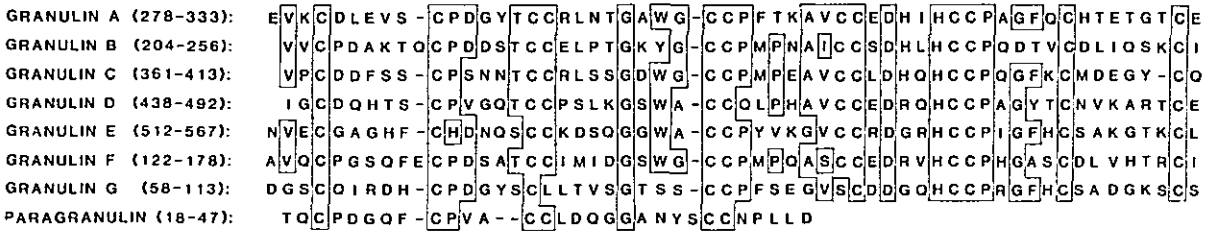

GRANULIN A (278-333):
EVKCDLEVS-CPDGYTCCRLNTGAWG-CCPFTKAVCCEDHIHCCPAGFQCHTETGTCE
GRANULIN B (204-256):
VVCPDAKTQCPDDSTCCELPTGKYG-CCPMPNAICCSDHLHCCPQDTVCDLIQSKCI
GRANULIN C (361-413):
VPCDDFSS-CPSNNTCCRLSSGDWG-CCPMPEAVCCLDHQHCCPQGFKCMDEGY-CQ
GRANULIN D (438-492):
IGCDQHTS-CPVGQTCCPSLKGSWA-CCQLPHAVCCEDRQHCCPAGYTCNVKARTCE
GRANULIN E (512-567):
NVECGAGHF-CHDNQSCCKDSQGGWA-CCPYVKGVCCRDGRHCCPIGFHCSAKGTKCL
GRANULIN F (122-178):
AVQCPGSQFECPDSATCCIMIDGSWG-CCPMPQASCCEDRVHCCPHGASCDLVHTRCI
GRANULIN G (58-113):
DGSCQIRDH-CPDGYSCLLTVSGTSS-CCPFSEGVSCDDGQHCCPRGFHCSADGKSCS
PARAGRANULIN (18-47):
TQCPDGQF-CPVA--CCLDQGGANYSCCNPLLD

Figure 3.13 Comparison of **(A)** rat grn repeats, **(B)** rat and human grn repeats. Numbers indicate percent amino acid identity between the grn repeats.

Rat Granulins							
G	100						
F	38	100					
B	37	52	100				
A	43	46	47	100			
C	43	49	52	60	100		
D	36	46	41	53	48	100	
E	49	42	38	41	45	44	100
	G	F	B	A	C	D	E

Rat Granulins

Rat Granulins							
G	**70**	41	40	46	45	38	41
F	40	**87**	50	50	46	50	45
B	38	53	**85**	50	53	41	37
A	40	47	48	**83**	54	50	50
C	45	49	54	60	**72**	48	46
D	40	51	43	55	52	**92**	48
E	47	41	40	46	40	43	**72**
	G	F	B	A	C	D	E

Human Granulins

is little homology in the inter-repeat sequences (Figure 3.8B). None of the grn repeats are flanked by a single basic or a pair of basic amino acids suggesting that cleavage at basic amino acids is unlikely to account for the processing of the grn precursor. However, like the human grn precursor, the rat precursor contains a loose consensus sequence, ΣX(K/R)XPΘ (where Σ and Θ are hydrophobic and aliphatic amino acids respectively, and X is any amino-acid) midway between some of the repeats (Figure 3.8B). It is possible that this consensus sequence may be important for the processing or folding of the rat and human grn precursor. There are no obvious transmembrane domains beyond the signal peptide (Figure 3.9B). There are 3 potential N-linked glycosylation sites, at Asn_{38}, Asn_{373} and Asn_{526}, suggesting that the rat grn precursor is a glycoprotein. The potential N-glycosylation site at Asn_{526} which is in GrnE, is the only potential glycosylation site between the rat and human grn precursor that is conserved.

3.3 Comparison of the Human and Rat Granulin Precursor cDNA and Protein.

The nucleotide sequence of rat grn precursor cDNA shows a 78.5% identity over the translated region and a 74.5% identity over the untranslated region with the human grn cDNA sequence. The polypurine/polypyrimidine tract in the 3' untranslated region is conserved between the human and rat cDNA. 12 of the 14 nucleotides in the 5' untranslated region of the rat cDNA are conserved in the human cDNA. The sequence between the grn repeats have predominantly replacement mutations (base substitutions that result in a change in amino acid), whereas the grn repeats contain predominantly silent mutations (base substitutions that do not cause

a change in the amino acid assignment). This suggests a selective pressure has operated to conserve the integrity of the grn repeats.

At the protein level, the rat grn precursor shows an identity of 75% with human progrn, with all 88 cysteine residues conserved between the human and rat grn precursor. If allowances are made for conservative amino acid substitutions, the conservation increases to 82%. The rat grn repeats are between 70 to 92% identical to the corresponding grn repeat in the human precursor, with the degenerate 10-cysteine grn G the least and grnD the most conserved repeat respectively (Figure 3.13).

3.4 Expression and Analysis of Recombinant Human Granulin.

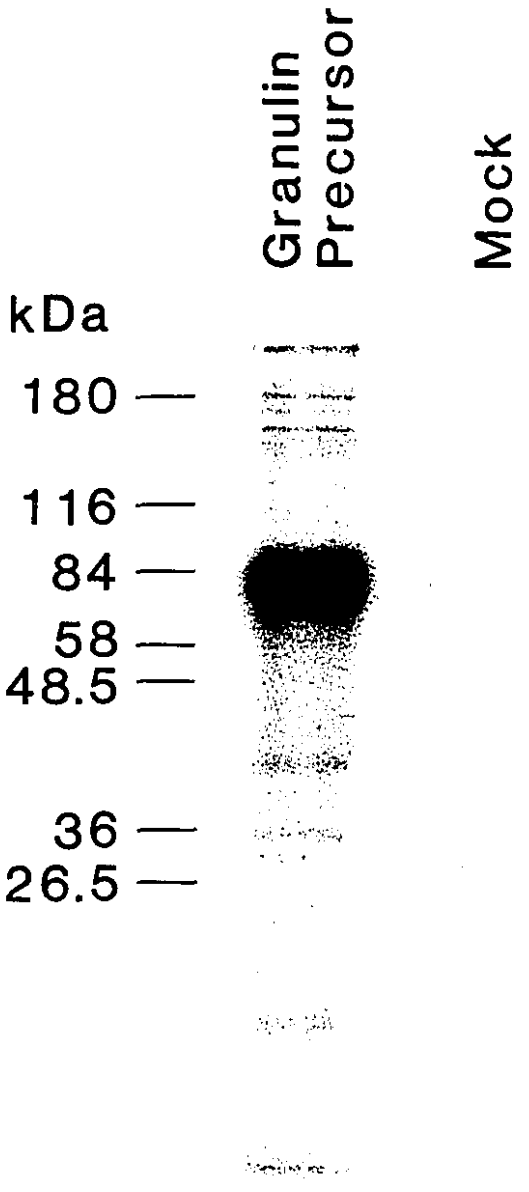
3.4.1 Expression of Recombinant Granulin Precursor.

To demonstrate that the cloned human grn precursor cDNA sequence codes for functional grns, the complete coding sequence of the grn precursor cDNA was cloned into a eukaryotic expression vector under the control of the SV-40 promoter (see Materials and Methods, section 2.10.1 for experimental details). The grn precursor-expression vector was transfected into COS-7 cells as described in Materials and Methods, section 2.11.1. At the same time, a second set of COS-7 cells were transfected with the expression vector without a cDNA insert; these will be referred to as the mock transfected cells. Advantage was taken of the abundance of cysteine residues in the grn precursor protein to metabolically label the proteins in the transfected and mock transfected cells with ^{35}S-Cysteine. The newly synthesized proteins can then be readily detected by autoradiography of an SDS-

PAGE analysis of the transfected cell extracts and supernatant. SDS-PAGE analysis of the supernatant of cells transfected with the expression vector containing the grn precursor cDNA showed an expressed recombinant protein of approximately 84 kDa, which was absent in the supernatant of mock transfected cells (Figure 3.14). No recombinant protein could be detected in the cell extract of grn precursor transfected cells. Additionally there was no difference in the pattern of protein expression between cell extracts of grn precursor transfected and mock transfected cells. This indicates efficient secretion of the recombinant protein by COS-7 cells. The observed 84 kDa recombinant protein is larger than the predicted human grn precursor of 63.6 kDa, most probably due to glycosylation. The observation of a 84 kDa product in the medium of cells transfected with the grn precursor cDNA indicate that COS-7 cells are unable to process the recombinant grn precursor protein into the smaller 6 kDa grn peptides.

To determine if the inability to process the grn precursor into the 6 kDa grn peptides is also true for other cell types, several human cell lines (A431, epidermoid carcinoma; A549, lung carcinoma; HL-60, promyelocytic-leukemia; U937, promonocytic-leukemia; Swei, B-lymphoblatoid), all of which express grn mRNA (see Chapter 4), were transfected with the expression vector containing the full length grn precursor cDNA. The proteins in the grn transfected and mock transfected cells were metabolically labelled with ^{35}S-Cysteine, and the cell extracts and supernatant analyzed for protein expression by SDS-PAGE. Interestingly, analysis of the medium of Swei cells transfected with the grn precursor cDNA revealed two expressed recombinant proteins of approximately 60 and 26 kDa respectively, which were

Figure 3.14 Expression of human recombinant grn precursor in COS-7 cells. COS-7 cells were transfected with pcEXV-3 containing the human grn precursor cDNA (grn precursor) or with pcEXV-3 alone (mock), and the newly synthesized proteins metabolically labelled with ^{35}S-Cys as described in Materials and Methods, sections 2.10.1 and 2.11.1 respectively. The labelled supernatant was dialysed against 0.1N acetic acid and an aliquot analyzed on a 8% polyacrylamide gel as outlined in Materials and Methods, section 2.5.2. The numbers on the left indicate protein size standards in kDa.


Granulin Precursor
Mock
kDa
180
116
84
58
48.5
36
26.5

absent in the supernatant of mock transfected cells (Figure 3.15). Whether the two smaller forms are a result of processing of the grn precursor during transit of the recombinant protein through the secretory pathway, or in the extracellular medium, is currently under investigation by other workers in our laboratory. No recombinant protein(s) could be detected in the cell extract of grn precursor transfected cells. Additionally, no difference in the pattern of protein expression was observed between cell extracts of grn precursor transfected and mock transfected Swei cells. This indicates the efficient secretion of the recombinant proteins by Swei cells. The observation of two recombinant products of 60 and 26 kDa in the medium of Swei cells and a single recombinant protein of 84 kDa in COS-7 cells transfected with the grn precursor cDNA respectively, suggest that Swei cells and COS-7 cells process the grn precursor differently (Figure 3.16).

SDS-PAGE analysis of the supernatant and cell extracts of A431, A549, HL-60, and U937 cells revealed no difference in the pattern of protein expression between grn precursor cDNA transfected and mock transfected cells. The absence of detectable recombinant protein is probably due to weak transcriptional activity of the SV40 promoter in these cells. Experiments are currently in progress by other workers in our laboratory, using several eukaryotic expression vectors under the control of different viral promoters, to study the processing of the grn precursor in various cell types e.g. fibroblasts, epithelial cells, myeloid cells, glial cells etc.

3.4.2 Expression of Recombinant Granulin-A.

Because COS-7 cells cannot process the grn precursor to release the mature

Figure 3.15 Expression of human recombinant grn precursor in Swei cells. pcEXV-3 containing the human grn precursor cDNA (Grn Precursor) or pcEXV-3 alone (mock) was transfected into Swei cells with Lipofectin™ as described in Materials and Methods, section 2.11.1. and the newly synthesized proteins metabolically labelled with ^{35}S-Cys. The labelled supernatant was analyzed by electrophoresis on a 10% polyacrylamide gel as described in Materials and Methods, section 2.5.2. Numbers on the left indicate protein size standards in kDa. The arrow indicates the dye front.

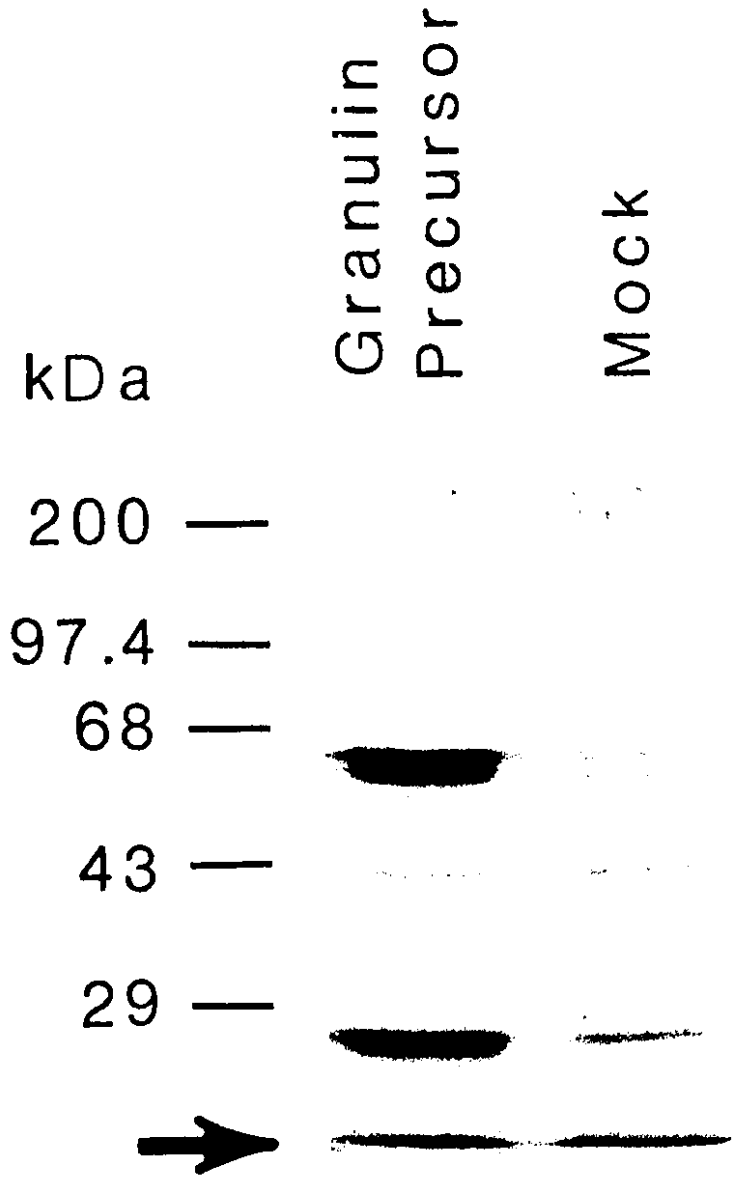
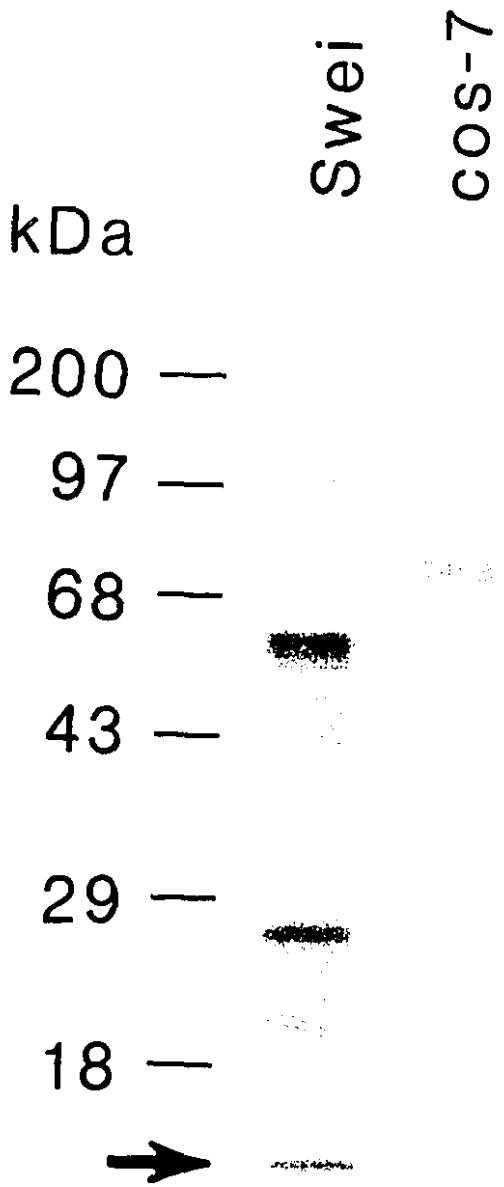

Granulin Precursor
Mock
kDa
200
97.4
68
43
29

Figure 3.16 Recombinant grn precursor is processed differently in COS-7 and Swei Cells. Supernatant from ^{35}S-Cys labelled COS-7 and Swei cells transfected with pcEXV-3 containing the grn precursor (see Materials and Methods, Sections 2.10.1 and 2.11.1) was analyzed on a 10% polyacrylamide gel (see Materials and Methods, section 2.5.2). The numbers on the left indicate protein size standards in kDa. The arrow indicates the dye front.

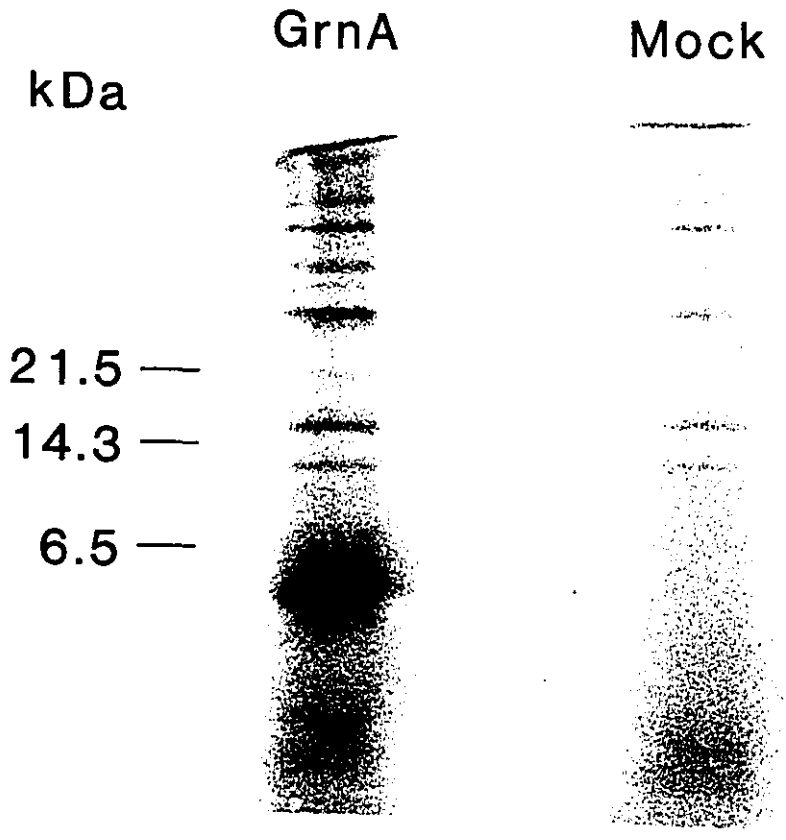

Swei
cos-7
kDa
200
97
68
43
29
18

grn peptides, a eukaryotic expression vector was constructed to allow for the production of recombinant grnA. This was achieved by linking the signal sequence of the Ly-6C peptide to the cDNA coding for the mature grnA. Ly-6C is a secreted GIP-anchored cell surface antigen found predominantly on neutrophils, monocytes and activated T-lymphocytes (222). The "Ly6C signal-grnA" hybrid was then inserted into a eukaryotic expression plasmid under the control of the SV40 promoter (see Materials and Methods, section 2.10.2 for experimental details). COS-7 cells were transfected with the grnA secretory expression plasmid and the medium metabolically labelled with ^{35}S-Cysteine to allow for the detection of expressed proteins. SDS-PAGE analysis of the supernatant from cells transfected with the grnA secretory plasmid showed an expressed product of about 6 kDa which was absent in the medium of mock transfected cells (Figure 3.17). SDS-PAGE analysis of cell extracts showed no visible differences between extracts from grnA and mock transfected cells, indicating efficient secretion of the recombinant protein. The 6 kDa expressed product is in agreement with the 6 kDa size of native grnA.

3.4.3 Recombinant Granulin Precursor but not Granulin-A is a Heparin-Binding Protein.

It is well known that heparin can act as a growth regulator by binding to and modulating the activity of many growth factors. Several examples can be cited, including acidic and basic FGF (95,96), HB-EGF-like growth factor (22), and amphiregulin (55). Recently, it was shown that soluble EGF-precursor is also a heparin binding protein (46). We therefore wanted to determine if the same were

Figure 3.17 Expression of human recombinant grn-A in COS-7 cells. COS-7 cells were transfected with pcEXV-3 containing the cDNA coding for mature grnA fused to the Ly6 leader sequence (grnA) (see Materials and Methods, section 2.10.2.) or with pcEXV-3 alone (mock), and the newly synthesized proteins metabolically labelled with ^{35}S-Cys as described in Materials and Methods, section 2.11.1. The labelled supernatants were passed through preconditioned Waters C_{18} SepPak cartridges and an aliquot analyzed on a 15% polyacrylamide gel as outlined in Materials and Methods, section 2.5.2. The numbers on the left indicate protein size standards in kDa.


GrnA
Mock
kDa
21.5
14.3
6.5

true for the grns.

The ability of the grn precursor to bind heparin affinity columns was determined by affinity chromatography. Partially purified recombinant grn precursor (fraction 33 of rpHPLC C4 column, Figure 3.18; see Materials and Methods, section 2.14 for experimental details) was applied onto a heparin column and eluted with increasing concentrations of NaCl. Because the recombinant grn precursor was radiolabeled with ^{35}S-Cysteine, eluates from the heparin column could be monitored by determining the counts from each fraction. This is shown in Figure 3.19. A peak of radioactivity elutes in fractions 31 to 37, corresponding to 0.42 to 0.54 M NaCl. Analysis of these fractions by SDS-PAGE confirms the presence of the recombinant grn precursor in the eluates. This result indicates that the grn precursor binds to heparin affinity columns, and is eluted from the heparin column with 0.5M NaCl. Using the same experimental approach, partially purified radiolabeled grnA was examined for its ability to bind heparin. Recombinant grnA binds very weakly to the heparin column and can be eluted with 0.25M NaCl. This salt concentration is below the physiological ionic strength, indicating that the very weak affinity of grnA for heparin is probably not physiologically significant.

3.4.4 Activity of Recombinant Granulin Precursor and Granulin-A.

Experiments on the physiochemical and biological properties of the recombinant human grn precursor and grnA are currently in progress and constitute a separate project. Experiments by other researchers in our laboratory indicate that recombinant grnA is biologically active as determined by its ability to inhibit the

Figure 3.18 RP-HPLC of recombinant grn precursor. Recombinant human grn precursor was partially purified by reversed-phase HPLC as described in Materials and Methods, section 2.13. Inset shows SDS-PAGE analysis (see Materials and Methods, section 2.5.2) of fractions 28 to 39. Numbers on right indicate protein size standards in kDa.

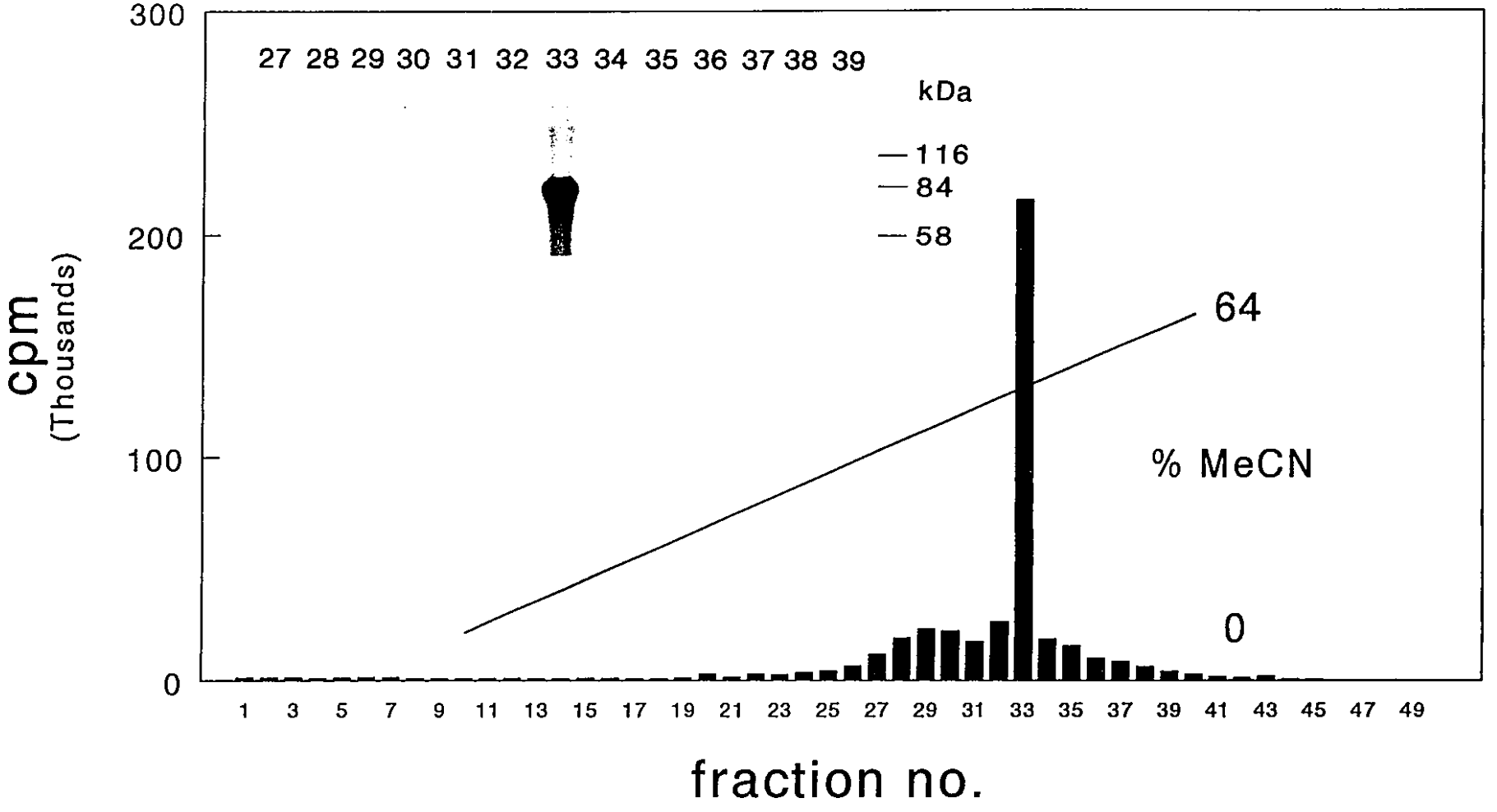
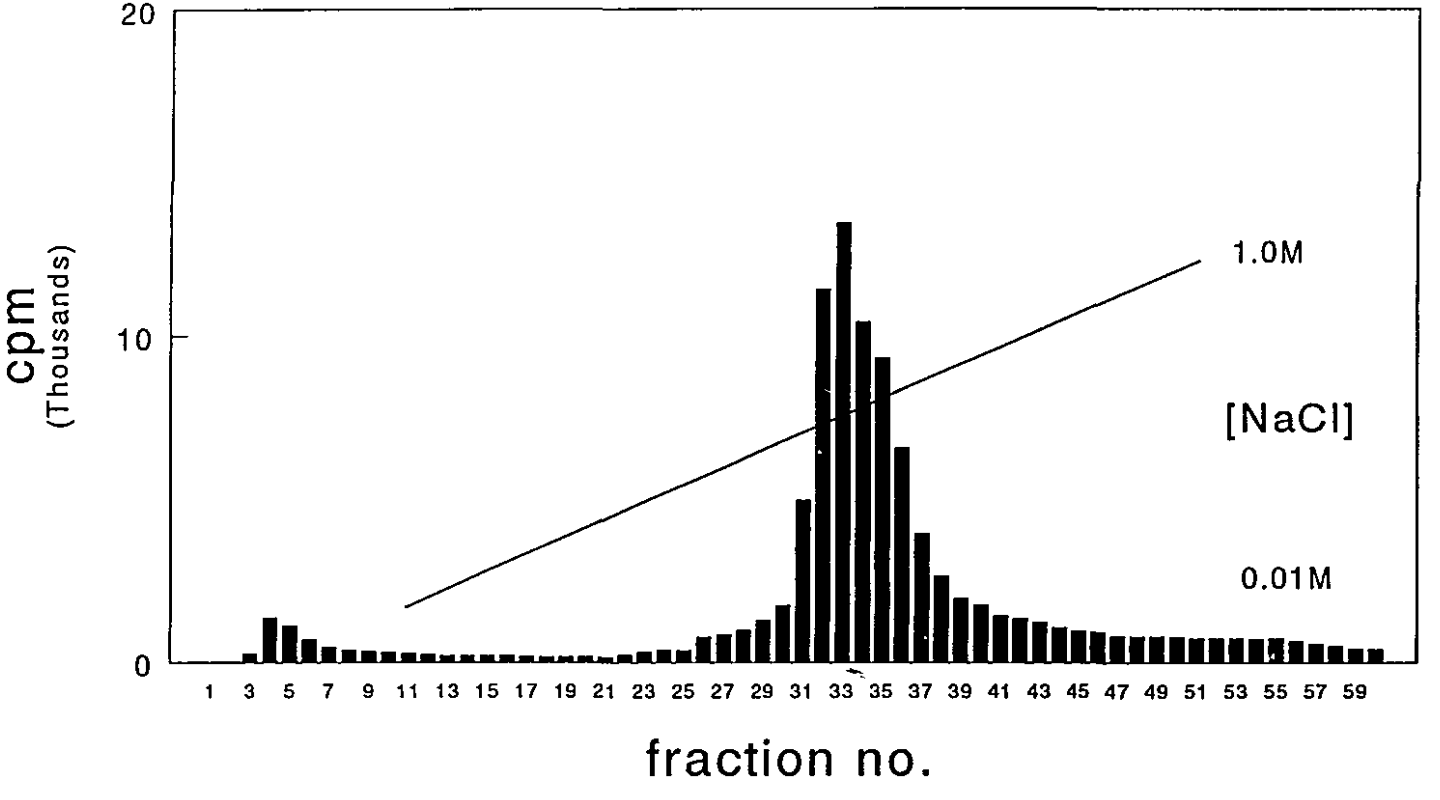

300
200
100
0
cpm
(Thousands)
27 28 29 30 31 32 33 34 35 36 37 38 39
kDa
116
84
58
64
% MeCN
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
fraction no.

Figure 3.19 Recombinant grn precursor is a heparin-binding protein. ^{35}S-Cys labelled recombinant grn precursor was applied to a heparin-sepharose column and eluted with a linear gradient of 0.01 to 1.0M NaCl as described in Materials and Methods, section 2.13. Fractions containing the recombinant grn precursor were determined by scintillation counting.


20
10
0
cpm
(Thousands)
1.0M
[NaCl]
0.01M
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
fraction no.

growth of A431 cells. Preliminary experiments indicate that recombinant grn precursor, but not grnA, is mitogenic for NIH-3T3 cells. The growth promoting effect of the recombinant grn precursor on NIH-3T3 cells is consistent with the findings of Zhou et.al., who recently demonstrated that the mouse grn precursor acts as a growth factor for NIH-3T3 cells (9). The ability to express a functionally active intact grn precursor will allow us to study the processing, secretion and biology of the grn precursor, which would prove invaluable towards understanding the physiological role of the grn gene products. Additionally, the ability to express biologically active recombinant grnA will provide a working system to express each individual grn repeat (e.g. grnB, grnC, grnD etc.), and will allow us to determine if the other grn peptides are biologically active.

DISCUSSION

3.1 Structure of the Human and Rat Granulin Precursor.

As a step towards understanding the function of the grns *in vivo*, we have isolated and sequenced the human and rat grn precursor cDNAs. The human and rat grn precursor consists of seven and one half cysteine-rich grn repeats arranged in tandem (Figure 3.11C). Some of these repeats have been isolated as discrete peptides; paragrn, grnA, grnB, grnC and grnD from the human (1), and epi1 and epi2 from the rat (2). The presence of all four known human grns and three grn-like sequences in a common precursor was unexpected. However, the iterative nature

of the grn precursor is not uncommon among proteins. Multiple sequential repeats of cysteine-rich domains is a recurrent motif among many regulatory proteins, particularly growth factors and adhesion molecules (42,43,239,240). The iterative nature of progrn draws comparisons with the epidermal growth factor (EGF) precursor. The EGF precursor consists of 9 EGF-like repeats arranged in tandem which is proteolytically processed to generate the biologically active EGF (42,43). However, unlike the EGF precursor, progrn has no transmembrane segment or cytoplasmic domain.

The high degree of homology between the human and rat grn precursor shows that the protein is highly conserved through vertebrate evolution. The isolation of grn-like peptides from teleost fish (8) and insect (4) further establish that these peptides have been conserved over larger evolutionary distances. The conservation of the grns over large evolutionary distances implies fundamental biological roles for these peptides. All 88 cysteine residues are conserved between the human and rat grn precursor implying that the cysteine residues are critical for maintaining the structure and/or function of the grns. Shoyab et al. have shown that reduction of epi1 and epi2 results in loss of the ability of epi1 and epi2 to inhibit the growth of A431 cells (2), confirming that some cysteines in disulfide linkage(s) are essential for biological activity. However, although the grns are disulfide bridged (1,2), there is no *a priori* reason to believe that all 88 cysteine residues in the precursor are involved in intramolecular disulfide bridging. Additionally, the presence of "free" cysteines may allow for the grn precursor to form dimer(s), either with another grn precursor molecule (homodimer), or with a different protein

(heterodimer). It will therefore be important to determine the extent of disulfide bridging in the grn precursor.

Computer analysis of the grn precursor predicts a protein with mixed secondary structure. The high content of cysteine-residues suggests that the grn peptides are compact proteins. The arrangement of the cysteine residues within each repeat is conserved and can be represented as $CX_{5-6}CX_5CCX_8CCX_6CCX_5CCX_5CX_{5-6}C$. Glycine residues recurrently occur between cysteine residues in each repeat and probably assist in peptide folding. Proline residues often follow the 2nd, 6th and 10th cysteine in each repeat and may confer stability to the compact structure of the grns. The sequence CPXG (where X is any amino acid) is frequently found in the grn repeats. This motif also occurs in other cysteine-rich proteins and has close to maximum ß-turn potential as predicted by the Chou and Fasman method (241).

Comparison of the repeats reveals that the sequence $CCXDX_2HCCP$ is common to all the 12-cysteine human and rat grn repeats. Interestingly, this sequence is also conserved in the locust grn (PMP-D1) (4), while in the three carp grns, the proline residue is substituted by an arginine (8). The significance of this conserved motif is not known but it is tempting to speculate that the sequence $CCXDX_2HCCP$ (e.g. residues 33 to 42 of human grnA) contributes significantly towards the function or folding of the peptide. A search of nucleic acid and protein databases using the BLAST E-mail server (229) offered by the National Library of Medicine, NIH, revealed no homology of this consensus motif to any sequences stored in the databases. However, a homology was observed between the motif

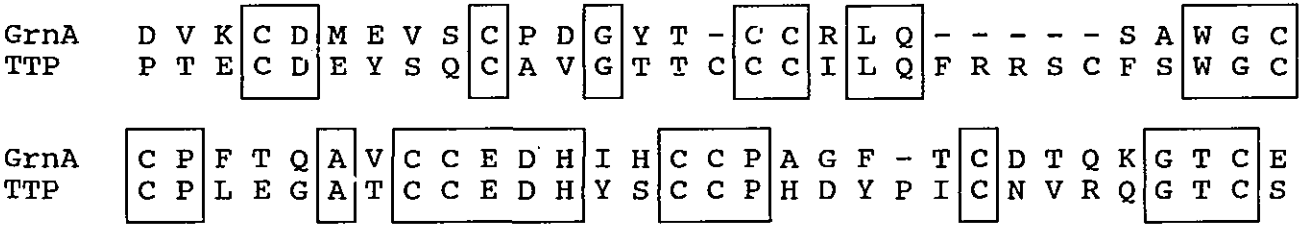
CCX_2HX_2C (e.g. residues 33 to 40 of human grnA) present in some of the grn repeats (grnA, B and C of the human and rat) and a consensus sequence within the active center of phospholipase A_2 (242). The function and/or evolutionary significance of this homology, if any, is not known.

Comparison of the consensus cysteine motif of the grn repeats with available protein and DNA sequence databases revealed an extended homology between the 12-cysteine motif of the grns and the C-terminal region of the deduced amino acid sequence of a tomato thiol protease gene induced by low temperature (38, and Figure 3.19). The significance of this homology is not known. The catalytic domain of the tomato thiol protease lies in the N-terminal region, and, while the purpose of the C-terminal region is not clear, it has been proposed that this domain may serve to regulate the protease activity by binding to metal ions (38). Interestingly, growth factor-like domains are found in other proteases such as urokinase plasminogen activator (uPA) and tissue plasminogen activator (tPA) (32).

Both the human and the rat grn precursors contain potential N-linked glycosylation sites with one potential glycosylation site conserved between the human and rat grn precursor. That the grn precursor is a glycoprotein is confirmed by microsequencing of the grn peptides (1). Additionally, Zhou et.al. have recently demonstrated that the mouse grn precursor is N-glycosylated (9). However it is not known if the grn precursor is O-glycosylated.

Our present understanding of the structural similarities and differences between the human and rat grns are insufficient to recognize all the significant structure-function relationships of these proteins. The determination of the structure

Figure 3.20 Alignment of human grnA with the C-terminal domain (amino-acids 236 to 297) of a tomato (*Lycopersicon esculentum*) thiol protease (TTP). A 47% identity in a 58 amino-acid overlap is observed.


GrnA D V K C D M E V S C P D G Y T - C C R L Q - - - - - S A W G C
TTP P T E C D E Y S Q C A V G T T C C C I L Q F R R S C F S W G C
GrnA C P F T Q A V C C E D H I H C C P A G F - T C D T Q K G T C E
TTP C P L E G A T C C E D H Y S C C P H D Y P I C N V R Q G T C S

of the carp and locust (PMP-D1) grn gene(s) and precursor proteins will prove invaluable towards this end, and perhaps advance our understanding of the *in vivo* functions of the grn gene products.

During the course of this project, two other groups reported the cDNA sequences of the human, rat, murine and guinea pig grn precursor (6,243) . Two rat precursor sequences were presented (243), one of which was identical to the sequence that we obtained. The second was a grn precursor isoform missing amino acids 393 to 470. This deletion fuses the amino and carboxyl termini of grn domains C and D respectively, creating a hybrid grn structure without causing a frame-shift. We did not detect this isoform but the rat C/D hybrid probably arises from the omission of an exon (see Chapter 5). While the mouse grn precursor shows similar conserved cysteine-motif repeats like the human and rat precursors, the guinea pig grn precursor (acrogranin) contains additional stretches of amino acids between the second and third repeats, and between the fourth and fifth repeats, and lacks part of the sequence in the fifth repeat (6). Interestingly, the deletion in the fifth repeat maps to an exon-intron junction in the human gene (see Chapter 5). Similarly, the additional sequence between the second and third repeats lie close to an exon-intron boundary in the human gene. This suggests that some of the length variations may be due to contraction and extension of exons at the intron/exon junction (intron-exon junctional sliding) respectively (244). The occurrence of intron-exon junctional sliding among members of a protein family is not uncommon. Analysis of gene structure and variation of protein sequence within gene families show that the intron-exon junction positions correspond with length variations within members of the protein

family (244).

There is greater intraspecies homology between equivalent grn repeats (e.g. rat A with human A, rat B with human grnB etc.) than between the individual domains of each protein (rat A with rat B,C,D,E,F, and G etc.) suggesting that the repeats are not functionally redundant. Although only four discrete peptides from the human precursor (grnA, grnB, grnC and grnD) and two discrete peptides from the rat precursor(epi1 and epi2) containing the cysteine repeat motif have been isolated thus far (1,2), the possibility that the remaining grn repeats are released from the precursor cannot be excluded. Grn D is the most conserved grn repeat showing 92% identity between the human and rat and is 100% conserved between the rat and the mouse (6,243). Interestingly, a comparison of the human and mouse EGF precursor (42,43,44) reveals that the biologically active mature EGF, like grnA and grnB, is not the most conserved EGF-repeat. The significance of the strict conservation of grnD is not known, but it strongly suggests that grnD may be functionally significant. The ability to express each individual grn repeat will allow us to determine if grnD or any of the other grn repeats (grns C,E,F and G) are biologically active.

A comparison of the grn precursor reveals that the structural homology extends throughout the precursor sequence which may imply a function for the intact precursor as well as the individual grn peptides. This is supported by the recent finding of an 88 kDa autocrine growth factor for a mouse teratoma derived highly tumorigenic cell line, that has been shown to be identical to the murine grn precursor (9). Additionally, acrogranin, a 67 kDa acrosomal glycoprotein of unknown function

isolated from guinea pig spermatogenic cells (5) was recently shown to be the guinea pig grn precursor (6). The finding of the grn precursor in spermatogenic cells further favours a different role for the intact grn precursor from the processed mature grn peptides. Alternatively, the precursor may be functionally silent but activated proteolytically in the extracellular milieu to give biologically active fragments, much as TGFß is stored in a latent form with the amino terminal fragment of its precursor (245), or basic FGF is sequestered to heparin-like molecules in the extracellular matrix (246). Interestingly, experiments with recombinant grn precursor indicate that the grn precursor, but not grnA, binds heparin. It therefore conceivable that the heparin sulfate proteoglycan-binding sites in the extracellular matrix of cells may be involved in the sequestration of the secreted grn precursor and in this manner regulate its biological activity.

3.2 Processing of the Granulin Precursor.

The grns were initially isolated from human leukocytes and rat bone marrow (1) and rat kidney (2) as discrete 6 kDa peptides. Subsequent reports indicate that the grns exist as multiple forms, ranging in size from 6-88 kDa. PC cells, a highly tumorigenic cell line derived from a mouse teratoma secrete an 88 kDa murine grn precursor protein (9). Equine neutrophils contain a 7.2 kDa peptide, eNAP-1, that is highly homologous to human grnE (3). Acrogranin, a 67 kDa acrosomal glycoprotein of unknown function that was recently identified as a product of the guinea pig grn precursor gene (6), is processed into smaller fragments of 62, 51, 39, and 22 kDa respectively during epididymal transit (5). TGF-e is a transforming

growth factor-like activity distinct from other known growth factors (89,90). A recent report of the amino-terminal protein sequence of bovine TGF-e shows that it is indistinguishable from grnA (7). However, bovine TGF-e has a size of 20-25 kDa (7), while human TGF-e has a reported size of 59 kDa (93). Additionally, several lower molecular weight forms of biologically active human TGF-e of 13-15 kDa and 6.5 kDa have been reported (94).

These results suggest that the processing of the grn precursor is complex and is probably dependent on the tissue and cell type, much as is the case for the EGF precursor, where, for example, the precursor is processed to generate mature EGF in the submaxillary gland, but not in the kidney although the EGF mRNA is highly expressed in the kidney (45). Additionally, it is conceivable that the various forms of the processed grn precursor as well as the intact precursor have overlapping and distinct biological functions *in vivo*.

That the processing of the grn precursor is dependent on cell type is further suggested by transfection experiments, where COS-7 cells are unable to process the grn precursor, while Swei cells appear to process the recombinant grn precursor to two smaller forms of approximately 60 and 26 kDa. Interestingly, these processed products of 60 kDa and 26 kDa are in close agreement with the 20-25 kDa bovine TGF-e and 59 kDa human TGF-e respectively. Both COS-7 and Swei cells express abundant grn mRNA (see Chapter 4), suggesting a similar fate for the endogenous grn precursor. The ability to express and detect the expressed grn gene product(s) provides an *in vitro* model to study the processing of the grn precursor in different cell types.

The mechanisms of processing of the grn precursor is not known. The processing of growth factors and related polypeptides can follow several pathways. Some, such as transforming growth factor-ß, are cleaved from their precursor at canonical dibasic sites (247); others, such as EGF are cleaved at monobasic sites (42,43); and many, such as transforming growth factor α (65), tumor necrosis α (248) and interleukin 1ß (249) are cleaved from their precursors at small non basic residues. Basic and dibasic potential cleavage sites are found in the human grn precursor flanking the repeats, but not in the rat grn precursor. This suggests that cleavage at basic sites is unlikely to be important in processing of the grn precursor to generate the 6 kDa grns. However, because the grns exist as multiple forms, it is possible that some of these forms may be derived from processing at basic sites. A consensus motif, ΣX(K/R)XPΘ, (where Σ and Θ are hydrophobic and aliphatic amino acids respectively, while X is any amino acid) is found between most of the grn repeats in the human and rat grn precursor respectively. This motif may be important for the processing or folding of the grn precursor. However, until the various grn gene products are better characterized, no definitive statement can be made regarding the mechanisms of processing of the grn precursor.

3.3 Constitutive vs Regulated Secretion of the Granulins.

The post translation pathway of the grn precursor is not completely known. However, in human granulocyte extracts, the grns copurified with a granular fraction free of cytoplasmic marker peptides (1). The granule location of the grns have been independently confirmed in horse neutrophils (3) and fish macrophage (8). In

addition, acrogranin has been immunolocalised to the acrosome of mature sperm (5,6). These results clearly indicate a granule localisation for the grn gene products. Additionally, several lines of indirect evidence point to a transit of the grn precursor through the rough endoplasmic reticulum/Golgi apparatus compartment: i) There is the presence of a probable signal sequence (234) ii) human grnC is glycosylated and glycosylation occurs exclusively in the rough endoplasmic reticulum and Golgi apparatus (250) iii) grns are disulfide bridged (1,2), implying exposure to the enzyme protein disulfide isomerase, which is located in the rough endoplasmic reticulum (251). Furthermore, transfection studies with the grn precursor cDNA show that COS-7 and Swei cells secrete recombinant grn into the culture medium. Since both cell lines express endogenous grn mRNA (see Chapter 4), this is likely to be physiologically relevant. Recently, Zhou et.al. purified the mouse grn precursor from the conditioned medium of PC cells, a highly tumorigenic cell line (9). This confirms that the grn gene products are secreted in some cells.

It is well established that proteins can be sorted in the Golgi apparatus into those bound for vesicular organelles e.g. lysosomes or secretory granules, and those destined for constitutive secretion or the plasma membrane (252,253). The grns appear to enter both vesicular organelles and constitutive secretory pathways depending on cell type. That the grns enter the former is suggested by the isolation of grns A, B, C, and D from the granular fraction of circulating leukocytes (1), and the immunolocalization of acrogranin to the acrosome of mature sperm (5,6). It should be noted that the neutrophil granules and the acrosomes are not secretory granules in the true sense, but rather are considered to be modified lysosomes. That

the grns enter a constitutive secretory pathway is suggested by their presence in the conditioned medium in the transfection experiments. Both COS-7 and Swei cells efficiently secrete recombinant grn gene products into the cell culture medium as shown by the absence of detectable recombinant grn gene products in the cell extracts of the transfected COS-7 and Swei cells. Swei cells express abundant grn mRNA (see Chapter 4), and MHC-class II antigens in Swei cells were shown to bind, *in vivo*, a fragment of the grn precursor (254), demonstrating that the endogenous grn mRNA is translated. This suggests that the secretion and processing patterns observed with recombinant grn precursor is also true for the endogenous grn precursor. Additionally, PC cells, a highly tumorigenic cell line secrete the mouse grn precursor into the conditioned medium (9).

Thus it is clear that the cumulative evidence based on work by other researchers (1,2,5,6,9) and the results reported here, prove that the grn precursor undergoes two distinct fates during transit. In some cells it is packaged and stored, whereas in other cells it is secreted intact, or partially cleaved, directly into the environment. The two different fates for the grn gene products i.e. storage in neutrophil granules and sperm acrosomes, and constitutive secretion, is similar to that observed with the EGF precursor. For example, in the salivary gland the EGF precursor is processed to the 6.2 kD mature EGF and stored in secretory granules (42,43), while the predominant form of EGF in body fluids and secretions is a 160-170 kDa soluble precursor form (46,51), suggesting that the EGF precursor is secreted in a constitutive fashion. The significance of the different fates for the grn gene products is not known, however it is clear that in order to elucidate the *in vivo*

functions of the grns, it will be important to understand the processing and secretion of the grn gene products.

In summary, the human grn precursor cDNA predicts a 593-residue glycoprotein of Mr 63.6, containing seven and one half grn repeats arranged in tandem. The grn repeats in the predicted rat grn precursor is structured in a similar fashion, and shows 75% identity, to the human grn precursor. Processing of the grn precursor appears to be tissue and cell type dependent. Secretion of the grn gene products appears to be both constitutive and regulated, with constitutive secretion of intact or partially processed precursor in some cells, and storage of more extensively cleaved peptides in vesicular organelles in other cell types. The availability of cloned grn cDNA and the ability to express recombinant grns will allow for the further characterization of the grn gene(s) and its products, which could lead to an understanding of the *in vivo* role(s) of the grn gene products.

CHAPTER 4

TISSUE DISTRIBUTION AND CELLULAR LOCALIZATION OF GRANULIN mRNA

INTRODUCTION

When the project started, the tissue distribution and cellular sites of synthesis of the grn gene were not known. Very little was known about the distribution of the grn peptides. Researchers in our laboratory have demonstrated the existence of human grns (grns A,B,C and D) in circulating leukocytes (1), and a rat grn (grn A) in bone marrow (1), while Shoyab et al. had isolated grn peptides (grnA/epi1 and grnB/epi2) from rat kidney (2). It was possible however, that the renal-derived grns/epis were not intrinsic to the kidney but were present through a passive mechanism, such as leukocyte entrapment. More recently, several groups have characterized grn peptides from different species; in each case the tissue source of the grn peptides were different (see Preface). It was therefore imperative to determine the tissue distribution and cellular localization of the grn gene mRNA and its products. More importantly, to understand the physiological role of the grn gene products, it is necessary to know the context of grn gene and peptide expression *in vivo*. Knowing which tissues and which cells within a given tissue express the grn gene may provide a basis to formulate hypotheses concerning the biological

functions of the grn gene products.

Because we did not have antibodies against the grn peptides at the time, it was not possible to elucidate the tissue and cellular distribution of the grn peptides. Although we recently obtained an antibody against the C-terminus of the intact human grn precursor, this antibody does not recognise all forms of the grn precursor, which is known to be processed differently depending on tissue and cell type (1,2,3,5,9 and see Chapter 3). Therefore a study on the levels and distribution of the grn gene products was not feasible. However, the availability of the cDNA of both the human and rat grn (see Chapter 3) provided the necessary probes to determine the tissue distribution and sites of cellular biosynthesis of the grn gene.

In this chapter we examine, by Northern blot analysis and *in situ* hybridization studies, the expression of the grn gene in cell lines and tissues. The *in situ* hybridization studies were carried out in collaboration with Dr. Adel Giaid at the Department of Pathology, Montreal General Hospital. A broad spectrum of cell lines were examined for grn gene expression by Northern blot analysis. We also determined, by Northern blot analysis, the distribution of grn gene expression in adult human, rat and rabbit tissues. *In situ* hybridizations were then performed on adult rat tissues to determine the cellular sites of biosynthesis of the grn gene. This chapter also presents Northern blot analyses of grn gene expression in human fetal tissue as an initial step to determine the possible involvement of the grn gene products in fetal physiology.

RESULTS

4.1 Northern Blot Analysis of Granulin Gene Expression.

4.1.1 Expression of Granulin mRNA in Cell Lines.

As the distribution of grn gene expression *in vitro* was not known at the start of the project, we looked for grn gene expression in various cell lines. The cells were maintained in the appropriate medium supplemented with 10% fetal bovine serum (FBS) in an atmosphere of 5% CO_2 at 37°C. Total cellular RNA was isolated using a modification of the acid guanidinium thiocyanate-phenol-chloroform extraction method initially described by Chomczynski and Sacchi (217), and is outlined in Materials and Methods, section 2.2.4. The RNA was denatured with glyoxal, fractionated by agarose gel electrophoresis, transferred to nylon membranes and hybridized with grn specific cDNA fragments as described in Materials and Methods, section 2.6.2.

The expression of grn mRNA in various cell lines is shown in Figure 4.1 and presented in Table 4.1. It should be noted that the tabulation of the grn gene expression is entirely subjective, and serves only to obtain an idea of the relative levels of grn gene expression in cell lines of different origin.

Because the grns were initially found in inflammatory cells and bone marrow (1), we expected grn gene expression in cell lines of myeloid origin e.g. HL60, U937 etc. We were surprised to find that grn gene expression was not restricted to myeloid cells, but was widespread in cell lines of diverse lineages, including cells of epithelial, myeloid, lymphoid, fibroblast and neural origin. Many cells in culture, of

Figure 4.1 Representative Northern blot analysis of grn gene expression in cell lines of diverse lineages. Grn mRNA expression was detected as outlined in Materials and Methods, section 2.6.2. See text (Chapter 4, section 4.1.1) for type and origin of cell line.

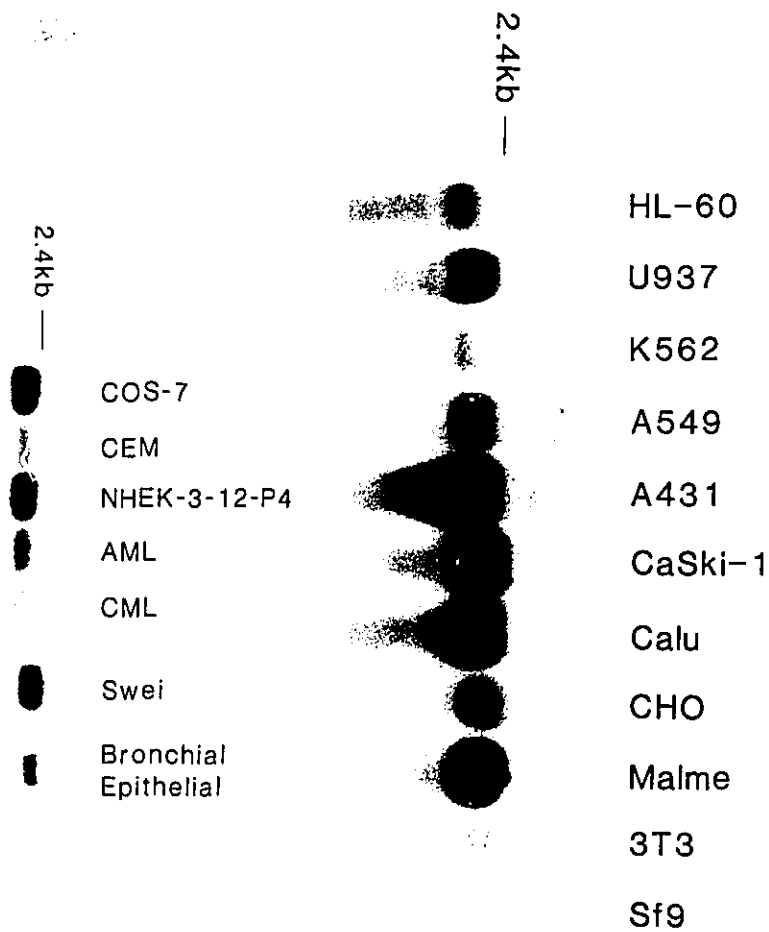

2.4kb
2.4kb
COS-7
CEM
NHEK-3-12-P4
AML
CML
Swei
Bronchial Epithelial
HL-60
U937
K562
A549
A431
CaSki-1
Calu
CHO
Malme
3T3
Sf9

Table 4.1 Distribution of grn gene expression in cells. The relative levels of gene expression is indicated by +/- (plus/minus) with - (minus) indicating undetectable expression, and + to ++++ indicating low (see Figure 4.1) to very high (see Figure 4.1) levels of gene expression respectively.

Type	Cell line	Origin	Granulin mRNA expression
myeloid	HL60	promyelocytic leukemia	++
	U937	promonocytic leukemia	+++
	K562	proerythroid leukemia	+
	KMOE	proerythroid leukemia	+
	CML	chronic myelogenous leukemia	+
	AML	acute myelogenous leukemia	++
lymphoid	Swei	B-lymphoblastoid	++++
	CEM	T-lymphoblastoid	+
epithelial	A431	epidermoid carcinoma, vulva	+++++
	Calu-1	epidermoid carcinoma, lung	++++
	CaSki	epidermoid carcinoma, cervical	++++
	SK-MES-1	squamous carcinoma, lung	++++
	A549	carcinoma, lung	+++
	CHO-K1	Chinese hamster ovary	+++
	Malme-3M	malignant melanoma	++++
fibroblast	NIH-3T3	Swiss mouse	+
	COS-7	kidney, monkey	++++
keratinocyte	NHEK-3.12-P4	normal human	+++
brain	SKNMC	neuroblastoma	++
	U373MG	glioblastoma, astrocytoma	++
	U87MG	glioblastoma, astrocytoma	+
	U118MG	glioblastoma	-
	C6	glial cell	-
muscle	SKUT-1	mixed mesodermal tumor	+
	SK-LMS-1	leiomyosarcoma, vulva	+
	L6	skeletal muscle myoblast, rat	-
insect	Sf9	ovary, fall armyworm	-

tumor or non-tumor origin, express grn mRNA. The grn mRNA transcript is approximately 2.3 kb in size, in close agreement with the length of the human and rat grn precursor cDNA (see Chapter 3).

4.1.1.1 Epithelial Cell Lines.

All cell lines of epithelial origin examined expressed abundant grn mRNA. A431 cells, a human cell line derived from an epidermoid carcinoma of the vulva (255), gave the strongest hybridization signal amongst all cell lines examined. Interestingly, A431 cells respond to the grn gene products (1,2), indicating the possibility of autoregulation. Very strong signals were also detected in cells derived from epidermoid carcinomas of the lung (Calu-1) and cervix (CaSki), and in SK-MES-1 cells, a cell line derived from a squamous carcinoma of the lung (255). Malme-3M cells, a malignant melanoma cell line (255) also expressed very high levels of grn mRNA. High levels of expression was also observed in A549 cells, a cell line of established from a carcinoma of a human lung (255). The grn mRNA is also expressed in a non-tumor epithelial cell line, CHO-K1 cells (255), which gave a strong hybridization signal. Interestingly, two grn mRNA species (2.3 and 2.5 kb) can be distinguished in some of the epithelial cells (Figure 4.2). The origin and significance of the second mRNA species is not known and requires further investigation. Plowman et. al. have reported the isolation of rat grn precursor cDNA with a 234 bp deletion (amino acids 393 to 470) (243), probably as a result of alternate splicing of an exon (see Chapter 5), thus alternate splicing could account for the second grn mRNA species observed in some epithelial cell lines. Other

Figure 4.2 Two grn mRNA species are detected in some epithelial cell lines. Grn mRNA expression was detected as outlined in Materials and Methods, section 2.6.2. See text (Chapter 4, section 4.1.1) for type and origin of cell line.

A431
A549
SK-MES-1
2.4kb —

possibilities include alternate splicing of other exons of the grn gene, or the use of a second downstream polyadenylation signal.

All the epithelial cell lines we examined initially were established secondary cell lines. Cells that can survive indefinitely in culture are not "normal". Therefore, we wanted to determine if normal epithelial cells express grn mRNA, equating primary cultured cells as a close *in vitro* approximation to normal cells *in vivo*. A blot, generously provided by Dr.Adel Giaid at the Montreal General Hospital, containing RNA from human bronchial epithelial cells which were maintained in primary culture, was probed for grn mRNA. The results show that these cells strongly express the grn gene (Figure 4.1), indicating that epithelial cells in primary culture express grn mRNA, and implying that some epithelial cells also express grn mRNA *in vivo*. This was later confirmed by *in situ* hybridization studies on rat tissue (see section 4.2).

Keratinocytes in culture can respond to the grn gene products (2). Epi1 stimulates the proliferation of murine keratinocytes *in vitro*, whereas epi2 inhibits the epi1-elicited growth of these cells (2). To determine if keratinocytes also express grn mRNA, we probed for grn gene expression in RNA isolated from NHEK-3-12-P4 cells (RNA kindly provided by Dr.R.Kremer, R.V.H), a non-transformed cell line established from normal human keratinocytes. The results show that keratinocytes in culture express grn mRNA (Figure 4.1) suggesting the possibility of autocrine or paracrine regulation.

4.1.1.2 Cells of Hemopoietic Derivation.

High levels of grn mRNA was detected in myelogenous leukemic cells of

promyelocytic (HL-60) and promonocytic (U937) lineages (255) (Figure 4.1, Table 4.1). The proerythroid-like leukemic cell lines, K562 and KMOE, gave weaker hybridization signals (Figure 4.1, Table 4.1). Cells of lymphoid origin also express grn mRNA (Figure 4.1, Table 4.1). Swei cells, a B-lymphoblastoid cell line expressed high levels of grn mRNA. MHC-class II antigens in Swei cells were shown to bind, *in vivo*, a fragment of the grn precursor (254), demonstrating that the expressed grn mRNA is translated. Expression was also detected, but at a lower level in CEM cells, a T-lymphoblastoid cell line.

It is possible that expression of the grn gene in myeloid and lymphoid cell lines in culture is due to the transformed state of these cells, or to the culture conditions. To address this, RNA was isolated from blood cells of patients with chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (AML) respectively, and probed for grn mRNA. The results presented in Figure 4.1 show that both CML and AML cells express grn mRNA, demonstrating that grn mRNA expression is not confined to myeloid and lymphoid cells grown in culture, but the gene is also expressed in leukemic cells *in vivo*.

The human grns were initially isolated from circulating leukocytes of healthy volunteers (1), and the cDNA subsequently isolated from a cDNA library constructed using the bone marrow of a healthy donor (see Chapter 3). These results demonstrate that the grn gene is also expressed in non-leukemic normal blood cells.

4.1.1.3 Fibroblasts.

Fibroblast cells in culture also express grn mRNA (Figure 4.1, Table 4.1).

COS-7 cells, a SV-40 (simian virus 40) transformed kidney fibroblast like cell line (255) express high levels of grn mRNA. In contrast, NIH-3T3 cells, a non-transformed mouse embryo fibroblast-like cell line showed lower grn gene expression. Whether the difference in gene expression between the two fibroblast-like cell lines is due to the state of the cells (transformed vs non-transformed) is not known. Interestingly, while COS cells are growth inhibited by the grn gene products epi1 and epi2, (2), Zhou et.al. have recently shown that the intact grn precursor (mouse) is mitogenic for 3T3 cells (9). Whether there is a correlation between the level of grn mRNA expression by the fibroblast-like cells and its response to the grn gene product(s) requires further investigation.

The expression of the grn gene in the above cell lines (epithelial, myeloid, lymphoid, and fibroblasts) indicates a widespread pattern of gene expression with no apparent specificity of expression *in vitro* i.e. tumor-derived vs normal cells and/or viral-transformed vs non-transformed cells in culture appear to have no effect on the absolute expression of grn mRNA.

4.1.1.4 Brain-derived Cells.

A grn-like peptide, PMP-D1 has recently been isolated from the neurosecretory *pars intercerebralis* of the locust (4), strongly suggesting that insect neural tissue express a grn-like gene. We therefore wanted to determine if the same were true for the vertebrates. Several brain-derived cell lines were examined for the expression of grn mRNA. Interestingly, we could detect grn gene expression, albeit at lower

levels than the epithelial and myeloid derived cell lines, in some cell lines of neural origin (Table 4.1). However, there is no apparent specificity of gene expression. For example, U373MG cells (human) of astrocytoma-glioblastoma origin (255) showed fairly strong grn mRNA expression while in U118MG cells (human), also of astrocytoma-glioblastoma origin (255) we could not detect any grn mRNA. SKNMC cells (human) of neuroblastoma origin (255) showed fairly strong grn mRNA expression, while lower levels were detected in U87MG cells (human), a cell line derived from a malignant glioma (255). No grn mRNA was detected in C6 cells, an astrocyte-like cell line derived from a rat glial tumor (255).

4.1.1.5 Muscle-derived Cells.

Muscle cell lines showed little or no expression of the grn gene (Table 4.1). SKUT-1, a cell line of a mixed mesodermal tumor of the uterus origin (255) and SK-LMS-1 a leiomyosarcoma of the vulva derived cell line (255) showed very little grn gene expression, while no hybridization signal could be detected in L6 cells, a non tumor rat skeletal muscle derived cell line (255).

When we were initially screening various cell lines for grn gene expression, all cells tested expressed grn mRNA. Desiring a negative control, we used RNA isolated from an insect cell line, Sf9. (Sf9 cells were kindly provided by Dr.F.Congote, R.V.H.). No grn gene expression could be detected in insect Sf9 cells.

To compare the results obtained using cell lines with the situation *in vivo*, we

examined the expression of grn mRNA in various rabbit, rat and human tissues.

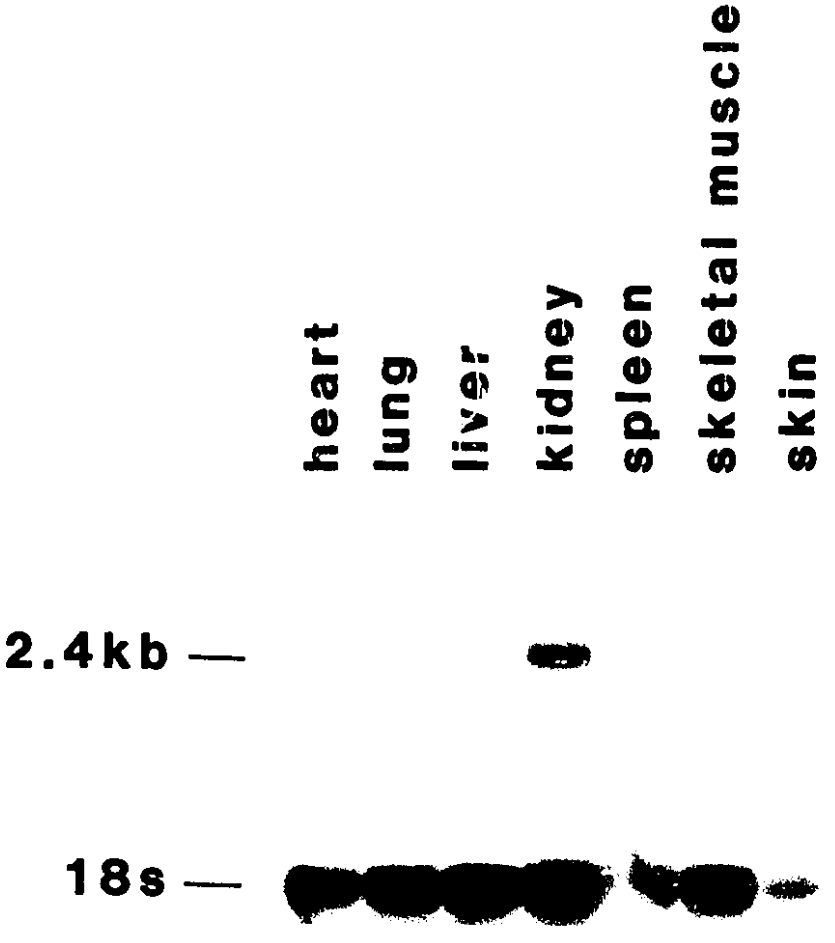
4.1.2 Expression of Granulin mRNA in Rabbit Tissues.

Various rabbit tissues were collected and immediately frozen in liquid nitrogen. The tissues were kept frozen at -70°C until needed. Total cellular RNA was isolated by the guanidinium phenol-chloroform extraction method as described in Materials and Methods, section 2.2.4. The RNA was denatured with glyoxal, transferred to nylon membranes following fractionation on an agarose gel and hybridized with a 1890 bp cDNA probe corresponding to nucleotides 6 to 1896 of the human grn precursor cDNA (see Materials and Methods, section 2.6.2 for experimental details). Very strong expression was observed with kidney RNA, and much weaker expression with spleen RNA. After prolonged exposure (24 days), hybridization was detected in total RNA from liver, lung, thymus, and brain, but not in skeletal muscle. The strong expression of grn mRNA in the kidney conclusively proves that the epi/grn peptides are intrinsically renal. We repeated the Northern blot analysis using tissue isolated from a different rabbit to see if the results were reproducible. A similar pattern of grn gene expression was observed, with very strong expression observed in the kidney, while most other tissues examined show weak expression (Figure 4.3). The same result was obtained when we probed rabbit RNA with a rat grn precursor cDNA (results not shown).

4.1.3 Expression of Granulin mRNA in Rat Tissues.

The distribution of grn gene expression in rat tissues is presented in Figures

Figure 4.3 Northern blot analysis of grn gene expression in rabbit tissues. 10μg of total RNA was loaded in each lane and the blot probed with RKL6 (top panel) and an 18S ribosomal RNA probe (bottom panel) as described in Materials and Methods, section 2.6.2.


heart
lung
liver
kidney
spleen
skeletal muscle
skin
2.4kb
18s

4.4 and 4.5. A transcript of 2.3 kb is expressed in tissue derived from all three embryonic germ layers; the ectoderm, endoderm and mesoderm. Interestingly, the abundance of grn mRNA in many rat tissues conflicted with Northern blot analysis of rabbit tissues, where although many tissues express the grn gene, its levels were low except in the kidney (Figure 4.3). Only one grn mRNA species was observed in all rat tissues examined, unlike some epithelial cell lines where two grn mRNA species were detected (Figure 4.2). Plowman et.al. reported a grn precursor cDNA sequence from rat kidney with a 234 bp deletion (243) in addition to a cDNA sequence identical to the rat grn precursor cDNA sequence presented in Chapter 3. However, we could only detect a single grn mRNA species of approximately 2.3 kb in Northern blot analysis of rat kidney RNA. This would suggest that the grn precursor mRNA with the 324 bp deletion is expressed at very low levels in the rat kidney.

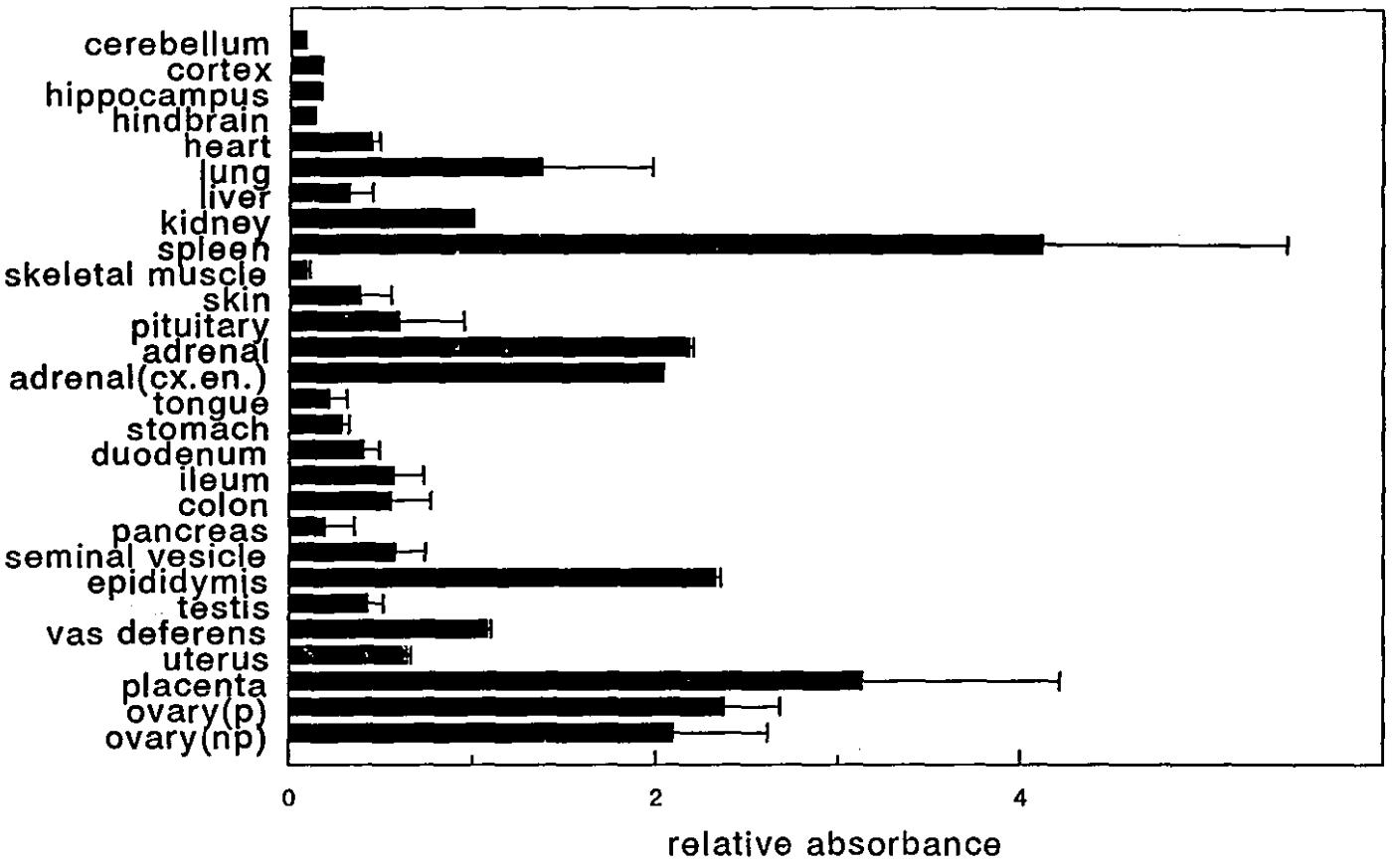
To obtain the relative levels of grn gene expression, densitometric readings were performed on the autoradiogram of the Northern blots of rat tissue RNA. To correct for loading discrepancy, the blots were reprobed with a 18S ribosomal RNA probe and the autoradiogram scanned by a densitometer. The corrected densitometer readings were then expressed relative to the intensity of kidney grn mRNA which was run as a standard in each of the blots scanned. To eliminate individual variation in gene expression, we performed two independent Northern blot analyses on RNA isolated from tissues of between two and six different rats. The averaged results are shown in Figure 4.5.

When the expression of grn mRNA in rat tissues was normalised as described

Figure 4.4 Representive Northern blot analysis of grn gene expression in rat tissues. 20μg total RNA was loaded in each lane and the blot probed with the cDNA insert of RKL28 (top panel) and an 18S ribosomal RNA probe (bottom panel) as described in Materials and Methods, section 2.6.2. (cortex enr.), (p) and (np) indicate adrenal cortex enriched, and ovaries from pregnant and non-pregnant rats respectively.

18s
2.4kb
tongue
stomach
duodenum
ileum
colon
pancreas
seminal vesicle
epididymis
testis
vas deferens
uterus
placenta
ovary (p)
ovary (np)
18s
2.4kb
cerebellum
cortex
hipoccampus
hindbrain
heart
lung
liver
kidney
spleen
skeletal muscle
skin
pituitary
adrenal
adrenal (cortex enr.)

Figure 4.5 Relative expression of grn gene in rat tissues. Following hybridization with the grn probe, the blots were reprobed with a 18S ribosomal probe to correct for loading discrepancy. Densitometric readings were obtained from autoradiograms of two to six separate Northern blot analysis of RNA isolated from tissues of different rats, and are given as the mean + SD. Northern analysis of RNA from cerebellum, cortex, hippocampus, hindbrain, and adrenal (cx.en.) was done once and do not have error bars. The expression is shown relative to the kidney which was used as an internal standard in all blots. (cx.en.), (p), and (np) indicate adrenal cortex enriched, and ovaries from pregnant and non-pregnant rats respectively.

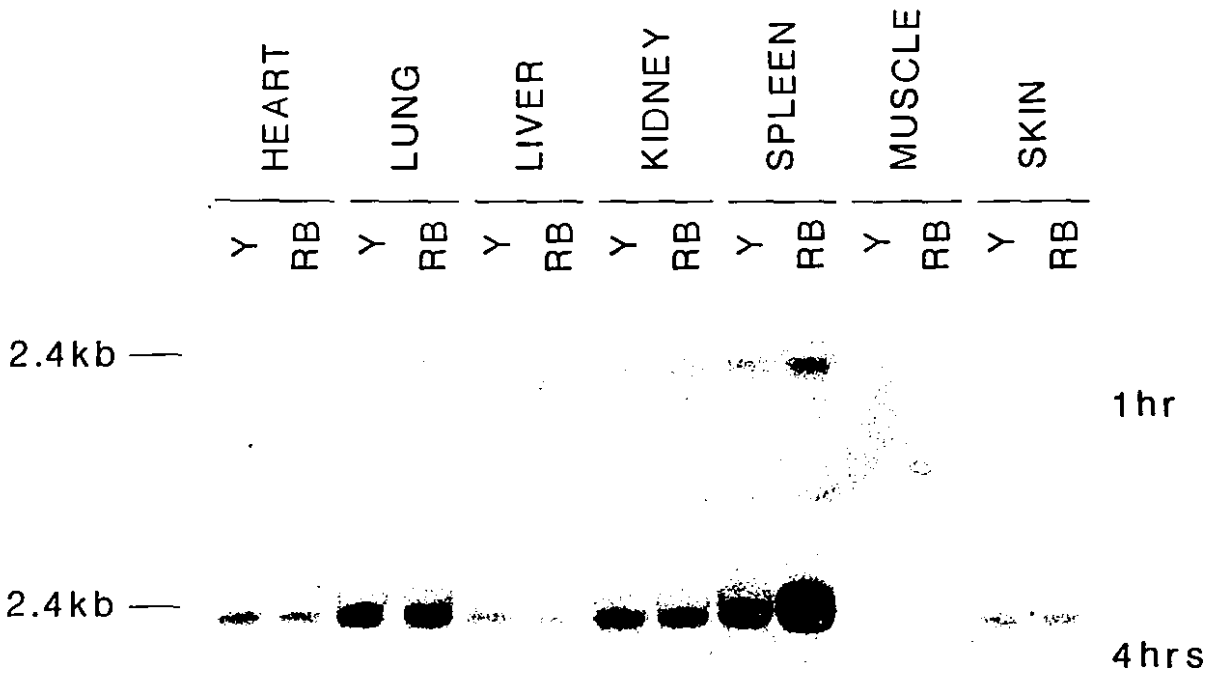

cerebellum
cortex
hippocampus
hindbrain
heart
lung
liver
kidney
spleen
skeletal muscle
skin
pituitary
adrenal
adrenal(cx.en.)
tongue
stomach
duodenum
ileum
colon
pancreas
seminal vesicle
epididymis
testis
vas deferens
uterus
placenta
ovary(p)
ovary(np)
0
2
4
relative absorbance

above, the spleen showed the strongest expression. The spleen contains abundant lymphoid and myeloid cells. Lymphoid and myeloid cells in culture express very high levels of grn mRNA (Figure 4.1, Table 4.1), which may account for the abundant grn mRNA in the spleen. Interestingly RNA isolated from the spleens of retired breeder rats (> 83 days) contain higher levels of grn mRNA than do spleens of younger rats (44-51 days) (Figure 4.6). No apparent difference in grn gene expression was observed between retired breeder and younger rat RNA from the heart, lung, liver, kidney, skeletal muscle, and skin (Figure 4.6). This is a preliminary observation requiring further experiments to provide a definitive result, but may suggest that grn gene expression is developmentally regulated in the spleen. The significance of this tissue specific age-related increase in grn gene expression is not known, and may be associated with the age-related decline in the immune system (256).

Other than the spleen, the highest levels of grn gene expression were observed in the placenta, reproductive tissues, and the adrenal. In the ovary no difference could be seen between expression of grn mRNA in non-pregnant rats and pregnant rats (19-20 days), suggesting that the hormones estrogen and progesterone probably do not regulate grn mRNA expression. However, further experiments are required for a definitive answer. In male reproductive tissue the highest expression is in the epididymis. Lower levels of expression were also observed in the seminal vesicles, vas deferens and testis.

In the adrenal, the relative level of grn mRNA expression was indistinguishable between whole adrenals and cortex enriched preparations. A subsequent Northern

Figure 4.6 Northern blot analysis of grn mRNA expression in tissues from young (44-51 day) and aged (>83 day) rats. 20 μg RNA was loaded in each lane and the blot probed with RKL6 as described in Materials and Methods, section 2.6.2. The autoradiograph exposure time is shown on the right. Y and RB indicate young and retired breeder respectively.


HEART
LUNG
LIVER
KIDNEY
SPLEEN
MUSCLE
SKIN
Y
RB
2.4kb
2.4kb
1hr
4hrs

blot analysis of RNA from the adrenal cortex and medulla showed no apparent difference in grn gene expression between the cortex and medulla (not shown). Relatively high levels of expression were observed in the lung and kidney and to a lesser extent in the skin, heart and liver. Expression was also observed in the gastrointestinal tract with the duodenum, ileum and colon showing similar hybridization signal intensity whereas the tongue and stomach gave weaker signals.

Grn mRNA expression was observed at lower levels in the CNS, with homogeneous distribution between the cerebellum, cortex, hippocampus and hind-brain. Higher levels were detected in the pituitary. There was no difference in the expression of the grn gene in extracts of whole 15, 19 and 20 day fetal brains and from the brains of adult retired breeders (Figure 4.7).

Extremely low levels of expression were observed in skeletal muscle, consistent with Northern blot analysis of muscle-derived cell lines (see section 4.1.1.5 and Table 4.1).

In summary, Northern blot analysis demonstrates a ubiquitous expression of grn mRNA in rodent tissues derived from all three embryonic germ layers; the endoderm, mesoderm, and ectoderm.

4.1.4 Expression of Granulin mRNA in Adult Human Tissues.

Post-mortem human tissue, obtained frozen from Dr. Giaid, was stored at -70°C prior to use. Northern blot analysis for grn mRNA expression was performed as described in Materials and Methods, section 2.6.2, and the results are presented in Figure 4.8. A similar pattern of grn gene expression was observed between the

Figure 4.7 Northern blot analysis of grn gene expression in fetal and adult rat brain. 20 µg RNA was loaded in each lane and the blot probed with RKL28 (top panel) and an 18S ribosomal RNA probe (bottom panel) as described in Materials and Methods, section 2.6.2. RB, AB, and FB indicate retired breeder (> 83 day) brain, adult brain (44-51 day), and fetal brain (20, 19, and 15 day) respectively.

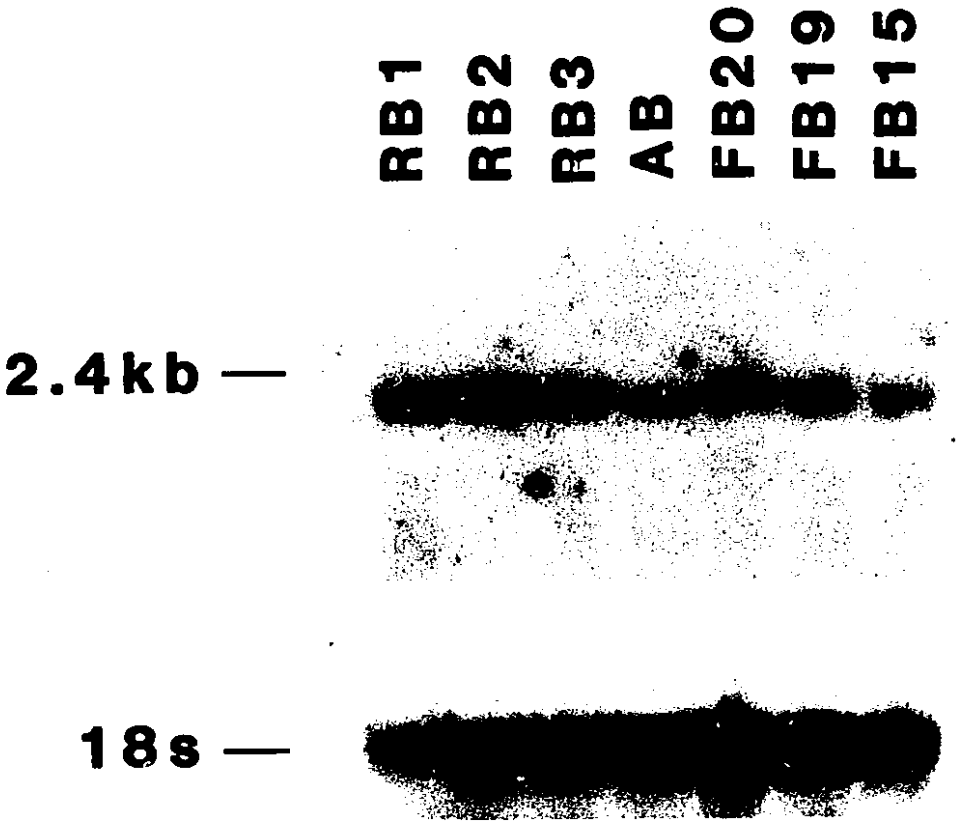

RB1
RB2
RB3
AB
FB20
FB19
FB15
2.4kb —
18s —

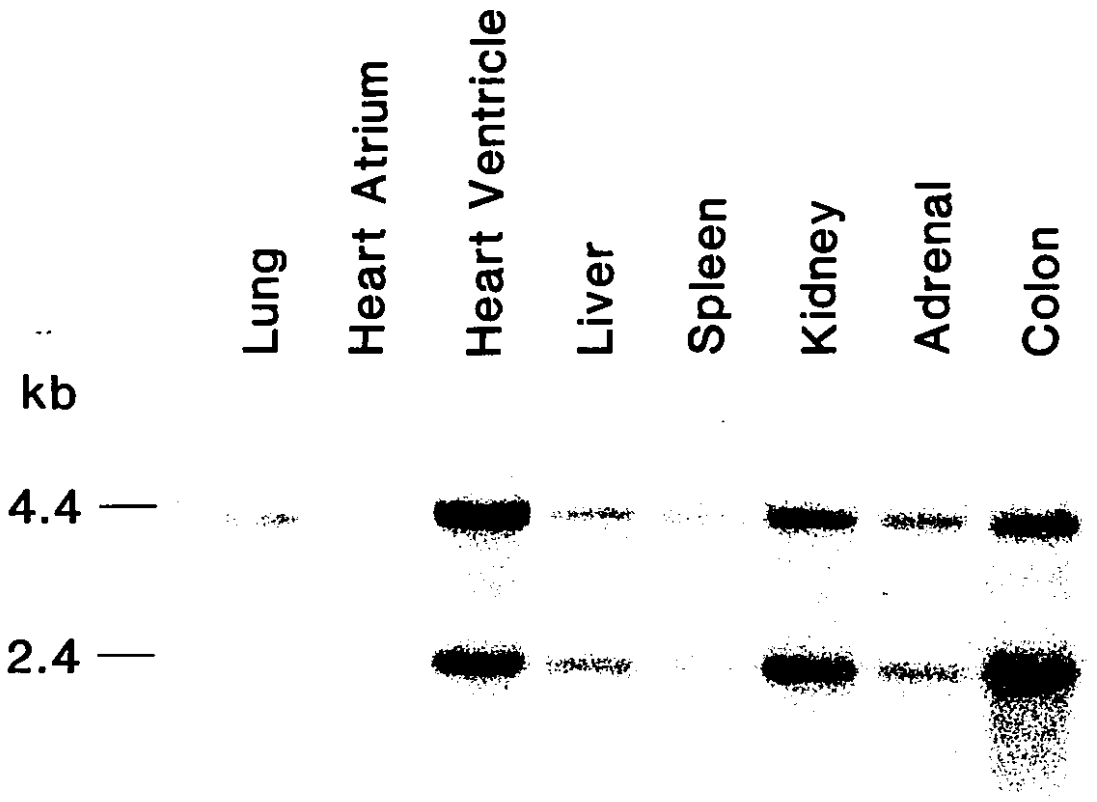
Figure 4.8 Distribution of grn gene expression in adult human tissues. 10 μg RNA was loaded in each lane and probed with a human grn cDNA restriction fragment corresponding to nucleotides 1225 to 1660 as described in Materials and Methods, section 2.6.2.

2.4kb
Heart
Lung
Liver
Kidney
Spleen
Intestine

human and rat tissue, with a 2.3 kb grn mRNA transcript detected in all adult human tissues examined. Expression was strongest in the lung, with the kidney and heart also showing strong hybridization signals. Grn gene expression could also be detected in the spleen and intestine. A weaker hybridization signal was detected in the liver.

When we examined a second set of post-mortem human tissue by Northern blot analysis, we were surprised to find two grn mRNA species, of 2.3 kb and 4.2 kb (Figure 4.9). In these tissues, the heart ventricle, adrenal, kidney and colon showed strong gene expression while weaker expression was observed in the spleen, liver and lung. A similar result showing two hybridizing species of 2.3 and 4.2 kb was obtained using different grn specific cDNA fragments corresponding to the amino, mid and carboxy portion of the grn precursor. This would suggest that the 4.2 kb hybridizing signal represents an authentic grn mRNA, and, could result from alternate splicing at the 5'end (see Chapters 3 and 5), incomplete processing of grn pre-mRNA, or the use of a second downstream polyadenylation site. Further experiments such as genomic Southern hybridization, and cloning and sequencing of the 4.2 kb transcript are required to clarify the origin and significance of this 4.2kb mRNA species. It should be noted, however, that only one set of post-mortem human tissues produced this anomalous result. All other cell lines and tissues including rabbit, rat, and human fetal tissue (see below), did not express detectable levels of the 4.2 kb grn transcript.

Figure 4.9 Northern blot analysis of adult human tissue indicate two grn mRNA species. 10 μg RNA was loaded in each lane and the blot probed with a human grn cDNA fragment corresponding to nucleotides 1225 to 1660 as described in Materials and Methods, section 2.6.2.

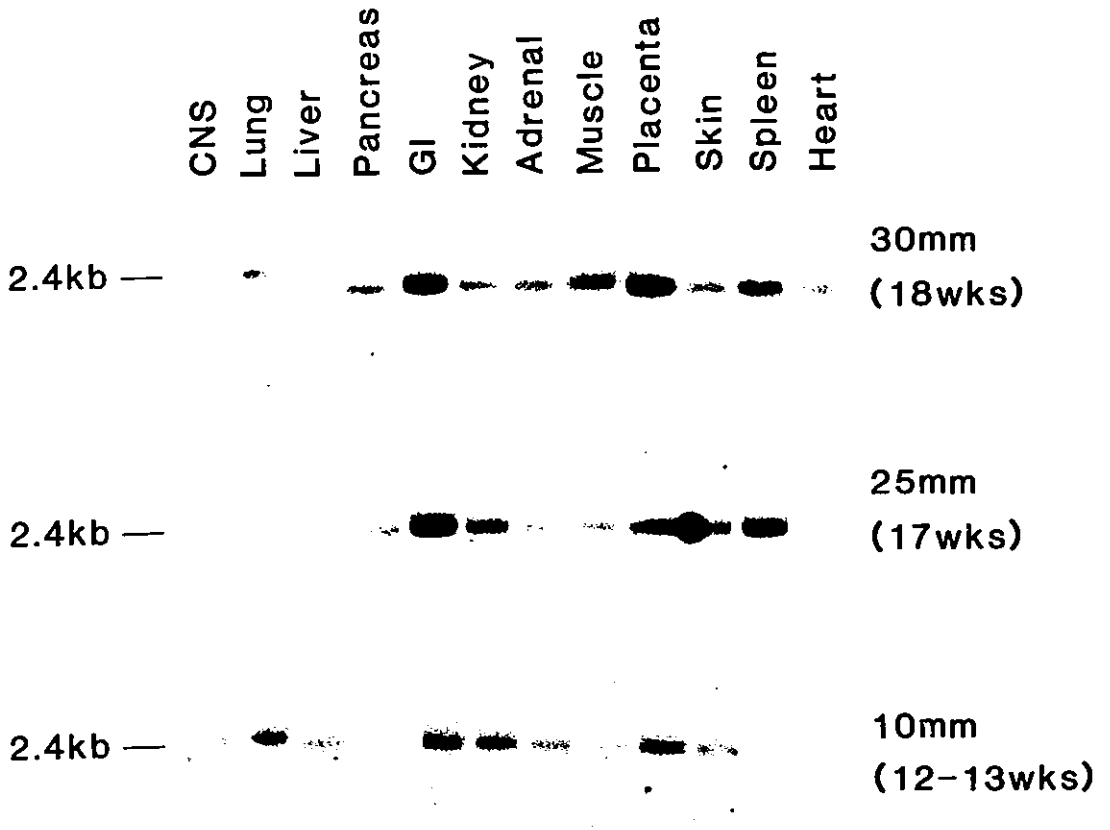

Lung
Heart Atrium
Heart Ventricle
Liver
Spleen
Kidney
Adrenal
Colon
kb
4.4
2.4

4.1.5 Expression of Granulin mRNA in Human Fetal Tissues.

Fetal tissue at various stages of gestation (12-13, 17, and 18 weeks respectively) were obtained frozen from Dr.A.Giaid and stored at -70°C prior to use. Total RNA was isolated and probed for grn mRNA as described in Materials and Methods, sections 2.2.4 and 2.6.2. The results are shown in Figure 4.10. All tissues examined express a single message of 2.3 kb with the placenta, spleen and GI tract showing strongest grn gene expression. Expression was also detected in the CNS, lung, liver, pancreas, kidney, adrenal, muscle, skin and heart. There was no apparent difference in grn gene expression with increasing gestational age. The distribution of grn mRNA expression in fetal tissues is similar to that of adult tissues, with both fetal and adult human tissue showing a ubiquitous pattern of grn gene expression. However, some differences are observed. The expression of grn mRNA in the GI tract appears more prominent in the fetus than the adult human and rat. Additionally, strong grn gene expression is observed in fetal muscle at 18 weeks gestational age. This is in contrast with the very low levels of grn mRNA expression in rat skeletal muscle (Figures 4.4 and 4.5) and muscle-derived cell lines (Table 4.1). The widespread grn gene expression in fetal tissues and the detection of grn mRNA as early as 12-13 weeks gestation, implicates the grn gene products as having an important function in fetal physiology. Plowman et.al. reported the presence of grn mRNA in embryonic carcinoma cell lines (243), further suggesting a role for the grn gene products in very early fetal development.

Unfortunately, because of the small number of samples as a result of the difficulty in obtaining post-mortem human adult and fetal tissue, it was not possible

Figure 4.10 Distribution of grn mRNA expression in human fetal tissues at various stages of gestation. 10 μg RNA was loaded in each lane and the blot probed with a cDNA fragment corresponding to nucleotides 1225 to 1660 of the human grn precursor cDNA as outlined in Materials and Methods, section 2.6.2. Numbers on right indicate length of feet in mm and corresponding gestational age.


CNS
Lung
Liver
Pancreas
GI
Kidney
Adrenal
Muscle
Placenta
Skin
Spleen
Heart
2.4kb —
30mm
(18wks)
2.4kb —
25mm
(17wks)
2.4kb —
10mm
(12–13wks)

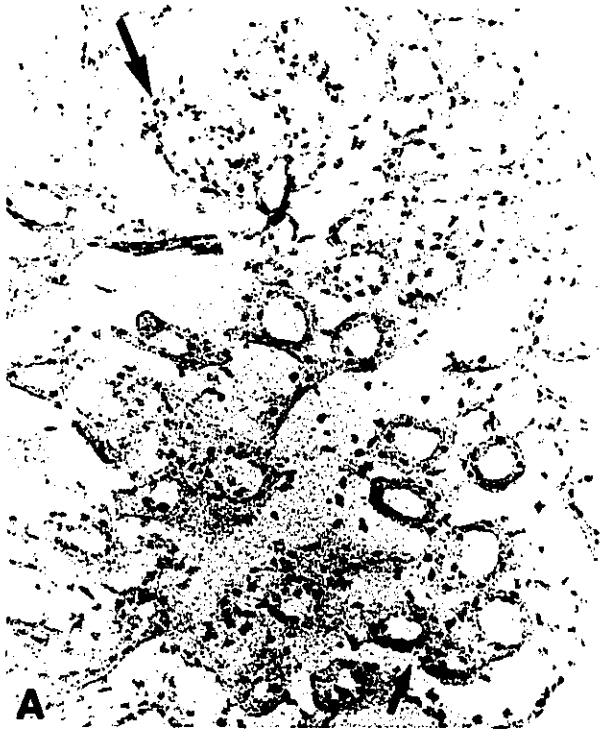
to quantitatively determine the expression of grn mRNA in human adult and fetal tissues.

4.2 Cellular Localization of Granulin Gene Expression by In Situ Hybridization.

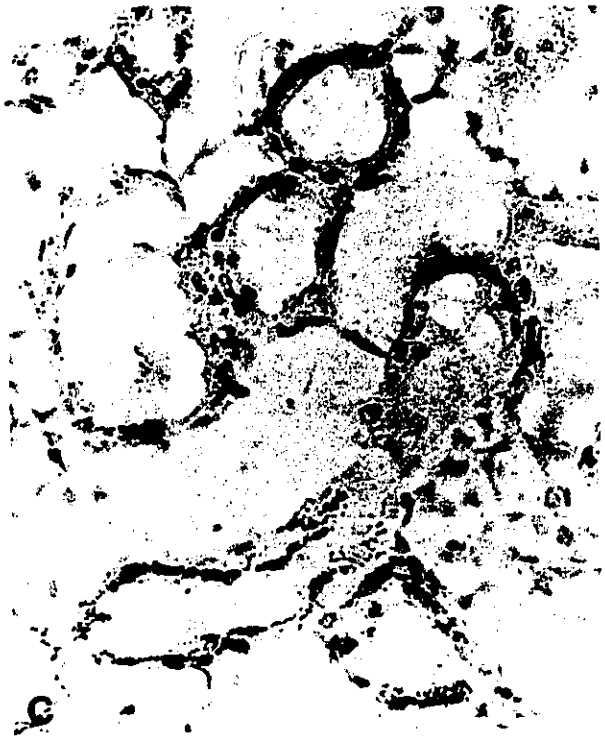
Northern blot analysis provides no indication of which cells within a given tissue express the grn gene. To address this question, we performed *in situ* hybridization analysis of rat tissues for grn mRNA expression, in collaboration with Dr.A.Giaid. The *in situ* hybridization analyses were performed as described in Materials and Methods, section 2.6.3, and the results are shown in Figures 4.11 and 4.12.

In rat kidney, grn mRNA was localized to the cortex region with few signals detected over the medulla. Hybridization signals were seen over epithelial cells of the proximal and distal tubules, and the Bowman's capsule (Figure 4.11). There was little significant hybridization signal over the glomerulosa tafts. In the liver, deposits of silver grains representing specific hybridization signals were primarily seen over scattered hepatocytes (Figure 4.11). Few signals were detected in other cell types in the liver. Scattered signals were seen over the respiratory epithelia (Figure 4.12). In situ hybridization of the spleen showed grn gene expression to be localized in lymphocytes (Figure 4.12). In the small intestine, hybridization signals were primarily seen over the intestinal epithelium with no apparent difference in hybridization intensity between the base of the villus and the tip (Figure 4.12). Scattered signals were also seen over the epithelia of the hair follicle and neuronal bodies of various nuclei of the hippocampus and brain stem (data not shown). Negative control

Figure 4.11 *In situ* hybridization analysis of grn mRNA expression. Kidney sections photographed under brightfield (**A** and **C**) and darkfield (**B** and **D**) conditions. Large arrows point to the glomerulus and small arrows to distal tubules respectively. A and B are at 250X magnification, C and D are at 500X magnification. (**E**) and (**F**) show brightfield and darkfield photographs respectively of a liver section (250X magnification) with arrows showing grn mRNA positive hepatocytes.


A


B


C


D

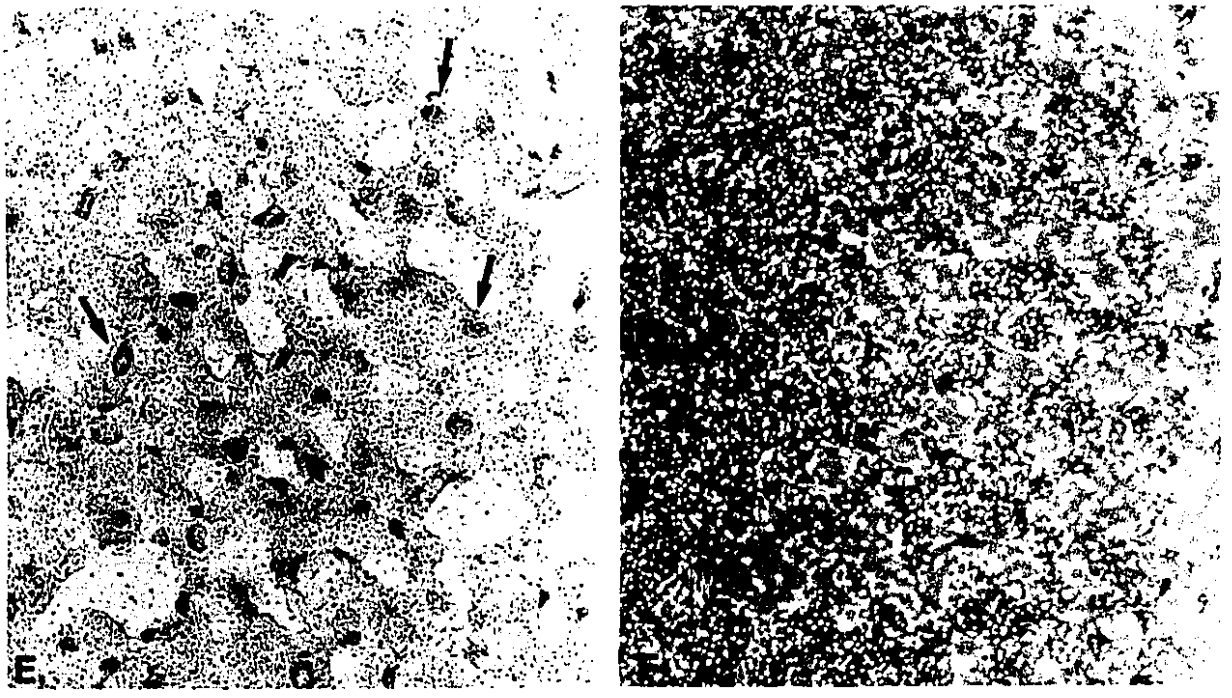
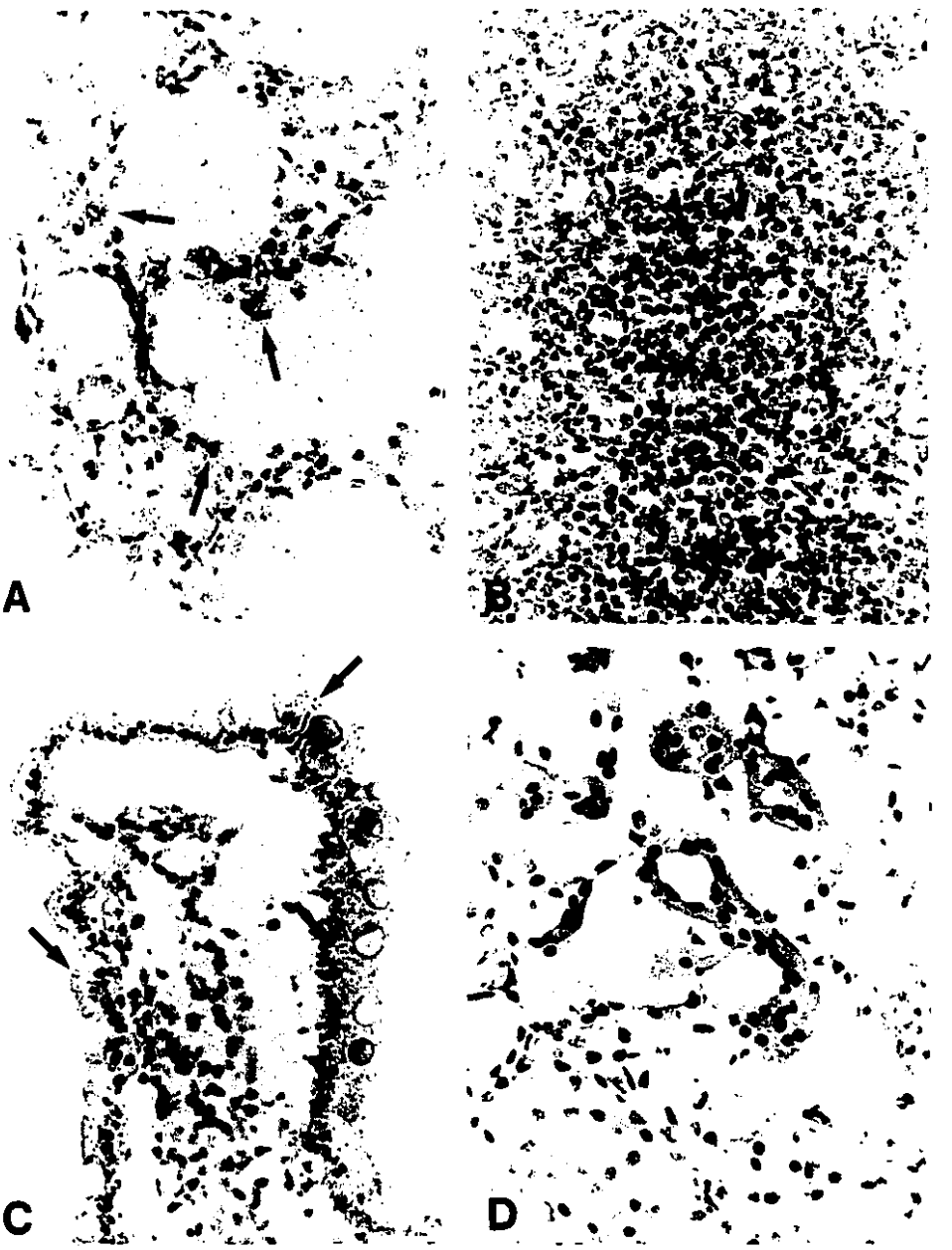


Figure 4.12 Expression of grn mRNA using *in situ* hybridization in sections of **(A)** rat lung with arrows showing respiratory epithelial cells, **(B)** spleen with arrows showing grn mRNA positive lymphocytes, and **(C)** the small intestine, with arrows showing the intestinal epithelium. Panel **(D)** represents a negative control showing a section of rat kidney hybridized with the grn sense probe. Magnification X250.


A
B
C
D

sections did not show any specific hybridization signals (Figure 4.12)

DISCUSSION

As an initial step towards understanding the physiological role of the grn gene products, we have studied the expression of the grn gene *in vitro* and *in vivo*. The Northern blot and *in situ* hybridization studies have provided valuable information to begin postulating the roles of grns *in vivo*. The possible physiological significance of the observed results is discussed below.

The abundant expression of grn mRNA in epithelial cells is particularly interesting, since many epithelial cells, including A431 cells, are growth modulated by, and some have been shown to have receptors for, the grn gene product(s) (2,82), raising the possibility of autocrine or paracrine regulation. Furthermore, many solid tumors are epithelial in origin, thus the propensity of transformed epithelial cells to express grn mRNA, and respond to the grn gene products, may have pathophysiological consequences. The possibility of autocrine and/or paracrine regulation is not restricted to epithelial cells but may also extend to other cell types. Grn mRNA is detected in fibroblast cells, which are also growth modulated by the grn gene product(s) (9). In addition, Zhou et.al. recently showed that PC cells, a highly tumorigenic cell line derived from a mouse teratoma, secretes and responds to the intact grn precursor (9) demonstrating that the grn gene product(s) may act as an autocrine growth factor for the producer cells.

We initially found grns in inflammatory cells and bone marrow (1), so the presence of grn mRNA in HL60 (promyelocyte) and U937 (promonocyte) was expected. The expression of the grn gene in two proerythroid cells, K562 and KMOE, and in the lymphoid cell lines, Swei and CEM, was unanticipated. Proerythroid cells and their progeny, and B- and T-cells are unlikely to be involved in regulating epithelial proliferation, suggesting other possible roles for the grn gene products. It is possible that the grn gene products may be involved in hematopoietic and immune functions, much as is the case with TGF-ß, which is widely expressed in myelopoietic and lymphopoietic cells, and shows pleiotropic immunoregulatory activities (86). Interestingly, the widespread occurrence of grn mRNA in myeloid cell-lines of diverse lineage contradicts the low representation (approximately 0.05 %) of grn message in the human bone marrow cDNA library. However, since these cell lines are transformed, care should be taken extrapolating the results to normal bone marrow. The apparent contradiction could also be due to stage-specific gene expression, induction of grn mRNA expression by factors in the cell culture medium, or inhibition of expression by factors in bone marrow.

Many cell lines of diverse lineages examined express grn mRNA. This ubiquitous pattern of grn gene expression in cultured cells differs from the localized grn gene expression *in vivo*. For example, human keratinocytes in culture express grn mRNA, which is not in accordance with *in vivo* findings, where, in the skin, grn gene expression was confined to the epithelia of the hair follicle with no significant grn mRNA expression in keratinocytes. Similarly, grn gene expression occurs in a restricted set of epithelial and myeloid cells *in vivo*, yet, cell culture experiments

indicate that all epithelial and myeloid cells express the grn gene. The basis of the discrepancy between the essentially universal expression of grn in cultured cells and its more localized expression *in vivo* is not clear. It is possible that grn gene expression *in vivo* is regulated by the extracellular environment or by factors secreted by surrounding cells. Thus the removal of the cells from their natural environment might contribute to the discrepancy between their behaviour in culture and *in vivo*. This is the case with TGF-ß1 where the expression of TGF-ß1 mRNA is strongly induced by the loss of cellular interactions with the ECM (257), possibly accounting for the well defined and limited expression of TGF-ß1 mRNA *in vivo* (69,70) in contrast to its synthesis by virtually all cells in culture (69,70). Therefore, caution should be taken when extrapolating from the results of grn gene expression in cultured cells to the situation *in vivo*.

The ubiquitous nature of grn gene expression in tissues is similar to the pattern of gene expression observed with some of the better studied growth factors, such as TGF-ß and bFGF (69,70,96). In the rat, grn gene expression is abundant in a number of tissues, especially in tissues that are rich in epithelia, such as the seminal vesicles, placenta, ovary and the adrenal. In contrast, tissues with little epithelium, such as the tongue, brain and skeletal muscle show correspondingly low grn gene expression. This is consistent with results from the cell lines where muscle- and brain-derived cell lines expressed little grn mRNA, and supports a role for the grn gene products in autocrine or paracrine maintenance of epithelial cells *in vivo*.

The abundance of grn mRNA in many rat and human tissues conflicts with grn mRNA expression in the rabbit, where although many tissues express the grn gene,

its levels were low except in the kidney. The biological rationale for the observed interspecies differences in grn gene expression is not clear but the high expression of grn mRNA in the kidney of all three species emphasizes a probable role for the grn gene products in renal function. It had been suggested that the grns extracted from the kidney (2) arose from leukocyte entrapment, but the *in situ* hybridization results conclusively show that renal epithelial cells express grn mRNA *in vivo*. Grn mRNA was detected in epithelial cells of the proximal and distal convoluted tubules and Bowman's capsule. The hybridization signal was low in the medullary region and in the glomerulus. The absence of hybridization in the epithelial cells of the medulla provides evidence for cellular specificity in grn gene expression among epithelial cells. In Northern blot analysis of cell lines grown in continuous culture, no specificity in grn gene expression among epithelial cells was observed, and although grn gene expression in the kidney is confined to epithelial cells, COS-7 cells, an SV-40 transformed kidney fibroblast-like cell line (255) express the grn gene *in vitro*. Grn/epi peptides have been purified from rat kidneys (2) demonstrating that the grn message is translated, and at least some of the precursor is cleaved to the smaller peptides. However, it is not known if the grn gene products enter the circulation, or are secreted directly into the lumenal space. It is possible that grns are involved in renal tissue growth and repair since epi1 stimulates anchorage independent growth of NRK cells in the presence of TGF-ß (2). The presence of grn mRNA in the tubules but not in the interstitium, is similar to the expression of the EGF gene which is also highly expressed in renal tubular epithelium (45). Under normal conditions the renal epithelium is not mitotically active (258) and the expression of the progrn gene in

these cells may serve nonmitogenic functions, as has been proposed for the EGF precursor protein (45).

In the intestine, rapid cell division occurs at the base of the villus while little cell division occurs at the tip of the villus (259). *In situ* hybridization studies show uniform grn mRNA expression in epithelia from the base to the tip of the villus, suggesting that the grn gene products may play a non-mitogenic role in the intestine.

The spleen and liver sections both show scattered staining for grn mRNA. The grn positive cells in the spleen appeared to be lymphocytes. Lymphocyte cell lines in culture express very high levels of grn mRNA. A fragment of the grn precursor is a major self-peptide bound to MHC-class II antigens in SWEI B-cells (254) demonstrating that the message is translated. We initially thought that grn gene expression in the liver might derive from reticuloendothelial cells since grns have been isolated from myeloid cells (1,3,254) and the grn gene is expressed in a number of myelogenous leukemic cell lines. This was not supported by the *in situ* hybridization results where grn positive cells in the liver do not line the sinusoids and have the appearance of hepatocytes, i.e. cells of epithelial origin, rather than reticuloendothelial cells.

The finding of grn gene expression in the brain, albeit at low levels, is intriguing. A grn-like peptide, PMP-D1, has recently been isolated from the neurosecretory *pars intercerebralis* of the locust (4). The evolutionary lines leading to insects and vertebrates diverged, at the latest, in the early Cambrian period, 570 million years ago (260). The occurrence of grn-like peptides in neural tissue across impressive evolutionary distances suggests that the grn gene products have

previously unsuspected roles in the brain, which, can at present only be guessed at; it may act as a mitogen or developmental hormone, or it may serve as a neurotransmitter or neuromodulator, or it may act as a maintenance factor. Interestingly, we could not detect a difference in grn gene expression between the developing adult and fetal brain, suggesting a non-mitogenic role for the grn gene products. However, since we only examined fetal brains at day 15 onwards, its role as a mitogen during the earlier stages of brain development cannot be excluded. It also conceivable that the grn gene products act as maintenance factors necessary for the continual survival of the brain cells, much as nerve growth factor enhances the survival of cholinergic neurons in the mammalian central nervous system (261). The function of the grns in the brain remain speculative at present, and elucidation of the roles of grn gene products will require a detailed analysis of the effects of the grns on neuronal and glial cells, as well as the knowledge of the post-translational fate of progrn in the vertebrate brain, and the neuroanatomic localization of both the grn mRNA and the translated protein.

The widespread expression of grn mRNA in fetal tissues implicates a role for the grn gene products in fetal physiology. The broad expression profile of grn mRNA is reminiscent of the ubiquitous mRNA distribution in human fetal tissue of the IGFs, peptide mitogens that have been implicated in fetal growth and development (147,148). Other growth factors that are widely expressed during embryogenesis include TGF-ß (69,70), bFGF (96), and TGF-α (68). Interestingly, unlike grn gene expression, TGF-α mRNA is selectively expressed in a restricted set of human fetal tissue, including nerve tissue, liver, adrenal, and kidney (68). During fetal

development, rapid cell division and differentiation occur. However, analysis of various fetal tissues showed that there was no apparent difference in grn gene expression with increasing gestational age. This may imply that the grn gene products may serve non-mitogenic functions during fetal development. It should be noted, however, that only one set of fetal tissue at each gestational stage was examined. Therefore more studies are required (better N-values) before a definitive statement can be made.

A limitation of all these studies is that measurements were made only at the mRNA level. Grn peptides have been isolated from hematopoietic tissue (1,3,8,254) and kidney (2). Whether the grn mRNA is translated in other tissues is not known at present. Additionally it is not known if the translated product is secreted from all cells. Tissue localisation of the grn gene products will complement the mRNA studies to further provide valuable information on the possible *in vivo* roles of the grn gene products. Progrn is differentially processed in various cell types (1,2,6,9 and see Chapter 3) and some of the products are mutually antagonistic in certain assay systems (2). The intact progrn has also been shown to act as an autocrine growth factor (9). It will therefore be essential to establish the post-translational processing of the precursor in each tissue. Additionally, in hypothesizing the role of the grn gene products *in vivo*, the localization and quantity of the grn receptor(s) should also be considered. It will therefore be imperative to determine the distribution and localization of the grn receptor(s).

In summary, the grn gene is strongly expressed in most, perhaps all, myeloid and epithelial cell lines *in vitro*, and at lower or undetectable levels in brain- and

muscle-derived cell lines. Grn gene expression can be detected in a restricted set of epithelial cells and lymphocytes *in vivo*. Although grn gene expression is also detected in fibroblasts and other non-epithelial cell types in culture, it appears to be much more restricted to epithelial and hemic cells *in situ*. The limited expression of the grn gene in epithelial and hemic cells implies cell specific functions for grn gene products, as well as cell specific regulation of grn gene expression *in vivo*. Several of the known biological properties of the grns resemble those of the EGF/TGF-α polypeptides. Whether this is physiologically significant is unclear but it is possible that the grns, or related products of the grn-gene, are involved in regulating cell growth. The grn gene product(s) may also serve a non-mitogenic function. It is becoming increasingly apparent that many growth modulating factors are multifunctional and the same may be for the grn gene product(s). The grn gene is expressed in fetal and adult tissues and is consistently abundant in those tissues that have a high content of epithelial cells. Based on work by other researchers (1,2,3,7,8,243,254) and the results reported in this chapter, the grn gene products are likely to be particularly significant in cells of the immune system and epithelial cells. The presence of grn gene expression in epithelial cells *in vivo* as shown here, the known effects of the grns/epis on epithelial cell growth (1,2,8,243), and the recently reported identification of a high affinity grn/epi receptor (Kd 2×10^{-10} M) on an epithelial cell line (82) suggest an important role for the grn gene product in autocrine or paracrine regulation of epithelial cells *in vivo*.

CHAPTER 5

CHROMOSOMAL LOCALIZATION, GENOMIC ORGANIZATION AND STRUCTURAL AND FUNCTIONAL ANALYSIS OF A PROMOTER OF THE HUMAN GRANULIN GENE.

INTRODUCTION

In this chapter, the structure of the human grn gene will be presented, followed by an analysis of the 5'-flanking region and promoter function. Also presented in this chapter is the chromosomal localization of the human grn gene. The work described in this chapter provides the foundation to address two important questions:

(i) *What regulates grn gene expression.*

The grn gene is expressed in a wide variety of tissue (see Chapter 4). *In vitro,* the grn gene is widely expressed in cell types of diverse lineages, including epithelial cells, myeloid and lymphoid cells and fibroblasts. However, its expression *in situ* is apparently restricted predominantly to hematopoietic and some epithelial cells. This restricted cell distribution suggests that grn gene expression is regulated *in vivo*. A primary control point in the regulation of gene expression is the initiation of mRNA synthesis. Transcription initiation is a complex process involving the interaction of *trans*-acting factors with specific *cis*-element DNA sequences frequently found in the

immediate vicinity of the transcription initiation site(s); i.e. the promoter region. Isolation of the 5'-flanking region of the human grn gene, identification of putative *cis*-acting response elements, and delineating a functional promoter of the grn gene represents an initial step towards understanding the molecular mechanisms regulating the cell specificity of grn gene expression and those governing grn gene expression in general.

(ii) *How did the grn gene evolve.*

Comparison of the human and rat grn precursor shows that the protein is highly conserved between human and rodent (see Chapter 3). The isolation of grn-like peptides from teleost fish (8) and locust (4) establishes that the grn gene products are also found in lower vertebrates and insects. These results clearly demonstrate that the grn gene has a long evolutionary history. To investigate the evolutionary origin of the cysteine-rich motifs that form the grn precursor (see Chapter 3), phylogenetic studies of its molecular structure are required. Therefore, knowledge of the exon/intron organization of the human grn gene is an initial step towards characterizing the evolutionary emergence of the grns.

RESULTS

5.1 Chromosome Localization of the Human Granulin Gene.

The polymerase chain reaction (PCR) was used to determine the chromosome

assignment of the human grn gene. Two independent grn specific primer pairs (see Materials and Methods, section 2.8 for experimental details) were used in the PCR with DNA from human-hamster somatic cell hybrids as the template. Human and hamster genomic DNA were used as the templates for positive and negative control reactions respectively. The PCR was performed as described in Materials and Methods, section 2.3.1, and the reaction products fractionated by agarose gel electrophoresis and transferred to nylon membranes. The membranes were then hybridized with grn specific probes complementary to nucleotide sequences lying between the primer pairs used in the PCR (see Methods and Material, section 2.8 for experimental details) and the results are shown in Figure 5.1. With both grn-specific primer pairs, Southern blotting of the PCR products revealed a single amplified product of the predicted size with human genomic DNA and with DNA from the somatic cell hybrid 811, which contains human chromosomes 8,17 and 18 (Figure 5.1, Table 5.1). No PCR product was obtained with hamster genomic DNA or with the control reaction (no DNA added). Localization of the grn gene to chromosomes 8 or 18 can be excluded because of the absence of hybridization in hybrids 803, 909, 967 and 1006 which contain chromosome 8, or hybrids 324, 734, 750, and 867 which contain chromosome 18 (Table 5.1). Therefore the grn gene must be on chromosome 17.

5.2.1 Isolation of Genomic Clones Containing the Protein Coding Region of the Granulin Gene.

To begin characterization of the grn gene structure, a commercial human

Figure 5.1 Southern Hybridization of PCR products. Conditions for the PCR are described in Materials and Methods, section 2.8. The amplified products were electrophoresed on a 1.5% agarose gel, transferred to nylon membranes (ZetaProbe, BioRad) by capillary blotting and hybridized with grn specific probes (Materials and Methods, section 2.6.1.). Conditions for the hybridizations and washes were essentially as described by the manufacturer (BioRad). **(A)** PCR with primer pair A and probed with a 21-mer oligonucleotide. **(B)** PCR with primer pair B and probed with a 1890-bp *Eco*RI/*Sac*I fragment. The description of primer pair A and B and the probes used are given in Materials and Methods, section 2.8.

B A

860
867
909
940
967
968
983
1099
1006
324
423
507
683
734
750
803
811
854
HUMAN
HAMSTER
CONTROL

Table 5.1 Segregation of Human Chromosomes in Human-Hamster Somatic Cell Hybrids. +/- indicate presence (+) or absence (-) of the noted chromosome in the cell hybrids. Percentage numbers are the percent of the cell population containing the noted chromosome. D - Deletion at p15.1-15.2.

Cell Line	Human Chromosome Content																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
324	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
423	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
507	-	-	+	-	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	40%	-	25%	-	+
683	-	-	-	-	+	-	-	-	-	-	+	45%	-	+	-	-	-	-	+	-	+	+	-	-
734	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
750	-	-	-	-	D	-	-	-	-	-	-	+	+	+	-	-	-	+	-	-	-	-	-	-
803	-	-	-	+	+	-	-	+	-	-	15%	-	-	-	-	-	-	-	-	-	-	+	+	-
811	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-
854	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
860	-	-	15%	-	+	+	-	-	-	15%	-	-	-	-	-	-	-	-	45%	-	+	-	-	-
867	60%	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	+	+	-	-	-	-	-
909	-	-	-	-	D	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-
940	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
967	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
968	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-
983	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1006	-	-	-	55%	+	-	+	+	-	-	-	-	+	-	+	-	-	-	+	-	+	-	-	-
1099	+	-	-	-	D	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	+	+	-	-
Hamster	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Human	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

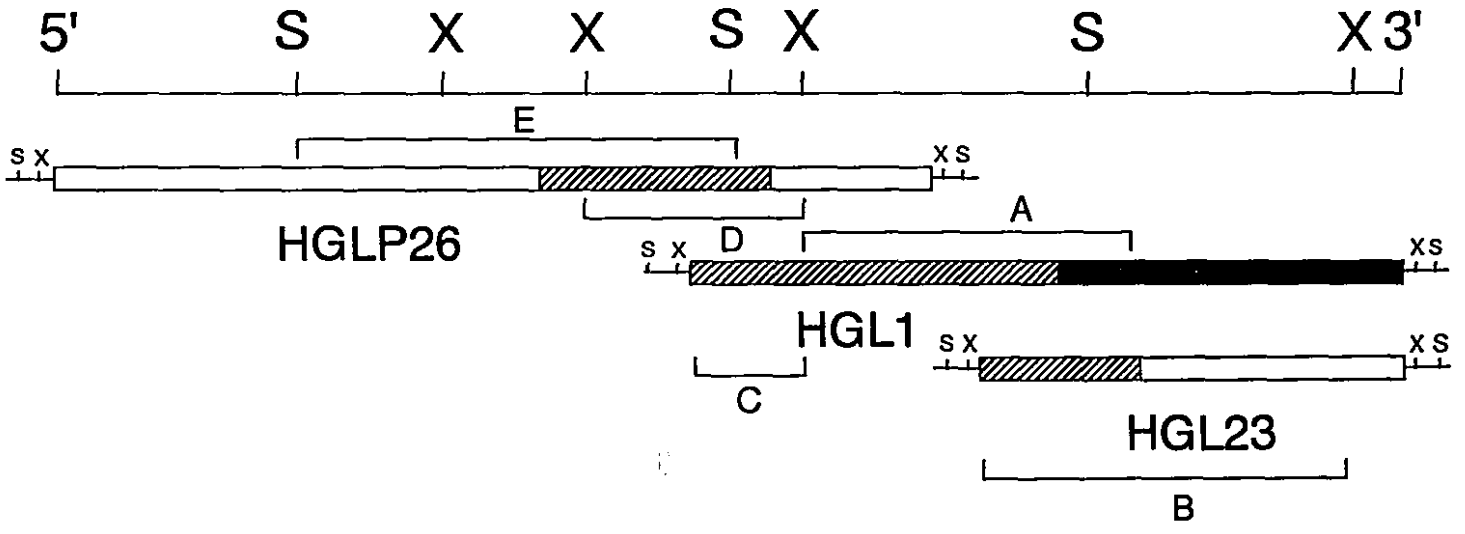
genomic library was screened with a 182-bp PCR derived genomic fragment corresponding to part of the grn A gene (see Chapter 3 section 3.1.1, and Figure 3.3). The choice of probe was governed by two factors: i) To maximize the stringency of the hybridization conditions and thereby eliminate non-specific hybridization, the probe should contain the exact nucleic acid sequence of the grn gene i.e. a homologous probe. ii) To increase the probability of isolating genomic clones containing the complete grn gene, the probe should correspond to the mid-portion of the grn precursor cDNA. The 182-bp PCR-derived genomic fragment corresponding to part of the grnA gene fits both criteria. The hybridization and washing conditions employed in screening the genomic library are described in Materials and Methods, section 2.4.3. Of approximately 45 000 clones screened, only one positively hybridizing clone, HGL1, was obtained. The DNA of HGL1 was isolated and digested with the restriction enzymes *Bam*HI, *Sac*I, *Sal*I and *Xho*I to determine the size of the DNA insert and to obtain a partial restriction map of the DNA insert. This restriction analysis suggested that HGL1 contained a DNA insert of approximately 13.5 kb. HGL1 DNA was then digested with the restriction enzyme *Xho*I, and the resulting fragments (DNA insert fragments of approximately 1200, 1300, 2300, 3200 and 5500 bp respectively) were fractionated by agarose gel electrophoresis and transferred to a nylon membrane. Restriction fragments containing the grn gene were then identified by Southern hybridization with a 1890-bp fragment corresponding to nucleotides 6 to 1896 of the human grn precursor cDNA, which covers the complete coding region of the grn precursor. Only one *Xho*I restriction fragment of 5500 bp, was recognised by the grn specific probe (Figure

5.2, fragment A). This fragment was then subcloned into Bluescript KSII and characterized by restriction analysis. The DNA sequence of the 5500bp *Xho*I fragment was then obtained by dideoxynucleotide sequencing according to the strategy described in Materials and Methods, section 2.4.3.

Analysis of the nucleotide sequence of the 5500bp *Xho*I fragment, and comparison with the human grn precursor cDNA sequence revealed that the *Xho*I fragment contained only the 5' portion of the grn gene. To determine if the DNA insert of HGL1 contained the 3'-end of the grn gene, HGL1 was probed with a *Eco*RI/*Kpn*I restriction fragment (fragment 12C, Figure 3.4) corresponding to nucleotides 1709 to 2099 of the 3'-end of the human grn precursor cDNA. No hybridization signal was obtained, indicating that the DNA insert of HGL1 did not contain the 3'-end of the protein coding region of the grn gene.

To obtain the 3'-end of the grn gene, the human genomic library was rescreened with a 390 bp cDNA restriction fragment corresponding to the 3'-end of the human grn precursor cDNA (Fragment 12C, Figure 3.4). One positively hybridizing clone, HGL23, with a DNA insert of approximately 10 kb was isolated. Restriction analysis of HGL23 suggested that the DNA insert of clone HGL23 overlapped the 5500 bp *Xho*I subclone of HGL1 by approximately 1400 bp (Figure 5.2). This was confirmed by nucleotide sequencing. A 8 kb *Xho*I restriction fragment of clone HGL23 recognized by the 3'-end grn cDNA probe (Figure 5.2, fragment B) was subcloned into Bluescript KSII, characterized by restriction analysis and partially sequenced to obtain the complete nucleotide sequence of the 3'-end of the grn gene.

Figure 5.2 Schematic and partial restriction map of overlapping genomic clones, HGL1, HGL23, and HGLP26 containing the protein coding region and 5' untranslated and flanking region of the grn gene. The hatched region indicates region sequenced. HGL1 contains a spurious 7 kb fragment at the 3' end indicated by the solid region. A, B, C, and D indicate fragments that were subcloned to obtain nucleotide sequence (see text). S and X indicate *Sac*I and *Xho*I respectively.


5'
S
X
X
S
X
S
X
3'
E
S X
X S
HGLP26
A
D
S X
X S
HGL1
S X
X S
C
HGL23
B

The entire protein coding region of the grn gene was therefore contained within two overlapping genomic clones; HGL1, from which was determined the sequence of the 5'-end of the grn gene, and, HGL23, from which the 3'-end of the grn gene sequence was obtained. The nucleotide sequence of the protein coding region of the human grn gene is shown in Figure 5.3.

5.2.2 Isolation of Genomic Clones Containing the 5'-Flanking Region of the Granulin Gene.

As an initial step towards understanding the molecular mechanisms involved in the regulation of grn gene expression, genomic clones containing the 5'-flanking region of the grn gene were obtained. Nucleotide sequencing of a 2.3 kb *Xho*I fragment of clone HGL1 (see section 5.2.1) revealed that it contained only 675 bp of 5'-flanking sequence of the human grn gene. To obtain genomic clones containing a greater portion of the 5'-flanking region, a human genomic library was probed with a 108 bp cDNA fragment corresponding to the 5' untranslated region of the grn gene (nucleotides -35 to -142). This fragment was obtained by a PCR based strategy described in Materials and Methods, section 2.3.3 and in Chapter 3, Section 3.1.3. Three clones, HGLP21, HGLP22 and HGLP26, were obtained from the human genomic library. Restriction mapping of HGLP21, HGLP22 and HGLP26 followed by Southern hybridization with two different grn specific cDNA fragments corresponding to i) the 5'-untranslated region (nucleotides -35 to -142), and ii) nucleotides 6 to 1896 of the grn precursor cDNA (spanning the complete coding region of the grn precursor), indicated that the three clones overlapped, and all three

Figure 5.3 Nucleotide sequence of the human grn gene. The nucleotide sequence is a composite of two overlapping clones, HGL1 and HGL23. HGL1 and HGL 23 overlap by about 1400 bp. Sequences of exons, intron/exon boundaries and 3' untranslated region are shown. The amino acid sequence is shown below the nucleotide sequence. The stop codon is indicated by #. The lengths of the introns are given in parentheses. IV refers to intervening sequence. The polyadenylation signal is underlined, and the nucleotide sequence resembling the consensus sequence frequently found downstream of the polyadenylation site is double underlined. The nucleotides are numbered according to the human cDNA sequence.


.:....IVO(~3000bp).....tactttgcagGCAGACCATGTGGACCCTGGTG 15
M W T L V
AGCTGGGTGGCCTTAACAGCAGGGCTGGTGGCTGGAACGCGGTGCCCAGATGGTCA 71
S W V A L T A G L V A G T R C P D G Q
GTTCTGCCCTGTGGCCTGCTGCCTGGACCCGGAGGAGCCAGCTACAGCTGCTGCC 127
F C P V A C C L D P G G A S Y S C C
GTCCCCTTCTGGtgagtgccc.....IV1(121bp).....gtctttctagGACAA 143
R P L L D K
ATGGCCCACAACACTGAGCAGGCATCTGGGTGGCCCCTGCCAGGTTGATGCCCACT 199
W P T T L S R H L G G P C Q V D A H
GCTCTGCCGGCCACTCCTGCATCTTTACCGTCTCAGGGACTTCCAGTTGCTGCCCC 255
C S A G H S C I F T V S G T S S C C P
TTCCCAGAGgtgagcgtgc.....IV2(115bp).....tgttccacagGCCGTGG 271
F P E A V
CATGCGGGGATGGCCATCACTGCTGCCCACGGGGCTTCCACTGCAGTGCAGACGGG 327
A C G D G H H C C P R G F H C S A D G
CGATCCTGCTTCCAAAGATCAGgtgcagctgg.....IV3(479bp).....cttg 349
R S C F Q R S
tcacagGTAACAACTCCGTGGGTGCCATCCAGTGCCCTGATAGTCAGTTCGAATGC 399
G N N S V G A I Q C P D S Q F E C
CCGGACTTCTCCACGTGCTGTGTTATGGTCGATGGCTCCTGGGGGTGCTGCCCCAT 455
P D F S T C C V M V D G S W G C C P M
GCCCCAGgtacaaatct.....IV4(100bp)....ttttcctcagGCTTCCTGCT 472
P Q A S C
GTGAAGACAGGGTGCACTGCTGTCCGCACGGTGCCTTCTGCGACCTGGTTCACACC 528
C E D R V H C C P H G A F C D L V H T
CGCTGCATCACACCCACGGGCACCCACCCCCTGGCAAAGAAGCTCCCTGCCCAGAG 584
R C I T P T G T H P L A K K L P A Q R
GACTAACAGGGCAGgtgaggaggt.....IV5(112bp).....cctcttccagTG 600
T N R A V
GCCTTGTCCAGCTCGGTCATGTGTCCGGACGCACGGTCCCGGTGCCCTGATGGTTC 656
A L S S S V M C P D A R S R C P D G S
TACCTGCTGTGAGCTGCCCAGTGGGAAGTATGGCTGCTGCCCAATGCCCAACgtga 708
T C C E L P S G K Y G C C P M P N
gtgagg.....IV6(234bp).....ccccactcagGCCACCTGCTGCTCCGATCA 728
A T C C S D H
CCTGCACTGCTGCCCCCAAGACACTGTGTGTGACCTGATCCAGAGTAAGTGCCTCT 784
L H C C P Q D T V C D L I Q S K C L
CCAAGGAGAACGCTACCACGGACCTCCTCACTAAGCTGCCTGCGCACACAGgtacc 836
S K E N A T T D L L T K L P A H T
agagg.....IV7(199bp).....tccctcacagTGGGGGATGTGAAATGTGACA 856
V G D V K C D
TGGAGGTGAGCTGCCCAGATGGCTATACCTGCTGCCGTCTACAGTCGGGGGCCTGG 912
M E V S C P D G Y T C C R L Q S G A W
GGCTGCTGCCCTTTTACCCAGgtaccaggtg.....IV8(79bp).....tctgcc 933
G C C P F T Q
ctagGCTGTGTGCTGTGAGGACCACATACACTGCTGTCCCGCGGGGTTTACGTGTG 985
A V C C E D H I H C C P A G F T C
ACACGCAGAAGGGTACCTGTGAACAGGGGCCCCACCAGGTGCCCTGGATGGAGAAG 1041
D T Q K G T C E Q G P H Q V P W M E K
GCCCCAGCTCACCTCAGCCTGCCAGACCCACAAGCCTTGAAGAGAGATGTCCCCTG 1097
A P A H L S L P D P Q A L K R D V P C
TGATAATGTCAGCAGCTGTCCCTCCTCCGATACCTGCTGCCAACTCACGTCTGGGG 1153
D N V S S C P S S D T C C Q L T S G
AGTGGGGCTGCTGTCCAATCCCAGAGgtatatggga.....IV9(219bp)..... 1179
E W G C C P I P E
cacccccagGCTGTCTGCTGCTCGGACCACCAGCACTGCTGCCCCCAGGGCTACA 1225
A V C C S D H Q H C C P Q G Y
CGTGTGTAGCTGAGGGGCAGTGTCAGCGAGGAAGCGAGATCGTGGCTGGACTGGAG 1281
T C V A E G Q C Q R G S E I V A G L E
AAGATGCCTGCCCGCCGGGCTTCCTTATCCCACCCCAGAGACATCGGCTGTGACCA 1337
K M P A R R A S L S H P R D I G C D Q
GCACACCAGCTGCCCGGTGGGGCAGACCTGCTGCCCGAGCCTGGGTGGGAGCTGGG 1393
H T S C P V G Q T C C P S L G G S W
CCTGCTGCCAGTTGCCCCATgtgagtgcct.....IV10(108bp).....ttccc 1413
A C C Q L P H
gccagGCTGTGTGCTGCGAGGATCGCCAGCACTGCTGCCCGGCTGGCTACACCTGC 1464
A V C C E D R Q H C C P A G Y T C
AACGTGAAGGCTCGATCCTGCGAGAAGGAAGTGGTCTCTGCCCAGCCTGCCACCTT 1520
N V K A R S C E K E V V S A Q P A T F
CCTGGCCCGTAGCCCTCACGTGGGTGTGAAGGACGTGGAGTGTGGGGAAGGACACT 1576
L A R S P H V G V K D V E C G E G H
TCTGCCATGATAACCAGACCTGCTGCCGAGACAACCGACAGGGCTGGGCCTGCTGT 1632
F C H D N Q T C C R D N R Q G W A C C
CCCTACCGCCAGgtcagtgcca.....IV11(85bp).....gaccatccagGGCG 1648
P Y R Q G
TCTGTTGTGCTGATCGGCGCCACTGCTGTCCTGCTGGCTTCCGCTGCGCAGCCAGG 1704
V C C A D R R H C C P A G F R C A A R
GGTACCAAGTGTTTGCGCAGGGAGGCCCCGCGCTGGGACGCCCCTTTGAGGGACCC 1760
G T K C L R R E A P R W D A P L R D P
AGCCTTGAGACAGCTGCTGTGAGGGACAGTACTGAAGACTCTGCAGCCCTCGGGAC 1816
A L R Q L L #
CCCACTCGGAGGGTGCCCTCTGCTCAGGCCTCCCTAGCACCTCCCCCTAACCAAAT 1872
TCTCCCTGGACCCCATTCTGAGCTCCCCATCACCATGGGAGGTGGGGCCTCAATCT 1928
AAGGCCCTTCCCTGTCAGAAGGGGGTTGTGGCAAAAGCCACATTACAAGCTGCCAT 1984
CCCCTCCCCGTTTCAGTGGACCCTGTGGCCAGGTGCTTTTCCCTATCCACAGGGGT 2040
GTTTGTGTGTGTGCGCGTGTGCGTTTCAATAAAGTTTGTACACTTTCTTAACAGTG 2090
TCTGATTTGCCGCCCCTGCCTGCCC 3022

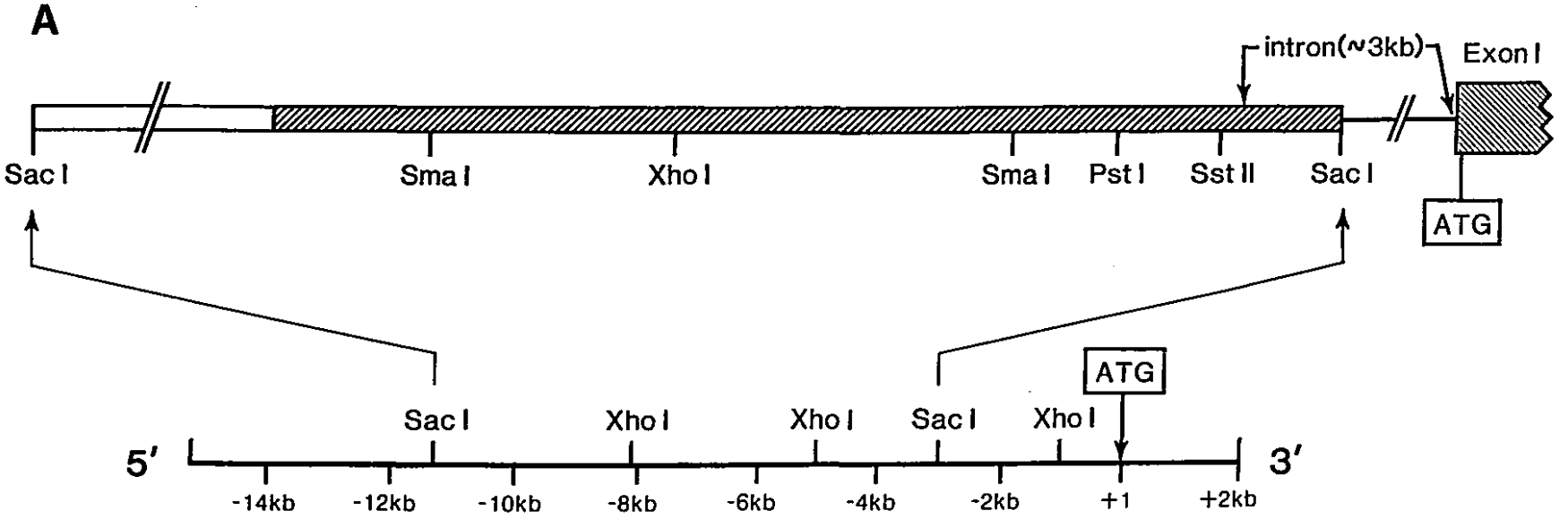
clones contained a substantial portion of the 5'-flanking region of the grn gene. One clone, HGLP26, with a DNA insert of approximately 17 kb, was then chosen for further analysis.

HGLP26 was digested with *Sac*I and *Xho*I to release DNA insert fragments of approximately 3.6, 5.0, and 8.3 kb for the *Sac*I restriction digest and 3.0, 3.2, 4.1, and 6.5 kb for the *Xho*I restriction digest as assessed by agarose gel electrophoresis. The restriction fragments were transferred to a nylon membrane and probed with two grn specific cDNA fragments corresponding to the 5'-untranslated region (nucleotides -35 to -142) and to nucleotides 6 to 1896 of the grn precursor cDNA respectively to obtain a partial restriction map of HGLP26 (Figure 5.2 and 5.4A). Restriction analysis suggested that the DNA insert of HGLP26 overlapped with the DNA insert of clone HGL1 (see section 5.2.1) by approximately 5.5 kb. Two overlapping restriction fragments from clone HGLP26, a 4.1 kb *Xho*I fragment and a 8.3 kb *Sac*I fragment (Figure 5.2, fragments D and E), recognized by the 5'-end grn cDNA probe used to screen the genomic library (nucleotides -35 to -142), was then subcloned into Bluescript KSII, characterized by restriction analysis and partially sequenced to obtain 2878 bp of the 5'-flanking region of the human grn gene (Figure 5.4B).

5.3.1 Organization of the Protein Coding Region of the Granulin Gene.

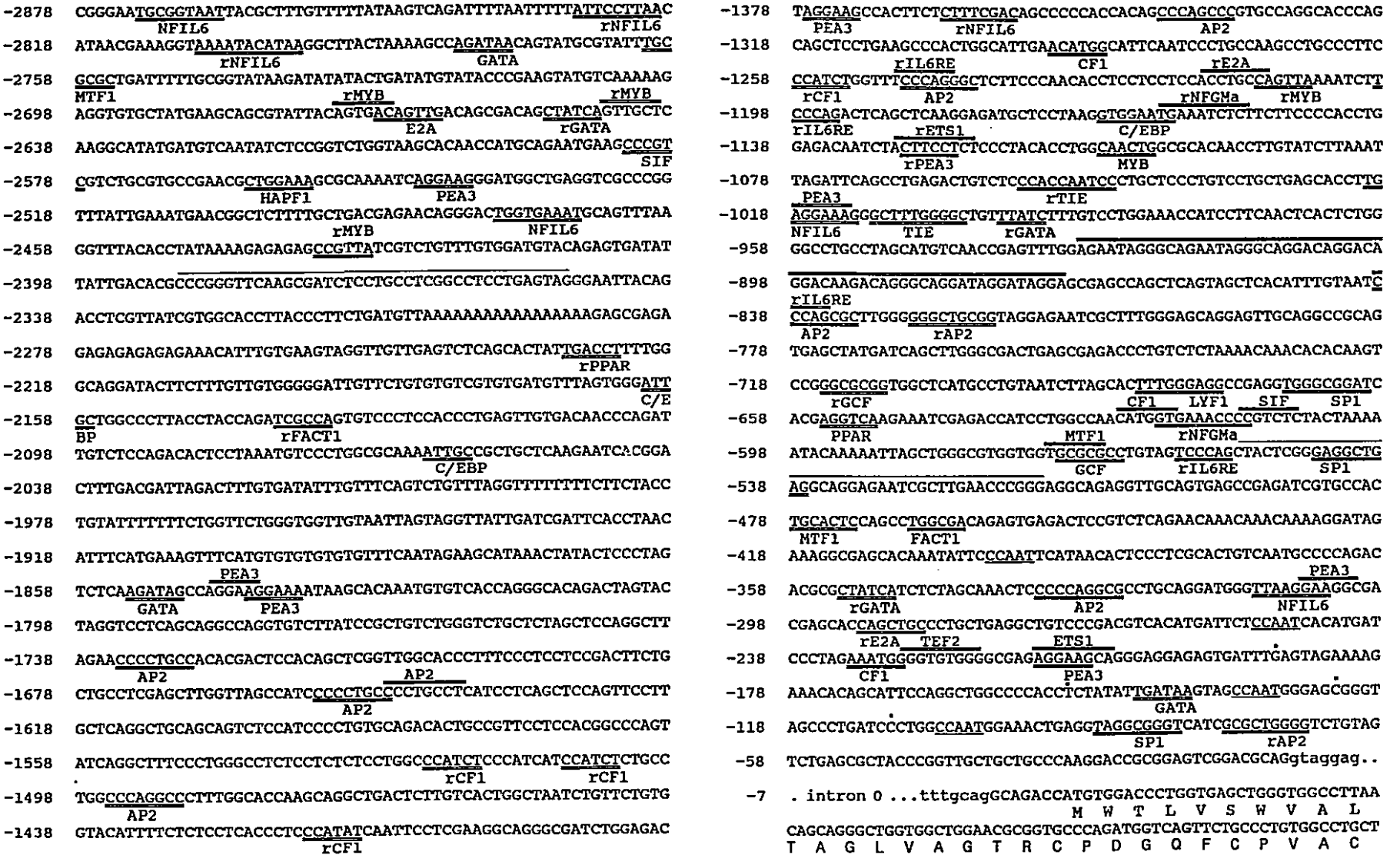
The complete nucleotide sequence of the protein coding region of the human grn gene, obtained from two overlapping genomic clones, HGL1 and HGL23, is shown in Figure 5.3. The protein coding region of the grn gene covers about 3700

Figure 5.4 Structure and nucleotide sequence of the 5' flanking and untranslated region of the human grn gene. **(A)** Schematic and partial restriction map of genomic clone (HGLP26) containing the 5'-flanking region and part of the coding region of the human grn gene. The ATG codon is boxed. Shown above HGLP26 is the *SacI* subclone used to obtain the 5'flanking sequence. The hatched area indicates region sequenced. **(B)** Nucleotide sequence of the 5' flanking region of the human grn gene. Shown also is part of the first intron (intron 0) and the initial 30 amino acids of the coding region. Nucleotide +1 denctes the A residue of the translation initiation codon (ATG), and residues preceding it are indicated by negative numbers. The putative transcription initiation sites determined by primer extension analysis are indicated by a ■. The sequences of potential regulatory elements are double underlined. (r) indicates putative response element in reverse complement. The four putative CCAAT boxes are underlined. The 11 bp direct repeat followed by eight 5 bp repeats in tandem are double overlined. The 41 bp anti-sense sequence is overlined. The homopurine/ homopyrimidine tracts are located at nucleotides -175 to -189; -196 to -215; -1116 to -1131; -1140 to -1156; -1204 to -1240; -1414 to -1433; -1527 to -1541; -1545 to -1552; -1680 to -1701; -1831 to -1845; -1966 to -1994; -2265 to -2303; -2435 to -2445.


A
intron(~3kb)
Exon I
Sac I
Sma I
Xho I
Sma I
Pst I
Sst II
Sac I
ATG
ATG
Sac I
Xho I
Xho I
Sac I
Xho I
5'
3'
-14kb
-12kb
-10kb
-8kb
-6kb
-4kb
-2kb
+1
+2kb

HGLP26

B


-2878 CGGGAATGCGGTAATTACGCTTTGTTTTTTATAAGTCAGATTTTAATTTTTATTCCTTAAC
NFIL6 rNFIL6
-2818 ATAACGAAAGGTAAAATACATAAGGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGC
rNFIL6 GATA
-2758 GCGCTGATTTTTGCGGTATAAGATATATACTGATATGTATACCCGAAGTATGTCAAAAAG
MTF1 rMYB rMYB
-2698 AGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTC
E2A rGATA
-2638 AAGGCATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGT
SIF
-2578 CGTCTGCGTGCCGAACGCTGGAAAGCGCAAAATCAGGAAGGGATGGCTGAGGTCGCCCGG
HAPF1 PEA3
-2518 TTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAAATGCAGTTTAA
rMYB NFIL6
-2458 GGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATAT
-2398 TATTGACACGCCCGGGTTCAAGCGATCTCCTGCCTCGGCCTCCTGAGTAGGGAATTACAG
-2338 ACCTCGTTATCGTGGCACCTTACCCTTCTGATGTTAAAAAAAAAAAAAAAAAGAGCGAGA
-2278 GAGAGAGAGAGAAACATTTGTGAAGTAGGTTGTTGAGTCTCAGCACTATTGACCTTTTGG
rPPAR
-2218 GCAGGATACTTCTTTGTTGTGGGGGATTGTTCTGTGTGTCGTGTGATGTTTAGTGGGATT
C/E
-2158 GCTGGCCCTTACCTACCAGATCGCCAGTGTCCCTCCACCCTGAGTTGTGACAACCCAGAT
BP rFACT1
-2098 TGTCTCCAGACACTCCTAAATGTCCCTGGCGCAAAATTGCCGCTGCTCAAGAATCACGGA
C/EBP
-2038 CTTTGACGATTAGACTTTGTGATATTTGTTTCAGTCTGTTTAGGTTTTTTTTCTTCTACC
-1978 TGTATTTTTTTCTGGTTCTGGGTGGTTGTAATTAGTAGGTTATTGATCGATTCACCTAAC
-1918 ATTTCATGAAAGTTTCATGTGTGTGTGTGTTTCAATAGAAGCATAAACTATACTCCCTAG
PEA3
-1858 TCTCAAGATAGCCAGGAAGGAAAATAAGCACAAATGTGTCACCAGGGCACAGACTAGTAC
GATA PEA3
-1798 TAGGTCCTCAGCAGGCCAGGTGTCTTATCCGCTGTCTGGGTCTGCTCTAGCTCCAGGCTT
-1738 AGAACCCCTGCCACACGACTCCACAGCTCGGTTGGCACCCTTTCCCTCCTCCGACTTCTG
AP2 AP2
-1678 CTGCCTCGAGCTTGGTTAGCCATCCCCCTGCCCCTGCCTCATCCTCAGCTCCAGTTCCTT
AP2
-1618 GCTCAGGCTGCAGCAGTCTCCATCCCCTGTGCAGACACTGCCGTTCCTCCACGGCCCAGT
-1558 ATCAGGCTTTCCCTGGGCCTCTCCTCTCTCCTGGCCCATCTCCCATCATCCATCTCTGCC
rCF1 rCF1
-1498 TGGCCCAGGCCCTTTGGCACCAAGCAGGCTGACTCTTGTCACTGGCTAATCTGTTCTGTG
AP2
-1438 GTACATTTTCTCTCCTCACCCTCCCATATCAATTCCTCGAAGGCAGGGCGATCTGGAGAC
rCF1
-1378 TAGGAAGCCACTTCTCTTTCGACAGCCCCCACCACAGCCCAGCCCGTGCCAGGCACCCAG
PEA3 rNFIL6 AP2
-1318 CAGCTCCTGAAGCCCACTGGCATTGAACATGGCATTCAATCCCTGCCAAGCCTGCCCTTC
rIL6RE CF1 rE2A
-1258 CCATCTGGTTTCCCAGGGCTCTTCCCAACACCTCCTCCTCCACCTGCCAGTTAAAATCTT
rCF1 AP2 rNFGMa rMYB
-1198 CCCAGACTCAGCTCAAGGAGATGCTCCTAAGGTGGAATGAAATCTCTTCTTCCCCACCTG
rIL6RE rETS1 C/EBP
-1138 GAGACAATCTACTTCCTCTCCCTACACCTGGCAACTGGCGCACAACCTTGTATCTTAAAT
rPEA3 MYB
-1078 TAGATTCAGCCTGAGACTGTCTCCCACCAATCCCTGCTCCCTGTCCTGCTGAGCACCTTG
PEA3 rTIE
-1018 AGGAAAGGGCTTTGGGGCTGTTTATCTTTGTCCTGGAAACCATCCTTCAACTCACTCTGG
NFIL6 TIE rGATA
-958 GGCCTGCCTAGCATGTCAACCGAGTTTGGAGAATAGGGCAGAATAGGGCAGGACAGGACA
-898 GGACAAGACAGGGCAGGATAGGATAGGACGAGCCAGCTCAGTAGCTCACATTTGTAATCC
rIL6RE
-838 CCAGCGCTTGGGGGGCTGCGGTAGGAGAATCGCTTTGGGAGCAGGAGTTGCAGGCCGCAG
AP2 rAP2
-778 TGAGCTATGATCAGCTTGGGCGACTGAGCGAGACCCTGTCTCTAAAACAAACACACAAGT
-718 CCGGGCGCGGTGGCTCATGCCTGTAATCTTAGCACTTTGGGAGGCCGAGGTGGGCGGATC
rGCF CF1 LYF1 SIF SP1
-658 ACGAGGTCAAGAAATCGAGACCATCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAA
PPAR MTF1 rNFGMa
-598 ATACAAAAATTAGCTGGGCGTGGTGGTGCGCGCCTGTAGTCCCAGCTACTCGGGAGGCTG
GCF rIL6RE SP1
-538 AGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGTGCCAC
-478 TGCACTCCAGCCTGGCGACAGAGTGAGACTCCGTCTCAGAACAAACAAACAAAAGGATAG
MTF1 FACT1
-418 AAAGGCGAGCACAAATATTCCCAATTCATAACACTCCCTCGCACTGTCAATGCCCCAGAC
PEA3
-358 ACGCGCTATCATCTCTAGCAAACTCCCCCAGGCGCCTGCAGGATGGGTTAAGGAAGGCGA
rGATA AP2 NFIL6
-298 CGAGCACCAGCTGCCCTGCTGAGGCTGTCCCGACGTCACATGATTCTCCAATCACATGAT
rE2A TEF2 ETS1
-238 CCCTAGAAATGGGGTGTGGGGCGAGAGGAAGCAGGGAGGAGAGTGATTTGAGTAGAAAAG
CF1 PEA3
-178 AAACACAGCATTCCAGGCTGGCCCCACCTCTATATTGATAAGTAGCCAATGGGAGCGGGT
GATA
-118 AGCCCTGATCCCTGGCCAATGGAAACTGAGGTAGGCGGGTCATCGCGCTGGGGTCTGTAG
SP1 rAP2
-58 TCTGAGCGCTACCCGGTTGCTGCTGCCCAAGGACCGCGGAGTCGGACGCAGgtaggag..
-7 . intron 0 ...tttgcagGCAGACCATGTGGACCCTGGTGAGCTGGGTGGCCTTAA
M W T L V S W V A L
CAGCAGGGCTGGTGGCTGGAACGCGGTGCCCAGATGGTCAGTTCTGCCCTGTGGCCTGCT
T A G L V A G T R C P D G Q F C P V A C

bp, with 12 exons and 11 introns. A twelfth intron of approximately 3 kb lies outside the open reading frame, in the 5' untranslated region, 7 bp upstream of the translation initiation codon. Postulated splice sites were determined by comparison with the human grn cDNA sequence. Splice acceptor and donor sequences (Table 5.2) agree with the "GT-AG" rule (262) and conform to splice site consensus sequence proposed by Mount (230), including a pyrimidine-rich region of approximately 11 nucleotides or greater found within 3 to 10 bp of the 3' splice site. At the 3' end of the grn gene, a single consensus polyadenylation signal, AATAAA, is present. Located 20 bp downstream of the polyadenylation site is the sequence AGTGTCTG; this sequence agrees closely with the consensus sequence YGTGTGYY (where Y is either pyrimidine) frequently found within about 25 bp downstream of the polyadenylation site (263). Thus, the human grn gene contains typical signals for mRNA splicing and polyadenylation (264,265).

In the coding region, 8 splice junctions occur between amino acid codons (type 0 introns), 3 occur after the first nucleotide of a codon (type I introns), and none occur after the second nucleotide of a codon (type II introns) (266). This can be compared to the average values of 41% type O, 36% type I, and 23% type II previously reported for vertebrate genes (267). Thus there is a clear preponderance of type 0 splice junctions in the grn gene. The introns within the protein coding region of the gene are small, ranging in size from 79 to 479 bp, with an average of 168 bp (Table 5.2). The exons vary in size, coding for 28 to 83 amino-acids.

Each 12-cysteine grn repeat is encoded by two exons, and in some instances, one exon codes for the carboxy-terminal of one repeat and the amino-terminal of an

Table 5.2 Size and splice junction sequences of exons and introns in the protein coding region of the human grn gene. Exon sequences are given in uppercase and intron sequences in lowercase. Amino acids with interrupted codons were assigned to the exon containing two of the three codon nucleotides.

[a] Consensus sequence from Mount (230)

[b] Intron type is according to Sharp (266). Type O indicates introns lying between two codons and type I indicates introns lying between the first and second nucleotides of a codon.

Exon	Amino Acids	5'donor (C,A)AG/gt(a,g)agt[a]	Intron	Length (bp)	3'acceptor (c,t)ag/G[a]	Type[b]
I	46	CTG/gtgagt	1	121	tag/G	O
II	42	GAG/gtgagc	2	115	cag/G	O
III	28	CAG/gtgcag	3	479	cag/G	I
IV	38	CAG/gtacaa	4	100	cag/G	O
V	45	CAG/gtgagg	5	112	cag/T	I
VI	37	AAC/gtgagt	6	234	cag/G	O
VII	42	CAG/gtacca	7	199	cag/T	I
VIII	33	CAG/gtacca	8	79	tag/G	O
IX	83	GAG/gtatat	9	219	cag/G	O
X	78	CAT/gtgagt	10	108	cag/G	O
XI	77	CAG/gtcagt	11	85	cag/G	O
XII	45					

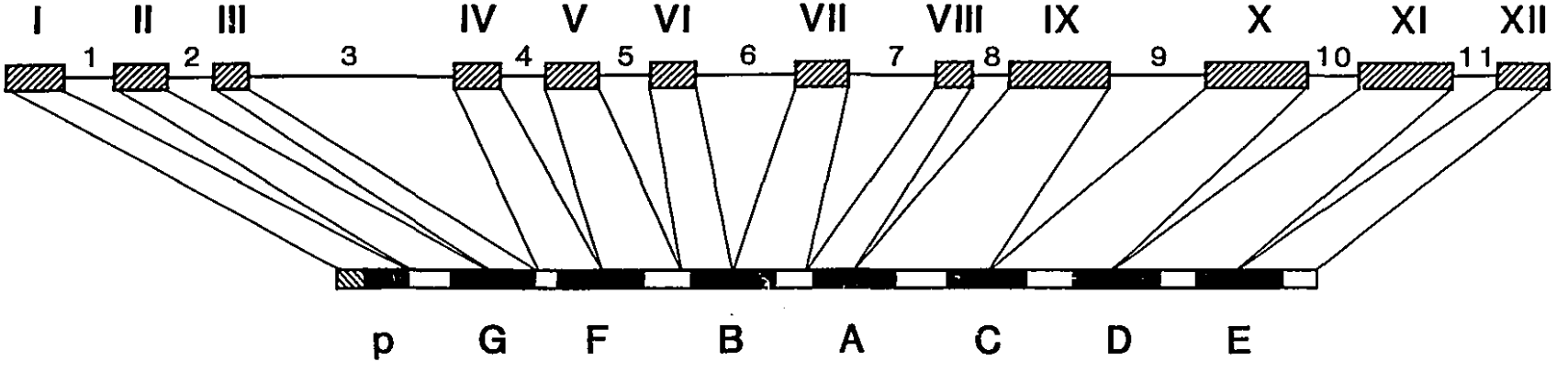
adjacent repeat (Figure 5.5). Exon I contains 7 bp of 5'-untranslated region and encodes the signal peptide and paragrn. Exon II and exon III encode grnG, while grnF is encoded by exons IV and V respectively. Exon VI and VII contain grnB. GrnA is encoded by exons VIII and IX, while grnC is encoded by exons IX and X /respectively. Exons X and XI contain grnD, and grnE is encoded by exons XI and XII respectively.

Type 0 introns (introns lying between two codons) bisect the translated grn protein repeats between the sixth and seventh cysteine (Figures 5.3 and 5.5). The position of these junctions is rigidly conserved in all seven grn repeats. The introns in the intervening regions between grn repeats are all type 1 (introns lying between the first and second nucleotide of a codon) and their position relative to the grn repeats is variable. Baba et.al. recently described the structural organization of the mouse grn gene (268). The intron/exon arrangement of the protein coding region of the mouse gene is identical to the human grn gene.

Three types of exons, which we have called α,ß and ßα can be recognised based on the peptide sequences that they encode and their type class (Figure 5.5, Table 5.2). α exons encode a six cysteinyl peptide sequence with the cysteine motif $CX_{5-6}CX_5CCX_8CCX_4$ corresponding to the amino-terminal half of a grn repeat, and have type 1 and 0 5' and 3' exonic boundaries respectively. ß exons encode a six cysteinyl peptide with a $X_2CCX_5CCX_5CX_{5-6}C$ array corresponding to the carboxyl terminus of the next grn repeat, with type 0 and 1 5' and 3' boundaries respectively. ßα exons encode a 12 cysteine peptide, $X_2CCX_5CCX_5CX_{5-6}CX_{25-30}CX_{5-6}CX_5CCX_8CCX_4$ comprising the carboxyl-terminal half of a grn repeat, an intervening spacer

Figure 5.5 Structural organization of the human grn gene. Exons are indicated by hatched boxes and introns by lines. Exons I, II, IV, VI and VIII are α-exons, exons III, V, VII and XII are ß-exons and exons IX, X and XI are ßα-exons (see section 5.3.1). Closed boxes represent the grn domains, open boxes represent the intervening spacer regions and the hatched box represents the signal sequence.

Granulin gene structural organization



Granulin precursor

sequence, and then the amino-terminal half of a grn repeat, with type 0 boundaries at both 5' and 3' ends. The 5' end of the protein coding region of the grn precursor consists of alternating α and ß exons. The pattern changes towards the 3' end of the gene where the ßα exons appear. The grn gene is therefore constructed of three types of exons arranged in the following manner: α-α-ß-α-ß-α-ß-α-ßα-ßα-ßα-ß, where α encodes the amino terminus and ß the carboxy terminus of a grn repeat (Figure 5.5).

5.3.2 Structural Analysis of the 5'-Flanking Region of the Human Granulin Gene.

The nucleotide sequence of the 5'-flanking and untranslated region of the human grn gene was obtained from two overlapping genomic clones, HGL1 and HGLP26 (see sections 5.2.1 and 5.2.2), and is shown in Figure 5.4B. Analysis of the 5'-flanking and untranslated sequence revealed several interesting structural features.

Firstly, there is no apparent TATA box. Although one sequence of TATAAA is present 2444 nucleotides upstream of the translation initiation codon, it is too far upstream to be involved in the initiation of transcription. In mammals, TATA boxes are usually located a short distance (25-30 bp) upstream of the transcription initiation site (269). A variant TATA sequence, TATATT is present at nucleotides -148 to -143 [the nucleotides of the 5' untranslated and flanking region are numbered such that -1 corresponds to the nucleotide preceding the translation initiation codon (ATG)], however it lies within the 5' untranslated region of the grn cDNA and most likely does not function as a TATA box. The absence of a canonical TATA box is consistent with the observation of multiple initiation sites in the grn gene (see section 5.4), since the

TATA box functions to ensure accurate initiation of transcription at a single site (270).

The 5'-flanking region contains 4 CCAAT boxes, at nucleotides -103 to -99; -133 to -129; -251 to -247; and -398 to -394. The CCAAT box at nucleotides -133 to -129 is conserved between the human and mouse genes (see section 5.3.3).

The grn gene promoter has a G+C content of 56.4% over 500 bp of 5'-flanking and untranslated sequence. This value is lower than the G+C content of some of the better characterized growth factor and growth factor receptor gene promoters. For example, a G+C content of 80% (nucleotides -262 to -1), 76%, and 78.6% (over the reported 1209 nucleotides of 5' sequence) is observed in the promoters of the TGF-ß1 (271), TGF-α (272), and PDGF-A chain (273) genes respectively, while the EGF receptor (274) and insulin receptor (275) gene promoters have a G+C content of 88% (nucleotides -540 to -1) and 77% (nucleotides -371 to +1) respectively. The human genome has an average G+C content of 40% (276).

A remarkable number of putative binding sites for transcriptional factors are present, in direct or in reverse orientation, in the 5'-flanking region of the human grn gene (Figure 5.4B). The locations of these putative response elements are presented in Table 5.3.

Interestingly multiple sets of potential binding sites for nuclear factors with only one mismatch to its consensus sequence (such sites are sometimes reported in the literature as possible binding sites for transcription factors) are present in the 5'-flanking region of the grn gene. For example, there are 25 AP1-like (13 complement, 12 reverse complement), 112 AP2-like (66 complement, 46 reverse complement), 40 SP1-like (18 complement, 22 reverse complement), 4 CRE/ATF-like

Table 5.3. Putative response elements in the 5' untranslated and flanking region of the human granulin gene.

Transcription Factor	Consensus (5'-3')	Location (nucleotide)	Comments	Reference
SP1	G/TG/AGGCG/TG/A G/AG/T	-87 to -79; -545 to -537; -668 to -660	Originally described as a specific factor required for SV-40 transcription, now known to play an important role in transcription of many viral and mammalian genes. Presence often correlated with absence of a TATA box in housekeeping genes and genes involved in growth control.	277,278
AP2	CCCA/CNG/CG/CG /C	-33 to -25; -74 to -67; -839 to -832; -1247 to -1240; -1341 to -1334; -1495 to -1488; -1648 to -1641; -1654 to -1647; -1734 to -1727.	A stimulatory transcription factor involved in the regulation of many eukaryotic genes. Developmentally regulated and induced by phorbol esters tumor promoters and cAMP.	279,280
PEA3	AGGAAG/A	-213 to -208; -308 to -303; -1018 to -1013; -1127 to -1122; -1379 to -1372; -1841 to -1836; -2544 to -2539.	Originally shown to bind the polyoma virus enhancer, and whose activity is regulated by phorbol esters and serum components and by a number of oncogenes such as v-*src*, polyoma middle-T, c-Ha-*ras*, v-*mos*, and v-*raf*.	281,282
CFI	ANATGG	-232 to -227; -626 to -621; -1258 to -1253; -1392 to -1387; -1415 to -1410; -1509 to -1504;	A ubiquitous nuclear factor initially shown to act on c-*myc*, IgH- and α-actin promoters.	283

Transcription Factor	Consensus (5'-3')	Location (nucleotide)	Comments	Reference
C/EBP	GTGGT/AT/AT/AG	-1167 to -1160; -2063 to -2059; -2161 to -2157	A nuclear factor presumed to have a role in energy metabolism and may have a role in regulating the balance between cell growth and differentiation.	284
FACT-1	TGGCGA	-466 to -461; -2138 to -2133.	A ubiquitous transcription factor that binds to the serum response element.	285
TEF-2	GGGTGTGG	-227 to -220	An enhancer factor that binds to the GT-IC motif on the SV40 promoter.	286
NF-IL6	TT/GNNGNAAT/G	-311 to -303; -1020 to -1012; -1363 to -1356; -2476 to -2468; -2828 to -2820; -2872 to -2864.	A nuclear factor involved in acute phase reactions, inflammation, and hematopoiesis.	287
IL6-RE	CTGGGA	-559 to -554; -840 to -835; -1199 to -1194; -1248 to -1243.	A well conserved sequence in cytokine-responsive regions of acute phase genes.	288
TIE	GNNTTGGNGN	-1010 to -1001; -1055 to -1046	A conserved motif in the promoter region of TGF-β1 inhibited genes.	289
MTF1	TGCA/GCNC	-478 to -472; -572 to -566; -2761 to -2755.	A transcription factor induced by heavy metals.	290

Transcription Factor	Consensus (5'-3')	Location (nucleotide)	Comments	Reference
GCF	C/GCGC/GC/GC/GC	-571 to -565; -715 to -709.	A negative regulator of the EGFR gene and other growth factor and growth factor receptor genes.	291
SIF	CCCGTC/A	-614 to -609; -2583 to -2578.	A ubiquitous nuclear factor that activates c-*fos* gene expression and is induced by c-*sis*/PDGF.	292
HAPF1	CTGGA/GAA	-2561 to -2555.	An IL-6 induced nuclear factor that cooperates with NF-IL6.	293
NFGMa	GA/GGA/GTTT/GCA T/C	-621 to -612; -1162 to -1153.	A nuclear protein that acts on cytokine gene promoters, and is induced by PMA and ConA.	294
MYB	T/CAACG/TG	-1107 to -1102; -1211 to -1206; -2434 to -2429; -2647 to -2642; -2667 to -2662.	A hematopoietic and tumor cell line specific factor that plays a critical role in cell proliferation and differentiation.	295
GATA	T/AGATAA/G	-143 to -138; -353 to -348; -997 to -992; -1853 to -1848; -2651 to -2646; -2780 to -2775.	The GATA family of transcription factors are involved in erythroid differentiation.	296

Transcription Factor	Consensus (5'-3')	Location (nucleotide)	Comments	Reference
PPAR	AGGTCA	-655 to -650; -2229 to -2224.	A member of the steroid hormone receptor superfamily that is thought to play a role in the development of liver tumors.	297
ETS-1	G/CA/CGGAA/TGT/C	-214 to -207; -1128 to -1121.	A nuclear oncoprotein that is believed to be a component of the cell signalling network.	298
LYF-1	C/TC/TTGGGAGA/G	-683 to -675	A lymphocyte specific transcriptional regulator that may be a member of the *Ets* family.	299
E2A	G/ACAGNTG	-291 to -285, -1218 to -1212; -2668 to -2662	A ubiquitous transcription factor that binds "E" boxes	300

(2 complement, 2 reverse complement) (301) and 78 NFIL6-like (40 complement, 38 reverse complement) potential binding sites respectively in the 5'-flanking region of the grn gene. These potential binding sites appear to be randomly distributed over the 2878 bp of 5'-flanking and untranslated region of the human grn gene.

In addition to the better characterized putative response elements present upstream of the translation initiation site, there are multiple copies of palindromic sequences as well as direct and inverted repeats in the 5'-flanking region. Such sequences have often been associated with enhancer and/or repressor activity of gene expression (302). Two noteworthy features are i) an adjacent 11-bp direct repeat, GAATAGGGCAG, followed by 8 repeats in tandem with the consensus sequence C/TAG/AGG/A occurring at nucleotides -928 to -871, and ii) an "antisense" sequence (5' CTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGG 3') at nucleotides -552 to -513 with a 93% match (38/41 nucleotides) to a 41 bp sequence (5' CCCGGGTTCAAGCGATCTCCTGCCTCGGCCTCCTTGAGTAG 3') at nucleotides -2388 to -2350. A search of nucleic acid and protein databases using the BLAST E-Mail server offered by the National Library of Medicine, NIH (229), revealed that the rat sarcoma virus v-*ras* oncogene contains a stretch of nucleotides, occurring in reverse complement, highly similar to the C/TAG/AGG/A repeats in the promoter of the grn gene. There are 7 1/2 repeats with the consensus sequence T/CCT/CTA/G occurring in the 5'-flanking region of the rat sarcoma virus v-*ras* oncogene (303). The significance of this motif in the v-*ras* oncogene is not known at present. The database search also revealed that the 41 bp "antisense" repeats present in the 5'-flanking region of the grn gene represents a member of the Alu family of repeated

sequences. Similar Alu sequences to the one found in the grn gene are present in approximately 2000 genes stored in the databases, including, the erythropoietin and tumor necrosis receptor genes (304,305), the transformation-associated protein p53 (306), and the breast cancer-associated antigen (DF3) gene (307). The Alu sequence family accounts for a minimum of 3-6% of the human genome (308). The function of the Alu dispersed repeats in the mammalian genome is not known at present.

Also denoted in the grn gene are 12 homopurine/ homopyrimidine tracts (i.e. regions composed of one strand containing primarily purine residues and the complementary strand containing primarily pyrimidine residues) randomly scattered along the 5'-flanking sequence (see legend of Figure 5.4). Homopurine/ homopyrimidine tracts have been shown to exhibit S1 nuclease sensitivity (309) and are recognized by specific transcription factors (310,311) involved in the regulation of other essential growth control genes like cKi-*ras* (312), EGF-R (313), and c-*myc* (314,315).

5.3.3 Comparison of the Human and Mouse Granulin Gene 5' Untranslated and Flanking Region.

A comparison of the 5' untranslated and flanking sequence of the human grn gene with the published mouse grn gene sequence (268) shows a 73% identity between the 5' untranslated and flanking regions of the human and mouse grn genes over a region of 300 bp 5' of the translation start codon (Figures 5.6 and 5.7). Analysis of the nucleotide sequence in this region reveals several putative binding

Figure 5.6 Sequence identity between the 5'-flanking and untranslated region of the human and mouse grn gene. Also shown is an intron in the untranslated region (intron 0). The translation initiation codon (ATG) is boxed. The numbering of nucleotides is based on the initiator codon with the first nucleotide preceding the A residue of the translation initiation codon corresponding to nucleotide -1. The conserved CCAAT box (nucleotide -129 to -133, human sequence) is shaded, and the almost conserved CCAAT boxes are overlined. Putative response elements that are conserved between the human and mouse are boxed and shaded, while almost conserved putative response elements are double over- and under-lined. The conserved polypyrimidine-polypurine stretches are boxed. The alignment of the 5'-flanking region of the human and mouse grn gene was performed using the program ALIGN (DNAstar Inc., Madison, WI)).

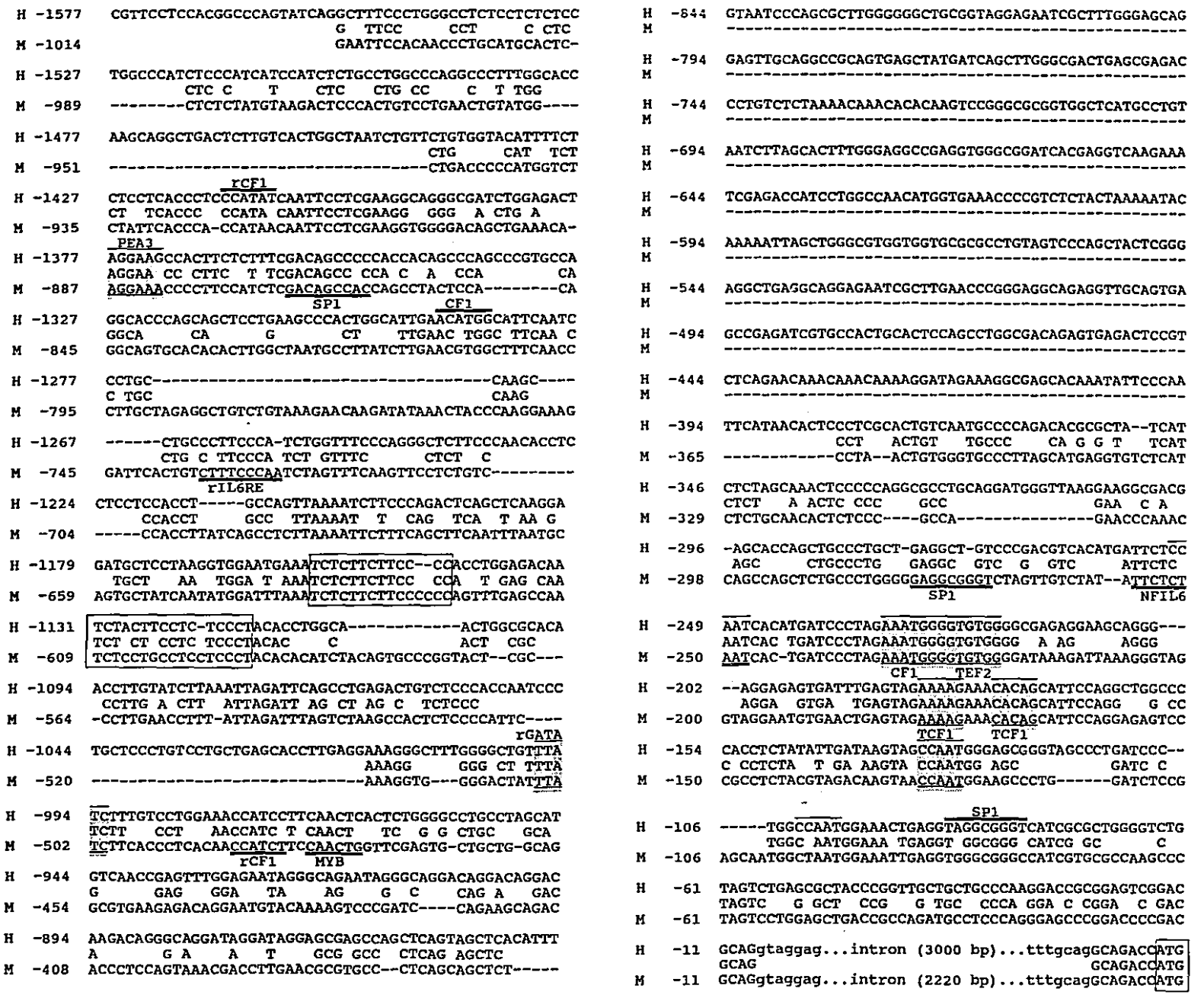
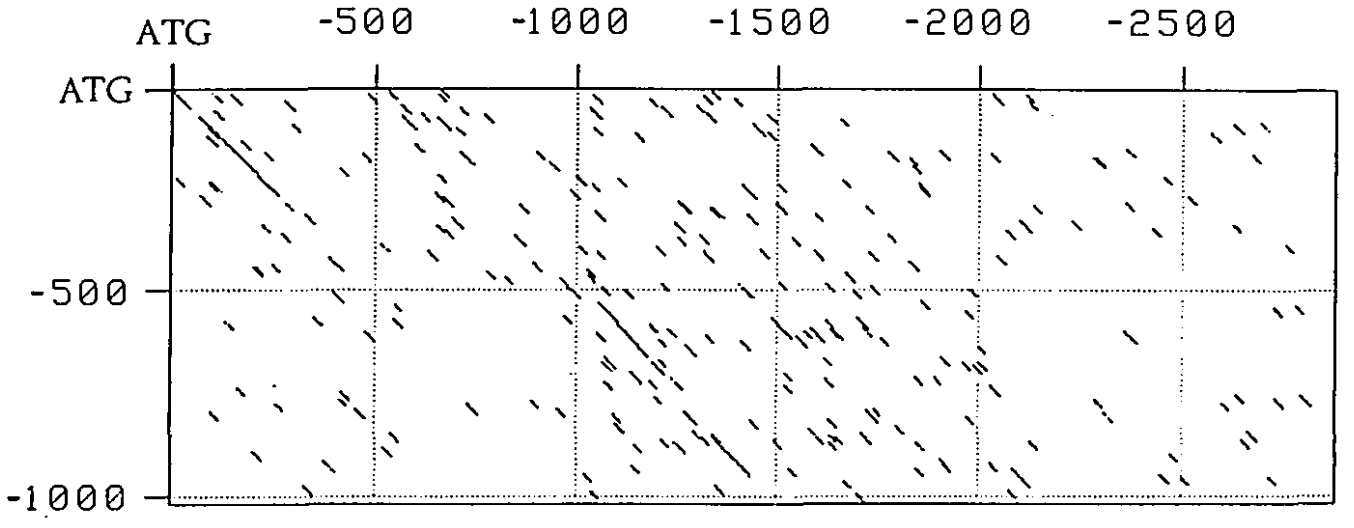

H -1577 CGTTCCTCCACGGCCCAGTATCAGGCTTTCCCTGGGCCTCTCCTCTCTCC
G TTCC CCT C CTC
M -1014 GAATTCCACAACCCTGCATGCACTC-
H -1527 TGGCCCATCTCCCATCATCCATCTCTGCCTGGCCCAGGCCCTTTGGCACC
CTC C T CTC CTG CC C T TGG
M -989 --------CTCTCTATGTAAGACTCCCACTGTCCTGAACTGTATGG----
H -1477 AAGCAGGCTGACTCTTGTCACTGGCTAATCTGTTCTGTGGTACATTTTCT
CTG CAT TCT
M -951 ----------CTGACCCCCATGGTCT
rCF1
H -1427 CTCCTCACCCTCCCATATCAATTCCTCGAAGGCAGGGCGATCTGGAGACT
CT TCACCC CCATA CAATTCCTCGAAGG GGG A CTG A
M -935 CTATTCACCCA-CCATAACAATTCCTCGAAGGTGGGGACAGCTGAAACA-
PEA3
H -1377 AGGAAGCCACTTCTCTTTCGACAGCCCCCACCACAGCCCAGCCCGTGCCA
AGGAA CC CTTC T TCGACAGCC CCA C A CCA CA
M -887 AGGAAACCCCTTCCATCTCGACAGCCACCAGCCTACTCCA--------CA
SP1 CF1
H -1327 GGCACCCAGCAGCTCCTGAAGCCCACTGGCATTGAACATGGCATTCAATC
GGCA CA G CT TTGAAC TGGC TTCAA C
M -845 GGCAGTGCACACACTTGGCTAATGCCTTATCTTGAACGTGGCTTTCAACC
H -1277 CCTGC----------CAAGC----
C TGC CAAG
M -795 CTTGCTAGAGGCTGTCTGTAAAGAACAAGATATAAACTACCCAAGGAAAG
H -1267 ------CTGCCCTTCCCA-TCTGGTTTCCCAGGGCTCTTCCCAACACCTC
CTG C TTCCCA TCT GTTTC CTCT C
M -745 GATTCACTGTCTTTCCCAATCTAGTTTCAAGTTCCTCTGTC---------
rIL6RE
H -1224 CTCCTCCACCT-----GCCAGTTAAAATCTTCCCAGACTCAGCTCAAGGA
CCACCT GCC TTAAAAT T CAG TCA T AA G
M -704 -----CCACCTTATCAGCCTCTTAAAATTCTTTCAGCTTCAATTTAATGC
H -1179 GATGCTCCTAAGGTGGAATGAAATCTCTTCTTCC--CCACCTGGAGACAA
TGCT AA TGGA T AAATCTCTTCTTCC CCA T GAG CAA
M -659 AGTGCTATCAATATGGATTTAAATCTCTTCTTCCCCCCAGTTTGAGCCAA
H -1131 TCTACTTCCTC-TCCCTACACCTGGCA----------ACTGGCGCACA
TCT CT CCTC TCCCTACAC C ACT CGC
M -609 TCTCCTGCCTCCTCCCTACACACATCTACAGTGCCCGGTACT--CGC---
H -1094 ACCTTGTATCTTAAATTAGATTCAGCCTGAGACTGTCTCCCACCAATCCC
CCTTG A CTT ATTAGATT AG CT AG C TCTCCC
M -564 -CCTTGAACCTTT-ATTAGATTTAGTCTAAGCCACTCTCCCCATTC----
rGATA
H -1044 TGCTCCCTGTCCTGCTGAGCACCTTGAGGAAAGGGCTTTGGGGCTGTTTA
AAAGG GGG CT TTTA
M -520 ----------AAAGGTG---GGGACTATTTA
H -994 TCTTTGTCCTGGAAACCATCCTTCAACTCACTCTGGGGCCTGCCTAGCAT
TCTT CCT AACCATC T CAACT TC G G CTGC GCA
M -502 TCTTCACCCTCACAACCATCTTCCAACTGGTTCGAGTG-CTGCTG-GCAG
rCF1 MYB
H -944 GTCAACCGAGTTTGGAGAATAGGGCAGGAATAGGGCAGGACAGGACAGGAC
G GAG GGA TA AG G C CAG A GAC
M -454 GCGTGAAGAGACAGGAATGTACAAAAGTCCCGATC----CAGAAGCAGAC
H -894 AAGACAGGGCAGGATAGGATAGGAGCGAGCCAGCTCAGTAGCTCACATTT
A G A A T GCG GCC CTCAG AGCTC
M -408 ACCCTCCAGTAAACGACCTTGAACGCGTGCC--CTCAGCAGCTCT-----
H -844 GTAATCCCAGCGCTTGGGGGGCTGCGGTAGGAGAATCGCTTTGGGAGCAG
M ----------
H -794 GAGTTGCAGGCCGCAGTGAGCTATGATCAGCTTGGGCGACTGAGCGAGAC
M ----------
H -744 CCTGTCTCTAAAACAAACACACAAGTCCGGGCGCGGTGGCTCATGCCTGT
M ----------
H -694 AATCTTAGCACTTTGGGAGGCCGAGGTGGGCGGATCACGAGGTCAAGAAA
M ----------
H -644 TCGAGACCATCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATAC
M ----------
H -594 AAAAATTAGCTGGGCGTGGTGGTGCGCGCCTGTAGTCCCAGCTACTCGGG
M ----------
H -544 AGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGA
M ----------
H -494 GCCGAGATCGTGCCACTGCACTCCAGCCTGGCGACAGAGTGAGACTCCGT
M ----------
H -444 CTCAGAACAAACAAACAAAAGGATAGAAAGGCGAGCACAAATATTCCCAA
M ----------
H -394 TTCATAACACTCCCTCGCACTGTCAATGCCCCAGACACGCGCTA--TCAT
CCT ACTGT TGCCC CA G G T TCAT
M -365 ----------CCTA--ACTGTGGGTGCCCTTAGCATGAGGTGTCTCAT
H -346 CTCTAGCAAACTCCCCCAGGCGCCTGCAGGATGGGTTAAGGAAGGCGACG
CTCT A ACTC CCC GCC GAA C A
M -329 CTCTGCAACACTCTCCC----GCCA----------GAACCCAAAC
H -296 -AGCACCAGCTGCCCTGCT-GAGGCT-GTCCCGACGTCACATGATTCTCC
AGC CTGCCCTG GAGGC GTC G GTC ATTCTC
M -298 CAGCCAGCTCTGCCCTGGGGGAGGCGGGTCTAGTTGTCTAT--ATTCTCT
SP1 NFIL6
H -249 AATCACATGATCCCTAGAAATGGGGTGTGGGGCGAGAGGAAGCAGGG---
AATCAC TGATCCCTAGAAATGGGGTGTGGGG A AG AGGG
M -250 AATCAC-TGATCCCTAGAAATGGGGTGTGGGGATAAAGATTAAAGGGTAG
CF1 TEF2
H -202 --AGGAGAGTGATTTGAGTAGAAAAGAAACACAGCATTCCAGGCTGGCCC
AGGA GTGA TGAGTAGAAAAGAAACACAGCATTCCAGG G CC
M -200 GTAGGAATGTGAACTGAGTAGAAAAGAAACACAGCATTCCAGGAGAGTCC
TCF1 TCF1
H -154 CACCTCTATATTGATAAGTAGCCAATGGGAGCGGGTAGCCCTGATCCC--
C CCTCTA T GA AAGTA CCAATGG AGC GATC C
M -150 CGCCTCTACGTAGACAAGTAACCAATGGAAGCCCTG------GATCTCCG
SP1
H -106 -----TGGCCAATGGAAACTGAGGTAGGCGGGTCATCGCGCTGGGGTCTG
TGGC AATGGAAA TGAGGT GGCGGG CATCG GC C
M -106 AGCAATGGCTAATGGAAATTGAGGTGGGCGGGCCATCGTGCGCCAAGCCC
H -61 TAGTCTGAGCGCTACCCGGTTGCTGCTGCCCAAGGACCGCGGAGTCGGAC
TAGTC G GCT CCG G TGC CCCA GGA C CGGA C GAC
M -61 TAGTCCTGGAGCTGACCGCCAGATGCCTCCCAGGGAGCCCGGACCCCGAC
H -11 GCAGgtaggag...intron (3000 bp)...tttgcagGCAGACCATG
GCAG GCAGACCATG
M -11 GCAGgtaggag...intron (2220 bp)...tttgcagGCAGACCATG

Figure 5.7 Dot matrix analysis of the 5' flanking sequence of the human (X axis) and mouse (Y axis) grn gene. The nucleotide sequence of the mouse 5'-flanking region was obtained from Baba et.al. (268). The numbering of nucleotides is based on the initiator codon. Each dot represents a 65% match (stringency) within a 20-nucleotide stretch (window size). The translation initiation site (ATG) is indicated. The dot matrix analysis was performed using the program DOTPLOT (DNAstar Inc., Madison, WI)


ATG
-500
-1000
-1500
-2000
-2500
ATG
-500
-1000

sites for transcription factors that are conserved or almost conserved between the two genes. These include partially conserved sites for the ubiquitous nuclear factors SP1 (nucleotides -87 to -79, human sequence) and NFIL6 (nucleotides -256 to -257), and entirely conserved binding sites for CF1 (nucleotides -232 to -226) as well as TCF1 (nucleotides -175 to -171, and -183 to -179), a T-cell transcription factor distally related to the *Ets* family of transcription factors (316) and TEF-2 (nucleotides -220 to -227). Also conserved within this region is a CCAAT box at nucleotides -133 to -129.

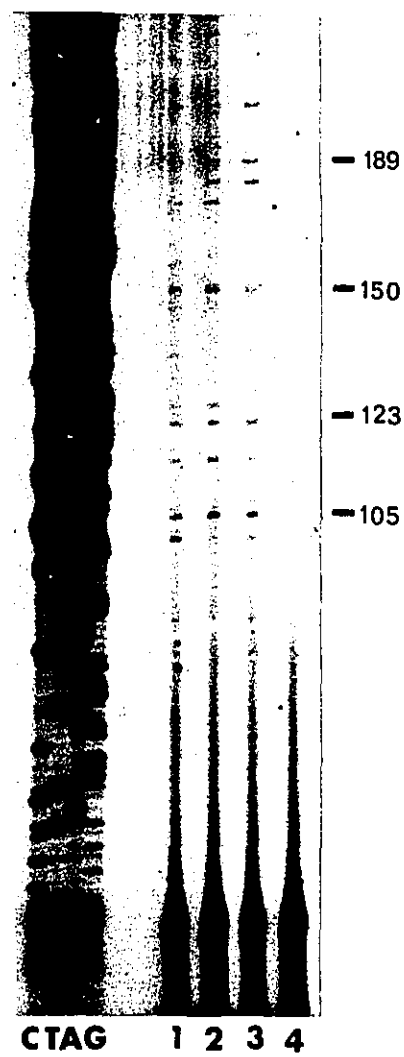
A second region of identity between the 5'-flanking sequence of the human and mouse genes occurs between nucleotides -950 to -1430 of the human grn gene and nucleotides -460 to -940 of the mouse grn gene respectively (Figures 5.6 and 5.7). Within this region, a conserved potential binding site for the ubiquitous transcription factor PEA3 is present (nucleotides -1377 to -1372) as well as two conserved short homopurine/homopyrimidine tracts.

Interestingly, while the 5'-flanking regions of both the human and mouse grn genes contain a remarkable number of sequences that could serve as potential binding sites for nuclear factors, most of these potential binding sites are not conserved between the human and mouse genes, and may represent regulatory sequences specific to the human or mouse promoter involved in the species specific expression of the grn gene (see Chapter 4). Alternatively, they may be spurious non-functional sequence elements. Further experiments involving DNase footprinting and gel mobility shift assays will be required to determine if some or all these sequence elements play a role in the expression and regulation of the grn gene.

5.4 Transcription Initiation Sites in the Human Granulin Gene.

In order to identify the probable transcription start site(s) of the human grn gene, primer extension reactions were employed. Two different primers, a 35-mer specific to the first coding exon (exon I; nucleotides 5 to 39 with respect to the translation start site), and a 17-mer specific to the 5'untranslated region (exon 0; nucleotides -35 to -51) were annealed to total RNA from human adult and fetal intestine and U937 cells, which express high levels of grn mRNA (see Chapter 4), and extended with reverse transcriptase (see Materials and Methods, section 2.9 for experimental details). The ^{32}P end- labelled extended products were separated by polyacrylamide gel electrophoresis and visualized by autoradiography. A representative primer extension reaction is shown in Figure 5.8. With both primers, multiple extension products were obtained. A control reaction using tRNA did not result in any extension products. A similar pattern of extended products were obtained over two separate experiments. The most abundant extension products had approximate (+/- 2 bp) lengths of 79, 116, 119, 125, 147, 168, 172, 205, and 253 bp with the 35-nt primer, and 71, 74, 82, 89, 92, 116, 147, 155, 176, 181, and 230 bp with the 17-nt primer. The 17-nt primer is complementary to a sequence 34 bp upstream of the sequence recognised by the 35-nt primer. Therefore, authentic grn-specific primer extended products obtained from the 35-nt primer should be 52 bp longer than the extended products obtained from the 17-nt primer. Accordingly, four extension products obtained with the 17-nt primer of lengths 74, 92, 116 and 155 bp, are in close agreement with those obtained with the 35-nt primer of lengths 125, 147, 168 and 205 bp. These results therefore suggest that the initiation site of

Figure 5.8 Representative primer extension analysis to determine transcription initiation sites of the human grn gene. Total RNA from human adult (lane 1) and fetal (lane 2) intestine, and U937 cells (lane 3), and tRNA (lane 4) was hybridized with a 17 nucleotide primer corresponding to nucleotides -35 to -51, and extended with reverse transcriptase as described in Materials and Methods, section 2.9. Primer-extended products were analyzed on a 7M urea/ 8% polyacrylamide gel. A sequencing ladder was generated from a *Sac*I subclone of HGLP26 (Figure 5.2, fragment E) using the same primer was run in parallel as a size standard. Transcription initiation sites that were also obtained with a different primer (see section 5.4) are marked. Numbers on right indicate the putative transcription start sites of the human grn gene.

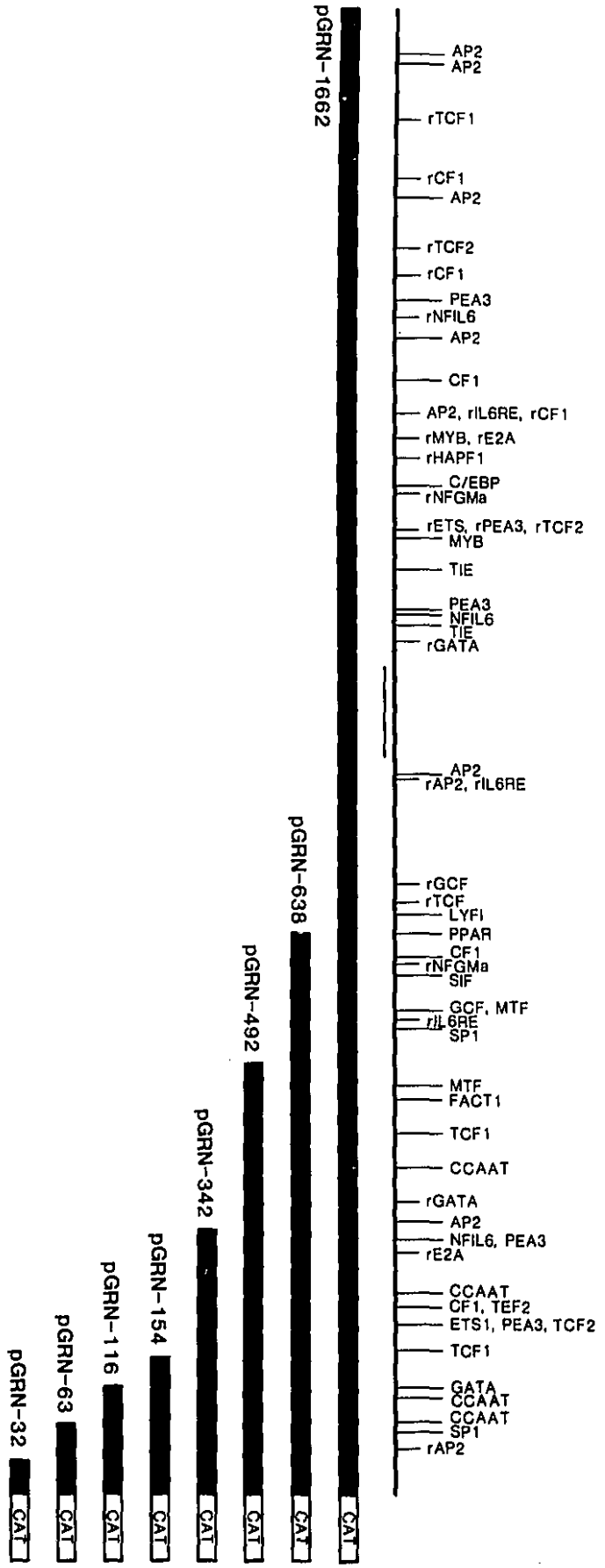

189
150
123
105
CTAG
1 2 3 4

grn gene transcription is heterogenous, occurring approximately 105, 123, 150, and 189 bp upstream of the initiator ATG codon. It should be noted that not all the extended products obtained with the 17-nt and 35-nt primers match. These extended products are probably a result of premature termination of the reverse transcriptase, or non-(grn)specific annealing of the primers to the RNA. Unfortunately, attempts to further confirm the transcription start sites by SI nuclease mapping were not successful. Because the grn gene initiates transcription from multiple sites, we have numbered the 5'-flanking region of the grn gene such that -1 refers to the nucleotide preceding the translation initiation codon (ATG).

5.5 Activity of the Human Granulin Gene Promoter.

To determine if the 5'-flanking region of the human grn gene could serve to direct gene transcription, several fragments containing varying lengths of the 5'-flanking region of the grn gene were cloned into the chloramphenicol acetyltransferase reporter plasmid, pBLCAT3. A similar plasmid, pBLCAT2, containing the thymidine kinase (TK) promoter was used as a positive control. All chimeric grn/CAT constructs contained the same 3' end (nucleotide -21, Figure 5.4B) but varied in length at the 5' end. Figure 5.9 depicts the structure of the 5'-flanking region of the human grn gene containing the putative response elements, and a schematic of the grn-CAT constructs used in the study of the 5'-flanking region of the human grn gene. The chimeric constructs are named according to the length of the cloned 5'-flanking region. It should be emphasized that the purpose of this study was to determine whether the 5'-flanking region of the grn gene could serve to drive

Figure 5.9 Schematic representation of the human Grn-CAT chimeric plasmids. The chimeric grn-CAT plasmids were generated as described in Materials and Methods, section 2.10.3. The relative positions of the CCAAT boxes and putative response elements are shown above the grn-CAT constructs. (r) indicates putative response element occurring in reverse orientation. Solid bar represents the eight bp tandem repeat (see text). All constructs contain a common 3'-end (nucleotide -21) and different 5'-ends. The name of the constructs reflect their length in bp.


pGRN-1662
pGRN-638
pGRN-492
pGRN-342
pGRN-154
pGRN-116
pGRN-63
pGRN-32
CAT
AP2
AP2
rTCF1
rCF1
AP2
rTCF2
rCF1
PEA3
rNFIL6
AP2
CF1
AP2, rIL6RE, rCF1
rMYB, rE2A
rHAPF1
C/EBP
rNFGMa
rETS, rPEA3, rTCF2
MYB
TIE
PEA3
NFIL6
TIE
rGATA
AP2
rAP2, rIL6RE
rGCF
rTCF
LYF1
PPAR
CF1
rNFGMa
SIF
GCF, MTF
rIL6RE
SP1
MTF
FACT1
TCF1
CCAAT
rGATA
AP2
NFIL6, PEA3
rE2A
CCAAT
CF1, TEF2
ETS1, PEA3, TCF2
TCF1
GATA
CCAAT
CCAAT
SP1
rAP2

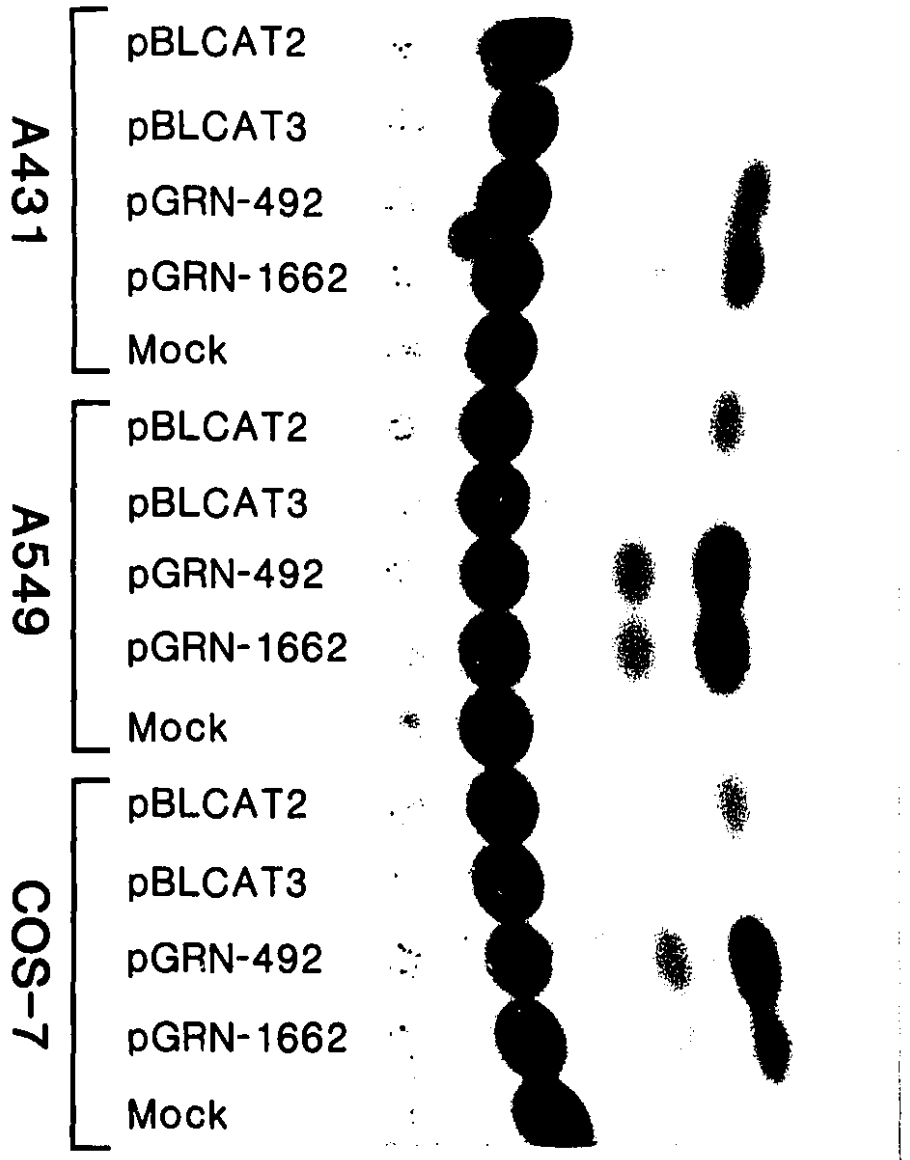
transcription, and to delineate a functional promoter of the human grn gene. It was not our intention to analyze the functional elements of the grn gene in complete detail. Therefore, the grn-CAT constructs were made to contain increasing lengths of 5'-flanking and untranslated region of the human grn gene.

Initial experiments using two constructs, pGRN-492 and pGRN-1662 (see Materials and Methods, sections 2.11.2 and 2.12 for experimental details) showed that the cloned 5'-flanking region of the human grn gene could drive the transcription of the bacterial chloramphenicol acetyltransferase gene in A431 (human epidermoid carcinoma), A549 (human lung carcinoma) and COS-7 (monkey kidney fibroblast-like) cells (Figure 5.10). This result confirms that the 5'-flanking region can serve as a functional promoter for the grn gene and that it can act in cells of different species and origin.

To further delineate the promoter of the human grn gene, two different human cell lines, one of epithelial origin (A549 lung carcinoma cells) and the other of hematopoietic origin (SWEI B-lymphoblastoid cells), were used for analysis of promoter activity based on their expression of high levels of grn mRNA (see Chapter 4, section 4.1.1). Eight different grn-CAT constructs, containing from 32 to 1662 bp of 5'-flanking sequence were used for the promoter studies (Figure 5.9).

The various chimeric grn/CAT constructs were transiently transfected into A549 and SWEI cells by the liposome mediated method, and CAT activity in the cell extracts determined 48 hrs post-transfection (see Materials and Methods, sections 2.11.2 and 2.12 for experimental details). A representative CAT analysis as well as the relative CAT activities of the grn-CAT constructs and the control plasmids,

Figure 5.10 The 5' flanking region of the human grn gene serves as a functional promoter in cells of different species and origin. Figure shows an autoradiogram of a TLC plate demonstrating CAT activity. The upper radioactive spots corresponds to di-acetylated chloramphenicol, the middle spots correspond to monoacetylated chloramphenicol and the lower spots to unreacted chloramphenicol. A431 (human epidermoid carcinoma of the vulva), A549 (human lung carcinoma), and COS-7 (monkey kidney fibroblast-like) cells were transfected with the plasmids pBLCAT2, pBLCAT3, pGRN-492, and pGRN-1662 and CAT activity was determined 48 hr post-transfection as described in Materials and Methods, sections 2.11.2 and 2.12. pBLCAT2 is a CAT plasmid containing the thymidine kinase (TK) promoter, while pBLCAT3 is a promoterless CAT construct. pGRN-492 and pGRN-1662 are CAT plasmids containing 492 and 1662 bp respectively, of the 5'-flanking region of the human grn gene. Mock shows CAT-activity in lysates of non-transfected cells.


A431
pBLCAT2
pBLCAT3
pGRN-492
pGRN-1662
Mock
A549
pBLCAT2
pBLCAT3
pGRN-492
pGRN-1662
Mock
COS-7
pBLCAT2
pBLCAT3
pGRN-492
pGRN-1662
Mock

pBLCAT2 and pBLCAT3, are shown in Figures 5.11 and 5.12 respectively.

In A549 cells, a lung carcinoma cell line, the strongest promoter activity was observed with pGRN-492 (Figure 5.12A). This construct contains the transcription initiation sites, CCAAT boxes as well as other structural features, including putative binding sites for several transcription factors such as SP1, AP2 and PEA3 (Figures 5.4B and 5.9). An increase in the length of the 5'-flanking region (pGRN-638) results in a slight decrease in CAT activity. Interestingly, located within this region (-513 to -659) is a putative response element for the GC factor, a negative regulator of growth factor and growth factor receptor genes (291). A further increase in the 5'-flanking region (pGRN-1662) results in a significant drop in transcriptional activity. This result suggests negative regulatory regions occur between nucleotides -659 and -1683. Interestingly, the eight 5 bp-repeat occurring in tandem similar to a sequence found in the 5'-flanking region of the rat sarcoma virus v-*ras* oncogene (303), lies within this region. Deleting the 5'-flanking region down to nucleotide -127 (pGRN-342, pGRN-154, and pGRN-116) results in a gradual decrease in CAT activity (Figure 5.12A). The transcriptional activity of pGRN-32 and pGRN-63 were not significantly different from the promoterless control plasmid, pBLCAT3 (Figure 5.12A). This is consistent with the observation that pGRN-32 and pGRN-63 do not contain any of the transcriptional start sites or any known putative response elements. The lack of CAT activity in pGRN-32 and pGRN-63, and demonstration of transcriptional activity in pGRN-116 and pGRN-154, would indicate that the 5'-flanking sequence up to nucleotides -137 to -175 is sufficient to confer grn promoter activity in A549 cells. This minimal region required for promoter activity contains 3 of the 4 transcriptional

Figure 5.11 Raw data of a representive experiment showing promoter activity in A549 cells. The transfection and CAT assays were performed as described in Materials and Methods, sections 2.11.2. and 2.12. respectively.

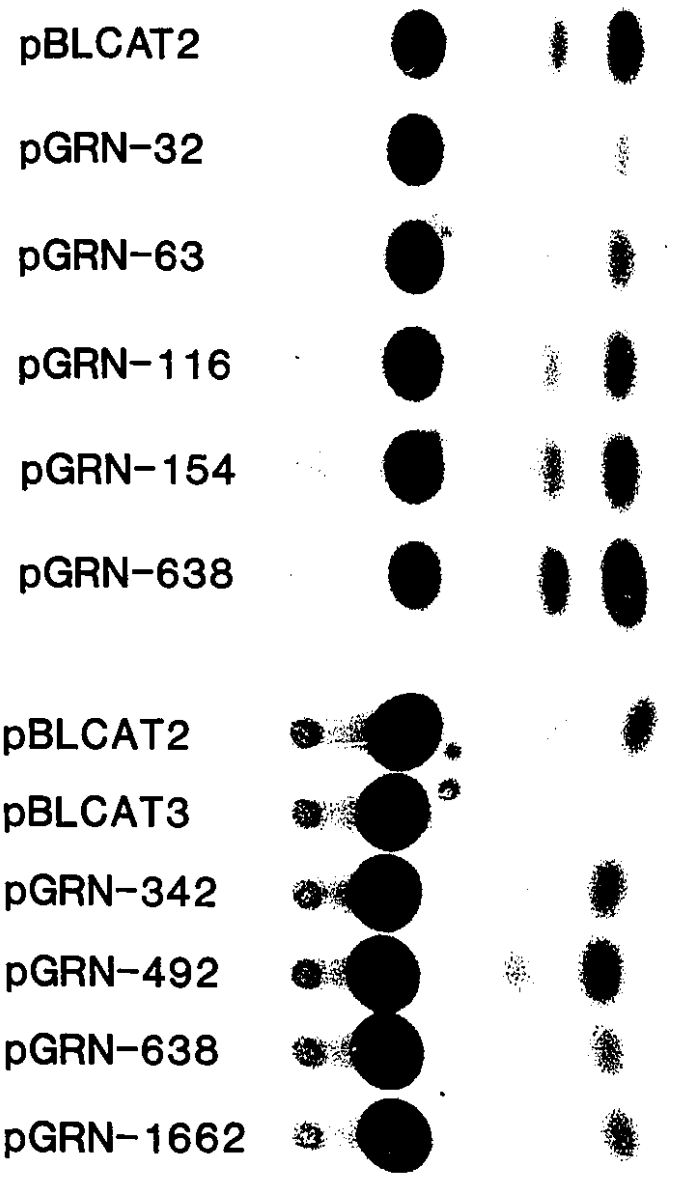
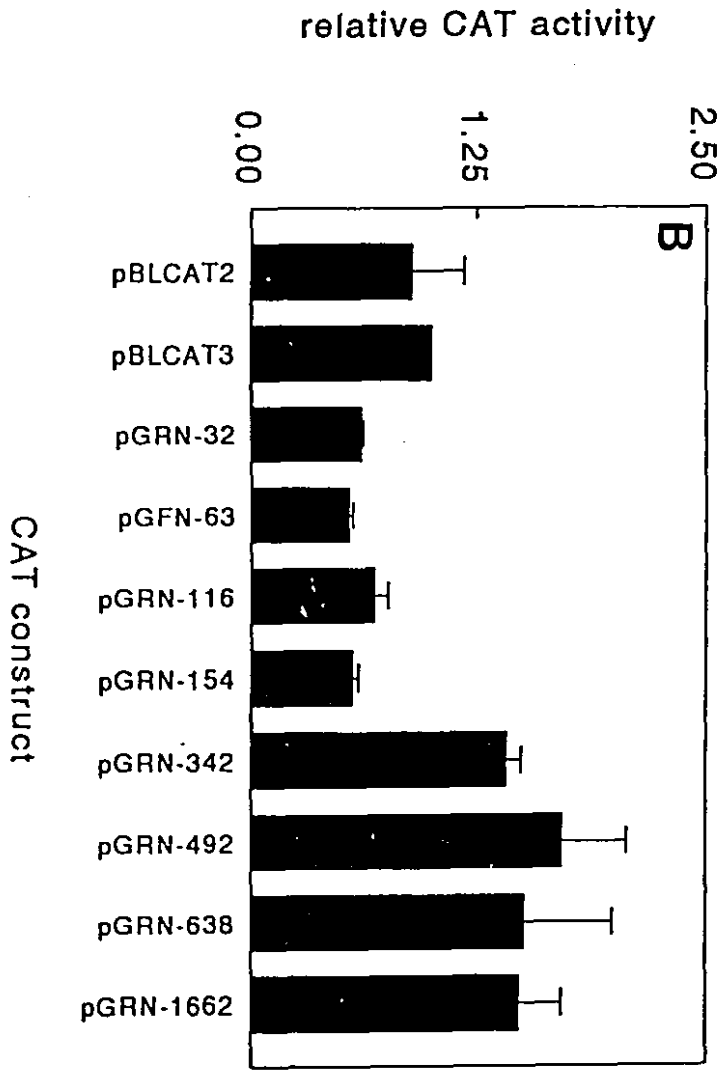
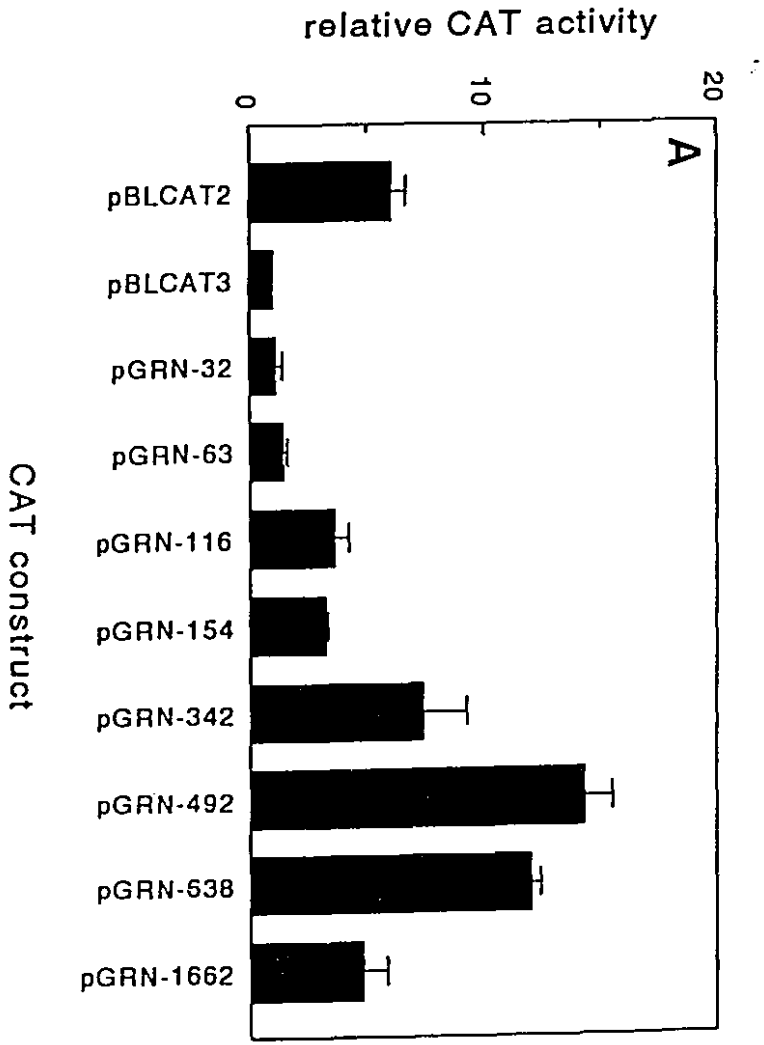

pBLCAT2
pGRN-32
pGRN-63
pGRN-116
pGRN-154
pGRN-638
pBLCAT2
pBLCAT3
pGRN-342
pGRN-492
pGRN-638
pGRN-1662

Figure 5.12 Graphical presentation of CAT activity in cell lysates of (A) A549 and (B) Swei cells. Cells were transfected with the grn-CAT chimeric plasmids indicated with the Lipofectin™ reagent as described in Materials and Methods, section 2.11.2., and CAT activity determined 48-52 hrs after transfection (Materials and Methods, section 2.12.). pBLCAT2 is a CAT plasmid containing the thymidine kinase (TK) promoter, while pBLCAT3 is a promoterless CAT construct. Values are given as a percentage of ^{14}C-chloramphenicol conversion per µg protein, relative to the CAT-activity of pBLCAT3. Bars represent the mean $\pm$ SD of three to four separate experiments. Note that the TK promoter containing pBLCAT2 does not give any significant CAT activity in Swei cells (see text for discussion).


relative CAT activity
0.00
1.25
2.50
B
pBLCAT2
pBLCAT3
pGRN-32
pGFN-63
pGRN-116
pGRN-154
pGRN-342
pGRN-492
pGRN-638
pGRN-1662
CAT construct


relative CAT activity
0
10
20
A
pBLCAT2
pBLCAT3
pGRN-32
pGRN-63
pGRN-116
pGRN-154
pGRN-342
pGRN-492
pGRN-638
pGRN-1662
CAT construct

initiation sites and 2 CCAAT boxes, and one putative binding site for Sp1 and AP2 respectively. The observation of increased CAT activity with pGRN-492 compared to pGRN-154 would suggest that there are enhancer elements occurring between nucleotides -175 and -513. This region contains several structural features, including two CCAAT boxes, 2 homopurine/homopyrimidine tracts, 2 putative PEA3 binding sites, and one potential AP2 response element.

To summarize, in A549 cells, 137 to 175 bp of upstream 5'-flanking and untranslated region is sufficient to confer grn gene transcriptional activity. However, structural elements that appear necessary for high levels of grn gene transcription are contained within nucleotides -175 and -513. Sequences upstream of nucleotide -659 appear to exert negative effects on grn gene transcription.

In Swei cells, a B-lymphoblastoid cell line, the constructs pGRN-342, pGRN-492, pGRN-638, and pGRN-1662 showed similar transcriptional activity, with all four constructs expressing higher CAT activity than the promoterless CAT construct, pBLCAT3 (Figure 5.12B) [It should be noted that the TK promoter containing construct, pBLCAT2, did not give any significant CAT activity in Swei cells. This is most likely because the thymidine kinase (TK) basal promoter is not active in Swei cells]. Nucleotides -363 to -1683 contain several structural features (Figures 5.4 and 5.9), including numerous putative response elements, direct and inverted repeats, and polypurine/ polypyrimidine tracts. Deletion of the 5'-flanking region of the grn gene to nucleotide -175 (pGRN-154) resulted in a loss of CAT activity, and all constructs deleted further (pGRN-116, pGRN-63, and pGRN-32) lacked any significant CAT activity (Figure 5.12B). This result suggests that sequences

downstream of nucleotide -175, which includes 3 of the 4 transcriptional start sites, 2 CCAAT boxes and one Sp1 and AP2 putative response element, are not sufficient for grn gene transcriptional activity in Swei cells. This result further suggests that the sequences required for minimal grn promoter activity in Swei cells are contained between nucleotides -175 and -363. Located within this region are several structural features, including a transcriptional initiation site, a CCAAT box, 2 PEA3 putative binding sites, a potential AP2 response element, and 2 polypurine/polypyrimidine tracts.

The pattern of transcriptional activity obtained with the grn-CAT constructs shows some differences between Swei and A549 cells. For example, comparison of the CAT activities obtained with pGRN492, pGRN-638, and pGRN-1662 in A549 cells showed a decrease in transcriptional activity with increasing length of the 5'-flanking region, while no apparent difference in transcriptional activity is observed between these constructs in Swei cells. Also, the minimal sequence required for basal promoter activity is different in A549 and Swei cells but in both cell lines there is a major increase in promoter activity between nucleotides -175 and -363, which suggests that key regulatory elements occur within this region. This would suggest that the 5'-flanking region of the grn gene contains response elements recognised by cell specific transcription factors, and that the expression of the grn gene is regulated in a cell specific fashion.

DISCUSSION

The discussion of the results is divided into two parts. The first part discusses the organization of the protein coding region of the grn gene. The second section discusses the structure of the 5'-flanking region and characterization of a functional promoter of the human grn gene.

5.1 Organization of the Protein Coding Region of the Human Granulin Gene.

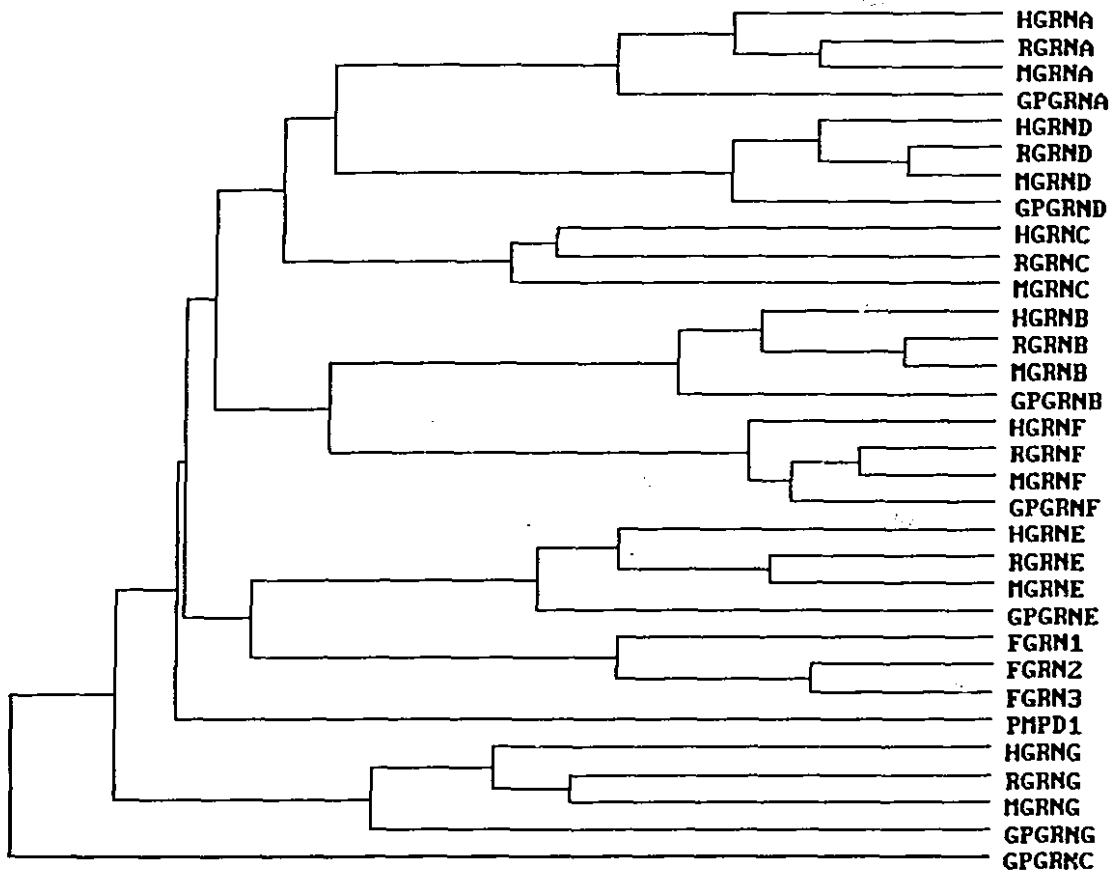
In proteins constructed as tandem repeats, exon boundaries and protein domains are often congruent. Many examples could be cited, including the epidermal growth factor precursor and related proteins (239,317,318), and adhesion molecules of the selectin class (319,320,321). Thus the construction of the grn repeat from two exons appears to go against a well established trend. The unusual exonic organization of the grn gene may allow for the formation of hybrid grn domains by alternate splicing. Single exon deletions of the ßα exons would generate the hypothetical grn hybrids AC, CD and DE without causing frameshifts in the nucleotide sequence. Other hybrids could be generated by the deletion of two or more exons provided identical type junctions are brought together. Hence alternate splicing of the grn gene may generate hybrid grn-like proteins with potentially new biological properties. There is evidence for two grn transcripts in some cell lines differing by about 200 bp (see Chapter 4 and Figure 4.2) which may be due to alternate splicing. Additionally Plowman et. al. reported a cDNA sequence predicting a grn isoform missing amino acids 393 to 470 (243), creating a grn C/D hybrid. The

borders of this deletion map exactly to the exon-intron junction of the human grn gene, and probably arises from the omission of an exon. It is suprising, however, that little significant alternate splicing was observed in Northern blot analysis of grn gene expression in various tissues and cell lines (see Chapter 4).

The exonic organization of a gene is intimately linked to its evolutionary history (322). A cluster analysis of the grn peptides from different species is shown in Figure 5.13. The grn repeats from corresponding positions on the different precursors (e.g. human grnA, rat grnA, mouse grnA, guinea pig grnA; human grnB, rat grnB, mouse grnB, guinea pig grnB; etc) segregate together in the cluster analysis. This suggests that the grn gene underwent a seven fold duplication prior to the evolutionary radiation of the mammals. Interestingly, the human grn gene contains predominantly type 0 introns (73%). Other genes with predominantly type 0 introns include those of type III collagen, ß-casein and the precursor of growth hormone, all encoding proteins which have been suggested previously to have evolved by exon duplication and/or exon recruitment (323,324,325).

The pattern observed in the grn gene may have arisen in several ways. The ancestral exon may have been of the ßα class which was then tandemly duplicated followed by the insertion of introns (326) to create the α and ß exons. The insertion of introns is clearly demonstrated within the trypsin family of serine proteases. Structural analysis of genomic sequences of members of the serine protease family reveals that unique introns have been inserted in the protease domains of tPA, uPA, chymotrypsin and elastase (327,328,329,330). It is suggested that new intron insertions tend to occur near the middle of pre-existing exons such that a gene

Figure 5.13 Dendrogram of the cluster analysis of the grn repeats. H, R, M, GP, and F indicate human, rat, mouse, guinea pig, and fish respectively. The mouse grn sequences were obtained from (243), the guinea pig grn sequences from (6), and the fish grn sequences from (8). The cluster analysis was performed using the program CLUSTAL (PC/GENE, IntelliGenetics, Mountain View, CA).


HGRNA
RGRNA
MGRNA
GPGRNA
HGRND
RGRND
MGRND
GPGRND
HGRNC
RGRNC
MGRNC
HGRNB
RGRNB
MGRNB
GPGRNB
HGRNF
RGRNF
MGRNF
GPGRNF
HGRNE
RGRNE
MGRNE
GPGRNE
FGRN1
FGRN2
FGRN3
PMPD1
HGRNG
RGRNG
MGRNG
GPGRNG
GPGRNC

evolves towards a uniformity of exon size (326,331). Interestingly insertion of an intron in a ßα-type exon of the grn gene will result in ß and α type exons of approximately equal size.

Alternatively, the α and ß exons may have arisen as separate units shuffled to form a primordial grn, and then duplicated as an [exon-intron-exon] unit, the ßα exons arising from a deletion of an intervening intron (332) bringing together the type 1 junctions of an α3' and a ß5' exon. Precise deletion of an entire intron has been reported for the rat insulin gene, resulting in the existence of two genotypic alleles (333,334). Similar observations have also been made for the mouse α-globin gene (335).

A third possibility is that a single primordial exon duplicated and the two copies then evolved to form the α and ß exons encoding a grn-like protein. Multiplication of this [exon-intron-exon] unit would generate a protein resembling the present grn precursor.

Many proteins, including extracellular matrix proteins, cell surface receptors, growth factor precursors, and enzymes of the blood-clotting and complement cascades are mosaic in nature, containing many copies of commonly occurring structural motifs. These motifs are defined by the conserved position of cysteine residues within the module. Examples of such motifs include the EGF-like repeats (42,43), the kringle structure (336), the short consensus repeat of many complement proteins (337), and the recently discovered trefoil motif/P domain (338). It will be interesting to determine whether grn related motifs occur in other proteins, and if so, whether they result from exon shuffling of exon type α, ß, ßα, [α-intron-ß] pairs, or

as combinations of these exonic units. Determining the exon-intron structure of the plant thiol proteases containing the grn consensus motif (38) may be a useful place to start.

In summary, the seven tandem repeats that make up the grn precursor mRNA are constructed from two exons. The construction of a tandem repeat from two regularly alternating exons is very unusual and has implications for the generation of novel grns by alternate splicing. The elucidation of the human grn gene organization provides the first step towards characterizing the evolutionary emergence of the grn gene. However, more phylogenetical studies of the grn gene structure are required. Grn-like peptides have been isolated from lower vertebrates (8) and insect (4). The cloning and characterization of their genes will permit the comparison of the cysteine-rich repeats of these genes, which may further help in understanding the evolutionary origins of the grn gene family. More importantly, analysis of the evolutionary history of the grn gene may well reveal possible *in vivo* functions for the grn gene products.

5.2 Characterization of the 5'-Flanking Region and a Promoter of the Human Granulin Gene.

To initiate an understanding of the mechanisms underlying the cell specific regulation of grn gene expression, the 5'-flanking region of the human grn gene was isolated and characterized. It should be emphasized that the purpose of this line of research was not to study the grn gene promoter in depth, but rather, to characterize the 5'-flanking region and to delineate a functional promoter of the human grn gene

as well as to outline regions of interest in the 5'-flanking region for future studies. This will provide valuable information as well as the necessary tools for an extensive study of the grn promoter and 5'-flanking region.

The grn gene 5'-flanking region contains several structural features common to promoters of genes with "housekeeping" or growth related functions. Housekeeping or cellular growth control gene promoters tend to have a high G+C content, contain one or more GC box elements representing binding sites for the cellular transcription factor Sp1, initiate transcription from many sites, and, most notably, lack appropriately positioned TATA and CCAAT boxes (339). The grn gene 5'-flanking and untranslated region has a G+C content of 56.4% over 500 bp upstream of the translation initiation site, contains several potential Sp1 binding sites, and shows multiple transcription initiation sites. There are 4 putative CCAAT boxes, but no typical TATA box situated within normal distances from the CCAAT boxes.

The lack of a TATA box but the presence of CCAAT boxes in the 5'-flanking region of the grn gene is different from other growth factor gene promoters. For example, the EGF (44), TGF-ß3 (340), HGF(341), and PDGF-A chain (273) gene promoters all contain a TATA box but do not have CCAAT boxes, while the promoters of TGF-α (272), and TGF-ß1 (271) lack both TATA and CCAAT boxes.

An interesting feature of the grn promoter is the absence of a canonical TATA box, which is usually present in genes that show a cell-specific pattern of expression (342). The absence of an appropriate TATA box is consistent with the observation of multiple transcription initiation sites in the grn gene, since the TATA box functions to ensure accurate initiation of transcription at a single site (270). Many genes have

been described that lack an appropriately positioned TATA box;only a few initiate transcription from a single site (272,275) while many show multiple initiation sites (271,343,344,345).

There are four CCAAT sequences in the 5'-flanking region of the grn gene, one of which is conserved between the human and mouse genes. Interestingly, all four CCAAT boxes are clustered around 400 bp of 5'-flanking and untranslated region upstream of the ATG translation initiation codon. CCAAT boxes are usually found, in eukaryotic genes, approximately 75 to 105 bp upstream of the transcription initiation site (238). In the case of the human grn gene, three of the four CCAAT boxes are located from 24 to 206 bp upstream of the transcription start sites, while a fourth CCAAT box is found downstream to the transcription initiation sites. The role of the CCAAT box in gene regulation is not completely understood. Several transcription factors have been identified that bind to the CCAAT box, these include CBP (284), CTF/NF-1 (346), CP1/2 (347) and CTY (348). Many of these factors are tissue specific, demonstrating that CCAAT sequence is not recognised by a single ubiquitous factor. The CCAAT motif is involved in transcription activation in a number of genes and in some instances is an important element in a variety of regulatory contexts. For example, a CCAAT box confers cell-type-specific regulation of the heat shock response on the *Xenopus* hsp70 gene (349) and mediates the transcriptional activation of the collagen promoter by TGF-ß (350). Thus it is possible that the CCAAT boxes in the 5'-flanking region of the grn gene may act in concert with other response elements present in the grn promoter to regulate the expression of the grn gene.

To determine if the 5'-flanking region of the grn gene can serve as a functional promoter, chimeric plasmids were constructed containing varying lengths of the 5'-flanking region of the grn gene fused to the bacterial chloramphenicol acetyltransferase (CAT) gene, and transfected into A549, A431, COS-7 and Swei cells, all of which express high levels of grn mRNA (see Chapter 4). The ability of the 5'-flanking region of the grn gene to drive transcription of the CAT reporter plasmid in these cells demonstrates that we have isolated a functional promoter of the human grn gene. The human grn gene promoter can also function in cells from other species. In addition, transfection of the grn-CAT constructs in epithelial- and lymphoid-derived cells results in a different pattern of transcriptional activity, suggesting that the 5'-flanking region of the grn gene contains both positive and negative response elements recognised by cell specific transcription factors. Analysis of the transcriptional activity of the various 5'-flanking deletion CAT constructs suggests that while 175 nucleotides of upstream 5'-flanking and untranslated sequence is sufficient to confer grn promoter activity in A549 cells, the structural elements necessary for high levels of grn gene expression, in both A549 and Swei cells, are located between nucleotides -175 to -513. Interestingly, this region is highly conserved between the human and mouse grn gene promoters.

Analysis of the 5'-flanking region of the human grn gene reveals several sequence features that have been associated with functional activity in promoters of other eukaryotic genes. Most of the putative response elements lie upstream of the sequence required for minimal promoter activity of the grn gene, suggesting that they are most probably involved in potentiating the basal levels of grn gene

expression, regulating cell specific grn gene expression, or act as repressors of transcription. The structural features present in the 5'-flanking region include potential binding sites for the transcription factors AP-1, AP-2, CREB, several cytokine and growth factor responsive transcription factors, as well as polypurine/polypyrimidine tracts. It is not known if any of these elements are functionally active, and thus the function of these elements in the context of grn gene expression remain speculative at present.

AP-1 was initially described as a transcription factor required for optimal activity of the human metallothionein II_A ($hMT\text{-}II_A$) promoter (357). It is composed of several proteins products of the jun and fos proto-oncogenes, and recognises the consensus sequence motif TGA(C/G)TCAG. The AP-1 response element is involved in mediating a transcriptional response to phorbol ester tumor promoters such as 12-O-tetradecanoyl-phorbol 13-acetate (TPA) and in mediating the induction of various genes in response to expression of transforming oncogene products (352,353). The 5'-flanking region of the human grn gene contains 25 sequences which show a 7 out of 8 match to the AP-1 consensus sequence, suggesting the possibility of grn gene regulation by fos and jun. This is interesting because the fos and jun proto-oncogene products have been associated with cellular processes involved in development, differentiation and neuronal function (352). Additionally, experiments on the possible induction of grn gene expression by phorbol ester tumor promoters are merited.

The 5'-flanking region of the grn gene contains numerous potential AP-2 binding sites; 9 sites showing a 100% match and 112 sites which show 8 out of 9

match (88%) to the consensus AP-2 element respectively. The AP-2 element is involved in mediating cAMP and phorbol ester induction of gene transcription (280). Another *cis*-acting element identified in cAMP-regulated genes is the CRE. The consensus CRE/ATF element was initially recognised as a response element that was able to confer cAMP responsiveness to the phosphoenolpyruvate carboxykinase gene (354) and has subsequently been shown to mediate cAMP induction of gene transcription in a wide variety of eukaryotic genes (355). The grn gene promoter contains 4 sequences which show a 7 out of 8 match to the CRE consensus sequence. While the CRE binding protein (CREB) is widely distributed in a variety of tissues and tissue culture cell lines (355), AP-2 is a cell specific factor (280). The significance of putative AP-2 elements and CREs in the grn promoter is not known. However, the presence of numerous putative *cis*-acting cAMP-dependent response elements in the 5'-flanking region of the grn gene suggests that cAMP may regulate grn gene expression.

Another interesting feature of the 5'-flanking region of the grn gene is the presence of consensus sequences implicated in cytokine and growth factor regulation of gene expression. Four copies of the hexanucleotide CTGGGA are found, in reverse complement, in the 5'-flanking region of the grn gene. This hexanucleotide (IL6-RE) was originally identified as the sequence element involved in the IL-6 induction of the rat α_2-macroglobulin gene and subsequently found to be well conserved in the cytokine-responsive regions of other acute-phase genes (288). In addition, there are 7 potential NFL-IL6 binding sites. NFL-IL6 was characterized as a nuclear factor recognising the IL-1 response element in the promoters of

numerous cytokine and acute-phase genes and is involved in acute phase reactions, inflammation and hematopoiesis (287). Also present in the 5'-flanking region of the grn gene are two sequences that match the consensus sequence of the TGF-ß1 inhibitory element (TIE). The TIE sequence is conserved in the promoter regions of other TGF-ß1 inhibited genes (289). The inhibition of gene expression by TGF-ß1 is mediated by binding of a fos-containing protein to this element (289). The contextual interactions between transforming growth factor-ß and the extracellular matrix and cells have been examined (85), implicating TGF-ß as an important mediator of tissue repair. The presence of putative cytokine/growth factor response elements in the grn promoter suggests that grn gene transcription may be regulated by IL-1, IL-6 and TGF-ß and implies a role for the grn gene product(s) in inflammation and wound repair. The isolation of the grns from inflammatory leukocytes (1), the widespread distribution of grn mRNA in hematopoietic cells (see Chapter 4) and the known growth modulating properties of the grn gene products on keratinocytes (2) and fibroblasts (2,9) also implicate the grn gene products in inflammation and tissue repair.

There are several purine/pyrimidine-rich domains within the 5'-flanking region of the grn gene, some of which are conserved between the human and mouse grn promoters. Polypurine / polypyrimidine tracts are often found in promoters of eukaryotic genes and frequently show S1-nuclease hypersensitivity (309). These regions have been shown to be functionally significant regulatory elements in the promoters of several essential growth control and housekeeping genes, such as cKi-*ras*, EGF-R, and c-*myc* (312,313,314,315). Polypurine/polypyrimidine regions may

deviate from a normal double-helical conformation (B-DNA) to form an intramolecular triple helix (H-DNA) (309). It is proposed that such conformations may potentiate basal levels of expression by providing access to transcription factors, facilitate protein-protein interactions, or inhibit stable chromatin assembly (309,356,357). The role, if any, of these polypyrimidine/polypurine tracts in the regulation of the grn gene remains to be determined.

Lastly, the 5' untranslated region (5'-UTR) of the human and mouse grn genes show a high degree of sequence conservation (73%) with several regions completely conserved between the two genes, suggesting that these conserved sequences may play an important, although still undefined, role in transcriptional or translational control of the grn gene. The ability of the 5'-UTR to regulate gene expression is exemplified by the ability of a portion of the 5'-UTR of the TGF-ß1 gene to inhibit, in a cell specific manner, translation of mRNA without affecting mRNA stability (358).

In summary, the 5'-flanking region of the human grn gene lacks a TATA box but has several CCAAT boxes, contains several copies of GC box elements and exhibits heterogeneity in transcription initiation sites; structural features shared by genes encoding proteins with "housekeeping" or growth control functions in the cell. The activity of the grn gene promoter is probably mediated by multiple factors as suggested by the remarkable number of putative *cis*-acting elements in the 5'-flanking region of the grn gene. Whether these elements are functional and act as general or tissue specific basal and/or inducible enhancers of grn gene transcription is not known. Further experiments including *in vitro* transfection studies, gel mobility shift, DNase footprinting assays, and mutagenesis studies are required to determine

if these elements are involved in the expression and regulation of the grn gene. The isolation and demonstration of promoter activity in the 5'-flanking region of the human grn gene provides the basis of further investigation of the transcriptional regulation of the grn gene.

SUMMARY AND CONCLUSIONS

Granulins (epithelins) are a novel family of cysteine-rich polypeptides with pleiotropic effects on cell growth *in vitro*. The objective of this PhD project was to characterize the grn family of peptides at the gene level, with respect to cDNA and genomic structures, gene expression and regulation. The findings of this research can be summarised as follows:

1. The human grn cDNA predicts a 593 amino acid precursor glycoprotein of M_r 63,600 containing seven and one-half grn repeats arranged in tandem. The rat grn precursor cDNA predicts a glycoprotein of 589 amino acids. The grn repeats in the predicted rat grn precursor is structured in a similar fashion to, and shows an overall identity of 75% with, the human grn precursor. The high conservation of the grn precursor may imply a function for the intact precursor as well as the individual grn peptides.

2. The grn precursor, but not grnA, is a heparin binding protein. It is therefore conceivable that the heparin sulfate proteoglycan-binding sites in the extracellular matrix of cells may be involved in the sequestration of the secreted grn precursor and in this manner regulate its biological activity.

3. Processing of the grn precursor appears to be tissue and cell type dependent. Secretion of the grn gene products appears to be both

constitutive and regulated, with constitutive secretion of intact or partially processed precursor in some cells, and storage of more extensively cleaved peptides in vesicular organelles in other cell types.

4. The grn gene is widely expressed in a variety of fetal and adult tissues derived from all three embryonic germ layers, and in cell lines of diverse lineages *in vitro*, including epithelial cells, lymphoid and myeloid cells, and fibroblasts. *In situ*, grn gene expression is restricted predominantly to hematopoietic and some epithelial cells. The limited expression of the grn gene in epithelial and hemic cells implies cell specific functions for the grn gene products as well as cell specific regulation of the grn gene expression *in vivo*.

5. The grn gene exists as a single copy in the human genome and is located on chromosome 17. The protein-coding region of the grn gene comprises 12 exons covering about 4 kb with each repeat encoded by two non-equivalent exons. The construction of a tandem repeat from two regularly alternating exons has implications for the generation of novel grns by alternate splicing, and in understanding the evolutionary origins of the 12 cysteine grn repeat.

6. The 5'-flanking region of the human grn gene lacks a TATA box but has several CCAAT boxes, contains several copies of GC box elements and exhibits heterogeneity in transcription initiation sites; structural features shared

by genes encoding proteins with "housekeeping" or growth control functions in the cell.

7. The 5'-flanking region of the human grn gene contains a remarkable number of putative response elements suggesting that transcription of the grn gene is probably mediated by multiple factors.

REFERENCES

1. **Bateman,A., Belcourt,D., Bennett,H.P.J., Lazure,C, Solomon,S.** 1990 Granulins, a novel class of peptide from leukocytes. *Biochem.Biophys.Res.Commun.* 173:1161-1168
2. **Shoyab,M., McDonald,V.L., Byles,C., Todaro,G., Plowman,G.D.** 1990 Epithelins 1 and 2: Isolation and characterization of two cysteine-rich growth modulating proteins. *Proc.Natl.Acad.Sci.USA.* 87:7912-7916
3. **Cuoto,M.A., Harwig,S.S.L., Cullor,J.S., Hughes,J.P., Lehrer,R.I.** 1992 Identification of eNAP-1, an antimicrobial peptide from equine neutrophils. *Infect.Immun.* 60:3065-3071
4. **Nakakura,N., Hietter,H., vanDorsselar,A., Luu,B.** 1992 Isolation and characterization of three peptides from the insect *Locusta migratoria*. *Eur.J.Biochem.* 204:147-153
5. **Anakwe,O.O., Gerton,G.L.** 1990 Acrosome biogenesis begins during meiosis: evidence from the synthesis and distribution of an acrosomal glycoprotein, acrogranin, during guinea pig spermatogenesis. *Biol.of Reprod.* 42:317-328
6. **Baba,T., Hoff,H.B., Nemoto,H., Lee,H., Orth,J., Arai,Y., Gerton,G.L.** 1993 Acrogranin, an acrosomal cysteine-rich glycoprotein, is the precursor of the growth modulating peptides, granulins, and epithelins, and is expressed in somatic as well as male germ cells. *Molec.Reproduc.Develop.* 34:233-243
7. **Parnell,P.G., Wunderlich,J., Carter,B., Halper,J.** 1992 Transforming growth factor e:Amino acid analysis and partial amino acid sequence. *Growth Factors* 7:65-72
8. **Belcourt,D., Lazure,C., Bennett,H.P.J.**, 1993 Isolation and primary structure of the three major forms of granulin-like peptides from hematopoietic tissues of a teleost fish (*Cyprinus carpio*). *J.Biol.Chem.* 268:9230-9237
9. **Zhou,J., Gao,G., Crabb,J.W., Serrero,G.** 1993 Purification of an autocrine growth factor homologous with mouse epithelin precursor from a highly tumorigenic cell line. *J.Biol.Chem.* 268:10863-69
10. **Minasian,E., Nicola,N.A.** 1992 A review of cytokine structures. *Protein Seq.*

Data Anal. 5:57-64

11. **Quesniaux,V.F.** 1992 Interleukins 9,10,11 and 12 and kit ligand: a brief overview. *Res.in Immunol.* 143:385-400
12. **Neta,R., Sayers,T.J., Oppenhein,J.J.** 1992 Relationship of TNF to interleukins. *Immunology Series* 56:499-566
13. **Sidhu,R.S., Bollon,A.P.** 1993 Tumor necrosis factor activities and cancer theraphy - a perspective. *Pharmacol.Therapeut.* 57:79-128
14. **Oppenheim,J.J., Zacharie,C.D.C., Mukaida,N., Matsushima,K.** 1991 Properties of the novel proinflammatory supergene "intercrine" cytokine family. *Ann.Rev.Immunol.* 9:617-648
15. **Begley,C.G.** 1993 Haemapoietic growth factors - from discovery to clinical application. *Medical J.Malaysia* 48:3-8
16. **Vihovac,R., Kusec,R., Jaksic,B.** 1993 Myeloid hemopoietic growth factors:review article. *Internat.J.Clin.Pharmacol.Therap.Toxicol.* 31:241-252
17. **Steward,W.S.** 1993 Granulocyte and granulocyte macrophage colony stimulating factors. *Lancet* 342:152-157
18. **Cohen,S., Elliott,G.A.** 1963 The stimulation of keratinization by a protein isolated from the submaxillary gland of the mouse. *J.Invest.Dermatol.* 40:1-5
19. **DeLarco,J.E., Todaro,G.J.** 1978 Growth factors from murine sarcoma virus-transformed cells. *Proc.Natl.Acad.Sci.USA.* 75:4001-4005
20. **Shoyab,M., Plowman,G.D., McDonald,V.L., Bradley,J.G., Todaro,G.J.** 1989 Structure and function of human amphiregulin:a member of the epidermal growth factor family. *Science* 243:1074-1076
21. **Ciccodicola,A., Dono,R., Obici,S, Simeone,A., Zollo,M., Persico,M.G.** 1989 Molecular characterization of a gene of the 'EGF family' expressed in undifferentiated human NTERA2 teratocarcinoma cells. *EMBO J.* 8:1987-1991
22. **Higashiyama,S., Abraham,J.A., Miller,J., Fiddes,J.C., Klagsbrun,M.** 1991 A heparin-binding growth factor secreted by macrophage-like cells that is related to EGF. *Science* 251:936-939
23. **Shing,Y., Christofori,G., Hanahan,D., Ono,Y., Sasada,R., Igarashi,I., Folkman,J.** 1993 Betacellulin - A mitogen from pancreatic beta-cell tumors.

Science 259:1604-1607

24. **Kimura,H., Fischer,W.H., Schubert,D.** 1990 Structure, expression and function of a schwannoma-derived growth factor. *Nature* 348:257-260

25. **Wen,D., Peles,E., Cupples,R., Suggs,S.V, Bacus,S.S, Luo,Y., Trail,G., Hu,S., Silbiger,S.M., Levy,R.C., Koski,R., Lu,H.S., Yarden,Y.** 1992 Neu differentiation factor: A transmembrane glycoprotein containing an EGF domain and an immunoglobulin homology unit. *Cell* 69:559-572

26. **Holmes,W.E., Sliwkowski,M.X., Akita,R.W., Henzel,W.J., Lee,J., Park,J.W., Yansura,D., Abadi,N., Raab,H., Lewis,G.D., Shepard,H.M., Kuang,W-J., Wood,W,I., Goeddel,D.V., Vandlen,R.L.** 1992 Identification of heregulin, a specific activator of $p185^{erbB2}$. *Science* 256:1205-1210

27. **Blomquist,M.C., Hunt,L.T., Barker,W.C.** 1984 Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. *Proc.Natl.Acad.Sci.USA.* 81:7363-7367

28. **Stroobant,P., Rice,A.P., Gullick,W.J., Cheng,D.J., Kerr,I.M., Waterfield,M.D.** 1965 Purification and characterization of vaccinia virus growth factor. *Cell* 42:383-393

29. **Massague,J., Pandiella,A.** 1993 Membrane-anchored growth factors. *Annu.Rev.Biochem.* 62:515-541

30. **Savage,C.R., Hash,J.H., Cohen,S.** 1973 Epidermal growth factor:location of disulfide bonds. *J.Biol.Chem.* 248:7669-7672

31. **Kohda,D., Shimada,I., Miyake,T., Fuwa,T., Inagaki,F.** 1989 Polypeptide chain fold of human transforming growth factor-α analogous to those of mouse and human epidermal growth factors as studied by two dimensional H-NMR. *Biochemistry* 28:953-959

32. **Appella,E. Webber,I.T., Blasi,F.** 1988 Structure and function of epidermal growth factor-like regions in proteins. *FEBS Lett* 231:1-4

33. **Tsuji,T., Okada,F., Yamaguchi,K., Nakamura,T.** 1990 Molecular cloning of the large subunit of transforming growth factor type ß masking protein and expression of the mRNA in various rat tissues. *Proc.Natl.Acad.Sci.USA.* 87:8835-8839

34. **Handford,P.A., Baron,M., Mayhew,M., Willis,A., Beesley,T., Brownlee,G.G., Cambell,I.D.** 1990 The first EGF-like domain from human factor IX contains a high affinity calcium binding site. *EMBO J.* 9:475-480

35. **Persson,E., Selander,M., Linse,S., Drakenberg,T., Ohlin,A-K., Stenflo,J.** 1989 Calcium binding to the isolated ß-hydroxyaspartic acid-containing epidermal growth factor-like domain of bovine faxtor-X. *J.Biol.Chem.* 264:16897-16904

36. **Ohlin,A-K., Linse,S., Stenflo,J.** 1988 Calcium binding to the epidermal growth factor homology region of bovine protein C. J.Biol.Chem. 263:7411-7417

37. **Sunnerhagen,M-S., Ullner,M., Persson,E., Teleman,O., Stenflo,J., Drakenberg,T.** 1992 How an epidermal growth factor (EGF)-like domain binds calcium. *J.Biol.Chem.* 267:19642-19649

38. **Schaffer,M.A., Fischer,R.L.** 1988 Analysis of mRNAs that accumulate in response to low temperature identifies a thiol protease gene in tomato. *Plant Physiol.* 87:431-436

39. **Laurence,D.J.R., Gusterson,B.A.** 1990 The epidermal growth factor:A review of structural and functional relationships in the normal organism and in cancer cells. *Tumor Biol.* 11:229-261

40. **Rutledge,B.J., Zhang,K., Bier,E., Jan,Y.H., Perrimon,N.** 1992 The drosophila *spitz* gene encodes a putative EGF-like growth factor involved in dorsal-ventral axis formation and neurogenesis. *Genes Dev.* 6:1503-1517

41. **Hill,R.J., Sternberg,P.W.** 1992 The gene *lin*-3 encodes an inductive signal for vulval development in *C.elegans. Nature* 358: 470-476

42. **Gray,A., Dull.T.J., Erlich,A.** 1983 Nucleotide sequencing of epidermal growth factor cDNA predicts a 128,000 molecular weight protein precursor. *Nature (London)* 303:722-725

43. **Scott,J.,Urdea,M., Quiroga,M., Sanchez-Pescador,R., Fong,N., Selby,M., Rutter,W., Bell,G.** 1983 Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science* 231:236-240

44. **Bell,G.I., Fong,N.M., Stempien,M.M., Wormsted,M.A., Caput,D., Ku,L., Urdea,M.S., Rall,L.B., Sanchez-Pescador,R.** 1986 Human epidermal growth factor precursor: cDNA sequence, expression *in vitro* and gene organization. *Nucleic Acids Res.* 14:8427-8446

45. **Rall,L.B., Scott,J., Bell,G.I.** 1985 Mouse prepro-epidermal growth factor synthesis by the kidney and other tissues. *Nature (London)* 313: 228-231

46. **Mroczkowski,B., Reich,M.** 1993 Identification of biologically active epidermal growth factor precursor in human fluids and secretions. *Endocrinology* 132:417-425

47. **Carpenter,G., Cohen,S.** 1975 Human epidermal growth factor and the proliferation of human fibroblasts. *J.Cell.Physiol.* 88:227-238

48. **Hollenberg,M.D., Cuatrecasas,P.** 1973 Epidermal growth factor:receptors in human fibroblasts and modulation of action by cholera toxin. *Proc.Natl.Acad.Sci.USA.* 70:2964-2968

49. **Gregory,H.** 1975 Isolation and structure of urogasterone and its relationship to epidermal growth factor. *Nature (London)* 257:325-327

50. **Burgess,A.W.** 1989 Epidermal growth factor and transforming growth factor-α. *British Med.Bull.* 45:401-424

51. **Fisher,D.A., Lakshmanan,J.** 1990 Metabolism and effects of epidermal growth factor and related growth factors in mammals. *Endocr.Rev.* 11:418-442

52. **Shoyab,M., McDonald,V.L., Bradley,J.G., Todaro,G.J.** 1988 Amphiregulin: A bifunctional growth modulating glycoprotein produced by the phorbol-12-myristate 13-acetate-treated human breast adenocarcinoma cell line MCF-7. *Proc.Natl.Acad.Sci.USA.* 85:6528-6532

53. **Plowman,G.D., Green,J.M., McDonald,V.M., et.al.** 1990 The amphiregulin gene encodes a novel epidermal growth factor related protein with tumor inhibitory activity. *Mol.Cell.Biol.* 10:1969-1981

54. **Johnson,G.R., Saeki,T., Gordon,A.W., Shoyab,M., Salomon,D.S.** 1992 Autocrine action of amphiregulin in a colon carcinoma cell line and immunocytochemical localization of amphiregulin in human colon. *J.Cell Biol.* 118:741-751

55. **Cook,P.W., Mattox,W.W., Keeble,M.R., Pittelkow,M.R., Plowman,G.D., Shoyab,M., Adelman,J.P., Shipley,G.D.** 1991 A heparin sulfate-regulated human keratinocyte autocrine factor is similar or identical to amphiregulin. *Mol.Cell.Biol.* 11:2547-2557

56. **Higashiyama,S., Lau,K., Besner,G.E., Abraham,J.A., Klagsbrun,M.** 1992 Structure of heparin-binding EGF-like growth factor:multiple forms, primary structure, and glycosylation of the mature protein. *J.Biol.Chem.* 267:6205-6212

57. **Dluz,S.M., Higashiyama,S., Damm,D., Abraham,J.A., Klagsbrun,M.** 1993 Heparin-binding epidermal growth factor-like growth factor expression in cultured fetal human vascular smooth muscle cells. *J.Biol.Chem.* 268:18330-18334

58. **Twardzik,D.R., Brown,J.P., Ranchalis,J.E., Todaro,G.J., Moss,B.** 1985 Vaccinia virus-infected cells release a novel polypeptide functionally related to transforming and epidermal growth factors. *Proc.Natl.Acad.Sci.USA.* 82:5300-5304

59. **Peles,E., Bacus,S.S., Koski,R.A., Lu,H.S., Wen,D., Ogden,S.G., Levy,R.B., Yarden,Y.** 1992 Isolation of the *neu*/HER-2 stimulatory ligand:a 44kd glyco protein that induces differentiation of mammary tumor cells. *Cell* 69:205-216

60. **DeLarco,J.E., Todaro,G.J.** 1978 Growth factors from murine sarcoma virus-transformed cells. *Proc.Natl.Acad.Sci.USA.* 75:4001-4005

61. **Todaro,G.J., DeLarco,J.E., Cohen,S.** 1976 Transformation by murine and feline sarcoma viruses specifically blocks binding of epidermal growth factor to cells. *Nature (London)* 264:26-31

62. **Derynck,R.** 1988 Transforming growth factor-α. *Cell* 54:593-595

63. **Massague,** 1983 Epidermal growth factor-like transforming growth factor:interaction with epidermal growth factor receptors in human placental membranes and A431 cells. *J.Biol.Chem.* 258:13614-13620

64. **Derynck,R., Roberts,A.B., Winkler,M.E., Chen,E.Y., Goeddel,D.V.** 1984 Human transforming growth factor-α:precursor structure and expression in *E.coli*. *Cell* 38:287-297

65. **Lee,D.L., Rose,T.M., Webb,N.R., Todaro,G.J.** 1985 Cloning and sequence analysis of a cDNA for rat transforming growth factor-α. *Nature (London)* 313:489-491

66. **King,L.E., Gates,R.E., Stoscheck,C.M., Nanney,L.B.** 1990 Epidermal growth factor/transforming growth factor alpha receptors and psoriasis. *J.Invest.Dermatol.* 95:10S-12S

67. **Sandgren,E.P., Luetteke,N.C., Palmiter,R.D., Brinster,R.L., Lee,D.C.** 1990 Overexpression of TGF-α in transgenic mice:induction of epithelial hyperplasia,pancreatic metaplasia, and carcinoma of the breast. *Cell* 61:1121-1135

68. **Yasui,W., Ji,Z-Q., Kuniyasy,H., Ayhan,A., Yokozaki,H., Ito,H., Tahara,E.** 1992 Expression of transforming growth factor alpha in human tissues:immunohistochemical study and Northern blot analysis. *Virchows Archiv.A Pathol.Anat.* 421:513-519

69. **Massague,J.** 1990 The transforming growth factor-ß family. *Annu.Rev.Cell.Biol.* 6:597-641

70. **Barnard,J.A., Lyons,R.M., Moses,H.L.** 1990 The cell biology of transforming growth factor-ß-b. *Bioch.Biophy.Acta.* 1032:79-87

71. **Sporn,M.B., Roberts,A.B.** 1992 Transforming growth factor-ß:recent progress and new challenges. *J.Cell.Biol.* 119:1017-1021

72. **Cate,R.L., Mattaliano,R,J., Hession,C., Tizard,R.,Farber,N.M., Cheung,A., Ninfa,E.G., Frey,A.Z., Gash,D.J.,Chow,E.P., Fisher,R.A., Bertonis,J.M., Torres,G.,Wallner,B.P., Ramachandran,K.L., Ragin,R.C., Manganaro,T.F.,Maclaughlin,D.T, Donahoe,P.K.** 1986 Isolation of the bovine and human genes for mullerian inhibiting substance and expression of the human gene in animal cells. *Cell* 45:685-698

73. **Padgett,R.W., St.Johnson,R.D., Gelbart,W.M.** 1987 A transcript from a Drosophila pattern predicts a protein homologous to the transforming growth factor-ß family. *Nature (London)* 325:81-84

74. **Wozney,J.M., Rosen,V., Celeste,A.J., Mitsock,L.M., Whitters,M.J., Kriz,R.W., Hewick,R.M., Wang,E.A** 1988 Novel regulators of bone

formation:Molecular clones and activities. *Science* 242:1528-1534

75. **Weeks,D.L., Melton,D.A.** 1987 A maternal mRNA to the vegetal hemisphere in *Xenopus* eggs codes for a growth factor related to TGF-beta. *Cell* 51:861-867

76. **Mason,A.J., Hayflick,J.S., Ling,N., Esch,F., Ueno,N., Ying,S-Y., Guilleman,R., Niall,H., Seeburg,P.H** 1985 Complementary DNA sequences of ovarian follicular fluid inhibin show precursor structure and homology with transforming growth factor-ß. *Nature (London)* 318:659-663

77. **Basler,K., Edlund,T., Jessell,T.M., Yamada,T.** 1993 Control of cell pattern in the neural tube:regulation of cell differentiation by dorsalin-1, a novel TGF-ß family member. *Cell* 73:687-702

78. **McPherron,A.C., Lee,S.J** 1993 GDF-3 and GDF-9 - 2 new members of the transforming growth factor-ß superfamily containing a novel pattern of cysteines. *J.Biol.Chem.* 268:3444-3449

79. **Kondaiah,P., Sands,M.J., Smith,J.M., Fields,A., Roberts,A.B., Sporn,M.B., Melton,D.A.** 1990 Identification of a novel transforming growth factor-ß (TGF-ß5) mRNA in *Xenopus laevis*. *J.Biol.Chem.* 265:1089-1093

80. **Harpel,J.G., Metz,C.N., Kojima,S., Rifkin,D.B.** 1992 Control of transforming growth factor-ß activity:latency vs. activation. *Prog.Growth Fact.Res.* 4:321-335

81. **Border,W.A., Ruoslahti,E.** 1992 Transforming growth factor-ß in disease:the dark side of tissue repair. *J.Clin.Invest.* 90:1-7

82. **Culouscou,J-M., Carlton,G.W., Shoyab,M.** 1993 Biochemical analysis of the epithelin receptor. *J.Biol.Chem.* 268:10458-10462

83. **Silberstein,G.B., Daniel,C.W.** 1987 Reversible inhibition of mammary gland growth by transforming growth factor-ß. *Science* 237:291-293

84. **Russell,W.E., Coffey,R.J., Ouellette,A.J., Moses,H.L.** 1988 Type ß transforming growth factor reversibly inhibits the early proliferative response to partial hepatectomy in the rat. *Proc.Natl.Acad.Sci.USA.* 85:5126-5130

85. **Noble,N.A., Harper,J.R., Border,W.A.** 1992 *In vivo* interactions of TGF-ß and extracellular matrix. *Prog.Growth Fact.Res.* 4:369-382

86. **Ruscetti,F., Varesio,L., Ochoa,A., Ortaldo,J.** 1993 Pleiotropic effects of

transforming growth factor-ß on cells of the immune system. *Ann.N.Y.Acad.Sci.* 685 488-500

87. **Wahl., Hunt,D.A., Wakefield,L.M., McCartney-Francis,N., Wahl,L.M., Roberts,A.B., Sporn,M.B.** 1987 Transforming growth factor type ß induces monocyte chemotaxis and growth factor production. *Proc.Natl.Acad.Sci.USA.* 84:5788-5792

88. **Shull,M.M., Ormsby,I., Kier,A.B., Pawlowski,S., Diebold,R.J., Yin,M., Allen,R., Sidman,C., Proetzel,G., Calvin,D., Annunziata,N., Doetschman,T.** 1992 Targeted disruption of the mouse transforming growth factor-ß1 gene results in multifocal inflammatory disease. *Nature (London)* 359:693-699

89. **Halper,J., Moses,H.L.** 1983 Epithelial tissue-derived growth factor-like polypeptides. *Cancer Res.* 43:1972-1979

90. **Halper,J., Moses,H.L.** 1987 Purification and characterization of a novel transforming growth factor. *Cancer Res.* 47:4552-4559

91. **Brown,C.A., Halper,J.** 1990 Mitogenic effects of transforming growth factor type-e on epithelial and fibroblastic cells - comparison with other growth factors. *Exp.Cell.Res.* 190:233-242

92. **Halper,J., Carter,B.J.** 1989 Modulation of growth of human carcinoma cells by heparin and growth factors. *J.Cell.Physiol.* 141:16-23

93. **Dunnington,D.J., Scott,R.G., Anzano,M.A., Greig,R.** 1990 Characterization and partial purification of human epithelial transforming growth factor. *J.Cell.Biochem.* 44:229-239

94. **Dunnington,D., Prichett,W., Moyer,M., Greig,R.** 1990 Identification of a low molecular weight form of epithelial transforming growth factor. *Life Sciences* 47:2059-2063

95. **Burgess,W.H., Maciag,T.** 1989 The heparin-binding (fibroblast) growth factor family of proteins. *Annu.Rev.Biochem.* 58:575-606

96. **Baird,A., Klagsbrun,M.** 1991 The Fibroblast Growth Factor Family - An overview. *Ann.N.Y.Acad.Sci.* 638:xi-xii

97. **Thomas,K.A., Riley,M.C., Lemmon,S.K., Baglan,N.C., Bradshaw,R.A.** 1980 Brain fibroblast growth factor:nonidentity with myelin basic protein fragments.

J.Biol.Chem. 255:5517-5520

98. **Gospodarowicz,D.** 1975 Purification of a fibroblast growth factor from bovine pituitary. *J.Biol.Chem.* 250:2515-2519

99. **Shing,Y., Folkman,J., Sullivan,R., Butterfield,C., Murray,J., Klagsbrun,M.** 1984 Heparin affinity:purification of a tumor-derived capillary endothelial cell growth factor. *Science* 223:1296-1299

100. **D'Amore,P.A., Klagsbrun,M.** 1984 Endothelial cell mitogens derived from retina and hypothalamus:biochemical and biological similarities. *J.Cell Biol.* 99:1545-1549

101. **Partanen,J., Vainikka,S., Alitalo,K.** 1993 Structural and functional specificity of FGF receptors. *Phil.Trans.R.Soc.Lond.B* 297-303

102. **Vlodavsky,I., Folkman,J., Sullivan,R., Fridman,R., Ishai-Michadi,R., Sasse,J., Klagsbrun,M.** 1987 Endothelial cell derived basic fibroblast growth factor : synthesis and deposition into subendothelial extracellular matrix. *Proc.Natl.Acad.Sci.USA.* 84:2292-2296

103. **Folkman,J., Klagsbrun,M., Sasse,J., Wadzinski,M., Ingber,D., Vlodavsky,I.** 1988 A heparin-binding angiogenic protein-basic fibroblast growth factor is stored within basement membrane. *Am.J.Pathol.* 130:393-399

104. **Jackson,A., Friedman,S., Zhan,X., Engkela,K.A., Forough,R., Maciag,T.** 1992 Heat shock induces the release of fibroblast growth factor-1 from NIH 3T3 cells. *Proc.Natl.Acad.Sci.USA* 89:10691-10695

105. **Mignatti,P., Morimoto,T., Riflan,D.B.** 1992 Basic fibroblast growth factor, a protein devoid of secretory signal sequence, is released by cells via a pathway independent of the endoplasmic reticulum-Golgi complex. *J.Cell.Physiol.* 151:81-93

106. **Schechter,J.E.** 1992 Is cellular disruption the mechanism of release of basic fibroblast growth factor from anterior pituitary gonadotropes? *Tissue Cell* 24:791-802

107. **Jaye,M., Howk,R., Burgess,W., Ricca,G.A., Chiu,I-M., Ravera,M.W., O'Brien,S.J., Modi,W.S., Maciag,T., Drohan,W.N.** 1986 Human endothelial cell growth factor:cloning, nucleotide sequence, and chromosome

localization. *Science* 233:541-545

108. **Abraham,J.A., Whang,J.L., Tumolo,A., Mergia,A., Friedman,J., Gospodarowicz,D., Fiddes,J.** 1986 Human basic fibroblast growth factor:nucleotide sequence and genomic organization. *EMBO J.* 5:2523-2528

109. **Florkiewicz,R.Z., Shibata,F., Barankiewicz,T., Baird,A, Gonzales,A-M., Florkiewicz,E., Shah,N.** 1991 Basic fibroblast growth factor gene expression. *Ann.N.Y.Acad.Sci.* 638:109-126

110. **Dickson,C., Smith,R., Brookes,S., Peters,G.** 1984 Tumorigenesis by mouse mammary tumor virus:proviral activation of a cellular gene in the common integration region *int*-2. *Cell* 37:529-536

111. **Wilkinson,D.G., Peters,G., Dickson,C., McMahon,A.P.** 1988 Expresssion of the FGF-related proto-oncogene *int*-2 during gastrulation and neurulation in the mouse. *EMBO J.* 7:691-695

112. **Dickson,C., Fuller-Pace,F., Kieffer,P., Acland,P., MacAllan,D., Petres,G.** 1991 Expression, processing and properties of *int*-2. *Ann.N.Y.Acad.Sci.* 638:18-26

113. **Sakamoto,H., Mori,M., Taira,M., Yoshida,T., Matsukawa,S., Shimizu,K., Sekiguchi,M., Terada,M., Sugimura,T.** 1986 Transforming gene from human stomach cancers and a noncancerous portion of stomach mucosa. *Proc.Natl.Acad.Sci.USA.* 83:3997-4001

114. **Yoshida,T., Sakamoto,H., Miyagawa,K., Little,P.F.R., Terada,M., Sugimura,T.** 1987 Genomic clone of *hst* with transforming activity from a patient with acute leukemia. *Biochem.Biophys.Res.Commun.* 142:1019-1024

115. **Yoshida,T., Miyagawa,K., Ogadiri,H., Sakamoto,H., Little,P.F.R., Terada,M., Sugimura,T.** 1987 Genomic sequence of hst, a transforming gene encoding a protein homologous to fibroblast growth factors and the *int*-2 encoded protein. *Proc.Natl.Acad.Sci. USA.* 84:7305-7309

116. **Delli-Bovi,P., Curatola,A.M., Kern,F.G., Greco,A.,Ittmann,M., Basilico,C.** 1987 An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a growth factor that is a member of the FGF family. *Cell* 50:729-737

117. **Taira,M., Yoshida,T., Miyagawa,H., Sakamoto,H., Terada,M., Sugimura,T.**

1987 cDNA sequence of human transforming gene *hst* and identification of the coding sequence required for transforming activity. *Proc.Natl.Acad.Sci.USA.* 84:2980-2984

118. **Yoshida,T., Sakamoto,H., Miyagawa,K., Sugimura,T., Terada,M.** 1991 Characterization of the *hst*-1 gene and its product. *Ann.N.Y. Acad.Sci.* 638:27-37

119. **Zhan,X., Bates,B., Hu,X., Goldfarb,M.** 1988 The human FGF-5 gene encodes a novel protein related to fibroblast growth factors. *Mol. Cell.Biol.* 8:3487-3498

120. **Goldfarb,M., Bates,B., Drucker,B., Hardin,J., Haub,O.** 1991 Expression and possible functions of the FGF-5 gene. *Ann.N.Y.Acad. Sci.* 638:38-52

121. **Marics,I., Adelaide,J., Raybaud,F., Mattei,M-G., Coulier,F. Planche,J., deLapeyriere, Birnbaum,D.** 1989 Characterization of the HST-related FGF.6 gene, a new member of the fibroblast growth factor gene family. *Oncogene* 4:335-340

122. **deLapeyriere,O., Rosnet,O., Benharroch,D., Raybaud,F., Marchetto,S., Planche,J., Galland,F., Mattei,M-G., Copeland,N.G., Jenkins,N.A., Coulier,F., Birnbaum,D.** 1990 Structure, chromosome mapping and expression of the murine FGF-6 gene. *Oncogene* 5:823-831

123. **Han,J-K., Martin,G.R.** 1993 Embryonic expression of FGF-6 is restricted to the skeletal muscle lineage. *Develop.Biol.* 158:549-554

124. **Coulier,F., Ollendorff,V., Marics,I., Rosnet,O., Batoz,M., Planche,J., Marchetto,S., Pebusque,M-J., deLapeyriere,O., Birnbaum,D.** 1991 The FGF6 gene within the FGF multigene family. *Ann.N.Y.Acad.Sci.* 638:53-61

125. **Rubin,J.S., Osada,H., Finche,P., Taylor,W.G., Rudikoff,S., Aaronson,S.A.** 1989 Purification and characterization of a newly identified growth factor specific for epithelial cells. *Proc.Natl. Acad.Sci.USA.* 86:802-806

126. **Finch,P., Rubin,J., Miki,T., Ron,D., Aaronson,S.A.** 1989 Human KGF is FGF-related with properties of a paracrine effector of epithelial cell growth. *Science* 245:752-755

127. **Aaronson,S.A., Bottaro,D.P., Miki,T., Ron,D., Finch,P.W., Fleming,T.P., Ahn,J., Taylor,W.G., Rubin,J.S.** 1991 Keratinocyte growth factor:a fibroblast

growth factor family member with unusual target cell specificity. *Ann.N.Y.Acad.Sci.* 638:62-77

128. **Tanaka,A., Miyamoto,K., Minamino,N., Takeda,M., Sato,B., Matsuo,H., Matsumoto,K.** 1992 Cloning and characterization of an androgen-induced growth factor essential for the androgen-dependent growth of mouse mammary carcinoma cells. *Proc.Natl.Acad.Sci.USA.* 89: 8928-8932

129. **Naruo,K., Seko,C., Kuroshima,K., Matsutani,E., Sasada,R., Kondo,T., Kurokawa,T.** 1993 Novel secretory heparin-binding factors from human glioma cells (glia-activating factors) involved in glial cell growth. *J.Biol.Chem.* 268:2857-2864

130. **Miyamoto,M., Naruo,K-I., Seko,C., Matsumoto,S., Kondo,T., Kurokawa,T.** 1993 Molecular cloning of a novel cytokine cDNA encoding the ninth member of the fibroblast growth factor family, which has a unique secretion property. *Mol.Cell.Biol.* 13:4251-4259

131. **Nakamura,T., Nawa,K., Ichihara,A.** 1984 Partial purification and characterization of hepatocyte growth factor from serum of hepatectomized rats. *Biochem.Biophys.Res.Commun.* 122 1450-1459

132. **Nakamura,T., Nawa,K., Ichihara,A., Kaise,N., Nishino,T.** 1987 Purification and subunit structure of hepatocyte growth factor from rat platelets. *FEBS Lett.* 224:311-318

133. **Gohda,E., Tsubouchi,H., Nakayama,H., Hirono,S., Sakiyama,O., Takahashi,K., Miyazaki,H., Hashimoto,S., Daikuhara,Y.** 1988 Purification and partial characterization of hepatocyte growth factor from plasma of a patient with hepatic failure. *J.Clin.Invest.* 81:414-419

134. **Zarnegar,R., Michalopoulos,G.** 1989 Purification and biological characterization of human hepatopoietin A, a polypeptide growth factor for hepatocytes. *Cancer Res.* 49:3314-3320

135. **Stoker,M., Gherardi,E., Perman,M., Gray,J.** 1987 Scatter factor is a fibroblast-derived modulator of epithelial cell mobility. *Nature (London)* 327:239-242

136. **Gherardi,E., Gray,J., Stoker,M., Perryman,M., Furlong,R.** 1989 Purification

of scatter factor, a fibroblast-derived basic protein that modulates epithelial interactions and movement. *Proc.Natl.Acad.Sci.USA.* 86:5844-5848

137. **Weidner,K.M., Arakaki,N., Hartmann,G., Vandekerckhove,J., Werdgart,S., Rieder,H** 1991 Evidence for the identity of human scatter factor and human hepatocyte growth factor. *Proc.Natl. Acad.Sci.USA.* 88:7001-7005

138. **Nakamura,T., Nishizawa,T., Hagiya,M., Seki,T., Shimonishi,M., Sugimura,A., Tashiro,K., Shimizu,S.** 1989 Molecular cloning and expression of human hepatocyte growth factor. *Nature (London)* 342: 440-443

139. **Tashiro,K., Hagiya,M., Nishizawa,T., Seki,T., Shimonishi,M., Shimizu,S., Nakamura,T.** 1990 Deduced primary structure of rat hepatocyte growth factor and expression of the mRNA in rat tissues. *Proc.Natl.Acad.Sci.USA.* 87:3200-3204

140. **Zarnegar,R., Muga,S., Rahlja,R., Michalopoulos** 1990 Tissue distribution of hepatopoietin-A : A heparin-binding polypeptide polypeptide growth factor for hepatocytes. *Proc.Natl.Acad.Sci.USA.* 87:1252-1256

141. **Matsumoto,K., Nakamura,T.** 1992 Hepatocyte growth factor:molecular structure, roles in liver regeneration, and other biological functions. *Crit.Rev.Oncog.* 3:27-54

142. **Montesano,R., Matsumoto,K., Nakamura,T., Orci,L.** 1991 Identification of a fibroblast-derived epithelial morphogen as hepatocyte growth factor. *Cell* 67:901-908

143. **Igawa,T., Matsumoto,K., Kanda,S., Saito,Y., Nakamura,T** 1993 Hepatocyte growth factor may function as a renotropic factor for regeneration in rats with acute renal injury. *Am.J.Physiol.* 265:F61-F69

144. **Jiang,W., Puntis,C.A., Nakamura,T., Hallett,M.B.** 1992 Neutrophil priming by hepatocyte growth factor, a novel cytokine. *Immunology* 77:147-149

145. **Mizuno,K., Higuchi,O., Ihle,J.N., Nakamura,T.** 1993 Hepatocyte growth factor stimulates growth of hematopoietic progenitor cells. *Biochem.Biophys.Res.Commun.* 194:178-186

146. **Froesch,E.R., Schmid,C., Schwander,J., Zapf,J.** 1985 Actions of insulin-like growth factors. *Ann.Rev.Physiol.* 47:443-467

147. **Cullen,K.J., Yee,D., Rosen,N.** 1991 Insulin-like growth factors in human malignancy. *Cancer Invest.* 9:442-454

148. **Daughaday,W.H., Rotwein,P.** 1989 Insulin-like growth factors I and II. Peptide, messenger ribonucleic acid and gene structures, serum and tissue concentrations. *Endocrine Rev.* 10:68-91

149. **Nagamatsu,S., Chan,S.J., Falkmer,S., Steiner,D.F.** 1991 Evolution of the insulin gene superfamily. *J.Biol.Chem.* 266:2397-2402

150. **Binoux,M., Hossenlopp,P., Hardouin,S., Lassarre,C., Gourmelen,M.** 1986 Somatomedin (insulin-like growth factors)-binding proteins:molecular forms and regulation. *Hormone Res.* 24:141-151

151. **Rinderknecht,E., Humbel,R.E.,** 1978 The amino-acid sequence of human insulin-like growth factor-I and its structural homology with proinsulin. *J.Biol.Chem.* 253:2769-2774

152. **Klapper,D.G., Svoboda,M.E., Van Wyk,J.J.** 1983 Sequence analysis of somatomedin-C: confirmation of identity with insulin-like growth factor-I. *Endocrinology* 112:2215-2217

153. **Han,V.K.M., Lund,P.K., Lee,D.C., D'Ercole,A.J.** 1988 Expression of somatomedin/Insulin-like growth factor messenger ribonucleic acids in the human fetus:identification,characterization, and tissue distribution. *J.Clin.Endocrin.Metabol.* 66:422-429

154. **Han,V.K.M., D'Ercole,A.J., Lund,P.K.** 1987 Cellular localization of somatomedin (Insulin-like growth factor) messenger RNA in the human fetus. *Science* 236:193-197

155. **Rinderknecht,E., Humbel,R.E.** 1978 Primary structure of human insulin-like growth factor-II. *FEBS Lett.* 89:283-286

156. **Deuel,T.F.** 1987 Polypeptide growth factors:roles in normal and abnormal cell growth. *Ann.Rev.Cell.Biol.* 3:443-492

157. **Deuel,T.F., Huang,J.S.** 1983 Platelet derived growth factor:purification, properties, and biological activities. *Prog.Hematol.* 13:201-221

158. **Ross,R., Raines,E.W., Bowen-Pope,D.F.** 1986 The biology of platelet-derived growth factor. *Cell* 46:155-169

159. **Balk,S.D.** 1971 Calcium as a regulator of the proliferation of normal, but not of transformed, chicken fibroblasts in a plasma containing medium. *Proc.Natl.Acad.Sci.USA.* 68:271-275

160. **Ross,R., Glomset,J., Kariya,B., Harker,L.** 1974 A platelet dependent serum factor that stimulates the proliferation of arterial smooth muscle cells *in vitro*. *Proc.Natl.Acad.Sci.USA* 71:1207-1210

161. **Betsholtz,C., Johnsson,A., Heldin,C-H., Westermark,B., Lind,P., Urdea,M.S., Eddy,R., Shows,T.B., Philpott,K., Mellor,A.L., Knott,T.J., Scott,J.** 1986 cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain and its expression in tumour cell lines. *Nature (London)* 320:695-699

162. **Josephs,S.F., Guo,C., Ratner,L., Wong-Staal,F.** 1984 Human proto-oncogene nucleotide sequences corresponding to the transforming region of simian sarcoma virus. *Science* 223:487-491

163. **Heldin,C-H., Westermark,B.** 1989 Platelet-derived growth factor:three isoforms and two receptor types. *Trends Genet.* 5:108-111

164. **Sasahara,M., Fries,J.W.U., Raines,E.W., Gown,A.M., Westrum,L.E., Frosch,M.P., Bonthron,D.T., Ross,R., Collins,T.** 1991 PDGF B-chain in neurons of the central nervous system, posterior pituitary, and in a transgenic model. *Cell* 64:217-227

165. **Yeh,H-J., Ruit,K.G., Wang,Y-X., Parks,W.C., Snider,W.D., Deuel,T.F.** 1991 PDGF A-chain gene is expressed by mammalian neurons during development and in maturity. *Cell* 64:209-216

166. **Li,Y-S., Milner,P.G., Chauhan,A.K., Watson,M.A., Hoffman,R.M., Kodner,C.M., Milbrandt,J., Deuel,T.F.** 1990 Cloning and expression of a developmentally regulated protein that induces mitogenic and neurite outgrowth activity. *Science* 250:1690-1694

167. **Milner,P.G., Shah,D., Veile,R., Donis-Keller,H., Kumar,B.V.** 1992 Cloning,nucleotide sequence, and chromosome localization of the human pleiotrophin gene. *Biochemistry* 31:12023-12028

168. **Nurcombe,V., Fraser,N., Herlaar,E., Heath,J.K.** 1992 MK: a pluripotential

embryonic stem-cell-derived neuroregulatory factor. *Development* 116:1175-1183

169. **Tsutsui,J., Kadomatsu,K., Matsubara,S., Nakagawara,A., Hamanoue,M., Takao,S., Shimazu,H., Ohi,Y, Muramatsu,T.** 1993 A new family of heparin-binding growth/differentiation factors:increased midkine expression in Wilms' Tumor and other human carcinomas. *Cancer Res.* 53:1281-1285

170. **Ishikawa,F., Miyazono,K., Hellman,U., Drexler,H., Wernstedt,C., Hagiwara,K., Usuki,K., Takaku,F., Risau,W., Heldin,C-H.** 1989 Identification of angiogenic activity and the cloning and expression of platelet-derived endothelial cell growth factor. *Nature (London)* 338:557-562

171. **Leung,D.W.,, Chaines,G., Kuang,W.J., Goddell,D.V., Ferrara,N.** 1989 Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science* 246:1306-1309

172. **Keck,P.J., Hauser,S.D., Krivi,G.** 1989 Vascular permeability factor, an endothelial cell mitogen related to PDGF. *Science* 246:1309-1312

173. **Quinn,T.P., Peters,K.G., DeVries,C., Ferrara,N., Williams,L.T.** 1993 Fetal liver kinase 1 is a receptor for vascular endothelial growth factor and is selectively expressed in vascular endothelium. *Proc.Natl.Acad.Sci.USA* 90:7533-7537

174. **Vaisman,N., Gospodarowicz,D., Neufeld,G.** 1990 Characterization of the receptors for vascular endothelial growth factor. *J.Biol.Chem.* 265:19461-19466

175. **Liotta,L., Mandler,R., Murano,G., Katz,D.A., Gordon,R.K., Chiang,P.K., Schiffmann,E.** 1986 Tumor cell autocrine motility factor. *Proc.Natl.Acad.Sci.USA.* 83:3302-3306

176. **Silletti,S., Raz,A.** 1993 Autocrine motility factor is a growth factor. *Biochem.Biophys.Res.Commun.* 192:446-457

177. **Burrow,C.R., Wilson,P.D.** 1993 A putative Wilms tumor-secreted growth factor activity required for primary culture of human nephroblasts. *Proc.Natl.Acad.Sci.USA.* 90:6066-6070

178. **Takeuchi,A., Hayashi,H., Naito,Y., Baba,T., Tamatani,T., Onzaki,K.** 1993

Human myelomonocytic cell line THP-1 produces a novel growth-promoting factor with a wide target cell spectrum. *Cancer Res.* 53:1871-1876

179. **Johnston,P.G., Rondinone,C.M., Voeller,D., Allegra,C.J.** 1992 Identification of a protein factor secreted by 3T3-L1 preadipocytes inhibitory for the human MCF-7 breast cancer cell line. *Cancer Res.* 52:6860-6865

180. **Lauber,B., Leuthold,M., Schmitter,D., Cano-Santos,J., Waibel,R., Stahel,R.A.** 1992 An autocrine mitogenic activity produced by a pleural human mesothelioma cell line. *Int.J.Cancer* 50:943-950

181. **Rao,L.S., Rao,S.G.A., Banerji,A.P.** 1989 A mitogenic factor from rat muscle. *Cancer Lett.* 45:27-34

182. **Perkel,V.S., Mohan,S., Herring,S.J., Baylink,D.J., Linkhart,T.A.** 1990 Human prostatic cancer cells, PC3, elaborate mitogenic activity which selectively stimulates human bone cells. *Cancer Res.* 50:6902-6907

183. **Rutka,J.T., Rosenblum,M.L., Stern,R., Ralston,H.J., Dougherty,D., Giblin,J., DeArmond,S.** 1989 Isolation and partial purification of growth factors with TGF-like activity from human malignant gliomas. *J.Neurosurg.* 71:875-883

184. **Reibman,J., Meixler,S., Lee,T.C., Gold,L.I., Cronstein,B.N., Haines,K.A., Kolasinski,S.L., Weismann,G.** 1991 Transforming growth factor ß1, a potent chemoattractant for human neutrophils, bypasses classic signal transduction pathways. *Proc.Natl.Acad.Sci.USA.* 88:6805-6809

185. **Ignotz,R.A., Massague,J.** 1986 Transforming growth factor-ß stimulates the expression of fibronectin and collagen and their incorporation into the extracellular matrix. *J.Biol.Chem.* 261:4337-4345

186. **Postlethwaite,A.E., Keski-Oja,J., Moses,H.L., Kang,A.H.** 1987 Stimulation of the chemotactic migration of human fibroblasts by transforming growth factor-ß. *J.Exp.Med.* 165:251-256

187. **Sporn,M.B., Roberts,A.B.** 1986 Peptide growth factors and inflammation, tissue repair, and cancer. *J.Clin.Invest.* 78:329-332

188. **Seppa,H., Grotendorst,G., Seppa,S., Schiffmann,E., Martin,G.R.** 1982 Platelet-derived growth factor is chemotactic for fibroblasts. *J.Cell.Biol.*

92:584-588

189. **Grotendorst,G.R., Seppa,H.E.J., Kleinman,H.K., Martin,G.R.** 1981 Attachment of smooth muscle cells to collagen and their migration toward platelet-derived growth factor. *Proc.Natl.Acad.Sci.USA.* 78:3669-3672

190. **Grant,M., Jerdan,J., Merimee,T.J.** 1987 Insulin-like growth factor-I modulates endothelial cell chemotaxis. *J.Clin. Endocrin.Metab.* 65:370-371

191. **Grant,M.B., Khaw,P.T., Schultz,G.S., Adams,J.L., Shimizuo,R.W.** 1992 Effects of epidermal growth factor, fibroblast growth factor, and transforming growth factor-ß on corneal cell chemotaxis. *Invest.Ophthalmol.Vis.Sci.* 33:3292-3301

192. **Brindle,N.P.J.** 1993 Growth factors in endothelial regeneration. *Cardiovas.Res.* 27:1162-1172

193. **Nishida,T., Tanaka,H., Nakagawa,S., Sabane,T., Awata,T., Manabe,R** 1984 Fibronectin synthesis by the rabbit cornea:effects of mouse epidermal growth factor and cyclic AMP analogs. *Jpn.J.Ophthalmol.* 196:196-202

194. **Schultz,G.S., White,M., Mitchell,R. Brown,G., Lynch,J., Twardzik,D.R., Todaro,G.J.** 1987 Epithelial wound healing enhanced by transforming growth factor-α and vaccinia virus growth factor. *Science* 235:350-352

195. **Fourtanier,A.Y., Courty,J., Muller,E., Courtois,Y., Prunieras,M., Barntault,D.** 1986 Eye-derived growth factor isolated from bovine retina and used for epidermal wound healing *in vivo. J.Invest.Dermatol.* 87:76-80

196. **Lynch,S.E., Nixon,J.C., Colvin,R.B., Antoniades,H.N.** 1987 Role of platelet-derived growth factor in wound healing. *Proc.Natl. Acad.Sci.USA.* 84:7696-7700

197. **Brown,G.L., Nanrey,L.B., Griffin,J. Cramer,A.B., Yancey,J.M., Curtsingerlll,L.J., Holtzin,L., Schultz,G.S., Jurkiewicz,M.J., Lynch,J.B.** 1989 Enhancement of wound healing by topical treatment with epidermal growth factor. *N.Engl.J.Med.* 321:76-77

198. **Grotendorst,G.R., Martin,G.R., Pencer,D.** 1985 Stimulation of granulation tissue formation by platelet-derived growth factor in normal and diabetic rats. *J.Clin.Invest.* 76:2323-2328

199. **Robson,M.C., Phillips,L.G., Thomason,A., Robson,L.E., Pierce,G.F.** 1992 Platelet-derived growth factor BB for the treatment of chronic pressure ulcers. *Lancet* 339:23-25170.

200. **Wang,X-F., Lin,H.Y., Ng-Eaton,E., Downward,J., Lodish,H.F., Weinberg,R.A.** 1991 Expression cloning and characterization of the TGF-ß type III receptor. *Cell* 67:797-805

201. **Rechler,M.M.** 1985 The nature and regulation of the receptors for insulin-like growth factors. *Annu.Rev.Physiol.* 47:425-442

202. **Yarden,Y., Ullrich,A.** 1988 Growth factor receptor tyrosine kinases. *Ann.Rev.Biochem.* 57:443-447

203. **Ullrich,A., Schlessinger,J.** 1990 Signal transduction by receptors with tyrosine kinase activity. *Cell* 61:203-212

204. **Kraus,M.H., Issing,W., Miki.T., Popescu,N.C., Aaronson,S.A.** 1989 Isolation and characterization of ERBB3, a third member of the ERB/epidermal growth factor receptor family:Evidence for overexpression in a subset of human mammary tumors. *Proc.Natl. Acad.Sci.USA* 86:9193-9197

205. **Bottaro,D.P., Rubin,J.S., Faletto,D.L., Chan,A,M-L., Kmiecik,T.E., VandeWoude,G.F., Aaronson,S.A.** 1991 Identification of the hepatocyte growth factor receptor as the c-*met* proto-oncogene product. *Science* 251:802-804

206. **Huff,J.L., Jelinek,M.A., Borgman,C.A., Lansing,T.J., Parsons,J.T.** 1993 The protooncogene c-*sea* encodes a transmembrane protein-tyrosine kinase related to the Met/hepatocyte growth factor/scatter factor receptor. *Proc.Natl.Acad.Sci.USA.* 90:6140-6144

207. **Eide,F., Lowenstein,D.H., Reichardt,L.F.** 1993 Neurotrophins and their receptors-current concepts and implications for neurologic disease *Exp.Neurol.* 121:200-214

208. **Barbacid,M.** 1993 Nerve growth factor:a tale of two receptors. Oncogene 8:2033-2042

209. **Lamballe,F., Klein,R., Barbacid,M.** 1991 *trk*C, a new member of the *trk* family of tyrosine protein kinases, is a receptor for neurotrophin-3 *Cell* 66:967-979

210. **Massague,J.** 1992 Receptors for the TGF-ß family. *Cell* 69:1067-1070

211. **Morgan,D.O., Edman,J.C., Standring,D.N., Fried,V.A., Smith,M.C., Roth,R.A., Rutter,W.J.** 1987 Insulin-like growth factor II receptor as a multifunctional binding protein. *Nature (London)* 329:301-307

212. **Lopez-Casillas,F., Wrana,J.L., Massague,J.** 1993 Betaglycan presents ligand to the TGF-ß signalling receptor. *Cell* 73:1435-1444

213. **Lobel,P., Dahms,N.M., Breitmeyer,J., Chirgwin,J.M., Kornfeld,S.** 1987 Cloning of the bovine 215 kDa cation-independent mannose 6-phosphate receptor. *Proc.Natl.Acad.Sci.USA* 84:2233-2237

214. **Sambrook,J., Fritch,E.F., Maniatis,T.** 1989 Molecular Cloning, A laboratory Manual, 2nd Ed. Cold Spring Harbour Laboratory Press.

215. **Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A., Struhl,K.** Current Protocols in Molecular Biology. Jonn Wiley and Sons Inc.

216. **Birnboim,H.C., Doly,J.** 1979 A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1517

217. **Chomczynski,P., Sacchi,N.** 1987 Single-step method of RNA isolation by acid guanidinium-thiocyanate-phenol-chloroform extraction. *Anal.Biochem.* 162:156-159

218. **Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B., Erlich,H.A.** 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491

219. **Laemmli,U.K.** 1970 Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature (London)* 227:680-685

220. **Sanger,F., Nicklen,S., Coulson,A.R.** 1977 DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.USA.* 74:5463-5467

221. **Miller,J., Germain,R.N.** 1986 Efficient cell surface expression of class II MHC molecules in the absence of associated invariant chain. *J.Exp.Med.* 164:1478-1489

222. **Palfree,R.G.E., Sirlin,S., Dumont,F.J., Hammerling,U.** 1988 N-terminal and cDNA characterization of murine lymphocyte antigen Ly-6C.2. *J.Immunol.*

140:305-310

223. **Luckow,B., Schutz,G.** 1987 CAT constructions with multiple unique restriction sites for functional analysis of eukaryotic promoters and regulatory elements. *Nuc.AcidsRes.* 15:13

224. **Behr,J.P., Demeneix,B.A., Loeffler,J.P., Perez,J.** 1989 Efficient gene transfer into mammalian primary endocrine cells with lipopolyamine coated DNA. *Proc.Natl.Acad.Sci.USA.* 86:6982-6986

225. **Demeneix,B.A., Kley,N., Loeffler,J.P.** 1990 Differentiation to a neuronal phenotype in bovine chromaffin cells is repressed by protein-kinase C and is not dependent on c-*fos* oncoproteins. *DNA Cell Biol.* 9:335-345

226. **Basthel,F., Boutillier,A.L., Giraud,P., Demeneix,B.A., Behr,J.P., Loeffler,J.P.** 1992 Gene regulation analysis by lipopolyamine-mediated gene transfer in primary neurons. *Methods Neurosci.* 9:291-312

227. **Gorman,C.M., Moffat,L.F., Howard,B.H.** 1982 Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol.Cell.Biol.* 2:1044-1048

228. **Lathe,R.** 1985 Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J.Mol.Biol.* 183:1-12

229. **Altschul,S.F., Gish,W., Miller,W., Myers,E.W., Lipman,D.J.** 1990 Basic local alignment search tool. *J.Mol.Biol.* 215:403-410

230. **Mount,S.M.** 1982 A catalog of splice junction sequences. *Nucleic Acids Res.* 10:459-472

231. The Qiagenologist Application Protocols, 3rd ed. 1990 Qiagen Inc., Chatsworth, CA

232. **Kozak,M.** 1984 Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. *Nucleic Acid Res.* 12:857-872

233. **Gyllensten,U.B.** and **Erlich,H.A.** 1988 Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc.Natl.Acad.Sci.USA.* 85:7652-7656

234. **von Heijne,G.** 1985 Signal sequences:the limits of variation. *J.Mol.Biol.* 184:99-105

235. **von Heijne,G.** 1983 Patterns of amino acids near signal-sequence cleavage sites. *Eur.J.Biochem.* 133:17-21

236. **Lazure,C., Seidah,N.G.,Pelaprat,D., Chretien,M.** 1983 Proteases and posttranslational processing of prohormones:a review. *Can.J. Biochem.Cell.Biol.* 61:501-515

237. **Kyte,J., Doolittle,R.F.** 1982 A simple method for displaying the hydropathic character of a protein. *J.Mol.Biol.* 157:105-132.

238. **von Heijne,G.** 1987 Sequence Analysis in Molecular Biology, Acadamic Press.

239. **Shimasaki,S., Koga,M., Esch,F., Cooksey,K., Mercado,M., Koba,A, Ueno,N., Ying,S.-Y., Ling,N., Guillemin,R.** 1988 Primary structure of the human follistatin precursor and its genomic organization. *Proc.Natl.Acad.Sci.USA.* 85:4218-4222

240. **Bevilacqua,M.P., Stengelin,S.,Gimbrone,M.A., Seed,B.** 1989 Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complementary regulatory proteins and lectins. *Science* 243:1160-1165

241. **Chou,P.Y., Fasman,G.D.** 1974 Conformational parameters for amino acids in helical, ß-sheets and random coil regions calculated from proteins. *Biochemistry* 13:211-221

242. **Gomez,F., Vandermeers,A., Vandermeers-Piret, M-C., Herzog,R., Rathe,J., Stievenart,M., Winand,J., Christophe,J.** 1989 Purification and characterization of five variants of phospholipase A_2 and complete primary structure of the main phospholipase A_2 variant in *Heloderma suspectum* (Gila Monster) venom. *Eur.J.Biochem.* 186:23-33

243. **Plowman,G.D., Green,J.M., Neubauer,M.G., Buckley,S.. McDonald,V.L., Todaro,G.J., Shoyab,M.** 1992 The epithelin precursor encodes two proteins with opposing activities on epithelial cell growth. *J.Biol.Chem.* 267:13073-13078

244. **Craik,C.S., Rutter,W.J., Fletterick,R.** 1983 Splice junctions: association with variation in protein structure. *Science* 220:1125- 1129

245. **Gentry,L.E., Nash,B.W.** 1990 The pro domain of pre-pro-transforming growth

factor β1 when independently expressed is a functional binding protein for the mature growth factor. *Biochemistry* 29:6851-6857

246. **Vlodavsky,I., Bashkin,P., Ishai-Michaeli,R., Chajek-Shaul,T., Bar-Shavit,R., Haimovitz-Friedman,A., Klagsbrun,M., Fuks,Z.** 1991 Sequestration and release of basic fibroblast growth factor. *Ann.N.Y.Acad.Sci.* 638:207-220

247. **Derynck,R., Jarrett,J.A., Chen,E.Y., Eaton,D.H., Bell,G.I., Assolan,R.K., Roberts,A.B., Sporn,M.B., Goeddel,D.V.** 1985 Human transforming growth factor-β complementary DNA sequence and expression in normal and transformed cells. *Nature (London)* 316:701-705

248. **Pennica,D., Nedwin,G.E., Hayflick,J,S., Seeburg,P.H., Derynck,R., Palladino,M.A., Kohr,W.J., Aggarwal,B.B., Goeddel, D.V.** 1984 Human tumour necrosis factor:precursor structure, expression and homology to lymphotoxin. *Nature (London)* 312:724-729

249. **March,C.J., Mosley,B., Larsen,A., Cerretti,D.P., Braedt,G., Price,V., Gillis,S., Henney,C.S., Kronheim,S.R., Grabstein,K., Conlon,P.J., Hopp,T.P., Cosman,D.** 1985 Cloning, sequence and expression of two distinct human interleukin-1 complementary DNAs. *Nature (London)* 315:641-647

250. **Hubbard,S.C., Ivatt,R.J.** 1981 Synthesis and processing of asparagine-linked oligosaccharides. *Ann.Rev.Biochem.* 50:555-583

251. **Freedman,R.B.** 1987 Folding into the right shape. *Nature (London)* 329:196-197

252. **Pugsley,A.P.** 1989 Protein Targetting (Academic, San Diego) pp126-168

253. **Burgess,T.L., Kelly,R.B.** 1987 Constitutive and regulated secretion of proteins. *Ann.Rev.Cell Biol.* 3:243-293

254. **Newcomb,J.R. Cresswell,P.** 1993 Characterization of endogenous peptides bound to purified HLA-DR molecules and their absence from invariant chain-associated-alphabeta-dimers. *J.Immunol.* 150:499-507

255. Catalogue of Cell lines and Hybrodomas, ed.7 1992, American Type Culture Collection, Rockville, MD.

256. **Li,D-D., Chien,Y-K., Gu,M-Z., Richardson,A., Cheung,H.T.** 1988 The age-

related decline in interleukin-3 expression in mice. *Life Sciences* 43:1215-1222

257. **Streuli,C.H., Schmidhauser,C., Kobrin,M., Bissell,M.J. Derynck,R.** 1993 Extracellular matrix regulates expression of the TGF-ß1 gene. *J.Cell Biol.* 120:253-260

258. **Baserga,R., Estensen,R.D., Peterson,R.O.** 1968 Cell proliferation in mouse kidney after the injection of a single injection of folic acid. *Lab.Invest.* 19:92-96

259. **Sleisenger,M.H., Fordtran,J.S.** eds. 1993 Gastrointestinal Disease:Pathophysiology, Diagnosis, and Management. W.B Saunders Company.

260. **McMenamin,M.A.S., McMenamin,D.L.S.** 1990 The emergence of animals, the cambrian breakthrough Columbia University Press.

261. **Lipton,S.A.** 1989 Growth factors for neuronal survival and process regeneration. *Arch.Neurol.* 46:1241-1248

262. **Breathnach,R., Chambon,P.** 1981 Organization and expression of eukaryotic split genes coding for proteins. *Annu.Rev.Biochem.* 50: 349-383

263. **McLauchlan,J., Gaffney,D., Whitton,J.L., Clements,J.B.** 1985 The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* 13:1347-1368

264. **Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S., Sharp,P.A.** 1986 Splicing of messenger RNA precursor. *Annu.Rev. Biochem.* 55:1119-1150

265. **Leff,S.E., Rosenfeld,M.G., Evans,R.M.** 1986 Complex transcriptional units:diversity in gene expression by alternate RNA splicing. *Annu.Rev.Biochem.* 55:1091-1117

266. **Sharp,P.A.** 1981 Speculations on RNA splicing. *Cell* 23:643-646

267. **Smith,M.W.** 1988 Structure of vertebrate genes:a statistical analysis implicating selection. *J.Mol.Evol.* 27:45-55

268. **Baba,T., Nemoto,H., Watanabe,K., Arai,Y., Gerton,G.** 1993 Exon/intron organization of the gene encoding the mouse epithelin/granulin precursor (acrogranin). *FEBS Lett.* 322:89-94

269. **Greenblatt,J.** 1991 Roles of TFIID in transcription initiation by RNA

polymarase II. *Cell* 66:1067-1070

270. **Smale,S.T., Baltimore,D.** 1989 The "Initiator" as a transcriptional control element. *Cell* 57:103-113

271. **Kim,S-J., Glick,A., Sporn,M.B., Roberts,A.B.** 1989 Characterization of the promoter region of the human transforming growth factor-ß1 gene. *J.Biol.Chem.* 264:402-408

272. **Jakobovits,E.B., Schlokat,U., Vannice,J.L., Derynck,R., Levinson,A.D.** 1988 The human transforming growth factor alpha promoter directs transcription initiation from a single site in the absence of a TATA sequence. *Mol.Cell.Biol.* 8:5549-5554

273. **Bonthron,D.T., Morton,C.C., Orkin,S.H., Collins,T.** 1988 Platelet-derived growth factor A chain: gene structure, chromosomal location, and basis for alternate mRNA splicing. *Proc.Natl.Acad. Sci.USA.* 85:1492-1496

274. **Ishii,S., Xu,Y., Roe,B.A., Merlino,G.T., Pastan,I.** 1985 Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. *Proc.Natl.Acad.Sci.USA.* 82:4920-4924

275. **Araki,E., Shimida,F., Uzawa,H., Mori,M., Ebina,Y.** 1987 Characterization of the promoter region of the human insulin receptor gene:evidence for promoter activity. *J.Biol.Chem.* 262:16186-16191

276. **Sober,H.A.** ed. 1968 CRC Handbook of Biochemistry:Selected data for molecular biology. The Chemical Rubber Co.

277. **Dynan,W.S.,Tjian,R.** 1983 Isolation of transcription factors that discriminate between different promoters recognised by RNA polymerase II. *Cell* 32:669-680

278. **Saffer,J.D., Jackson,S.P., Annarella,M.B.** 1991 Developmental expression of Sp1 in the mouse. *Mol.Cell.Biol.* 11:2189-2199

279. **Williams,T. Tjan,R** 1991 Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. *Genes Dev.* 5:670-682

280. **Imigawa,M., Chiu,R., Karin,M.** 1987 Transcription factor AP-2 mediates induction by two different signal-transduction pathways:protein kinase C and cAMP. *Cell* 51:251-260

281. **Martin,M.E., Piette,J., Yaniv,M., Tang,W.J., Folk,W.R.** 1988 Activation of the polyomavirus enhancer by a murine activator protein 1 (AP-1) homolog and two contigous proteins. *Proc.Natl.Acad.Sci.USA.* 85:5839-5843

282. **Wasylyk,C., Flores,P., Gutman,A., Wasylyk,B.** 1989 PEA3 is a nuclear target for transcription activation by non-nuclear oncogenes. *EMBO J.* 8:3371-3378

283. **Riggs,K.J., Merrell,K.T., Wilson,G., Calame,K.** 1991 Common factor 1 is a transcriptional activator which binds in the c-myc promoter, the skeletal α-actin promoter, and the immunoglobin heavy chain enhancer. *Mol.Cell.Biol.* 11:1765-1769

284. **Umek,R.M., Friedman,A.D., McKnight,S.L.** 1991 CCAAT-enhancer binding protein:a component of a differentiation switch *Science* 251:288-292

285. **Lee,C.Q., Yun,Y., Hoeffler,J.P., Habener,J.F.** 1990 Cyclic AMP responsive transcriptional activation of CREB-327 involves interdependent phosphorylated subdomains. *EMBO J.* 9:4455-4465

286. **Davidson,I., Xiao,J.H., Rosales,R., Staub,A., Chambon,P.** 1988 The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. *Cell* 54:931-942

287. **Akira,S., Isshiki,H., Sugita,T., Tanabe,O., Kinoshita,S., Nishio,Y., Nakajima,T., Hirano,T., Kishimoto,T.** 1990 A nuclear factor for IL-6 expression (NF-IL6) is a member of the C/EBP family. *EMBO J.* 9:1897-1906

288. **Hattori,M., Abraham,L.J., Northemann,W., Fey,G.H.** 1990 Acute-phase reaction induces a specific complex between hepatic nuclear proteins and the interleukin 6 response element of the rat α_2-macroglobulin gene. *Proc.Natl.Acad.Sci.USA.* 87:2364-2368

289. **Kerr,L.D., Miller,D.B., Matrisian,L.M.** 1990 TGF-ß1 inhibition of transin/stromelysin gene expression is mediated through a *fos* binding sequence. *Cell* 61:267-278

290. **Culotta,V.C., Hamer,D.H.** 1989 Fine mapping of a mouse metallothionein gene metal response element. *Mol.Cell.Biol.* 9:1376-1380

291. **Kitadai,Y., Yamazaki,H., Yasui,W., Kyo,E., Yokozaki,H., Kajiyama,G., Johnson,A.C., Pastan,I., Tahara,E.** 1993 GC factor represses transcription

of several growth factor/receptor genes and causes growth inhibition of human gastric carcinoma cell lines. *Cell Growth Differen.* 4:291-296

292. **Wagner,B.J., Hayes,T.E., Hoban,C.J., Cochran,B.H.** 1990 The SIF binding element confers sis/PDGF inducibility onto the c-*fos* promoter. *EMBO J.* 9:4477-4484

293. **Majello,B., Arcone,R., Toniatti,C., Ciliberto,G.** 1990 Constitutive and IL-6 induced nuclear factors that interact with the human C-reactive protein promoter. *EMBO J.* 9:457-465

294. **Shannon,M.F., Gamble,J.R., Vadas,M.A.** 1988 Nuclear proteins interacting with the promoter region of the human granulocyte/macrophage colony stimulating factor gene *Proc.Natl.Acad.Sci.USA.* 85:674-678

295. **Ness,S.A., Marknell,A., Graf,T.** 1989 The v-*myb* oncogene product binds to and activates the promyelocyte-specific *mim*-1 gene. *Cell* 59:1115-1125

296. **Orkin,S.H.** 1990 Globin gene regulation and switching:circa 1990 *Cell* 63:665-672

297. **Issemenn,I., Green,S.** 1990 Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. *Nature (London)* 347:645-650

298. **Wasylyk,C., Gutman,A., Nicholson,R., Wasylyk,B.** 1991 The c-*ets* oncoprotein activates the stromelysis promoter through the same elements as several non-nuclear proteins. *EMBO J.* 10:1127-1134

299. **Lo,K., Landau,L.R., Smale,S.T.** 1991 LyF-1, a transcription regulator that interacts with a novel class of promoters for lymphocyte specific genes. *Mol.Cell.Biol.* 11:5229-5243

300. **Murre,C., Schonleber, McCaw,P., Baltimore,D.** 1989 A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, myoD, and myc proteins. *Cell* 56:777-783

301. **Roesler,W.J., Vandenbark,G.R., Hanson,R.W.** 1988 Cyclic AMP and the induction of eukaryotic gene transcription. *J.Biol.Chem.* 263: 9063-9066

302. **Davidson,E.H., Jacobs,H.T., Britten,R.J.** 1983 Very short repeats and coordinate induction of genes. *Nature* 301:468-470

303. **Rasheed,S., Norman,G.L., Heidecker,G.** 1983 Nucleotide sequence of the rasheed rat sarcoma virus oncogene: new mutations. *Science* 221:155-157

304. **Noguchi,C.T., Bae,K.S., Chin,K., Wada,Y., Schechter,A.N., Hankins,W.D.** 1991 Cloning of the human erythropoietin receptor gene. *Blood* 78:2548-2556

305. **Garret,M., McHendry-Rinde,B., Spickofsky,N., Margolskee,R.F.** 1992 Isolation of a clone which induces expression of the gene encoding the human tumor necrosis factor receptor. *Gene* 215-222

306. **Lamb,P., Crawford,L.** 1986 Characterization of the human p53 gene. *Mol.Cell.Biol.* 6:1379-1385

307. **Abe,M., Kufe,D.** 1993 Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene. *Proc.Natl.Acad.Sci.USA.* 90:282-286

308. **Jelinek,W.R., Schmid,C.W.** 1982 Repetitive sequences in eukaryotic DNA and their expression. *Ann.Rev.Biochem.* 51:813-844

309. **Wells,R.D., Collier,D.A., Hanvey,J.C., Shimizu,M., Wohlrab,F.** 1988 The chemistry and biology of unusual DNA structures adopted by oligopurine/oligopyrimidine sequences. *FASEB J.* 2:2939-2949

310. **Wang,Z.Y., Duell,T,K.** 1992 An S1 nuclease-sensitive homopurine/homopyrimidine domain in the PDGF A-chain promoter contains a novel binding site for the growth factor-inducible protein EGR-1. *Biochem.Biophys.Res.Commun.* 188:433-439

311. **O'Neill,D., Bornschlegel,K., Flamm,M., Castle,M., Bank,A.** 1991 A DNA-binding factor in adult hematopoietic cells interacts with a pyrimidine-rich domain upstream from the human δ-globin gene. *Proc.Natl.Acad.Sci.USA.* 88:8953-8957

312. **Hoffman,E.K., Trusko,S.P., Murphy,M., George,D.L.** 1990 An S1 nuclease-sensitive homopurine/homopyrimidine domain in the cKi-*ras* promoter interacts with a nuclear factor. *Proc.Natl.Acad.Sci.USA* 87:2705-2709

313. **Johnson,A.C., Jinno,Y., Merlino,G.T.** 1988 Modulation of epidermal growth factor receptor proto-oncogene transcription by a promoter site sensitive to S1 nuclease. *Mol.Cell.Biol.* 8:4174-4184

314. **Postel,E.H., Mango,S.E., Flint,S.J.** 1989 A nuclease hypersensitive element of the human c-*myc* promoter interacts with a transcription initiation factor. *Mol.Cell.Biol.* 9:5123-5133

315. **Hay,N., Bishop,J.M., Levans,D.** 1987 Regulatory elements that modulate expression of human c-*myc*. *Genes Dev.* 1:659-671

316. **Van deWetering,M., Oosterwegel,M., Dooijes,D., Clevers,H.** 1991 Identification and cloning of TCF-1, a T lymphocyte specific transcription factor containing a sequence specific HMG box. *EMBO J.* 10:123-132

317. **Sudhof,T.C., Russel,D.W., Goldstein,J.C., Brown,M.S., Sanchez-Pescador,R., Bell,G.I.** 1985 Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science* 228:893-895

318. **Delgadillo-Reynoso,M.G., Rollo,D.R., Hursh,D.A., Raff,R.A.** 1989 Structural analysis of the uEGF gene in the sea urchin strongylocentrotus purpuratus reveals more similarity to vertebrate than to invertebrate genes with EGF-like repeats. *J.Mol.Evol.* 29: 314-327

319. **Johnston,G.I., Bliss,G.A., Newman,P.J., McEver,R.P.,** 1990 Structure of the human gene encoding granule membrane protein-140, a member of the selectin family of adhesion receptors for leukocytes. *J.Biol.Chem.* 265:21381-21385

320. **Ord,D.C., Ernst,T.J., Zhou,L-J., Rambaldi,A., Spertini,O., Griffin,J., Tedder,T.F.** 1990 Structure of the gene encoding the human leukocyte adhesion molecule-1 (TQ1,Leu-8) of lymphocytes and neutrophils. *J.Biol.Chem.* 265:7760-7767

321. **Collins,T., Williams,A., Johnston,G.I., Kim,J., Eddy,R., Shows,T., Gimbrone,M.A., Bevilacqua,M.P.** 1991 Structure and chromosomal location of the gene for endothelial-leukocyte adhesion molecule 1. *J.Biol.Chem.* 266:2466-2473

322. **Patthy,L.** 1987 Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* 214:1-7

323. **Yamada,Y., Liau,G., Mudryj,M., Obici,S., DeCrombrugghe,B.** 1984 Conservation of the sites for one but not another class of exons in two chick

collagen genes. *Nature (London)* 310:333-337

324. **Jones,W.K., Yu-Lee,L.Y., Clift,S.M., Brown,T.L., Rosen,J.M.** 1985 The rat casein multigene family:fine structure and evolution of the ß-casein gene. *J.Biol.Chem.* 260:7042-7050

325. **Barta,A., Richards,R.I., Baxter,J.D., Shine,J.** 1981 Primary structure and evolution of the rat growth hormone gene. *Proc.Natl.Acad.Sci.USA.* 78:4867-4871

326. **Rogers,J.** 1985 Exon shuffling and intron insertion in serine protease genes. *Nature (London)* 315:458-459

327. **Ny,T., Elgh,F., Lund,B.** 1984 The structure of the human tissue-type plasminogen activator gene:correlation of intron and exon structures to functional and structural domains. *Proc.Natl.Acad. Sci.USA.* 81:5355-5359

328. **Nagamine,Y.** 1984 cDNA and gene nucleotide sequence of porcine plasminogen activator. *Nucleic Acids Res.* 12:9525

329. **Bell,G.I. Quinto,C., Quiroga,M., Valenzuela,P., Craik,C.S., Rutter,W.J.** 1984 Isolation and sequence of a rat chymotrypsin-ß gene. *J.Biol.Chem.* 259:14265-14270

330. **Swift,G.H., Craik,C.S., Stary,J.S., Quinto,C., Lahaie,R.G., Rutter,W.J., MacDonald,R.J.** 1984 Structure of the two related elastase genes expressed in the rat pancreas. *J.Biol.Chem.* 259: 14271-14279

331. **Lonberg,N., Gilbert,W.** 1985 Intron/exon structure of the chicken pyruvate kinase gene. *Cell* 40:81-90

332. **Blake,C.** 1983 Exons-present from the beginning? *Nature (London)* 306:535-537

333. **Cordell,B., Bell,G.I., Tischer,E., DeNoto,F.M., Ullrich,A., Pictet,R., Rutter,W.J., Goodman,H.M.** 1979 Isolation and characterization of a cloned rat insulin gene. *Cell* 18:533-543

334. **Lomedico,P., Rosenthal,N., Efstratiadis,A., Gilbert,W., Kolodner,R., Tizard,R.** 1979 The structure and evolution of the two nonallelic rat preproinsulin genes. *Cell* 18:545-558

335. **Nishioka,Y., Leder,A., Leder,P** 1980 Unusual a-globin-like gene that has

cleanly lost both globin intervening sequences. *Proc.Natl.Acad.Sci.USA.* 77:2806-2809

336. **Trexler,M., Patthy,L.** 1983 Folding autonomy of the kringle 4 fragment of human plasminogen *Proc.Natl.Acad.Sci.USA.* 80:2457-2461

337. **Kristensen,T., Ogata,R.T., Chung,L.P., Reid,K.B.M., Tack,B.F.** 1987 cDNA structure of murine C4b-bnding protein, a regulatory component of the serum complement system. *Biochemistry* 26:4668-4674

338. **Hoffmann,W., Hauser,F.** 1993 The P-domain or trefoil motif:a role in renewal and pathology of mucous epithelia? *Trends Biochem.Sci.* 18:239-243

339. **Mitchell,P.J., Tjian,R.** 1989 Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245:371-378

340. **Lafyatis,R., Lechleider,R., Kim,S-J., Jakowlew,S., Roberts,A.B., Sporn,M.B.** 1990 Structural and functional characterization of the transforming growth factor-ß3 promoter. *J.Biol.Chem.* 265:19124-19136

341. **Okajima,A., Miyazawa,K., Kitamura,N.** 1993 Characterization of the promoter region of the rat hepatocyte-growth-factor/scatter-factor gene. Eur.J.Biochem. 218:113-119

342. **Johnson,P.F., McKnight,S.L.** 1989 Eukaryotic transcriptional regulatory proteins. *Annu.Rev.Biochem.* 58:799-839

343. **Reynolds,G.A., Basu,S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L., Luskey,K.L.** 1984 HMG CoA reductase:a negatively regulated gene with unusual promoter and 5' untranslated regions. *Cell* 38:275-285

344. **Ueda,K., Pastan,I., Gottesman,M.M.,** 1987 Isolation and sequence of the promoter region of the human multidrug resistance (P-glycoprotein) gene. *J.Biol.Chem.* 262:17432-17436

345. **Ishii,S., Merlino,G.T., Pastan,I.** 1985 Promoter region of the human harvey *ras* proto-oncogene:similarity to the EGF receptor proto-oncogene promoter. *Science* 230:1378-1381

346. **Jones,K.A., Kadonaga,J.T., Rosenfeld,P.J., Kelly,T.J., Tjian,R.** 1987 A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48:79-89

347. **Chodose,L.A., Baldwin,A.S., Carthew,R.W., Sharp,P.A.** 1988 Human CCAAT binding proteins have heterologous subunits. *Cell* 53:11-24

348. **Hooft van Huijsduijnen,R.A.M., Bollekens,J., Dorn,A., Benoist,C., Mathis,D.** 1987 Properties of a CCAAT-box binding protein. *Nucleic Acids Res.* 15:7265-7282

349. **Bienz,M.** 1986 A CCAAT Box confers cell-type-specific regulation on the *Xenopus* hsp70 gene in oocytes. *Cell* 46:1037-1042

350. **Rossi,P., Karsenty,G., Roberts,A.B., Roche,N.S., Sporn,M.B., deCrombrugghe** 1988 A nuclear factor-1 binding site mediates the transcriptional activation of a type 1 collagen promoter by transforming growth factor-ß. *Cell* 52:405-414

351. **Lee,H., Haslinger,A., Karin,M., Tjian,R.** 1987 Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature (London)* 325:368-372

352. **Curran,T., Franza,B.R.** 1988 Fos and Jun:The AP-1 connection. *Cell* 55:395-397

353. **Angel,P., Hattori,K., Smeal,T., Karin,M.** 1988 The *jun* proto-oncogene is positively autoregulated by its product, jun/AP-1. *Cell* 55:875-885

354. **Short,J.M., Wynshaw-Boris,A., Short,H.P., Hanson,R.W.** 1986 Characterization of the phosphoenolpyruvate carboxykinase (GTP) promoter regulatory region. *J.Biol.Chem.* 261:9721-9726

355. **Roesler,W.J., Vandenbark,G.R., Hanson,R.W.** 1988 Cyclic AMP and the induction of eukaryotic gene transcription. *J.Biol.Chem.* 263: 9063-9066

356. **Christophe,D., Cabrer,B., Bacolla,A., Targovnik,H., Pohl,V., Vassart,G.** 1985 An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream of the thyroglobulin gene. *Nucleic Acids Res.* 13:5127-5144

357. **Htun,H., Dahlberg,J.E.** 1988 Single strands, triple strands, and kinks in H-DNA. *Science* 241:1791-1796

358. **Kim,S-J., Park,P., Koeller,D., Kim,K.Y., Wakefield,L.M., Sporn,M.B., Roberts,A.B.** 1992 Post-transcriptional regulation of the human transforming growth factor-ß1 gene. *J.Biol.Chem.* 267: 13702-13707