MOLECULAR EVOLUTION AND ORIGIN OF TWO PEPTIDE SUPERFAMILIES

John E. McRory
B.Sc. (Honours), University of Victoria, 1990

A Dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in the Department of Biology

We accept this thesis as conforming to the required standard

Dr. N.M. Sherwood, Supervisor (Dept. of Biology)

Dr. G.O. Mackie, Departmental Member (Dept. of Biology)

Dr. F. Choy, Departmental Member (Dept. of Biology)

Dr. T.W. Pearson, Outside Member (Dept. of Biochemistry)

Dr. Penny Swanson, External Examiner
National Marine Fisheries Service, USA

Supervisor: Dr. Nancy M. Sherwood

Abstract

In mammals the glucagon and insulin superfamilies each include a number of structurally-related hormones. The origin of each superfamily has been the subject of debate for many years resulting in an hypothesis that each superfamily arose from a single ancestral gene encoding a bioactive molecule. During evolution of the vertebrates, this gene is thought to have duplicated and changed to encode the existing family members.

The glucagon superfamily is composed of nine members that have similar intron/exon structure, amino acid sequences and gene length. Two neuropeptides in this family, growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP) are of special interest in this thesis. In addition to GRF and PACAP, the superfamily is composed of vasoactive intestinal peptide (VIP), peptide histidine methionine (PHM), secretin, glucose-dependent insulin-inducing peptide (GIP), glucagon, and glucagon-like peptide (GLP)-I and -II. This thesis presents nucleotide sequence data from a protochordate (tunicate; *Chelysoma productum*), bony fish (catfish; *Clarias macrocephalus*) and bird (chicken; *Gallus domesticus*) to help in the interpretation of the evolutionary steps in the glucagon superfamily.

Molecular biological techniques were used in isolating the *grf/pacap* cDNA clones from tunicate, catfish and chicken and the genes from tunicate and chicken. It was observed that two family members, GRF and PACAP, are encoded by the same gene in tunicate, catfish and chicken, unlike in mammals where the two peptides are encoded by two separate genes. Therefore, the duplication that gave rise to the two genes must have occurred after the divergence of the reptilian/avian lineage from the mammalian lineage about 250 million years ago. The tunicate, a sister group to amphioxus and the vertebrates, but a taxon that evolved before amphioxus, contains one distinct gene that encodes GRF_{1-27}/$PACAP_{1-27}$ and a second gene encoding $glucagon_{1-27}$/VIP_{1-27}. These four peptides are short compared to their counterparts in mammals, but the biologically active core is present. The two tunicate cDNA clones have high nucleotide sequence identity (80%) suggesting a recent gene duplication. In addition, a partial gene was isolated for each cDNA. The gene organization shows that GRF_{1-27} and $PACAP_{1-27}$ are each on separate, but adjacent, exons in one gene; likewise $glucagon_{1-27}$ and VIP_{1-27} are on separate exons in the second gene. The degree of identity among the four exons suggests that two tunicate genes resulted from an exon duplication followed by a complete gene duplication. These data suggest that two ancestral tunicate genes are the progenitors from which the existing vertebrate superfamily arose.

The evolution of another group of peptides in the insulin superfamily was investigated also. The presence of a distinct insulin and insulin-like growth factor (IGF) within the protochordates (tunicates) was demonstrated using molecular techniques. This is the first report of a true insulin and IGF from an invertebrate. The amino acid sequence of insulin in tunicates is 64% identical to the amino acid sequence in human insulin, whereas tunicate IGF is 59% identical to both human IGF-I and IGF-II. The tunicate clones encode amino acids that have been shown in mammals to be essential for receptor binding, for determination of tertiary structure and for formation of disulfide linkages. Both mRNAs were found to be expressed in the protochordate brain, neural gland, heart and intestine by use of a reverse transcriptase/polymerase chain reaction (RT/PCR). *In situ* analysis confirmed that the *insulin* and *igf* mRNA synthesis occurs in neurons of the tunicate brain. The widespread expression pattern and high sequence identity between tunicate insulin and IGF (87%) may reflect their common origin.

It is clear that protochordates are a nodal point in the evaluation of two important peptide superfamilies. At least six hormones (GRF, PACAP, VIP, glucagon, insulin and IGF) identified in mammals come from a 600 million year lineage in which the peptides have become more distinct from each other in primary structure, length and tissue location.

Examiners:

Dr. N.M. Sherwood, Supervisor (Dept. of Biology)

Dr. G.O. Mackie, Departmental Member (Dept. of Biology)

Dr. F. Choy, Departmental Member (Dept. of Biology)

Dr. T.W. Pearson, Outside Member (Dept. of Biochemistry)

Dr. Penny Swanson, External Examiner
National Marine Fisheries Service, USA

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Abbreviations

aa	amino acids
ACTH	adrenocotricotropin hormone
AMP	adenosine monophosphate
ATG	a codon of nucleotide bases that initiate translation
bp	base pairs
c	chicken
cAMP	cyclic (adenosine 3',5'-cyclic monophosphate)
cDNA	complementary deoxyribonucleic acid
c f	catfish
CNS	central nervous sytem
Denhardt's solution	1% Ficoll, 1% polyvinylpyrrolidone, 1% bovine serum albumin
DIG	digoxigenin
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytocine 5'-triphosphate
dGTP	2'-deoxyguanine 5'-triphosphate
dTTP	2'-deoxythreonine 5'-triphosphate
dNTP	2'-deoxyribonucleoside 5'-triphosphate
E	embryonic day after fertilization
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
FSH	follicle-stimulating hormone
GH	growth hormone peptide
gh	growth hormone nucleotide sequence
GIP	glucose-dependent insulinotropic polypeptide (old name-gastric inhibitory polypeptide)
GLP	glucagon-like peptide
GnRH	gonadotropin-releasing hormone
GRF	growth hormone-releasing hormone
grf	growth hormone-releasing hormone nucleotide sequence
grf/pacap	growth hormone-releasing hormone/pituitary adenylate cyclase-activating polypeptide nucleotide sequence
HPLC	high-pressure liquid chromatography
IGF	insulin-like growth factor
igf	insulin-like growth factor nucleotide sequence
ILP	insulin-like peptide
i r	immunoreactive

IRR	insulin-related receptor
LH	luteinizing hormone
LHRH	luteinizing hormone-releasing hormone
LIRP	locust insulin-related peptide
αMSH	melanocyte-stimulating hormone
mRNA	messenger RNA
MOPS	3-[N-morpholino]propanesulfonic acid
NPY	neuropeptide Y
PACAP	pituitary adenylate cyclase-activating polypeptide
pacap	pituitary adenylate cyclase-activating polypeptide nucleotide sequence
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PHI	peptide histidine isoleucine
PHM	peptide histidine methionine
poly A+	RNA with adenosine tail
PRL	prolactin
PRP	PACAP-related peptide
PTH	parathyroid hormone
PTHrP	parathyroid hormone-related peptide
r	rat
RIA	radioimmunoassay
RNA	ribonucleic acid
RNase	ribonuclease
RT	reverse transcriptase
s	salmon
SSC	sodium chloride/sodium citrate
SDS	sodium dodecyl sulfate
Taq	*Thermus aquaticus* DNA polymerase
T_3	triiodothyronine
TAE	tris acetate EDTA
TBE	tris borate EDTA
TE	Tris EDTA
TGF	transforming growth factor
TRH	thyrotropin-releasing hormone
U	units
UTR	untranslated region
VIP	vasoactive intestinal peptide
vip	vasoactive intestinal peptide nucleotide sequence

Acknowledgments

I would like to extend all my appreciation to all that helped me in completing this thesis. There are to many to list, so here in my habit of briefness, is a shortened list for all who I deeply extend my appreciation: Robin , Kris, Dave P., Sandra, Dave L. and Imogen. However, I do want to thank the two people who taught me most how to survive in this crazy world. The first person was my supervisor and friend Nancy Sherwood who taught me how to exist in academia, had faith in my research abilities and most importantly for giving me a chance to prove myself. The second person I would like to acknowledge is my dad. My dad taught me how to solve my problems and most importantly, never give up, just back up and go at the problem a different way. He used to say, and I know now it was not a new saying, that nothing is impossible with curiosity and a good head on your shoulders. Thanks for everything folks. :>

"I am a great believer in luck, and I find the harder I work the more I have of it."
Thomas Jefferson (1743-1826)

CHAPTER 1

General introduction

In the middle of the last century, Claude Bernard proposed that some organs function by releasing their products into the circulation for transport to their effective destination (in: Robin 1979). This idea initiated the search for and definition of what we now refer to as hormones. Molecules of the classical endocrine system are defined as molecules that are produced by cells within a gland and/or organ and released into the bloodstream. These molecules act on distant, usually well-defined, target cells that express specific receptors either on the cell surface or within the cell.

Present day knowledge of hormone actions show that the targets and functions are far more varied than previously suspected. The boundaries between one path of communication and another are not concrete and some hormones may display alternative actions at different points in the life cycle of an organism, or even in different tissues. No longer are hormones thought to be produced in a single distinct organ and have a single function. Rather, hormones are now known to be produced in tissues other than traditional endocrine organs. Hormones may have roles during embryogenesis, growth and development long before the traditional target organ is functional.

The synthesis of hormones outside of their traditional production sites provided the first evidence for their changing roles. The secretions of the pancreatic islets, the thyroid, the adrenal and the pituitary were considered to be strictly-defined hormones. However, in the last few decades it has become clear that many hormones including glucagon, somatostatin and insulin are produced outside the pancreas and immunohistochemistry has been used to show their presence in invertebrates that lack a pancreas. Other examples of hormones that are produced in tissues external to the initial source include: triiodothyronine (T_3), which is found in the embryo long before the onset of embryonic thyroid function (Evans 1982); neurotransmitters and neuromodulators, which are present in both the adrenal glands and the brain; and adrenocorticotropin (ACTH) and growth hormone-releasing hormone (GRF), which are produced not only in the pituitary and hypothalamus, respectively, but are also produced in the placenta and gonads (Campbell and Scanes 1992).

Most hormones (factors) associated with growth also appear to be produced by tissues other than the traditional endocrine glands; and originally, peptide growth factors were referred to as "tissue" growth factors. For example, epidermal growth factor (EGF) was found in the submaxillary gland (Carpenter and Cohen 1979); insulin like-growth factors I and II (IGF-I and IGF-II) were extracted from the liver (Jones and Clemmons 1995); and transforming growth factors (TGFs) were originally purified

from neoplastic cells (Sporn *et al.*1986). However, it soon became evident that these particular growth factors were produced in many other sites in both developing and fully developed organisms. Therefore, one could speculate that a growth factor synthesized by cells outside a specialized gland is more likely to act on neighboring cells, thereby exhibiting a paracrine mode of action. Such an action may be quite different from its initial role of an endocrine agent in that it is not released into the blood. A growth factor may also act on the cell from which it is released, thereby acting in an autocrine fashion. Thus, by finding the production of growth factors outside the glands of traditional synthesis, we are discovering a variety of "new roles" for "old hormones".

Growth factors may have different roles in embryo and adult

One fascinating growth stage of vertebrates is embryogenesis. During this period, most types of intercellular communication are across small distances. Paracrine and autocrine factors may be advantageous for the embryo because their effects are short ranged and provide guidance for local growth. The process of embryonic induction is beginning to be understood as a result of the structural similarity of many developmental factors to known adult forms of the same growth factors. Perhaps the only difference between a substance used both as an embryonic inducer and an adult hormone is that in the

earlier case it is used in a very localized manner, such as in a gradient pattern, and in the latter situation it is distributed throughout the organism via the blood (i.e. IGF-II has differential expression and function in the embryo and adult).

It is worth emphasizing that hormones expressed not only early in embryogenesis, but also during fetal and adult life, may have a functional role and mode of action that changes as the organism develops. One can envision that a multifunctional peptide may act early in life either by paracrine, autocrine and/or juxtacrine modes to influence differentiation and growth, whereas later, its function may change to an endocrine function regulating metabolic steps. Comparative studies provide answers as to the origin of hormones, mechanisms of cellular interactions and functional adaptations that may be replicated in mammals. Observing hormones and growth factors from the point of view of both ontogeny and phylogeny gives scientists the best chance to understand the complexities of a full grown, functioning organism.

Growth factors have novel roles and nontraditional origins

Our view of growth factors has evolved considerably in the last few years. We have found that growth factors initially thought to be produced in one specific organ may in fact be produced in many other tissues and, in some cases, their functions are different in the new sites of production. This is

demonstrated by the human parathyroid hormone-related peptide (PTHrP) (Vasavada *et al.* 1993). Initially, the PTHrP and its cDNA were isolated from malignant tumors associated with the syndrome of humoral hypercalcemia of malignancy. However, PTHrP bears similarity to the parathyroid hormone (PTH) sequence and is believed to have arisen from a gene duplication. Initially, native PTHrP was shown to be produced from the parathyroid chief cell, but it soon became evident that PTHrP was produced also in a wide variety of cells and tissues such as the epidermis, central nervous tissue, and preterm myometrium. Thus, the role of PTHrP may change from one tissue to another. Within the parathyroid and epidermis, PTHrP appears to regulate calcium translocation, whereas in the lactating breast PTHrP affects calcium signaling (Vasavada *et al.* 1993). All actions appear to occur through the same receptor. However, the cell specific response is determined by the cell type to which PTHrP binds. The two actions of one growth factor are determined by the intracellular pathways of the cells involved. These steps are controlled by many factors.

One growth factor is growth hormone-releasing factor (GRF)

The central nervous system (CNS) and the environment interact directly and indirectly to affect the rate of growth of an organism. Although there are many factors that stimulate

growth, the CNS is the primary initiator of the growth cascade and subsequent maturation of the organism.

In 1960 Seymour Reichlin provided early evidence that the brain was responsible for the control of growth. His evidence that rats with hypothalamic lesions grew less well than the control animals led him to conclude that a factor within the brain controlled growth (Reichlin 1960a,b). Then, in 1964 rat hypothalamic extracts were shown to contain a substance that specifically causes the release of growth hormone from rat pituitary cells *in vitro* (Deuben and Meites 1964); this evidence of a growth hormone-releasing factor (GRF) initiated the search to isolate the hypothalamic releasing factor. Initially, several different groups isolated and partially purified a growth hormone-releasing hormone, but an amino acid sequence was not determined (Dhariwal *et al.* 1965: Frohman *et al.* 1971; Schally *et al.* 1971; Stachura *et al.* 1972). It was not until 1982 that the primary sequence of growth hormone-releasing factor was determined almost simultaneously by two groups (Guillemin *et al.* 1982; Rivier et al. 1982).

Vertebrates contain distinct GRF peptides

Approximately 14 years have passed since a growth hormone-releasing factor (GRF) was isolated and sequenced from a human pancreatic tumour. Rivier and co-workers (1982) found a 40-amino-acid GRF peptide with a free carboxy terminus within their tumour extract. In addition to a 40-amino-acid form,

Guillemin *et al.* (1982) found a 44-amino-acid, amidated GRF peptide as well as a 37-amino-acid peptide from a different single pancreatic tumour. In 1984 the hypothalamic form of GRF was sequenced and found to be identical to the pancreatic tumour sequence (Ling *et al.* 1984). In the subsequent years, the primary sequence of GRF from 12 vertebrate species has been identified. These 12 GRF sequences are from the human, rat (Speiss *et al.* 1983), mouse (Frohman *et al.* 1989; Suhr *et al.* 1989), hamster (Ono *et al.* 1994), cow, goat (Esch *et al.* 1983; Brazeau *et al.* 1984), sheep (Brazeau *et al.* 1984), pig (Böhlen *et al.* 1983), carp (Vaughan *et al.* 1992), salmon (Parker *et al.* 1993), catfish (McRory *et al.* 1995) and chicken (McRory *et al.* 1996). The peptide structures were determined by either isolating and sequencing the peptide or by deducing the amino acids from isolated cDNA or genomic clones. Of the 12 GRF sequences known, the peptide was purified by protein chemistry from 7 of the species and was deduced from the cDNA or gene sequence for the human, rat, mouse, chicken (Chapter 2), salmon and catfish (Chapter 4).

GRF peptides have distinct functions and locations

GRF directly stimulates growth hormone release through a distinct GRF receptor. The release of GH has been studied in a wide variety of mammals such as human, goat, rat and mouse. In addition, the effectiveness of mammalian GRF to release GH has been investigated in a number of nonmammalian species.

Human GRF has been shown to stimulate GH release in birds (Scanes and Harvey 1984), amphibians (*Rana perezi*) (Malagon *et al.* 1991), goldfish (Peter *et al.* 1984) and rainbow trout (Luo and McKeown 1989). Rat GRF stimulated GH secretion from hatchling and adult turtle pituitaries *in vitro* (Denver and Licht 1989; Denver and Licht 1991). Marchant and Peter (1989) observed that other neuropeptides such as gonadotropin-releasing hormone (GnRH) act as potent GH-releasing factors in goldfish. This observation casts some doubt on whether teleosts have a specific GRF. However, the isolation and sequencing of carp GRF showed that nonmammalian vertebrates (teleosts) do have a native GRF. Synthetic carp GRF, like its mammalian counterparts, potentiates the release of pituitary GH *in vitro* (Vaughan *et al.* 1992) .

GRF is expressed in extrahypothalamic tissues

The biological role of GRF at one level is centered around the release of GH. Immunoreactive GRF also has been observed in the placenta (Baird *et al.* 1985; Meigan *et al.* 1988) and, to a lesser extent, in the gastrointestinal tract, pancreas and adrenal (Christophe 1993). The identity and role of these extra-hypothalamic GRF-like peptides is somewhat controversial. Within the placenta, both high and low molecular weight GRF-immunoreactive peptides have been detected (Baird *et al.* 1985; Meigan *et al.* 1988). The low molecular weight placental GRF-like peptide co-eluted on the HPLC with hypothalamic $GRF_{1\text{-}43OH}$

providing indirect evidence that the structures are identical. Direct evidence that the placental GRF peptide structure is identical to the hypothalamic form was obtained by isolation of the *grf* cDNA from the placenta. Placental GRF appears to have an expression pattern that is developmentally regulated during gestation as placental *grf* mRNA is abundant from mid-pregnancy to birth (Suhr *et al.* 1989). It is interesting that along with the high *grf* mRNA levels in the placenta are elevated *gh* mRNA levels in the fetal pituitary. This occurs at a stage of development when neither the pituitary nor the median eminence portal capillaries have developed. This has led researchers to believe that fetal GH is under extrahypothalamic control, perhaps placental GRF.

GRF may also be involved in reproduction, by an autocrine, paracrine and/or endocrine path. GRF has been reported in human Leydig cells (Moretti *et al.* 1990), human ovary (Moretti *et al.* 1990; Bagnato *et al.* 1992), human follicular fluid (Moretti *et al.* 1989) and catfish and chicken ovary and testis (McRory *et al.* 1995; McRory *et al.* 1996). As in the placenta, the GRF peptide in the gonads is identical to the hypothalamic form. The mammalian *grf* mRNAs produced in the testis and ovary are a similar size, approximately 1750bp, but are significantly larger than the 700bp hypothalamic form. This increase in mRNA size is due to the extended 5'- and 3'-untranslated regions, the function of which is unknown.

The regulation of the GRF gene in the gonads may not be the same as in the brain. There is evidence that the testicular *grf* mRNA transcript is under the control of a gene promoter located about 10Kb upstream of the transcription start site (Berry and Pescovitz 1988). One function for GRF is the control of follicular maturation. Within the ovary, GRF has been shown to bind to rat granulosa cells and stimulate FSH-induced cAMP production and cell proliferation (Moretti *et al.* 1990). GRF also stimulates bovine granulosa cell proliferation, but does not affect progesterone production *in vitro* (Spicer and Stewart 1996). Therefore, the gonadal *grf* gene may be under the control of a different promoter and the different mRNA lengths may represent another example of tissue-specific regulation. This implies that gonadal *grf* mRNA is transcribed independently from the hypothalamic and placental *grf* mRNA.

Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) is a neuropeptide related to GRF

An unexpected discovery in mammals was the presence of a peptide structurally related to GRF. This peptide was named pituitary adenylate cyclase-activating polypeptide (PACAP) because of its ability to stimulate adenylate cyclase and hence increase cAMP in pituitary cell cultures. In 1989 Miyata isolated and purified a 38-amino-acid form of PACAP from sheep hypothalami. In the following year another shorter form encoding 27 amino acids was found to be identical in

comparison to the first 27 amino acids of $PACAP_{1-38}$ (Miyata *et al.* 1990). To date, a cDNA encoding PACAP has been isolated from human (Kimura *et al.* 1990; Ohkuba *et al.* 1992), sheep (Kimura *et al.* 1990), rat (Ogi *et al.* 1990), chicken (McRory *et al.* 1996), salmon (Parker *et al.* 1993) and catfish (McRory *et al.* 1995) brains. The gene encoding PACAP has been isolated from human (Hosoya *et al.* 1992), chicken (McRory *et al.* 1996) and salmon (Parker *et al* 1996).

The location and effects of PACAP suggest that one function is to stimulate release of pituitary hormones. PACAP immunoreactive fibers have been detected in the supraoptic and paraventricular nuclei and in both the external and internal zones of the median eminence (Köves *et al.* 1991). *In vitro* studies showed that PACAP caused an increase in the release of not only GH, but also of several other pituitary hormones (Hart *et al.* 1992). This effect may result from the action of PACAP to stimulate the accumulation of intracellular cyclic AMP from dispersed rat anterior pituitary cells (Propato-Mussafiri *et al.* 1992). Also, PACAP stimulated GH and PRL release from GH_3 tumour cells (Propato-Mussafiri *et al.* 1992) and enhanced the effect of gonadotropin-releasing hormone on LH release (Culler and Paschall, 1991). PACAP receptors in the pituitary are specific for PACAP and are distinct from those for GRF. It is PACAP type II receptors that bind PACAP on both somatotrophs and gonadotrophs (Murakami *et al.* 1995).

PACAP has an extrahypothalamic location and role

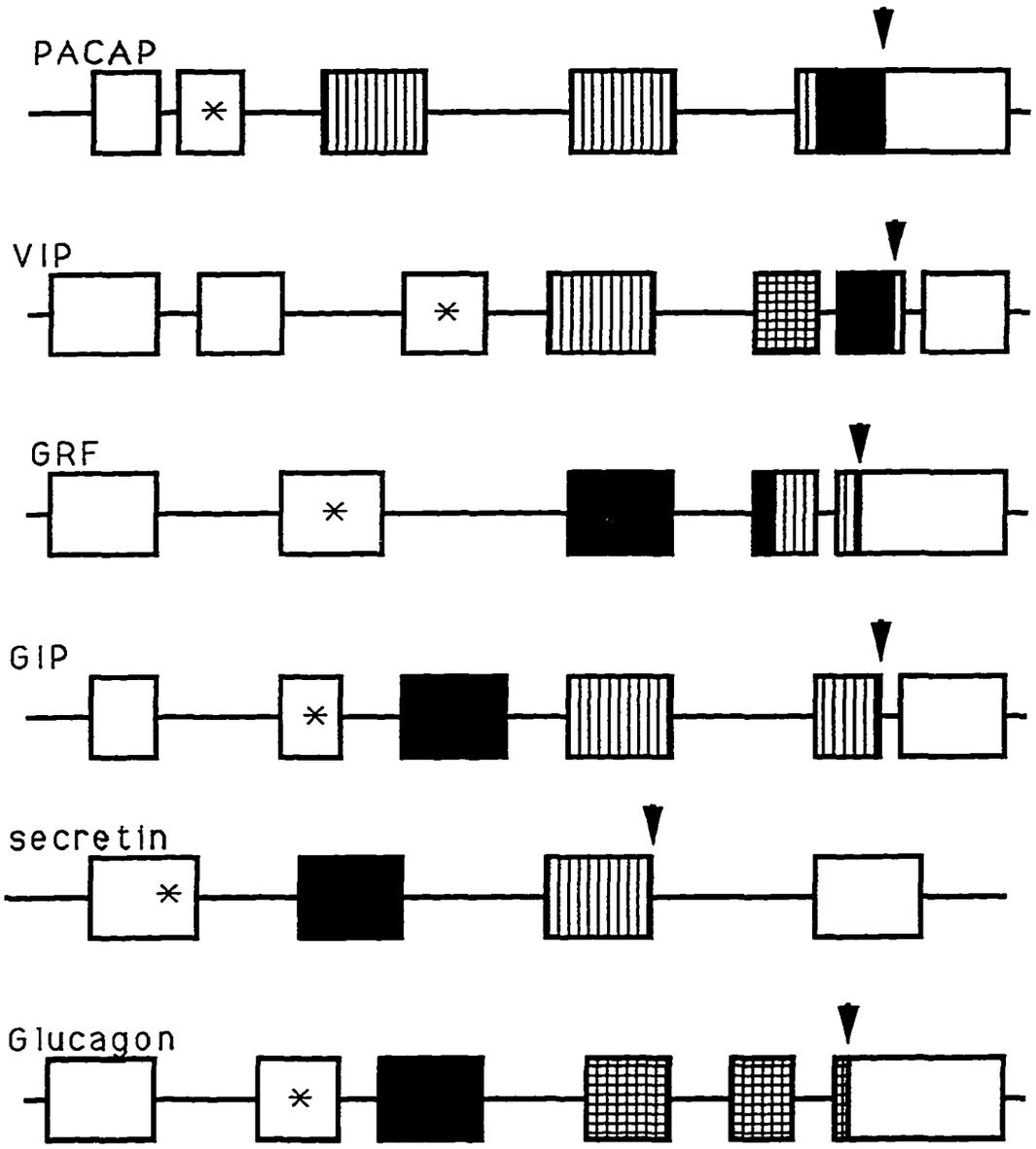
PACAP is assumed to have functions outside of the pituitary because it is detected by antisera in many locations: brain, testis, ovary, adrenal, pancreas, gastrointestinal tract, and respiratory tract (Köves *et al.* 1990; Gottschall *et al.* 1990; Uddman *et al.* 1991; Arimura 1992). In addition to the role of PACAP as a releaser of pituitary hormones, evidence suggests that PACAP acts as an early growth factor during development. PACAP has been shown to promote neuroblast growth in the cerebral cortex, cerebellar granule cells and sympathetic ganglia of fetal mice (DiCicco-Bloom 1992). Further evidence that PACAP is a potential growth factor is based on the following evidence: PACAP promotes proliferation of mouse primordial germ cells (Pesce *et al.* 1996); PACAP is produced in specific tumour cells; PACAP receptors are detected on human glial cell tumours (Robberecht *et al.* 1994); and PACAP, like GRF, is expressed early in rat brain development with levels peaking at birth (Masuo *et al.* 1994).

PACAP and GRF are members of the glucagon superfamily

The glucagon superfamily includes several families of peptides isolated from a wide assortment of animals. The family members found in humans are glucagon (Thomsen *et al.* 1972), secretin (Carlquist *et al.* 1985), vasoactive intestinal peptide (VIP) (Itoh *et al.* 1983), glucose-dependent insulin-

inducing peptide (GIP) (Brown and Dryburgh 1971) and pituitary adenylate cyclase-activating polypeptide (PACAP) (Miyata et al 1989) (Fig 1.1). The genes of the superfamily members all have similar intron/exon structure and encode similar amino acid sequences. GRF and PACAP are two neuropeptides of the glucagon superfamily that stimulate the release of pituitary GH. In mammals PACAP and PACAP-related peptide (PRP) are encoded on one gene, whereas GRF is on a separate gene. In nonmammalian vertebrates a gene encoding GRF alone has not been identified. Rather, GRF, the major candidate for the release of mammalian GH within mammalian systems, appears to be encoded in the PACAP gene in birds (McRory *et al.* 1996)

Figure 1.1: Schematic diagram of nine human glucagon superfamily members. The exons are shown by boxes and the introns are shown by lines. The bioactive peptide products of each gene are shown by a black box. In two genes, there is more than one bioactive product; PHM is encoded on the exon immediately 5' to the VIP exon and glucagon-like peptides are encoded on the exons 3' to glucagon (cross hatch). The cryptic peptide exons are shown by a box with vertical lines. The exons encoding 5' and 3'-untranslated regions are white boxes and the signal peptide exons have a star in the box. The arrow indicates the stop codon to signal the end of translation.


PACAP
VIP
GRF
GIP
secretin
Glucagon

and fish (Parker *et al.* 1993; McRory *et al.* 1995).

Insulin belongs in its own superfamily

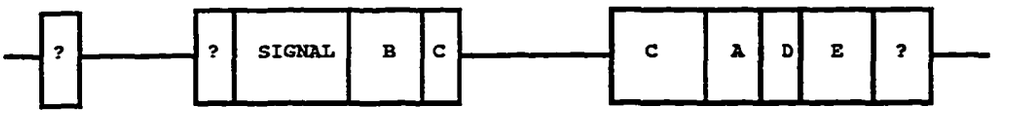
Another hormone family that is present in all vertebrates and is essential for growth and metabolism is the insulin superfamily consisting of insulin, insulin-like growth factor-I (IGF-I) and IGF-II (Figure 1.2). The structure of the insulin molecule has been highly conserved during vertebrate evolution (Chance *et al.* 1968; Blundell *et al.* 1975; Cutfield *et al.* 1979; Chan *et al.* 1981; Bajaj *et al.* 1983; Pollack *et al.* 1987). At the present time the primary sequence of insulin from over 50 vertebrate species is known. In addition, the sequence of the preproinsulin gene and cDNA from over 40 species has been determined (Genbank database).

Insulin is one of the key hormonal integrators of growth and metabolism and plays a similar role in all vertebrates. In the absence of insulin, severe metabolic imbalances occur because of the failure of many cells in the body to utilize glucose and amino acids normally. In humans the inability to metabolize glucose leads to diabetes mellitus associated with glucosuria, ketonuria, growth arrest, and a negative nitrogen balance. This ultimately leads to death from acute metabolic acidosis. Hence, the classic description of the body "melting down into urine" describes an acute condition of diabetes.

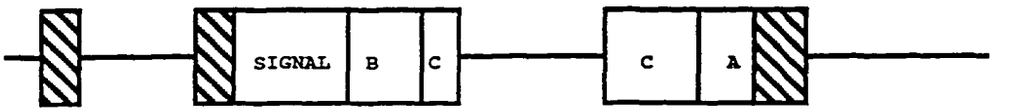
The discovery of insulin by Banting and Best in 1922 provided the first and only means to treat diabetes. In addition, insulin

Figure 1.2: Schematic for the gene structure of amphioxus insulin-like peptide, human insulin, insulin-like growth factor-I and insulin-like growth factor-II. Boxes with a question mark indicate a sequence was not found but a sequence is speculated to exist. Other regions are designated B, C, A, D, and E for their respective portion of each peptide. Boxes with diagonal lines indicate 5' and 3' untranslated regions.

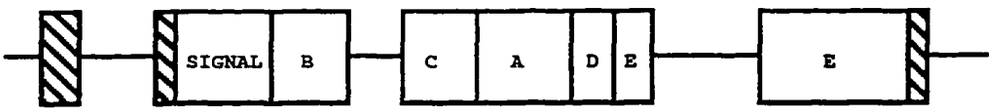
AMPHIOXUS INSULIN-LIKE PEPTIDE PARTIAL GENE



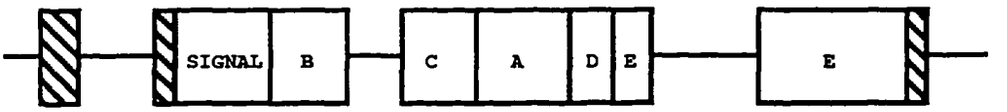
INSULIN GENE



INSULIN-LIKE GROWTH FACTOR-I GENE



INSULIN-LIKE GROWTH FACTOR-II GENE



was the first protein whose complete primary structure became known. Indeed identification of amino acid sequence substitutions in the molecule from different species helped to lay the conceptual foundations for the genetic code and for the molecular basis of evolution (Sanger, 1959).

The 6000 Dalton insulin molecule consists of two short peptide chains linked by two disulfide bridges. In almost all tetrapods the A chain is 21 amino acids in length and the B chain is 30 amino acids in length. The two chain hormone is derived from an intermediate precursor, proinsulin, which consists of the B and A chains linked to a connecting peptide, known as the C-peptide by adjacent pairs of dibasic residues. However, the initial translation product of the insulin mRNA is preproinsulin, which contains a N-terminal signal peptide of 24 amino acids in addition to the proinsulin. Preproinsulin is synthesized in the pancreatic ß cells located within the cell clusters known as the islets of Langerhans, which are dispersed throughout the pancreas in most vertebrates. A rapid cleavage of preproinsulin to proinsulin occurs in the rough endoplasmic reticulum (Palzelt *et al.* 1978), and fully folded, oxidized proinsulin is transported to the Golgi where it is packaged into storage granules along with a converting protease (Steiner *et al.* 1984). During the formation and maturation of the secretion granules, proinsulin is cleaved to liberate insulin and the C-peptide. The conversion process is not completely efficient because a small amount of proinsulin (about 1-2%) can be

detected in plasma extracts of several different species (Chance *et al.* 1968; Melani *et al.* 1970; Nolan *et al.* 1971). Insulin is stored in secretion granules and is liberated by several stimuli, although glucose is a predominant factor (Hedeskov 1980). Glucose not only stimulates the Ca^{2+}-dependent release of stored insulin, but also increases the synthesis of insulin mRNA (Neilsen *et al.* 1985; Mommsen and Plisetskaya 1991).

Insulin-like growth factor is also a member of the insulin superfamily

IGF-I is important for proper animal growth, tissue development, and differentiation (Jones and Clemmons 1995; Daughaday and Rotwein 1989; Froesch *et al.* 1985). IGF-I belongs to the insulin superfamily which includes insulin and IGF-II in a narrow sense, but includes relaxin, molluscan insulin-related hormone, and insect prothoracicotropic hormone in the broader sense. The release of IGF-I predominantly from the liver is mainly under the control of pituitary GH, as evidenced by an increased number of IGF-I transcripts in the liver and other tissues after administration of GH (Bichell *et al.* 1992; Foyt *et al.* 1992). It is IGF-I rather than GH that is thought to promote cell division and differentiation in extrahepatic tissues (Romagnolo *et al.* 1992). The *igf-I* cDNA sequence has been reported from a variety of animals including human (Jansen *et al.* 1983), rat (Casella *et al.* 1987; Murphy *et*

al. 1987), pig (Tavakkol *et al.* 1988), cow (Wong *et al.* 1989; Francis *et al.* 1988), chicken (Kajimoto and Rotwein 1989), frog (Kajimoto and Rotwein 1990), trout (Shamblott and Chen 1992), coho salmon (Cao *et al.* 1989; Duguay *et al.* 1992), chinook salmon (Wallis and Devlin 1993) and hagfish (Nagmatsu *et al.* 1991). Also, amphioxus has a cDNA sequence similar to both insulin and IGF-I, and thus may be ancestral to both peptides (Chan *et al.* 1990). The sequence identity of IGF-I among all species examined is very high; the identity between the human IGF-I protein sequence and that of salmon (Cao *et al.* 1989; Duguay *et al.* 1992), frog (Kajimoto and Rotwein 1989), and chicken (Kajimoto and Rotwein 1990) is 80%, 81% and 86%, respectively. Within mammals and lower vertebrates, IGF-I is produced predominately in the liver, but lower concentrations are found in muscle, spleen, fat, heart, testis, brain and kidney. To date, only hagfish have IGF-I expression that is limited to the liver (Nagamatsu *et al.* 1991).

IGF may be a nervous system growth factor

In addition to the role of IGF-I in the growth of tissue, IGF-I is believed to be important in the growth and development of the central nervous system (CNS). It is not known if IGF-I acts on the brain in an autocrine and/or paracrine manner or crosses the blood-brain barrier to act as an endocrine agent. Roles such as regulating neuronal and glial function (Baskin *et al.*1988), acting as a neurotrophic factor (Schwartz *et al.* 1992) and

modulating synaptic transmission (Schwartz *et al.* 1992) have been reported. Both full length and truncated forms of IGF-I have been purified from human fetal brain tissue (Sara *et al.* 1986). The truncated form of human fetal brain IGF-I, which lacks the first three amino acids, also has been isolated from porcine uterus (Ogasawara *et al.* 1989) and bovine colostrum (Francis *et al.* 1988); all have an amino acid sequence identical with the human liver form. The tripeptide Gly-Pro-Glu, which originates from the N-terminal of the intact IGF-I peptide, is a product of post-translational processing and functions within the CNS (Sara *et al.* 1986). Compared to the full length molecule, the truncated IGF-I lacking the tripeptide is 5-10 times more potent in its ability to stimulate DNA and protein synthesis and in competing for binding sites on the brain membrane (Sara *et al.* 1989). It has been suggested that the tripeptide itself may act as a neuroactive peptide because it has a potent stimulatory action on the release of acetylcholine (Sara *et al.* 1989).

Rationale

The research in my thesis addresses several major questions related to the molecular evolution of hormones. The first question concerns the evolutionary origin of specific hormones. Are specific hormones present only in vertebrates or are they also present in invertebrates? What is the oldest phylogenetic group where a specific hormone can be identified? The second

question is the mechanism by which hormone evolution occurred. Has a hormone family expanded in number in more recently evolved species and is the mechanism by exon and/or gene duplication? Are base substitutions or changes in exon/intron boundaries important? How rapid and stable are the changes? The third question concerns the evolution of the biological function. This is a broad and difficult question. My research strategy is to simply ask what portion of the known biologically active core of the hormone is conserved in evolution. The fourth question concerns whether an evolutionary change occurs in the tissue where the mRNA for a specific hormone is expressed. Are hormones expressed in many tissues in phylogenetically older animals? Does regulation of the control of expression increase in recently evolved animals? Are there novel tissues of expression that have been overlooked because we thought that each hormone would be expressed in a single organ? Finally, my research addresses the idea that some hormones may have novel functions. I specifically examine whether brain hormones that are known to release hormones from the pituitary gland may also have a role in the early development of the brain.

The second and third chapters of my thesis are an examination of the *grf/pacap* gene structure as an indication of where the *pacap* gene duplicated to encode two distinct genes, one encoding GRF and the other PACAP. To investigate this aspect of my thesis, I studied the *grf/pacap* gene expression in chicken tissues. Both GRF and PACAP are still encoded on the

same gene. However, expression of the chicken gene proved interesting because the gene undergoes exon sliding and skipping. Alternative expression of the mRNA in embryos and adults results in up to three different GRF peptides and one PACAP from the single gene. The observation that the two peptides PACAP and GRF are encoded on the same gene in birds suggests the gene duplication that gave rise to independent *pacap* and *grf* genes occurred in an ancestral mammalian lineage.

The fourth chapter of my thesis concerns the identification of the DNA that encodes GRF and PACAP in fish The Thai catfish *grf/pacap* cDNA shows the same gene organization as the chicken cDNA in which both GRF and PACAP are expressed in one gene. In addition to isolating the cDNA, I determined that tissue expression for the mRNA is in the testis, ovary, and intestine. This expression pattern is similar to the 2 mammalian and avian genes.

The fifth chapter of my thesis concerns isolating and sequencing two cDNAs from the tunicate. The first cDNA encoding tunicate $PACAP_{1-27}$ had 96% amino acid identity with all PACAP peptides, except chicken. This cDNA also encoded a GRF-like peptide with 59% amino acid identity with human GRF. The other clone was similar in nucleotide sequence to the *grf/pacap* cDNA, but encoded different peptides whose identity and exon location suggested that both glucagon and VIP are the result of a gene duplication. However, the timing of this duplication event is unknown.

In the sixth and seventh chapters of my thesis, I determined the sequence and tissue expression of hormones that are members of another superfamily. Chapter six of the thesis concerns isolating, sequencing and determining the expression of two different *igf-I* cDNAs from the Thai catfish (*Clarias macrocephalus*) and chapter seven concerns insulin and IGF from a protochordate (*Chelyosoma productum*). One form of IGF was unique because its expression was restricted to the brain, unlike the ubiquitous IGF that is expressed predominately in the liver and to a lesser extent in most other tissues. Isolation of the catfish brain-specific IGF was serendipitous, but exciting because expression of *igf* mRNA specifically within the CNS has never been shown. Brain-specific IGF supports the idea that IGF may be involved in the maintenance and growth of the nervous system.

Insulin and IGF have been flagships for understanding the structure of peptide hormones and their genes. I decided to investigate the presence of insulin and IGF in the protochordates (tunicates) because they are a sister group of the present day chordates who last shared a common ancestor about 600 million years ago. It is commonly believed that an ancient neotenous relative of tunicates gave rise to a chordate ancestor. Tunicates, because of their position on the evolutionary scale, provide a excellent organism in which to investigate the origin of insulin and IGF. I isolated from tunicate and sequenced two cDNAs encoding distinct insulin and

IGF. This result was unexpected because previous research showed a single hybrid *insulin/igf* cDNA was present in amphioxus, another sister group to the chordates, but one that evolved after the tunicates. My finding of a distinct insulin and IGF in a tunicate suggests that the ancestry of this superfamily precedes the vertebrates. Thus, the amphioxus insulin-like peptide is not the ancestor to both insulin and IGF but rather is most likely an IGF and suggests the amphioxus insulin exists but has not been isolated.

It is now evident that protochordates encode hormones that have sequence and structural identity to their mammalian counterparts. I have isolated and sequenced 4 tunicate cDNAs from two different hormone superfamilies that potentially encode six different peptides. With such a high degree of identity, it is obvious that the glucagon and insulin gene superfamilies precede the tunicates by a much longer time interval than previously thought.

References

Arimura A 1992 Pituitary adenylate cyclase-activating polypeptide (PACAP): discovery and current status of research. Regulatory Peptides 37: 287-304.

Bagnato A, Moretti C, Ohnishi J, Frajese G, Catt KJ 1992 Expression of the growth hormone-releasing hormone gene and its peptide product in the rat ovary. Endocrinology 130: 1097-1102.

Baird A, Wehrenberg WB, Böhlen P., Ling N 1985 Immunoreactive and biologically active growth hormone-releasing factor in rat placenta. Endocrinology 117: 1598-1601.

Bajaj M, Blundell TL, Pitts JE, Woods SP, Tatnell MA 1983 Dogfish insulin. European Journal of Biochemistry 135: 535-542.

Baskin DG, Wilcox, Figlewicz DP, Dorsa DM 1988 Insulin and insulin-like growth factos in the CNS. Trends in Neuroscience 11:107-111.

Berry SA and Pescovitz OH 1988 Identification of a rat GRF-like substance and its messenger RNA in rat testis. Endocrinology 123: 661-663.

Bichell DP, Kikuchi K, Rotwein P 1992 Growth hormone rapidly activates insulin-like growth factor I gene transcription in vivo. Molecular Endocrinology 6: 1899-1905.

Blundell TL, Wood SP 1975 Is the evolution of insulin Darwinian or due to selectively neutral mutations. Nature 257: 197-203

Böhlen P, Wehrenberg WB, Esch F, Ling N, Brazeau P, Guillemin R 1984 Rat hypothalamic growth hormone releasing factor: isolation, sequence analysis and total synthesis. Biochemical Biophysical Research Communications 125: 1005-1012.

Brazeau P, Böhlen P, Esch F, Ling N, Wehrenberg WB, Guillemin R 1984 Growth hormone-releasing factor from ovine and caprine hypothalamus: Isolation, sequence analysis and total synthesis. Biochemical Biophysical Research Communications 125: 606-614.

Brown JC and Dryburgh JR 1971 A gastric inhibitory polypeptide. II. The complete amino acid sequence. Canadian Journal of Biochemistry 49: 867-872.

Campbell RM, Scanes CG 1992 Evolution of the growth hormone-releasing factor (GRF) family of peptides. Growth Regulation 2: 175-191.

Cao Q-P, Duguay SJ, Plisetskaya E, Steiner DF, Chan SJ 1989 Nucleotide sequence and growth hormone-regulated expression of salmon insulin-like growth factor-I mRNA. Molecular Endocrinology. 3: 2005-2010.

Carlquist M, Jörvall H, Forssmann W-G, Thulin L, Johansson C, and Mutt V 1985 Human secretin is not identical to the porcine/bovine hormone. IRCS Medical Science 13: 217-218.

Carpenter G and Cohen S 1979 Epidermal growth factor. Annual Review of Biochemistry. 48: 193-216.

Casella SJ, Smith EP, Van Wykm JJ, Joseph DR, Hynes MA, Hoyt EC, Lund PK 1987 Isolation of rat testis cDNAs encoding an insulin-like growth factor-I precursor. DNA and Cell Biology 6: 325-330.

Chan SJ, Cao Q, Steiner DF 1990 Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. Proceedings of the National Academy of Science (USA) 87: 9319-9323.

Chan SJ, Kwok SCM, Steiner D 1981 The biosynthesis of insulin: some genetic and evolutionary aspects. Diabetes Care 4: 4-10.

Chance RE, Ellis RM, Bromer WW 1968 Porcine proinsulin; characterization and amino acid sequence. Science 161: 165-167.

Christophe J 1994 Type I receptores for PACAP (a neuropeptide even more important than VIP?). Biochemical and Biophysical Acta 1154: 183-199.

Culler MD, Paschall CS 1991 Pituitary adenylate cyclase activating polypeptide (PACAP) potentiates the gonadotropin-releasing activity of luteinizing hormone-releasing hormone. Endocrinology 129: 2260-2262.

Cutfield JF, Curfield SM, Dodson EJ, Dodson GG, Emdin SF, Reynolds CD 1979 Structure and biological activity of hagfish insulin. Journal of Molecular Biology. 132: 85-100.

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 Reviews 10: 68-91.

Denver RJ, Licht P 1989 Neuropeptides influencing in vitro pituitary hormone secretion in hatchling turtles. Journal of Experimental Zoology 251: 306-315.

Denver RJ, Licht P 1991 Several hypothalamic peptides stimulate thyrotropin and growth hormone secretion by adult turtle pituitary glands. Comparative Biochemistry and Physiology 100: 603-606.

Deubin RR, Meites J 1964 Stimulation of pituitary hormone release by a hypothalamic extract in vitro. Endocrinology 74: 408-414.

Dhariwal APS, Krulich L, Katz SH, McCann SM 1965 Purification of growth hormone-releasing factor. Endocrinology 77: 932-936.

DiCicco-Bloom EM 1992 Pituitary adenylate cyclase activating polypeptide (PACAP) potently stimulates mitosis, neurogenesis and survival in cultured rat sympathetic neuroblasts. Regulatory Peptides 37: 319.

Duguay SJ, Park LK, Samadpour M, Dickhoff WW 1992 Nucleotide sequence and tissue distribution of three insulin-like growth factor I prohormones in salmon. Molecular Endocrinology 6: 1202-1210

Esch F, Böhlen P, Ling N, Brazeau P, Guillemin R 1983 Isolation and characterization of the bovine hypothalamic growth hormone releasing factor. Biochemical Biophysical Research Communications 117: 772-779.

Evans RM, Birnberg NC, Rosenfeld MG. 1982 Glucocortocoids and thyroid hormones transcriptionally regulate growth hormone gene expression. Proceedings of the National Academy of Science (USA) 79: 7659-7663.

Foyt HL, Lanau F, Wolochack M 1992 Effect of growth hormone on levels of differentially processed insulin-like growth factor I mRNAs in total and polysomal mRNA populations. Molecular Endocrinology 6: 1881-1898.

Francis GL, Upton FM, Ballard FJ, McNeil KA, Wallace JC. 1988 Insulin-like growth factors I and 2 in bovine colostrum: sequence and biological activities compared to a potent truncated form. Journal of Biochemistry 251: 95-103.

Froesch ER, Schmidt C, Schwander J, Zapf J 1985 Actions of insulin-like growth factors. Annual Review of Physiology 47: 443-467.

Frohman LA, Maran JW, Dhariwal APS 1971 Plasma growth hormone respones to intrapituitary injections of growth hormone releasing factor (GRF) in the rat. Endocrinology 88: 1483-2488.

Frohman MA, Downs TR, Chomczynski P, Frohman LA 1989 Cloning and characterization of mouse growth hormone

releasing hormone (GRF) complementary DNA: Increase GRF messenger RNA levels in the growth hormone deficient lit/lit mouse. Molecular Endocrinology 3: 1529-1536.

Gottschall PE, Tatsuno I, Miyata A, Arimura A 1990 Characterization and distribution of binding sites for the hypothalamic peptide, pituitary adenylate cyclase activating polypeptide. Endocrinology 127: 272-277.

Guillemin R, Brazeau P, Böhlen P, Esch F, Ling N, Wehrenberg W 1982 Growth hormone releasing factor from a human pancreatic tumor that caused acromegaly. Science 218: 585-587.

Hart GR, Gowing H, Burrin JM 1992 Effects of a novel hypothalamic peptide, pituitary adenylate cyclase activating polypeptide, on pituitary hormone release in rats. Journal of Endocrinology 134: 33-41.

Hedeskov CJ 1980 The mechanisms of glucose induced insulin secretion. Physiological Reviews 60:442-509.

Hosoya M, Kimura C, Ogi K, Ohkubo S, Miyamoto Y, Kugoh H, Shimizu M, Onda H, Oshimura M, Arimura A, Fujino, M 1992 Structure of the human pituitary adenylate cyclase-activating polypeptide gene. Biochemical Biophysical Acta. 1129: 199-206.

Itoh N, Obata K, Yanaihara N, and Okamoto H 1983 Human preprovasoactive interstinal polypeptide contains a novel PHI-27-like peptide, PHM-27. Nature 304: 547-549.

Jansen M, van Schaik FMA, Ricker AT, Bullock B, Woods DE, Gabbay KH, Nussbaum AL, Sussenbach JS, Van den Brande JL Sequence of cDNA encoding human insulin-like growth factor I precursor. 1983 Nature 306: 609-611.

Jones JI, Clemmons DR 1995 Insulin-like growth factors and their binding proteins: biological actions. Endocrine Reviews 16: 3-34.

Kajimoto Y, Rotwein P 1989 Structure and expression of a chicken insulin-like growth factor I precursor. Molecular Endocrinology 3: 1907-1913.

Kajimoto Y, Rotwein P 1990 Evolution of insulin-like growth factor I (IGF-I): Structure and expression of an IGF-I precursor from *Xenopus laevis.* Molecular Endocrinology 4: 217-226.

Kimura C, Ohkubo S, Ogi K, Hosoya M, Itoh Y, Onda H, Jiang L, Dahl RR, Stibbs HH, Arimura A, Fujino 1990 A novel peptide which stimulates adenylate cyclase: Molecular cloning and characterization of the ovine and human cDNAs. Biochemical Biophysical Research Cummunications 166: 81-89.

Köves K, Arimura A, Gorcs TG, Somogyvari-Vigh A 1991 Comparative distribution of immunoreactive pituitary adenylate cyclase activating polypeptide and vasoactive intestinal polypepetide in rat forebrain. Neuroendocrinology 54: 159-169.

Köves K, Arimura A, Somogyvari-Vigh A, Vigh S, Miller J 1990 Immunohistochemical demonstration of a novel

hypothalamic peptide, pituitary adenylate activating polypeptide, in the ovine hypothalamus. Endocrinology 127: 264-271.

Ling N, Esch F, Böhlen P, Brazeau P, Wehrenberg W, Guillemin R 1984 Isolation, primary structure and synthesis of human hypothalamic somatocrinin: growth hormone releasing factor. Proceedings of the National Academy of Science (USA) 81: 4302-4306.

Luo D, McKeown BA 1989 Immunohistochemical detection of a substance resembling growth hormone-releasing factor in the brain of the rainbow trout (*Salmo gairdneri*) Experientia 45: 577-580.

Malagon NM, Castaño JP, Dobado-Berrios PM, Garcia-Navarro S, Garcia-Navarro F 1991 Human pancreatic growth hormone-releasing factor (1-44) stimulates GH cells in an anuran amphibian (*Rana perezi*). General Comparative Endocrinology 84: 461-469.

Marchant TA, Peter RE 1989 Hypothalamic peptides influencing growth hormone secretion in the goldfish, Carassius auratus. Fish Physiology and Biochemistry 7: 133-139.

Masuo Y, Tokito F, Matumoto Y, Shimamooto N, Fujino M 1994 Ontogeny of pituitary adenylate cyclase activating polypeptide (PACAP) and its binding sites in the rat brain. Neuroscience letters 170: 43-46.

McRory JE, Parker DB, Ngamvongchon S, Sherwood NM 1995 Sequence and expression of cDNA for pituitary adenylate

cyclase activating polypeptide (PACAP) and growth hormone-releasing hormone (GHRH)-like peptide in catfish. Molecular and Cellular Endocrinology 108: 169-177.

McRory JE, Parker RL, Sherwood NM 1996 Expression and alternative processing of a chicken gene encoding both growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP). Accepted for publication in DNA and Cell Biology.

Meigan G, Sasaki A, Yoshinaga K 1988 Immunoreactive growth hormone releasing hormone in rat placenta. Endocrinology 123: 1098-1102.

Melani F, Rubenstein AH, Oyer P, Steiner DF 1970 Identification of proinsulin and C-peptide in human serum by a specific immunoassay. Proceedings of the National Academy of Science (USA) 67: 148-155.

Miyata A, Arimura A, Dahl RR, Minamino N, Uehara A, Jian L, Culler MD, Coy DH 1989 Isolation of a novel 38 residue-hypothalamic polypeptide which stimulates adenylate cyclase in pituitary cells. Biochemical and Biophysical Research Communication 164: 567-574.

Miyata A, Jiang L, Dahl RD, Kitada C, Kubo K, Fujino M, Minamino N, Arimura A 1990 Isolation of a neuropeptide corresponding to the N-terminal 27 residues of a pituitary adenylate cyclase activating polypeptide with 38 residues. Biochemical and Biophysical Research Communication. 170: 643-648.

Mommsen TP, Plisetskaya EM 1991 Insulin in fishes and agnathans: History, structure and metabolic regulation. Reviews in Aquatic Sciences 4: 225-259.

Moretti C, Fabbri A, Gnessi L 1990 Immunohistochemical localization of growth hormone releasing hormone in human gonads. Journal of Endocrinological Investigations 13: 301-305.

Murakami Y, Koshimura K, Yamauchi K, Nishiki M, Tanaka J, Furuya H, Miyake T, Kato Y 1995 Pituitary adenylate cyclase activating polypeptide (PACAP) stimulates growth hormone release from GH3 cells through type II PACAP receptors. Regulatory Peptides 56: 35-40.

Murphy LJ, Bell GI, Duckworth ML, Friesen HG 1987 Identification, characterization, and regulation of a rat complementary deoxyribonucleic acid which encodes insulin-like growth factor-I. Endocrinology 121: 684-691.

Nagamatsu S, Chan SJ, Falkmer S, Steiner DF 1991 Evolution of the insulin superfamily. Journal of Biological Chemistry 266: 2397-2402.

Neilson DA, Welsh M, Casadaban MJ, Steiner DF 1985 Control of insulin gene expression in pancreatic β-cells and in an insulin-producing cell line RIN-5F cells. I) Effects of glucose and cyclic AMP on the transcription of insulin mRNA. Journal of Biological Chemistry 303: 11234-11239.

Nolan C, Margoliash E, Peterson JD, Steiner DF 1971 The structure of bovine proinsulin. Journal of Biological Chemistry 246: 2780-2795.

Ogasawara M, Karey KP, Marquardt H, Sirbasku DA 1989 identification and purification of truncated insulin-like growth factor I from porcine uterus. Evidence for high biological potency. Biochemistry 28: 2710-2721.

Ogi K, Kimura c, Onda H, Arimura A, Fujino M 1990 Molecular cloning and characterization of cDNA for the precursor of rat pituitary adenylate cyclase avtivating polypeptide (PACAP). Biochemical Biophysical Research Communications 173: 1271.

Ohkubo S, Kimura C, Ogi K, Okazaki K, Hosoya M, Onda H, Miyata A, Arimura A, Fujino M 1992 Primary structure and characterization of the precursor to human pituitary adenylate cyclase activating polypeptide. DNA and Cell Biology 11: 21-30.

Ono M, Miki N, Demura H, Tadokoro K, Nagafuchi S, Yamada M 1994 Molecular cloning of cDNA encoding the precursor for hamster growth hormone-releasing factor. DNA Sequence 5: 93-102.

Palzelt C, Larecgue AD, Duguid JR, Carrol RJ, Keim P 1978 Detection and kinetic behaviour of preproinsulin in pancreatic islets. Proceedings of the National Academy of Science (USA) 75: 1260-1264.

Parker DB, Coe IR, Dixon GH, Sherwood NM 1993 Two salmon neuropeptides encoded by one brain cDNA are structurally related to members of the glucagon superfamily. European Journal of Biochemistry 215: 439-448.

Parker DB, Powers ME, Swanson P, Rivier J, and Sherwood NM. 1996 Exon skipping in the gene encoding pituitary adenylate cyclase-activating polypeptide (PACAP) in salmon alters the expression of two hormone that stimulate growth hormone release. Endocrinology Accepted..

Pesce M, Canipari R, Ferri G-L, Siracusa G, De Felici M 1996 Pituitary adenylate cyclase-activating polypeptide (PACAP) stimulates adenylate cyclase and promotes proliferation of mouse primordial germ cells. Development 122: 215-221.

Peter RE, Nahorniak CS, Vale WW, Rivier JE 1984 Human pancreatic growth hormone-releasing factor (hpGRF) stimulates growth hormone release in goldfish. Journal of Experimental Zoology 231: 161-163.

Pollack HG, Kimmel JR, Hamilton JW, Rouse JB, Ebner KE, Lance V, Rawitch AB 1987 Isolation and structures of alligator gar insulin and pancreatic polypeptide. General and Comparative Endocrinology 67: 375-383.

Propato-Mussafiri R, Kanse SM, Ghatei MA, Bloom SR 1992 Pituitary adenylate cyclase activating polypeptide releases 7B2, adrenocorticotrophin, growth hormone and

prolactin from the mouse and rat clonal pituitary cell lines AtT-20 and GH3. Journal of Endocrinology 132: 107-113.

Reichlin S 1960a Growth and the hypothalamus. Endocrinology 67: 760-773.

Reichlin S 1960b Thyroid function, body temperature regulation and growth in rats with hypothalamic lesions. Endocrinology 66: 340-354.

Rivier J, Spiess J, Thorner M, Vale W 1982 Characterization of growth hormone releasing factor from a human pancreatic islet tumor. Nature 300: 276-278.

Robberecht P, Woussen-Colle MC, Vertongen P, De Neef P, Hou X, Salmon I, and Brotchi J 1994 Expression of pituitary adenylate cyclase-activating polypeptide (PACAP) receptors in human glial cell tumors. Peptides 15: 661-665.

Robin ED 1979 In *Claude Bernard and the Internal Environment.* Marcel D., ed. Academic Press. pp. 1-15.

Romagnolo D, Akers RM, Wong EA, Boyle PL, McFadden TB, Turner JD 1992 Overexpression of ovine insulin-like growth factor-I stimulates autonomous autocrine or paracrine growth in bovine mammary-derived epithelial cells. Molecular Endocrinology 6: 1774-1780.

Sanger F 1959 Chemistry of insulin. Science 129: 1340-1344.

Sara VR, Carlsson-Skwirut C, Andersson C, Hall E, Sjogren B, Holmgren A, Jornvall H 1986 Characterization of somatomedins from human fetal brain: Identification of a

variant form of insulin-like growth factor I. Proceedings of the National Academy of Science (USA). 83: 4904-4907.

Sara VR, Carlsson-Skwirut C, Bergman T, Jornvall H, Roberts PJ, Crawford M, Hakansson LN, Civalero I, Nordberg A 1989 Identification of GLY-PRO-GLU (GPE), the aminoterminal tripeptide of insulin-like growth factor I which is truncated in brain, as a novel neuroactive peptide. Biochemical and Biophysical Research Communications 165: 766-771.

Scanes CG, Harvey S 1984 Stimulation of growth hormone secretion by human growth hormone releasing factor (hpGRF) and thyrotropin releasing hormone (TRH) in anesthetized chickens. General and Comparative Endocrinology 56: 198-203.

Schally AV, Baba Y, Nair RMG 1971 The amino acid sequence of a peptide with growth hormone releasing activity isolated from the porcine hypothalamus. Journal of Biological Chemistry 246: 6647-6650.

Schwartz MW, Figlewicz DP, Baskin DG, Woods SC, Porte D 1992 Insulin in the brain: hormonal regulation of energy balance. Endocrine Reviews 13: 387-414.

Shamblott MJ, Chen TT 1992 Identification of a second insulin-like growth factor in a fish species. Proceedings of the National Academy of Science (USA) 89: 8913-8917

Shur ST, Rahal JO, Mayo KE 1989 Mouse growth hormone releasing hormone: Precursor structure and expression in brain and placenta. Molecular Endocrinology 3: 1693-1700.

Speiss J, Rivier J, Vale W 1983 Characterization of rat hypothalamic growth hormone-releasing factor. Nature 303: 532-534.

Spicer LJ, Stewart RE 1996 Interaction among bovine somatotrophin, insulin and gonadotropins on steroid production by bovine granulosa and thecal cells. Journal of Dairy Science 79: 813-820.

Sporn MB, Roberts AB, Wakefield LM, Associan RK. 1986 Transforming growth factor: Biological function and chemical structure. Science 233:523-534.

Stachura ME, Dhariwal APS, Frohman LA 1972 Growth hormone synthesis and release in vitro: effects of partially purified ovine hypothalamic extract. Endocrinology 91: 1071-1078.

Steiner DF, Docherty K, Carrol R 1984 Golgi granule processing of peptide hormone and neuropeptide precursors (a minireview) Journal of Biochemistry 24: 121-130.

Suhr ST, Rahal JO, Mayo KE 1989 Mouse growth hormone-releasing hormone: precursor structure and expression in brain and placenta. Molecular Endocrinology 3: 1693-1700.

Tavakkol A, Simmen FA, Simmen RCM 1988 Porcine insulin-like growth factor-I (pIGF-I): complementary deoxyribonucleic acid cloning and uterine expression of messenger ribonucleic acid encoding evolutionarily conserved IGF-I peptides. Molecular Endocrinology 2: 674-681.

Thomsen J, Kristiasen K, Brunfeldt K and Sundby F 1972 The amino acid sequence of human glucagon. FEBS Letters 21: 315-319.

Uddman R, Luts A, Arimura A, Sundler F 1991 Pituitary adenylate cyclase-activating polypeptide (PACAP), a vasoactive intestinal peptide (VIP)-like peptide in the respiratory tract. Cell and Tissue Research 265: 197-201.

Vasavada RC, Wysolmerski JJ, Broadus AE, Philbrick WM 1993. Identification and characterization of a GC-rich promoter of the human parathyroid hormone-related peptide gene. Molecular Endocrinology 7: 273-282.

Vaughan JM, Rivier J, Speiss J, Peng C, Chang JP, Peter RE, Vale W 1992 Isolation and characterization of hypothalamic growth-hormone releasing factor from common carp, *Cyprinus carpio*. Neuroendocrinology 56: 539-549.

Wallis AE, Devlin RH 1993 Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. Molecular Endocrinology 7: 409-422.

Wong EA, Ohlsen SM, Godfredson JA, Dean DM, Wheaton JE 1989 Cloning of ovine insulin-like growth factor-I cDNAs: heterogeneity in the mRNA population. DNA sequence 8: 649-657.

CHAPTER 2

Gene organization and expression of a chicken gene encoding both growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP).

A version of this chapter has been accepted by DNA and Cell Biology for publication: Expression and alternative processing of a chicken gene encoding both growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP), John E. McRory, Robin L. Parker and Nancy M. Sherwood

Summary

The chicken growth hormone-releasing factor (*grf*) gene was isolated, sequenced and characterized. In addition, three different mRNAs were isolated from juvenile and adult brain. The first cDNA encoded GRF_{1-46}, the second cDNA encoded GRF_{1-43} due to a sliding intron-exon boundary and the third cDNA skipped exon four and encoded GRF_{33-46}. We also determined that juvenile chicken *grf/pacap* mRNA is expressed in the brain and gonads, but not in the pituitary, heart, liver, kidney, crop, small intestine, large intestine, eye, and muscle. This gene is interesting in terms of evolution because another neuropeptide, pituitary adenylate cyclase-activating polypeptide (PACAP) is encoded within the same *grf* gene in chicken, but on different genes in mammals. We showed that these 2 neuropeptides are encoded also in the same cDNA in fish (chapter 4). The present evidence from chicken suggests a *grf/pacap* gene duplication may have occurred in stem mammals.

Introduction

Growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP), two members of the glucagon superfamily, are neuropeptides that stimulate the release of pituitary growth hormone (GH). In mammals, PACAP and PACAP-related peptide (PRP) are encoded on one gene, whereas GRF is on a separate gene. In nonmammalian vertebrates a gene encoding only GRF has never been identified. Rather, GRF, the major candidate for the release of mammalian GH within mammalian systems, appears to be encoded in the *pacap* gene, at least in fish (Parker *et al.* 1993; McRory *et al.* 1995).

The location and physiological role of PACAP are diverse and not fully understood (Rawlings and Hezareh 1996). Recent studies report that PACAP caused an increase in the release of GH, adrenocorticotropic hormone (ACTH) and luteinizing hormone (LH) (Hart *et al.* 1992). Further studies with cultured cells showed that PACAP stimulated follicle stimulating hormone (FSH) and melanocyte stimulating hormone (α-MSH) secretion from pituitary cells, GH and prolactin (PRL) from GH_3 tumour cells (Propato-Mussafiri *et al.* 1992) and enhanced the effect of gonadotropin-releasing hormone (GnRH) on LH and FSH release from rat anterior pituitary cells (Culler and Paschall 1991). PACAP receptors in the pituitary are distinct from those of GRF; PACAP binds to somatotrophs and gonadotrophs via PACAP type II receptors (Murakami *et al.* 1995).

Although a chicken GRF has never been reported, human GRF has been shown to initiate the release of GH from the pituitary of chicken (Scanes and Harvey 1984), turtle (Denver and Licht 1989, 1991), frog (Malagon *et al.* 1991), goldfish (Peter *et al.* 1984), and rainbow trout (Luo and McKeown 1989; Luo *et al.* 1990). In vitro, rat and human GRF stimulate GH release from both rat (Rivier *et al.* 1982; Vale *et al.* 1983) and chicken pituitary cells (Perez *et al.* 1987).• However, there is still controversy as to how avian GH release is controlled.

Many factors influence the release of GH from the avian pituitary. Thyrotropin-releasing hormone (TRH) appears to be physiologically important as a GH-releasing factor in birds (Harvey 1990). TRH is a modified tripeptide which, in addition to stimulating release of thyrotropin, has been shown to influence GH release in mammals, birds, fish, reptiles, and amphibians. Among vertebrates, the effect of TRH on GH release varies depending on species and physiological condition. For instance, TRH has been shown to stimulate GH release in hypothyroid rats and in many human pathological conditions. However, TRH did not affect GH release in healthy individuals (Guisti *et al.* 1986). Although TRH has little effect on conscious adult birds, it has been shown to stimulate GH release in immature and anaesthetized birds (Harvey 1990). Whereas TRH may play a role in stimulating GH release in birds, evidence indicates that regulation of GH release in birds is, in fact, similar to that in other vertebrates. For instance, rat and

human GHRH stimulate GH release from both rat pituitary cells (Rivier *et al.* 1982) and chicken pituitary cells (Perez *et al.* 1987). The release of GH from chicken pituitary cells was enhanced by the presence of dibutyryl cAMP, suggesting that cAMP acts as a second messenger in avian GH release, just as it does in mammals. These findings provide support for the presence of both GRF receptors on chicken pituitary somatotrophic cells and for a GRF molecule in the chicken brain.

To investigate the physiological roles of GRF and PACAP in the chicken, we isolated and sequenced the chicken *grf/pacap* gene and three different cDNAs. To determine the tissue distribution of *grf/pacap* mRNA, we used tissues from juvenile chickens; the tissues included brain, gonads, pituitary, heart, liver, kidney, crop, small intestine, large intestine, eye, and muscle. Exon sliding and exon skipping were analyzed by reverse transcriptase/polymerase chain reaction (RT/PCR) and subsequent sequencing of the isolated PCR products.

Materials and Methods

Identification of cDNA and gene using PCR

Chicken (*Gallus domesticus*) brains were removed, placed immediately in liquid nitrogen and stored at -80°C. Total RNA was extracted with an acidic guanidinium thiocyanate method (Chomczynski and Sacchi 1987), followed by purification of poly A+ rich mRNA on two consecutive oligo dT_{12-18} columns. Single stranded cDNA was synthesized with 10µg poly A+ rich mRNA, 2mM oligo dT_{20} (primer E), 5 X Superscript buffer, 2mM dNTP, 10mM DTT, 5U RNA guard (Pharmacia), and 200 U RT Superscript (BRL) to a final volume of 25µl. The reaction was heated to 42°C for 1.25hr and terminated by increasing the temperature to 95°C for 10 min.

Amplification was performed in a 50µl volume with 0.2µg cDNA, 5U Taq, 1x Taq buffer (Promega), 200mM dNTP's, 2.5mM $MgCl_2$, and 20 pmol of primers D (PA-1) (5'-catgtttggacagaacaca acgtgagcg) and F (3'-PA) (5'-cattcggatgggatcttcacggatag). The reaction was carried out for 35 cycles at 94°C for 1 min, 45°C for 1.5 min, 72°C for 1.5 min and for a 5.3 min extension at 72°C. Amplified bands were cloned into pBluescript KS+ (Stratagene), electroporated into XL-1 competent cells, and prepared for sequencing with an alkaline hydrolysis method (Birnboim 1983). Both strands were sequenced with [α-^{35}S] dATP using the USB Sequenase chain termination method (Sanger

et al. 1977) and CircumVent thermal cycle sequencing kit. (New England Biolabs). All sequencing gels were 6% polyacrylamide/7M urea wedge gels, dried under vacuum at 80°C and exposed to Kodak XAR-5 film for 12-24h.

Amplification of the 5' end

A modified version of Frohman's (1988) RACE protocol was utilized to amplify the 5' end of the chicken *grf/pacap* cDNA. To amplify the 5' end, 1μg Poly A+ mRNA was mixed with 10pmol primer D, and 7μl DEPC treated water to a final volume of 10μl, heated at 65°C for 5min, and then cooled on ice. Single stranded cDNA was synthesized with the above mRNA/primer mixture, 5μl Superscript buffer, 10mM dNTP, 10mM DTT, 5U RNA guard (Pharmacia), and 200 U RT Superscript (BRL) to a final volume of 25μl. The reaction was heated to 42°C for 1.25h and terminated by increasing the temperature to 95°C for 6min. The first strand synthesis was concentrated to 12.5 μl, of which 10μl was extended with dATP, 1μl water and 1μl TdT enzyme (BRL). PCR conditions were identical to the above except for the use of primers D and E' (oligo dT_{20}).

Tissue expression using reverse transcriptase/PCR

Brain, ovary/oviduct, testis, pituitary, heart, liver, kidney, crop, small intestine, large intestine, eye, and the muscle were removed from 25-day-old chickens and extracted in TriZol (BRL). Complementary DNA was synthesized from 1μg of total

RNA using 200U avian reverse transcriptase (H^- RT Superscript, BRL), 10mM DTT, 0.5mM each dNTP, 50U RNA guard, 2mM primer E, and 1X H^- RT buffer for a total reaction volume of 20µl. The reaction proceeded for 90min at 41°C followed by 10 min at 90°C. PCR amplifications were done with 0.5µl of newly transcribed single stranded cDNA from each tissue, 5U Taq DNA polymerase, 1x Taq buffer (Promega), 0.2mM each dNTP, 0.4mM of primers A (5'-gagccccgcccgtgcttaccgcag) and D (Fig. 2.1), and 2.5mM $MgCl_2$ in a 50µl reaction for 35 cycles (94°(1')-55°(1.5')-72°(1.75').

Identification of *grf/pacap* gene using genomic library screening

A total of 10^6 pfu from the chicken genomic library (Stratagene) were screened with the 294bp PCR cDNA fragment (primers D/F). Duplicate nylon membrane (BioRad) lifts were prehybridized at 50°C in 6x SSC, 5x Denhardt's solution, 0.5% SDS and 30mg/ml sea urchin DNA (blocking DNA) for 4 h. The hybridization solution, consisting of 6x SSC, 0.5% SDS, and 100mg blocking DNA, was added to the [α-^{32}P] dCTP (Dupont) labeled probe ($2.4x10^7$ cpm/ml) and incubated at 50°C overnight. The membranes were washed under high stringency (0.1xSSC/0.1%SDS) for 50 min at 50°C, then exposed to Kodak XAR-5 film for 7 days at -80°C.

A positive clone was subjected to three additional rounds of screening to purify the clone. The insert was removed from the

phage DNA with Sac1 and subcloned into pBluescript KS (Stratagene). Three of the four Sac1 subclones were subjected to nested deletions (Pharmacia) on both strands followed by sequencing of the two strands, as to manufacturer's instruction.

Gene copy number by Southern blot analysis

Chicken liver DNA was ground and digested with proteinase K (Sigma) overnight at 55°C. The DNA was purified with three subsequent extractions using phenol:chloroform:isoamyl alcohol alcohol (24:24:1) and one additional extraction with chloroform:isoamyl alcohol (24:1). The DNA was dialyzed against TE (pH 8) overnight. 10μg DNA was digested with either EcoRI, Sac1, HindIII, PvuII, and KpnI and electrophoresed through a 1% agarose gel, in duplicate. The DNA was transferred to the alkaline Zeta-Probe GT membrane as to manufacturer's (BioRad) specification. Prehybridization was in 7% SDS, 0.25M $NaHPO_4$, pH 7.2, and 1mM EDTA at 65°C for 15min. Hybridization was in fresh prehybridization solution plus the randomly primed α-^{32}P[dCTP] labeled probe (294bp PCR product) for 17hr at 65°C. The hybridized membranes were rinsed under low stringency with a solution containing 5%SDS, 40mM $NaHPO_4$ and 1mM EDTA. The membranes were then washed for 45min at 65°C with fresh solution. The wash solution was raised to high stringency by changing the SDS concentration: 1% SDS, 40mM $NaHPO_4$, and 1mM EDTA. The membranes were washed twice for 45min at 65°C

with fresh solution. After washing, the membranes were sealed in plastic and exposed 24hr to Kodak Biomax film with intensifying screens at -80°C.

Results

Organization of the chicken *grf/pacap* gene

A single band of 294bp resulted from the first DNA amplification of chicken RNA/cDNA. This cDNA fragment was used to screen the chicken genomic library. One million clones were screened to produce a single plaque that hybridized to the probe. The lambda clone of approximately 12500bp, produced 4 fragments when digested with Sac1 (Figure 2.1). These smaller fragments were purified and subcloned into pBluescript KS (subclones 1.8, 3.1, 3.2, and 4.4).

Subclones 1.8, 3.1 and 3.2 contained 6526bp of a chicken *pacap* gene (Figure 2.2); clone 4.4 consisted of approximately 5Kb of 3' flanking region and therefore was omitted. Subclone 1.8 (1682bp) contained exons 3, 4, and 5 encoding the cryptic peptide, GRF, and PACAP, respectively. Exon 3 has 134bp and contains part of the cryptic peptide and the nucleotide reading frame that encodes a dibasic processing site (Lys-Arg) between the cryptic and GRF peptide. Exon 4 has 96 nucleotides that code for the initial 32 amino acids of chicken GRF. The final portion of GRF on exon 5 shows chicken GRF is unique in that it is 46 amino acids, the longest known GRF. On exon 5 immediately downstream of the coding region of chicken GRF_{1-46} is the coding region for $PACAP_{1-38}$ separated from GRF by a Lys-Arg processing site. The chicken $PACAP_{1-38}$ is identical to the mammalian form except for the second amino acid, which has an isoleucine substituted for an alanine. Clone3.1 (2160bp)

Figure 2.1: Illustration of the size, subclone orientation and organization of the gene and cDNA encoding the chicken GRF/PACAP prepropeptide. Exons deduced from the nucleotide sequence of the chicken *grf/pacap* cDNA are boxed. The order of the exons, as determined by the cDNA sequence, are 5'UTR, signal peptide, cryptic peptide, GRF, and PACAP followed by the 3'UTR. Below the gene is a schematic map of the chicken *grf/pacap* cDNA clone showing the primer hybridization sites. Primer E, and primer F were used for 3' RACE reactions and primer E' and D were used for the 5' RACE reaction. Primers A and D were used to detect *grf/pacap* tissue expression. Primer combinations A-D, B-D and C-D were used to detect *grf/pacap* mRNA in the chick embryo (Chapter 3)

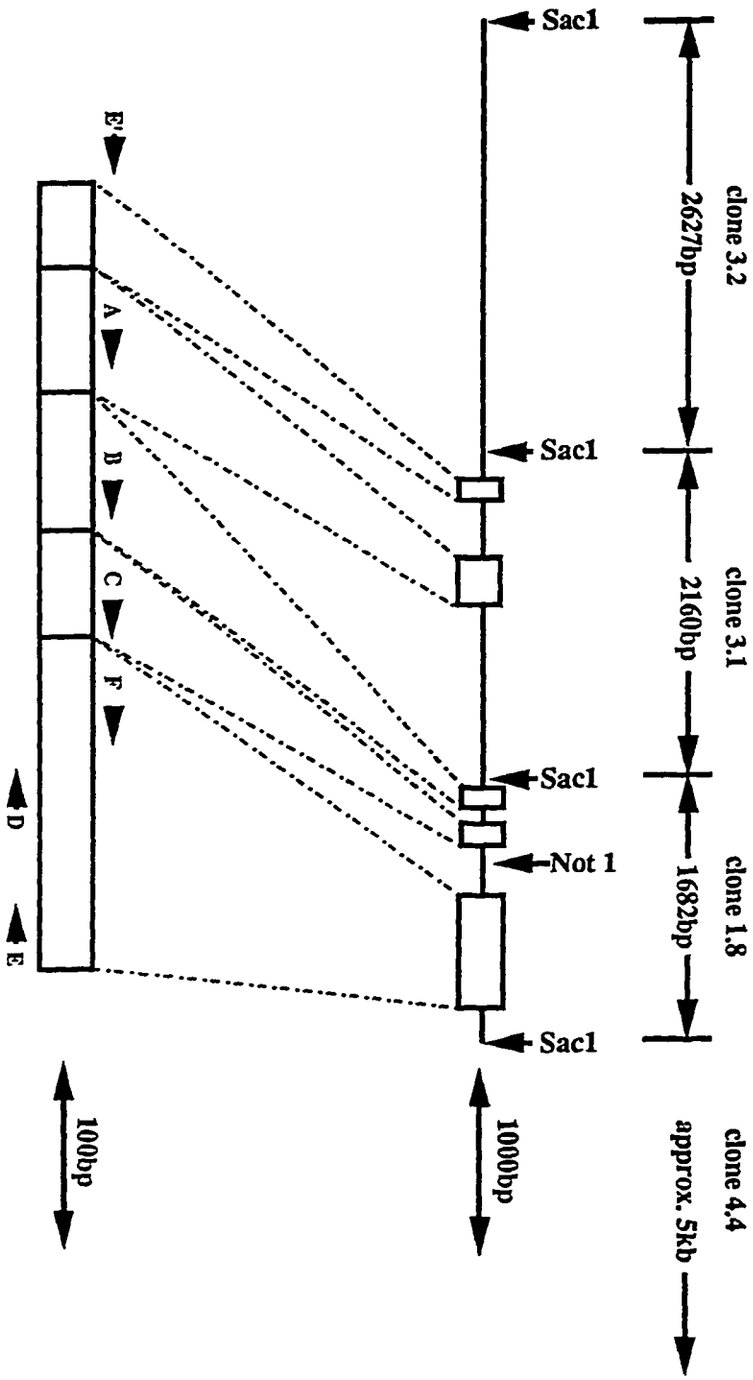

Sac1
Sac1
Sac1
Not 1
Sac1
E'
A
B
C
F
D
E
100bp
1000bp
clone 3.2
2627bp
clone 3.1
2160bp
clone 1.8
1682bp
clone 4.4
approx. 5kb

Figure 2.2: Nucleotide sequence of the chicken *grf/pacap* gene. Nucleotides comprising subclones 1.8, 3.1, and 3.2 are shown along with the intron/exon boundaries and the 5'- and 3'-flanking regions. The translated amino acid sequence is shown in the single letter code below the nucleotide sequence of coded exons and both sequences are numbered on the right. The nucleotide numbering begins at the beginning of the clone, whereas the amino acid numbering begins at the initiating methionine. All exons are in bold capital letters with an arrow to show the transcription start site. The first exon is composed completely of 5'UTR. The central portion of introns 2 and 4 is omitted to save space, but the complete sequence will be submitted to Genbank. The intron-exon boundary for nucleotides encoding GRF_{1-46} is shown (+). The other splice site, 9bp toward the 3' end, is shown by the double symbol (++). Within the promoter, the CAAT and TATAAA sequence motif have been underlined.

```
   1 tttttcccacagtaacaattctggttgaaataataaaaaggattatttttggatatgttaaatactgaaattttgattt  80
  81 ttggacttctggtgtaatttttttttcctgggggttccttgctaccaagtgtaagtatattattgacttttgaatccgatgg 160
 161 gcttttagaaaaaggagttaatttatatatttggggtgtttctctgagatatttcactccatgaaaacagatttcttcta 240
 241 agcctcagcaagacttgagatcaccctaaaatgtatgcattgttcttgttttccataagcatccttgagttagccttctc 320
 321 tgatgtttaaatgtagggaggtgaccagaatttgctctgagacacaatgacaaggaaggtatgagcacagaattacagaa 400
 401 agggaaaaaatacatattctactcagataaaataaagctgttgtcaataacatgatttattcaaactcctcatctatggg 480
 481 aagtaagtaactctgttctgaatatatctactgtctataactcgacagatcagttctgcagttcgtgttctggaccaggg 560
 561 attgcgtggggtgcactgggcacaaccacatctttgatttctttcttcaaattctcagtgtagaatgtttttatattacc 640
 641 tcaaaagcttgagatacaaggtaaatacaggggtataggtttaaagttctgtgtaagtgtggggaaaatatctccacatc 720
 721 acagagtttggagaaaaggcaatctgcatttgctgatgtgcacatacaaatttctatgggttcttcatgctacattaaaa 800
 801 gcttcacaaggaatctttccccaacttccagcgttgattagtgaacagcgctattagtcattacactcaatatacttggt 880
 881 gtccttccttgattttcctgaagaagcagcagtagggagagctgaatcacgagttttcctcatatttttcttaatgaaaa 960
 961 atcatttgcactgctgttctggaacacagcaccaggtaatgtcaacacattagaaagatgacaaaagttcccaatcacgg 1040
1041 gtctcgagtcagctccctttagtgaggttaattgagcttccaattcgaatatagtgagtcgtattacgcgcgctcactgg 1120
1121 ccgtcgttttacaacgtcgtgactgggaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgcca 1200
1201 gctggcgtatagcgaagaggcccgcaccgatcgcccaacccaacagttgcgcagcctgaatggcgaatgggacgcgccct 1280
1281 gtagcgctcattaagcgcggcgggaagctctaaatcgggggctcctttaggttccgattagtgctttacggcacctcga 1360
1361 ccccaaaaaaacttgattagggtgatggttcacgtagtggcctacgcctgatagacggtttttttcgccctttgacgtt 1440
1441 ggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatt 1520
1521 tataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaa 1600
1601 atattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggggcccctattgtttatttttctaatacattc 1680
1681 aaatatgtatccgctcatgagacaataaccctgataaatgcttcagtaatattgcataatttacaatggcgaagaaaact 1760
1761 atacatcccccccagacaatctagggtgtctctctacctccaagagatcgatgtcaagagatcgcctagaacatcgtat 1840
1841 tatctctttccggagaggcaagtatacaagaagaattaactcaatgtgaagaatacatcttagagggattatacgatga 1920
1921 ctcctaccctctgatctgggcttcttcctatcgcgcactctgagagaagcaacgagaattcagccctttataagagtgt 2000
2001 gattttagccaaaatgaattccatcccaccatcccaaaagtcggatactgtgggttcagtaccgtgcatcgatcgatcg 2080
2081 atggagcatgcatcgatgctagcatgcatgctagctagctattctttatttttttttttgtatgctagataaagctttcg 2160
2161 gtttggattttacgacgtcttgcatacgcactgaacttgacattgacatctctttgcacacctaaactaaaaataattaa 2240
2241 gcccattactcatcctcagagaaccaaggtaatcaaggagtctttgcaactgagcaaaagctgtccttcagaaggaatct 2320
2321 ttccccaacttccagcgttgattagtgaacagcgctattagtcattacactcaatatacttggtgtccttccttgatttt 2400
2401 cctgggggaagaagcaggagtagggggagaactgaatcacgagttttcctcatatttttcttaatgaaaaagctcattaat 2480
2481 ggatgctgttctggaacacaggaccaggtaatgtcaaaacattaaaaagatgacaaaagttcccaaggacggatctcgag 2560
2561 tcgactcacaaccagaggttgccaggacattgcaggactgaactttctctttcaaccctccaatctattttgagctctcc 2640
2641 agaggaaatgggatttgtccacagtaaccactgtggctgagataataaactgcattactctctggatatgtttaaaatac 2720
2721 tgagactttaatatgatttttgcatctcagtgtaatttcttttccgatgctagcatcgatgctagctgggcgggttatca 2800
2801 tgcccaacgtcgtagctgtgataaaaaaaaatgacgtctgtgctgtagctgatcgatgcatgcatgaataaaaaaagtgt 2880
```

```
2881 gtgactccgtgctgatgctgtgcttggggctttcttgctaccaagtgtaagtgctatgtgagttgcagcttcgcatttgc 2960
                CAAT-box                          TATAA-box
2961 agactcctatgggcaattttttagaaaaaggagttaatttaatataaatttggggtgtttctctgaagatatttcactcca 3040

3041 cagtgaaaacagatttcttctaagcctcagGCGAATATTGACAGCCCCCCTTTTTTTTCCTTTATTTGTCGAGTCGATTC 3120
3121 CCTAACCACCCAACAACTCTCTGCGCTTCTGCGCCTTCTTCATCCTTGCCCAGCGGAAAAGCCGGGAGCCCTTTGACTCT 3200
3201 TTCGGCCGCAACTTGGGGAGATAGCTCTATTTTTCCCCCCTCCTCTCTGGGGTTTTTCTCCTTTTTCCTCTCTCCCTTTC 3280
3281 CCTTCCGCAGCCACACGCTCTCAGTGCCGGGTGTCACAgtgtgtaaatcaagacttgaggatcaccctaaggtgtatgcc 3360
3361 ttgttcttgtttcagtagtacagagtgaatgaaaaaccactggataagcatgttgagttagcttctctgatttgggtgta 3440
3441 ggagtgacaagaatttgctctgagacacagGTTTC ATG AGT GGC AAT GTG TAT AAA ACG CTC TTA ACC  3508
   1                                 M   S   G   N   V   Y   K   T   L   L   T    11

3509 CTC CTG GTC TAT GGA TTA ATA ATG CAT TGC AAC GTC TAC TGC TCA CCC GAC CGT TGG ACT  3568
  12 L   L   V   Y   G   L   I   M   H   C   N   V   Y   C   S   P   D   R   W   T    31

3569 CCA GTA CCC GGC GCT AAG gtgagtctgtcagtgcaatatgctactctcacatcaggctctgtgtcacaagtcat 3642
  32 P   V   P   G   A   K                                                           37
3643 ctgccaatctatcagtgctgttaagtggaattactgagtaggtgcttggcccaccaaggctgagaatccagctgcagtgg 3722
3723 atcagcccatctacccccctgcacacacgtgtggattcaccccatcccctgccaaccctgccaccccatgctgccccacac 3802
                              ------------INTRON 2-------------
4763 aggaatgactcctcctttgctactcttatttccactgtatggggttaagaagactcgtcacgctgggctgagcactggag 4842
4843 cgagctcgctccgtcccgcacggtcccgcggcggggacggggctggggacgtccggctgagcccgcccgtgcttaccgca 4922
4923 g CTG GAG GAG GAG GTA TAC GAC GAG GAC GGG AAT ACC CTA CAG GAC TTC GCA CTA CGA     4980
  38   L   E   E   E   V   Y   D   E   D   G   N   T   L   Q   D   F   A   L   R      56

4981 GCA GGA GCC CCT GGG GGT GGC GGG CCG CGC CCG CGC TGG GGC AGG TGT ACG GCG CTG TAC  5040
  57 A   G   A   P   G   G   G   G   P   R   P   R   W   G   R   C   T   A   L   Y    76

5041 TAC CCG CCG GGA AAG AG gtgacagaggggcgccggatagggccgggggggggagggggggaatgggaaacctaag 5114
  77 Y   P   P   G   K   R                                                           82
5115 ggccccccgggggaggccgggaaatatcgtaatttccgccccacctgggctgcgcgagcgggggagggggggtggggagggag 5194
5195 ggcgcctcggggatgggcgctgacgggccgtgccccggcag G CAC GCC GAT GGG ATC TTC AGC AAA GCC  5263
  83                                          H   A   D   G   I   F   S   K   A    91

5264 TAC AGG AAA CTC CTG GGC CAG CTG TCC GCA AGA AAT TAC CTG CAC TCC CTG ATG GCC AAG  5323
  92 Y   R   K   L   L   G   Q   L   S   A   R   N   Y   L   H   S   L   M   A   K   111

5324 CGG GTC GG gtaagggctgcggcgggacgggagcgaacaaagcgcggcgcgcggcggccggggcggggcggcccattc 5400
 112 R   V   G                                                                       114
5401 tccccgcggtgctctgccggaacgagagaggcggccgcacccggggctcggcgtccctcccgcggggcagccccgggtgg 5480
                              ------------INTRON 4------------        +           ++
5641 cgggttgggtcgggcccgggagggcccctcctgatggttgtgtccttctcggtgctttgcag C GGT GCC AGC AGC 5715
 115                                                               G   A   S   S   118

5716 GGC CTG GGG GAC GAG GCG GAA CCG CTC AGC AAG CGC CAC ATA GAC GGC ATC TTC ACG GAC  5775
 119 G   L   G   D   E   A   E   P   L   S   K   R   H   I   D   G   I   F   T   D   138

5776 AGC TAC AGC CGC TAC CGG AAA CAA ATG GCT GTC AAG AAA TAC TTA GCG GCC GTC CTG GGG  5835
 139 S   Y   S   R   Y   R   K   Q   M   A   V   K   K   Y   L   A   A   V   L   G   158

5836 AAA AGG TAT AAA CAA AGA GTT AAA AAC AAA GGA CGC CGA GTA GCG TAT TTG TAG gatgagca 5897
 159 K   R   Y   K   Q   R   V   K   N   K   G   R   R   V   A   Y   L   *           176
5898 accgccgctgccgtgcgtagtcctgagagagagagagagagagagagagagattgagagagagagagagagagagaga 5977
5978 cccaaccaccccaacccaaacaaaagtcatttccaaagtgacggaacgaccgccgctcccgtgttccccaaacatgtatt 6057
6058 tatgtataagtaagccattaaatgaataatattttgataataatatggttttcttttgtacgaaagcacagatctacttt 6137
6138 gtggaccaatccttgagttatatatgagatagaatatatatatataatactgctactaaagagcgattcttcataccaag 6217
6218 ctgcaccaggacgagagttcgcctgagctgttagtttttatagaaaacaaatagacgaaaaaaaaaaaaaagacaatcac 6297
6298 cgcttccaacagcgctcctatttttgtaacggaaacgaaaagggcactgtttttattgccacgggggcgaacacctcagt 6377
6378 tctcaccgtgtgcgctgtgatagggagggctcacgcagcaggggtccccccggcctcgatctctctctctatttccccca 6457
6458 ccccccccttttttttttttccctttgattccggtcctatccgtatcagtcctcctcagagcgatgag             6525
```

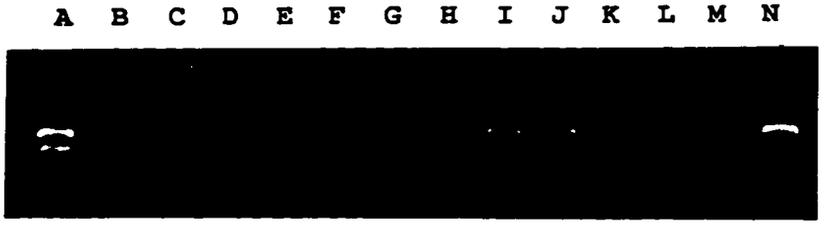
contained exon 1 that encodes the 5'-untranslated region (UTR) (194bp), an intron (142bp) and exon 2 that encodes the signal peptide and a portion of the cryptic peptide (114bp). Clone 3.1 also contained 344bp of regulatory region. Within the 344bp was a CAAT region at position 2974bp, and a TATAA region at position 3002bp; these nucleotides are consensus regulatory regions. Intron 2 of 1337bp was the longest intron and introns 3 of 178bp and 4 of 371bp contained an unusually high G/C content and numerous $(G)_n$ repeats. Clone 3.2 (2627bp) was exclusively a putative promoter region.

Expression of the *grf/pacap* gene in different tissues

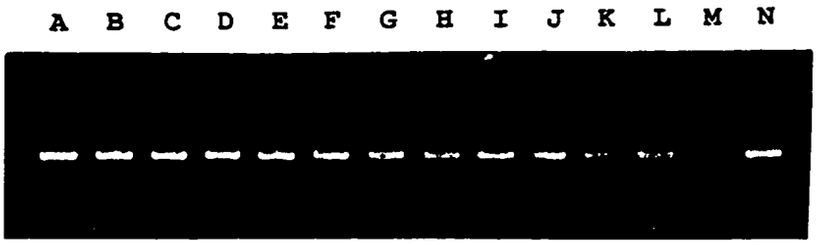
In juveniles at 25 days after hatching, chicken *grf/pacap* mRNA was detected not only in the brain, but also in tissues external to the brain using a RT/PCR method. Expression of *grf/pacap* mRNA was detected within the brain (lane A), ovary/oviduct (lane I) and testis (lane J) of the chicken (Figure 2.3A). Transcripts for *grf/pacap* were not detected within the pituitary, heart, liver, kidney, crop, small intestine, large intestine, eye, or the muscle. Two bands were amplified from the brain mRNA using the RT/PCR method. These two bands along with the single bands from the ovary/oviduct and testis were purified and sequenced to verify the PACAP sequence. The longest band in the brain and single band in gonads contained all exons, whereas the smaller band in the brain lacked exon 4. The cDNA appeared to be of good quality as determined by the PCR

Figure 2.3: Reverse transcriptase assay to detect *grf/pacap* mRNA in various chicken tissues. **(A)** Tissue cDNA samples used for reactions were as follows: lane A, brain; lane B, pituitary; lane C, heart; lane D, liver; lane E, kidney; lane F, crop; lane G, small intestine; lane H large intestine; lane I, ovary/oviduct; lane J, testis; lane K, eyeball; lane L, muscle; lane M, negative control (no DNA); lane N, positive control. **(B)** Tubulin detected by the RT/PCR assay to determine the quality of mRNA/cDNA for the assay in (A). The lanes contained the mRNA/cDNA as above except lane N which contained the tubulin clone as the positive control. The band size was 294bp as expected from primers PA1 and 3'PA.

A



B

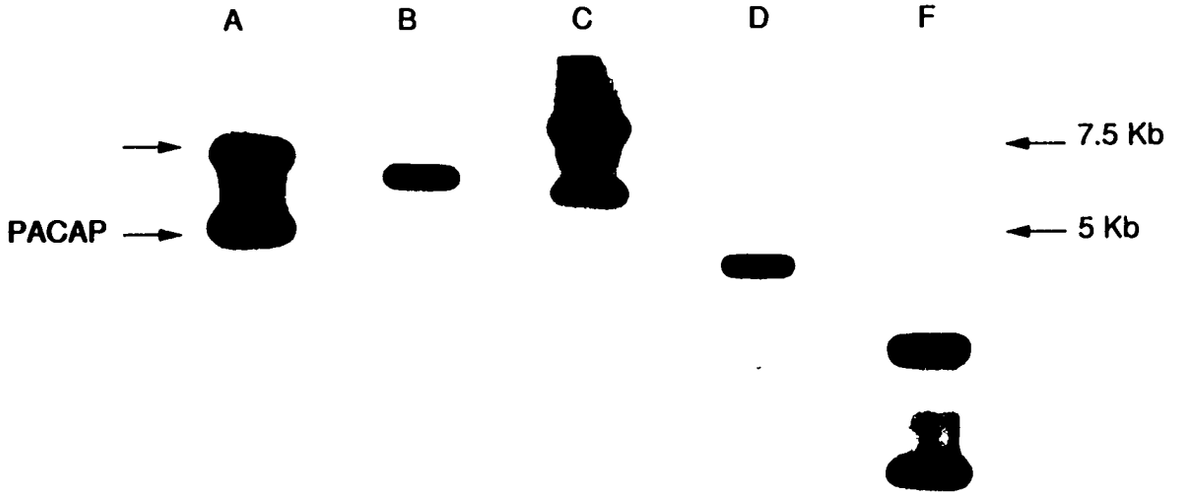


products obtained with tubulin primers (Figure 2.3B).

Determination of gene copy number using Southern blot analysis

Southern blot analysis of chicken genomic DNA using the 294bp PCR cDNA fragment as a DNA probe revealed two bands (Figure 2.4). All five genomic DNA restriction digests had two areas hybridizing to the cDNA probe. Both bands appeared when low and high stringency washes were applied to the membrane and no other bands appeared with low stringency washes.

Figure 2.4: Southern blot analysis of chicken genomic DNA using the 294bp *pacap* probe. 10μg DNA was digested with one of the following: EcoRI (lane A); Sac1 (lane B); HindIII (lane C); PvuII (lane D); and KpnI (lane F). The digest was electrophoresed through a 1% agarose gel. Approximate molecular sizes of 7.5Kb and 5Kb are shown on the right.


A
B
C
D
F
PACAP
7.5 Kb
5 Kb

Discussion

Gene organization implies a gene duplication in ancestral mammals

I have isolated from a chicken (c) genomic library a clone that encodes both a GRF peptide and PACAP. This is the first report of an avian GRF. Both PACAP and GRF belong to the glucagon superfamily in which the members have similar intron/exon organization and sequence identity. This organization in which both peptides are encoded on the same gene is similar to that in fish, but unlike mammals, which have two genes encoding each peptide separately.

The association of the chicken *grf/pacap* gene with members of a superfamily including glucagon, secretin, GRF, and vasoactive intestinal peptide (VIP) is illustrated by a high sequence identity between cPACAP and cVIP (Talbot *et al.* 1995). In the mRNA sequences, the nucleotides encoding the cPACAP region have 80% identity with the nucleotides encoding the cVIP region. This high degree of identity is likely to explain the two bands in each lane of the Southern blot (Figure 8) because our 294bp probe encoded cPACAP. A common ancestral gene may have given rise to PACAP, GRF and VIP, although the point in time at which the genes duplicated and diverged is still unknown (Campbell and Scanes 1992; Sherwood *et al.* 1994). The presence of both the cGRF molecule and cPACAP in a single gene implies that a gene duplication occurred after the ancestral mammals split from the reptilian-avian line and not

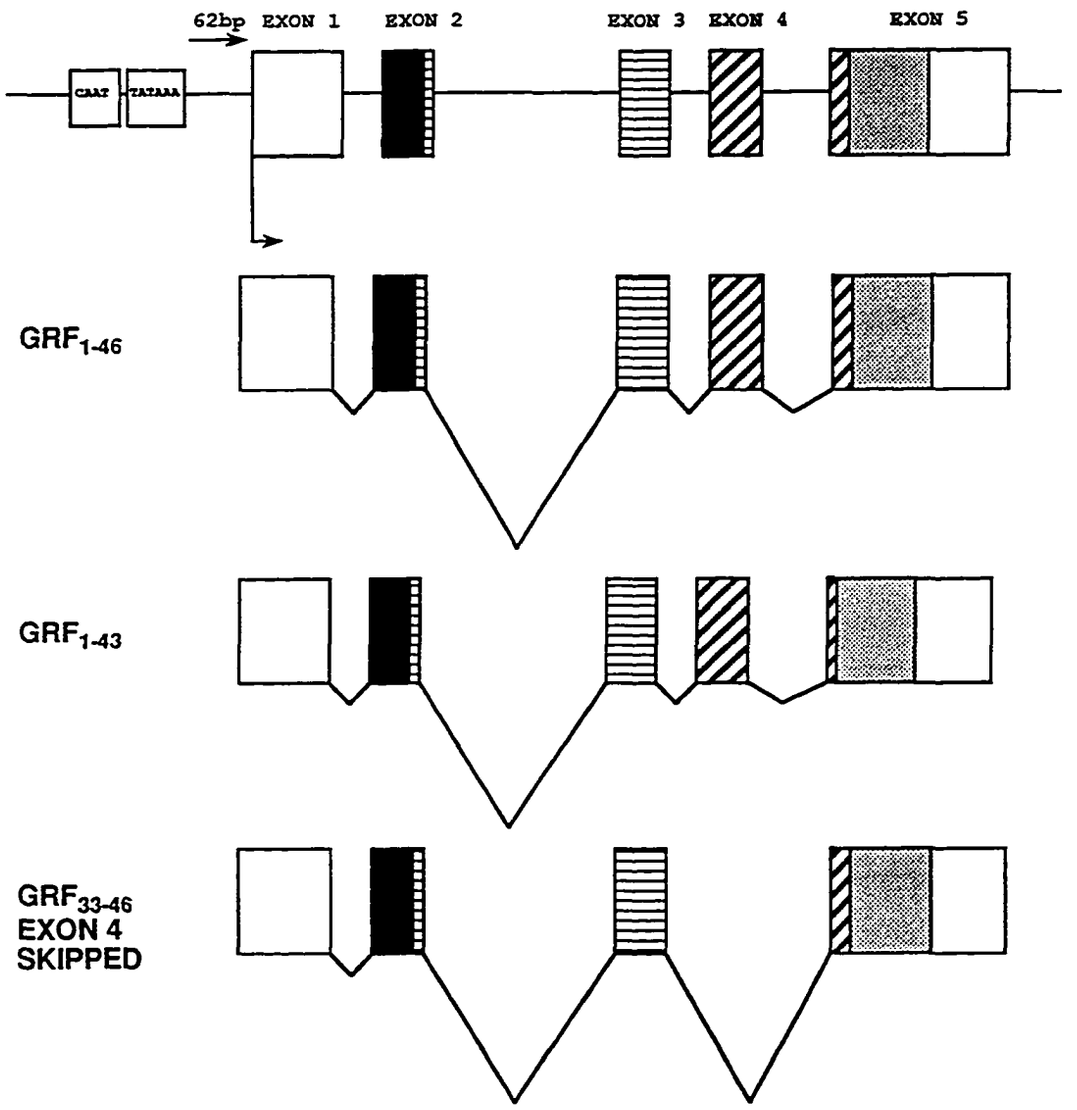
before the emergence of the bony fish as previously thought (Campbell and Scanes 1992).

There is a high degree of identity of cPACAP to other members within the PACAP family. The portion of DNA encoding the PACAP protein has nucleotides that are 92% identical to the human gene. The deduced PACAP amino acid sequence is 97% identical to the human sequence with only one change at position 2 where an isoleucine in chicken PACAP is substituted for an alanine in human PACAP. In contrast, the cGRF peptide compared to GRF in other species has only 42% amino acid identity to human, 47% to rat and 76% to carp GRF (Vaughan *et al.* 1992). This divergence among species is not surprising in view of the relatively low sequence identity of 68% between human and rat GRF.

Exon sliding and skipping produces 3 different mRNAs

The organization of the chicken *grf/pacap* gene is similar to both the human *grf* and *pacap* genes and to other members of the glucagon superfamily. The chicken *grf/pacap* gene is composed of 5 exons and as sequenced, encompassed 6525bp. All 5 exon locations and intron/exon boundaries were confirmed by isolating cDNA clones from the 5' and 3' RACE reactions with adult brain cDNA (Figure 2.5). However, in sequences of the PCR fragments, we observed that the intron/exon boundary between exons 4 and 5 has considerable variation (Fig. 2.5). The first

Figure 2.5: Schematic diagram of alternative splicing for the chicken *grf/pacap* gene to produce three different mRNA's. The gene is shown on top with the three different mRNAs shown below. The first mRNA was the most abundant and encodes GRF_{1-46}. The second mRNA has the splice site 9bp downstream encoding GRF_{1-43}. The third encodes a mRNA that skips exon four and therefore omits the nucleotides that encode GRF_{1-32}.


62bp
EXON 1
EXON 2
EXON 3
EXON 4
EXON 5
CAAT
TATAAA
GRF1-46
GRF1-43
GRF33-46
EXON 4
SKIPPED

boundary occurs at position 5703bp; the second boundary slides 9bp downstream to position 5712; and the third boundary shows that exon 4 is lacking altogether. At both splice sites nine bases apart, proper consensus splice sites exist. Therefore, the chicken *grf/pacap* mRNA transcript has splice donor sites that encode a 43-amino-acid GRF. The acceptor site was also shown to slide 9bp upstream to encode a GRF of 46 residues. The intron nucleotides at the 5' splice site of intron 4, GG:GT(A) and the last 4 nucleotides of intron 4 at both 3' splice sites (NCAG:C), closely match the splice site consensus sequences as found in vertebrates (Padgett *et al.* 1986; Green 1991). This sliding mechanism has not been reported for this family of peptides.

The function of exon sliding is not known other than to encode two GRFs of different length with, potentially, two different functions. Recent evidence suggests that human GRF_{1-44} may affect somatotroph differentiation in the embryonic chicken pituitary (Porter *et al.* 1995) and the development of chick neuroblasts and their neurotransmitters (Kentori and Vernadakis 1990). These effects on early brain development and GH-releasing somatotrophs in the pituitary may reflect an early role of GRF_{1-43} and/or GRF_{1-46} in avian systems.

The third mRNA transcript synthesized is a cDNA for chicken *grf/pacap* that lacks exon 4 and hence, the nucleotides encoding GRF_{1-32}. The critical part of the peptide is thought to be contained in this sequence as mammalian GRF_{1-29} is the core

required for full biological activity (Ling *et al.* 1984). The importance of GRF during development is implied by the absence of exon skipping in the embryo prior to hatching. The physiological effect of increasing the ratio of PACAP to GRF after hatching is not known. This skipping of exon four has been reported for cDNAs from three other family members (Parker *et al.* 1993; Seungkwon *et al.* 1995; Talbot *et al.* 1995). The salmon *grf/pacap* cDNA has exon skipping only in the brain and, like the chicken gene, omits exon 4 encoding GRF_{1-32}. In contrast, the *vip/phi* gene in chickens and turkeys does not skip exon 4 in the brain, but does skip this exon in the optic nerve and small intestine to produce a cDNA lacking peptide histidine isoleucine (PHI). In addition, alternative splicing of exons has been reported to occur in a variety of other pre-mRNAs such as bovine growth hormone (Dirksen *et al.* 1994), fibronectin (Huh and Hynes 1994), neural cell adhesion molecule (Tacke and Goridis 1991) and ß-tropomyosin (Guo *et al.* 1991). Nucleotides within the introns and exons of these different genes are known to affect the rate and effectiveness of intron removal. These elements, when mutated, have been shown to alter the effectiveness of small nuclear ribonucleoprotein particles (snRNPs) and serine-arginine (SR) proteins to remove or skip exons (Zahler *et al.* 1992; Watakabe *et al.* 1993). The chicken *grf/pacap* gene has sequences that affect pre-mRNA splicing. Introns 3 and 4 and exons 4 and 5 contain factors such as purine rich motifs (Dirksen *et al.* 1994; Xu *et al.* 1993) and intron 4 has

TGCATG elements (Huh and Hynes 1994); both of these are known to alter exon/intron removal. Therefore, these nucleotide motifs in conjunction with the splice site proteins may explain the skipping of exon 4.

Tissue expression of chicken *grf/pacap* mRNA occurs in brain and gonads

This is the first report that a mRNA encoding both the chicken GRF and PACAP peptides is expressed in the brain and gonads. The expression differs between these two tissues. Both a full length mRNA and shorter (lacking exon 4) mRNA are expressed in the adult brain, whereas only the full length mRNA is detected in the gonads. Transcripts for *grf/pacap* mRNA expression were not present in pituitary, heart, liver, kidney, crop, small intestine, large intestine, eye/optic nerve, or muscle. In evolutionary terms, this expression pattern is similar to the expression of the separate mammalian *grf* and *pacap* genes. In mammals *grf* mRNA is expressed in the hypothalamus, testis, ovary, and placenta, whereas mammalian PACAP is expressed in the same tissues in addition to the gastrointestinal tract (Arimura 1992). However, we did not detect *grf/pacap* mRNA in the chicken small or large intestine, a site known to produce PACAP in mammals and catfish.

The importance of the vertebrate GRF and PACAP in gonads is not yet clear. GRF is known to act directly on the ovary to promote follicular maturation (Moretti *et al.* 1990) and

immunoreactive PACAP was shown to be in spermatids, spermatogonia and primary spermacytes suggesting a role as an autocrine regulator in developing germ cells (Hannibal and Fahrenkrug 1995).

The similar tissue distribution of GRF and PACAP mRNA transcripts further supports the idea that the separate mammalian *grf* and *pacap* genes resulted from a duplication of an ancestral *grf/pacap* gene in vertebrates. The encoding of GRF and PACAP on one gene in chickens implies that the gene duplication that led to encoding the two hormones separately must have occurred in evolution in mammals after reptiles and birds diverged from the ancestral vertebrate stem line.

In conclusion, I have identified the chicken *grf/pacap* gene and the resulting three different cDNAs. Exon sliding occurs in embryo and adult transcripts resulting in mRNA encoding GRF_{1-43} or GRF_{1-46}, and $PACAP_{1-38}$. Exon skipping occurs only in juvenile and adult brains to produce GRF_{33-46} and $PACAP_{1-38}$. A plethora of alternatively spliced genes have been reported and the mechanism of alternative splicing is now beginning to be understood.

References

Arimura A 1992. Pituitary adenylate cyclase activating polypeptide (PACAP): discovery and current status of research. Regulatory Peptides 37: 287-304.

Birnboim M 1983 A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymology 100: 243-255.

Campbell RM, Scanes CG 1992 Evolution of the growth hormone-releasing factor (GRF) family of peptides. Growth Regulation 2: 175-191.

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

Culler MD Paschall CS 1991 Pituitary adenylate cyclase activating polypeptide (PACAP) potentiates the gonadotropin-releasing activity of luteinizing hormone-releasing hormone. Endocrinology 129: 2260-2262.

Denver RJ, Licht P 1989 Neuropeptides influencing in vitro pituitary hormone secretion in hatchling turtles. Journal of Experimental Zoology 251: 306-315.

Dirksen WP, Hampson RK, Sun Q, Rottman FM 1994 A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. Journal of Biological Chemistry 269: 6431-6436.

Frohman MA 1988 Rapid production of full length cDNAs from rare transcripts: amplification using a single gene-specific

oligonucleotide primer. Proceedings of the National Academy of Science (USA) 85: 8998-9002.

Green MR 1991 Biochemical mechanisms of constituitive and regulated pre-mRNA splicing. Annual Reviews of Cell Biology 7: 559-599.

Guisti M, Lomeo A, Mazzocchi G, Torre, R, Monachesi M, Sessarego P, Attanasio R, and Giordano G 1986 Effect of simultaneous administration GHRH(1-40) and TRH on GH, PRL, and TSH release in normal man. Journal of Endocrinological Investigations 9: 497-501.

Guo W, Mulligan GJ, Wormsley S, Helfman DM 1991 Alternative splicing of ß-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. Genes and Development 5: 2096-2107.

Hannibal J, Fahrenkrug J 1995 Expression of pituitary adenylate cyclase activating polypeptide (PACAP) gene by rat spermatogenic cells. Regulatory Peptides 55: 111-115.

Hart GR, Gowing H, Burrin JM 1992 Effects of a novel hypothalamic peptide, pituitary adenylate cyclase activating polypeptide, on pituitary hormone release in rats. Journal of Endocrinology 134:33-41.

Harvey, S 1990 Thyrotrophin-releasing hormone: a growth hormone-releasing factor. Journal of Endocrinology 125: 345-358.

Huh GS Hynes RO. 1994 Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. Genes and Development 8: 1561-1574.

Kentori S, Vernadakis A 1990 Growth hormone-releasing hormone and somatostatin influence neuronal expression in developing chick brain. II Cholinergic neurons. Brain Research 512: 297-303.

Ling N, Baird A, Wehrenberg WB, Ueno W, Munegun T, Brazeau P 1984 Synthesis and in vitro bioactivity of C-terminal deleted analogs of human growth hormone-releasing factor. Biochemical and Biophysical Research Communications 123: 854-861.

Luo D, McKeown BA 1989 Immunohistochemical detection of a substance resembling growth hormone-releasing factor in the brain of the rainbow trout (*Salmo gairdneri*) Experientia 45: 577-580.

Luo D, McKeown BA, Rivier J, Vale W 1990 In vitro response of rainbow trout (*Oncorhynchus mykiss*) somatotrophs to carp growth hormone-releasing factor (GRF) and somatostatin. General and Comparative Endocrinology 80: 288-298.

Malagon NM, Castaño JP, Dobado-Berrios PM, Garcia-Navarro S, Garcia-Navarro F 1991 Human pancreatic growth hormone-releasing factor (1-44) stimulates GH cells in an anuran amphibian (*Rana perezi*). General and Comparative Endocrinology 84:461-469.

McRory JE, Parker DB, Ngamvongchon S, Sherwood NM 1995 Sequence and expression of cDNA for pituitary adenylate cyclase activating polypeptide (PACAP) and growth hormone-releasing hormone (GHRH)-like peptide in catfish. Molecular and Cellular Endocrinology 108: 169-177.

Moretti C, Bagnato A, Frajese G, Catt K 1990 Receptor-mediated actions of growth hormone releasing factor on granulosa cell differentiation. Endocrinology 127: 2117-2126.

Murakami Y, Koshimura K, Yamauchi K, Nishiki M, Tanaka J, Furuya H, Miyake T, Kato Y 1995 Pituitary adenylate cyclase activating polypeptide (PACAP) stimulates growth hormone release from GH3 cells through type II PACAP receptors. Regulatory Peptides 56: 35-40.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA 1986 Splicing of messenger RNA precursors. Annual Reviews of Biochemistry 55: 1119-1150.

Parker DB, Coe IR, Dixon GH, Sherwood NM 1993 Two salmon neuropeptides encoded by one brain cDNA are structurally related to members of the glucagon superfamily. European Journal of Biochemistry 215: 439-448.

Perez FM, Malamed S, Scanes CG 1987 Growth hormone secretion from chicken adenohypophyseal cells in primary culture: effects of human pancreatic growth hormone-releasing factor, thyrotropin-releasing hormone, and

somatostatin on growth hormone release. General and Comparative Endocrinology 65: 408-414

Peter RE, Nahorniak CS, Vale WW, Rivier JE 1984 Human pancreatic growth hormone-releasing factor (hpGRF) stimulates growth hormone release in goldfish. Journal of Experimental Zoology 231: 161-163.

Porter TE, Couger GS, Dean CE, Hargis BM 1995 Ontogeny of growth hormone (GH)-secreting cells during chicken embryonic development: Initial somatotrophs are responsive to GH-releasing hormone. Endocrinology 136: 1850-1856.

Propato-Mussafiri R, Kanse SM, Ghatei MA, Bloom SR 1992 Pituitary adenylate cyclase activating polypeptide releases 7B2, adrenocorticotrophin, growth hormone and prolactin f from the mouse and rat clonal pituitary cell lines AtT-20 and GH3. Journal of Endocrinology 132: 107-113.

Rawlings SR, Hezareh M 1996 Pituitary adenylate cyclase-activating polypeptide (PACAP) and PACAP/vasoactive intestinal polypeptide receptors: action on the anterior pituitary gland. Endocrine Reviews 17: 4-29.

Rivier J, Spiess J, Thorner M, Vale W 1982 Characterization of growth hormone releasing factor from a human pancreatic islet tumor. Nature 300: 276-278.

Sanger F, Nicklen S, Coulson AR 1977 DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science (USA) 74: 5463-5467.

Scanes CG, Harvey S 1984 Stimulation of growth hormone secretion by human growth hormone releasing factor (hpGRF) and thyrotropin releasing hormone (TRH) in anesthetized chickens. General and Comparative Endocrinology 56: 198-203.

Seungkwon Y, Silsby JL, Farris J, Foster DN, El Halawani ME 1995 Tissue-specific alernative splicing of turkey preprovasoactive intestinal peptide messenger ribonucleic acid, its regulation and correlation with prolactin secretion. Endocrinology 136: 2602-2610.

Sherwood NM, Parker DB, McRory JE, Lescheid DW 1994 Molecular evolution of growth hormone-releasing hormone and gonadotropin-releasing hormone. In: Fish Physiology, N.M. Sherwood and C. L. Hew (eds). 13: 3-66 Academic Press, Orlando.

Tacke R, Goridis C. 1991 Alternative splicing in the neural cell adhesion molecule pre-mRNA: Regulation of exon 18 skipping depends on the 5'-splice site. Genes and Development 5: 1416-1429.

Talbot RT, Dunn IC, Wilson PW, Sang HM, Sharp P 1995 Evidence for alternative splicing of the chicken vasoactive intestinal peptide gene transcript. Journal of Molecular Endocrinology 15: 81-91.

Vale W, Vaughan J, Yamamoto G, Speiss J, Rivier J 1983 Effects of synthetic human pancreatic (tumor) GH releasing factor and somatostatin, triiodothyronine, and

dexamethasone on GH secretion in vitro. Endocrinology 112: 1553-1555.

Vaughan JM, Rivier J, Speiss J, Peng C, Chang JP, Peter RE, Vale W 1992 Isolation and characterization of hypothalamic growth-hormone releasing factor from common carp, *Cyprinus carpio*. Neuroendocrinology. 56: 539-549.

Watakabe A, Tanaka K Shimura Y 1993 The role of exon sequences in splice site selection. Genes and Development 7: 407-418.

Xu R, Teng J, Cooper TA 1993 The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Molecular and Cellular Biology 13: 3660-3574.

Zahler AM, Lane WS, Stolk JA, Roth MB 1992 SR proteins: a conserved family of pre-mRNA splicing factors. Genes and Development 6: 837-847.

Chapter 3

Embryonic expression pattern of the chicken growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP) gene.

Summary

GRF is a neuropeptide releasing factor that is known to regulate GH secretion. In this study evidence is presented to suggest that GRF may have a role in early brain development. Chicken *grf* mRNA is expressed throughout the embryonic period, long before it functions as a releasing factor. In the developing chick embryo, foreshortened *grf* mRNA is expressed. In the first three days of development, transcripts containing, exons 4 and 5, but not 1-3 are detected. On the fourth day of development, transcripts containing only exons 3-5 are detected and beginning on the fifth day of embryonic development, full length mRNA (exons 1-5) is detected. In chicken, *grf* transcripts produced in the brain are processed in an age-specific manner into three different mRNAs that encode three different proteins. The first mRNA encodes full length GRF of 46 amino acids (aa) and the second mRNA encodes a shorter GRF (43aa) as a result of exon sliding at the splice site of the intron4:exon 5 boundary. In juveniles and adults, an additional third mRNA encodes GRF33-46 because exon 4 is skipped. The early and continued expression of *grf/pacap* mRNA in the developing brain suggests the GRF and PACAP neuropeptides have an important role in early CNS development.

Introduction

In addition to the role of PACAP as a releaser of pituitary hormones, preliminary evidence shows that PACAP promotes neuroblast proliferation and neurite outgrowth in the cerebral cortex, cerebellar granule cells and sympathetic ganglia of fetal mice (DiCicco-Bloom 1992). Further evidence that PACAP is a potential growth factor is based on the following evidence: PACAP promotes proliferation of mouse primordial germ cells (Pesce *et al.* 1996); PACAP is produced in specific tumour cells; PACAP receptors are detected on human glial cell tumours (Robberecht *et al.* 1994); and PACAP is expressed early in rat brain development with levels peaking at birth (Masuo *et al.* 1994). Meanwhile, human GRF has been shown to affect somatotroph differentiation in the embryonic chicken pituitary (Porter *et al.* 1995) and to accelerate the development of chick neuroblasts and synthesis of neurotransmitters (Kentori and Vernadakis 1990).

To investigate if GRF and PACAP have a role in early brain development in the chicken, the expression of the chicken *grf/pacap* mRNA was investigated. Brain samples were taken every day from day 1 (24h) until day 21 (hatching). Using RT/PCR and several primers directed to different regions of the gene, I detected *grf/pacap* mRNA expression in the developing embryo. Embryonic expression of mRNA could be detected as early as day 1, but closer analysis revealed a change in transcription to the adult pattern on embryo day 5.

Materials and Methods

Developmental expression using RT/PCR

Four dozen eggs were placed in a 41°C humidified incubator and beginning on day 1 and at every 24 h thereafter, eggs were removed, the embryo dissected from the yolk and total RNA extracted with TriZol (BRL). Four whole embryos were removed and extracted for each of the first 3 days, whereas only two whole embryos were dissected and extracted on days 4-6. For days 7-10, one whole head was extracted and from days 11-21, one brain was dissected from the skull and extracted. For each sample, single stranded cDNA was synthesized from 10μg total RNA, 2mM primer E, 5μl Superscript buffer, 2mM dNTP, 10mM DTT, 5U RNA guard (Pharmacia), and 200 U RT Superscript (BRL) to a final volume of 25μl. The nucleotide sequence for the primers used in this experiment are given in chapter 2 or below and the hybridization site is shown in Fig 2.1. The reaction was heated to 42°C for 1.5hr and terminated by increasing the temperature to 95°C for 10min.

Amplification was performed in a 50μl volume with 0.2μg cDNA, 5U Taq, 1x Taq buffer (Promega), 200mM dNTP's, 2mM $MgCl_2$, and 0.2pmol of each primer set A/D, B/D (primer B: 5-ggaataccctgcaggacttcgc) or C/D (primer C: 5'-caaagcctacagga aactcctgggcc). Primer hybridization sites are shown in Fig. 2.1. The reaction was carried out for 35 cycles at 94°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min. PCR reaction products were electrophoresed through a 3% agarose gel.

Determination of *grf/pacap* mRNA length in embryos at 1-5 days of development

To determine the approximate mRNA lengths, a cDNA extension reaction was performed using poly A+ mRNA from embryos of 1-5 days of age. 10μg poly A-rich mRNA from each day was combined with 1mM primer D and heated to 90°C for 10 min. The DNA was cooled to 57°C for 20min and placed on ice. Added to each of the DNA/primer mixtures was a solution consisting of 5μl Superscript buffer, 2.5mM dATP, dTTP, and dGTP, 1.25mM of each dCTP and [α-^{32}P]dCTP, 10mM DTT, 5U RNA guard (Pharmacia), and 200 U RT Superscript (BRL) for a final volume of 25μl. The mixtures were incubated at 42°C for 1.5hr. The reverse transcription reaction was terminated at 90°C for 10min. The mixtures were chilled on ice, treated with the addition of 10U RNase H (Sigma) and incubated at 37°C for 1hr. The reactions were PEG (2.5M Nacl, 5M PEG) precipitated on ice for 1.5hr. The PEG solution was spun at 10,000g, the pellet was washed in 75% ethanol and rehydrated in water. The solution was isopropanol precipitated, washed with 75% ethanol, rehydrated in water and the resulting DNA solution was passed through NAP-5 columns (Pharmacia). The solution was isopropanol precipitated, washed with 75% ethanol, rehydrated with 10μL water and electrophoresed through a 6% polyacrylamide gel under reducing conditions. Radioactive bands were visualized by autoradiography of dried gels using

Kodak Biomax film and intensifying screens during 2 days at -70°C.

Identification of cDNA by library screening

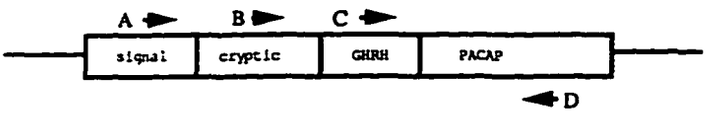
A total of 10^6 pfu from a 13-day-old chicken embryo cDNA library (gift from Dr. B. Ranscht, La Jolla Cancer Research Foundation) was screened with the 294bp fragment produced by PCR, as described in Chapter 2. Prehybridization and hybridization were done as above, except a 55°C hybridization temperature was used. The membranes were washed under high stringency (0.1xSSC/0.1%SDS) for 30 min at 55°C, then exposed to Biomax (Kodak) film for 2 days at -80°C. Four of the 15 positive clones from the first round of screening were selected for two additional rounds of screening. The inserts were removed from the phage with EcoRI and subcloned into EcoRI cut pBluescript KS (Stratagene). Both strands of the four clones were sequenced.

Results

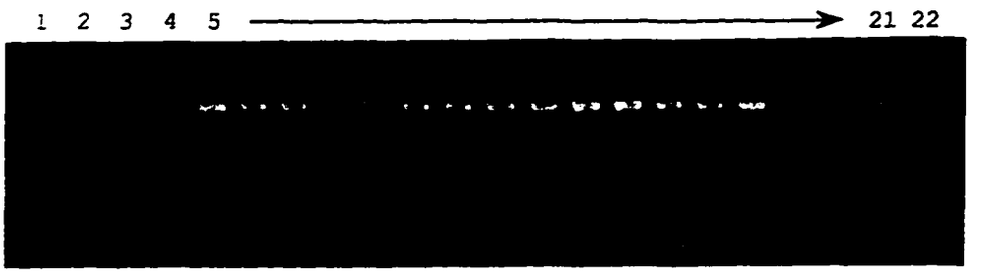
Developmental expression of the chicken *grf/pacap* gene

Expression of full length chicken *grf/pacap* mRNA was first detected in the 5-day-old embryonic brain with a sensitive RT/PCR method. However, during early development of the chick embryo, we were able to demonstrate loose transcription of the gene (Figures 3.1-3.3). Using primers C and D, we detected a partial *grf/pacap* mRNA consisting of exons 4 and 5 in chicks after 24hr of incubation (day 1). Transcripts encoding exons 2 and 3 were not detected on days 1 to 3 of development using primers A and D between exons 2 and 5 or primers B and D between exons 3 and 5. (Figures 3.1 and 3.2). The incomplete mRNA transcript is also seen with the cDNA elongation reaction (Figure. 3.3). On days 1-3, expression of exons 4 and 5 was detected as seen by the size of the cDNA extension reaction for those days. On day 4, expression of exons 3, 4, and 5 was detected as seen by the size of the bands produced with the day 4 cDNA extension reaction and using primers B and D. On day 5 and continuing till hatching at day 21, a complete chicken *grf/pacap* mRNA was produced as detected with primers A and D (Figures 3.1 and 3.2) and as shown by the cDNA extension reaction band length (Figure 3.3).

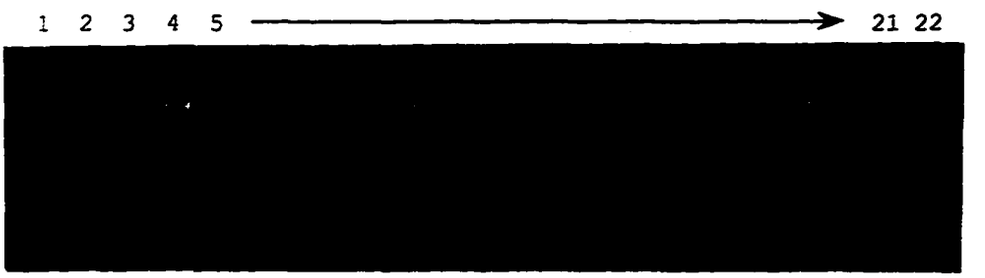
Figure 3.1: Reverse transcriptase assay to detect *grf/pacap* mRNA in developing chick embryos. mRNA was extracted from embryos each day until hatching (lanes 1-21). Lane 22 is the negative control (no DNA). The chicken *grf/pacap* cDNA schematic shows the sites of hybridization of the different primers. The first panel shows complete *grf/pacap* mRNA synthesis beginning on day 5 as detected with primers A and D. The second panel show exons 3-5 (primers B and D) are detected beginning on day 4 and the third panel shows incomplete synthesis of the *pacap* mRNA beginning on day 1 as detected with primers C and D.



Primers A/D (exons 2-5)



Primers B/D (exons 3-5)



Primers C/D (exons 4-5)

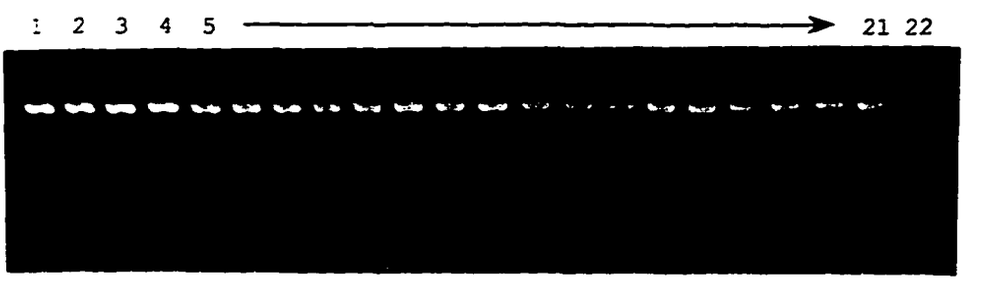


Figure 3.2: Reverse transcriptase assay to demonstrate loose transcription of the chicken *grf/pacap* gene. Day 1 to 5 cDNA/mRNA was amplified with one of three sets of primers. Lane A indicates primers C and D were used, lane B indicates B and D were used; and lane C indicates primers A and D were used. Approximate molecular sizes are shown on the right.

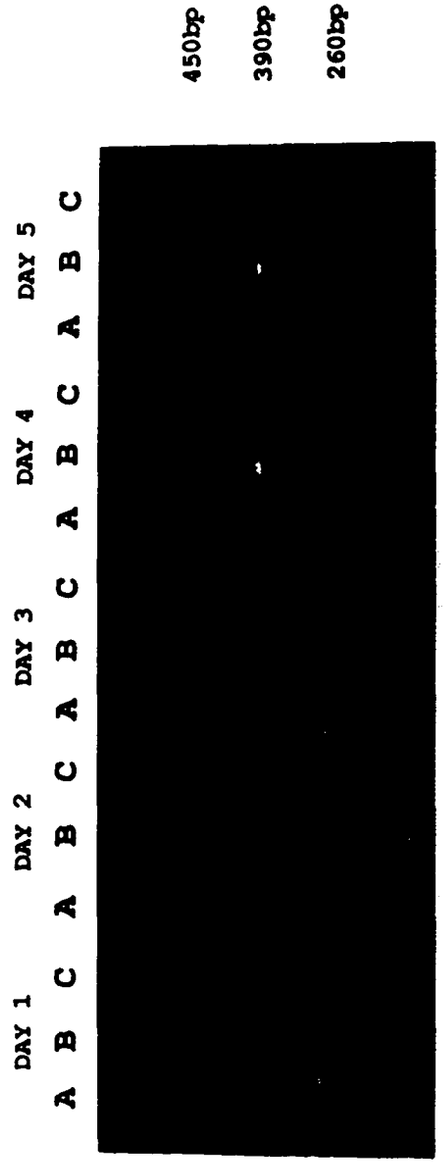
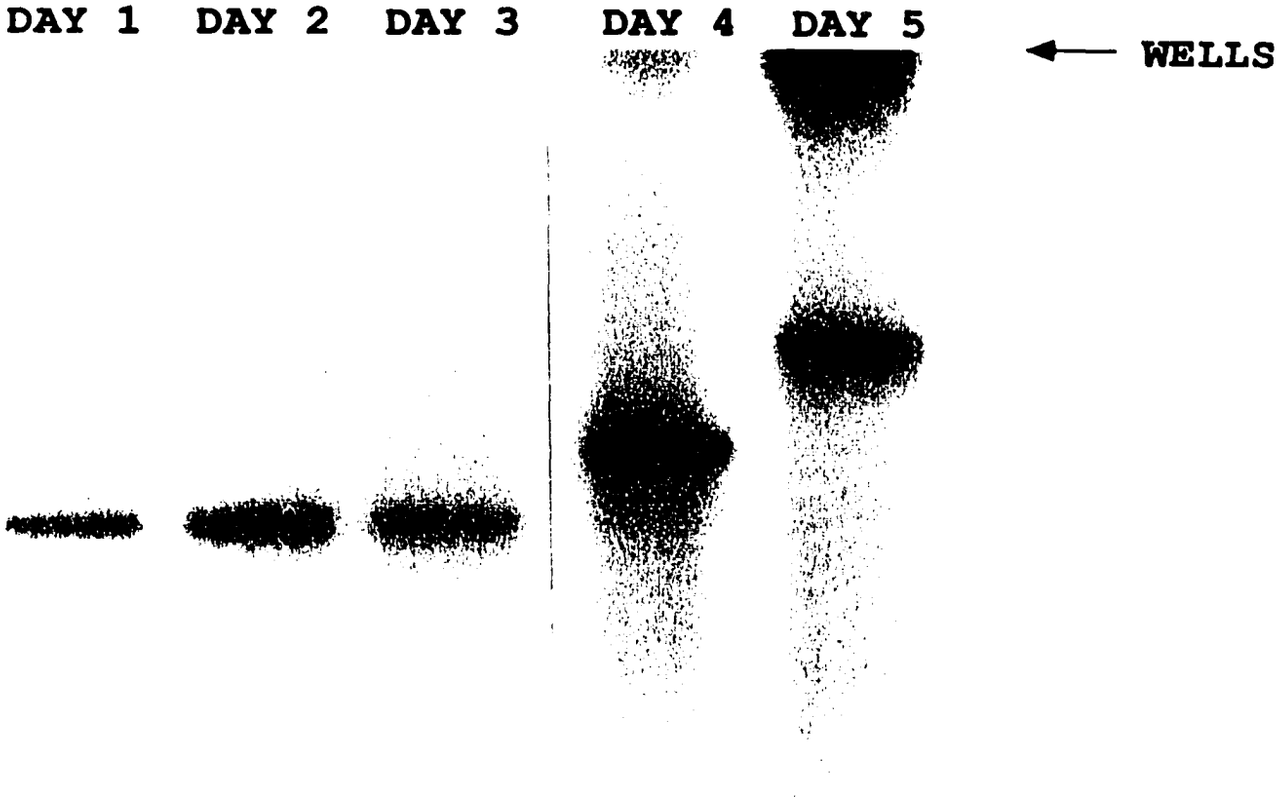

450bp
390bp
260bp
DAY 1
DAY 2
DAY 3
DAY 4
DAY 5
A B C A B C A B C A B C A B C

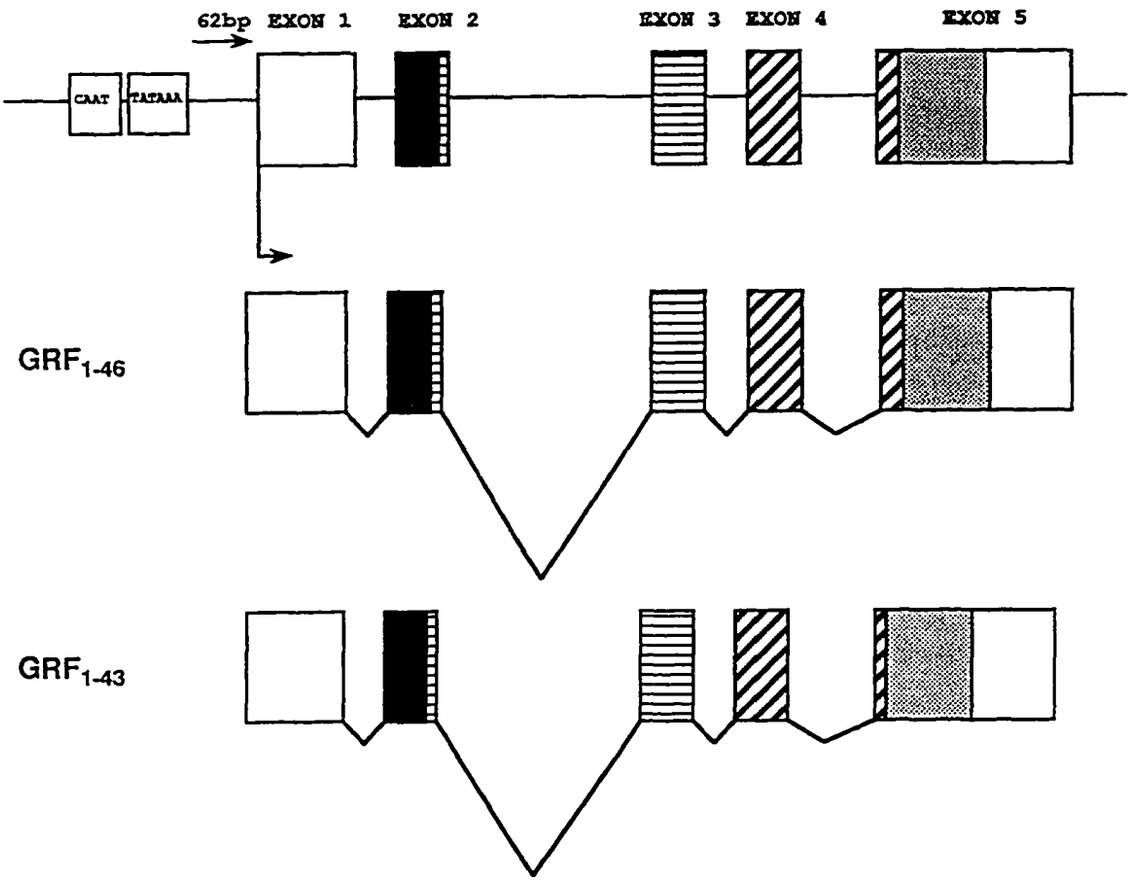
Figure 3.3: Determination of *grf/pacap* mRNA length from 1-5 day-old chick embryos. Day 1-5 mRNA was reverse transcribed in the presence of [α-^{32}P]dCTP and primer D. The cDNA containing the α-^{32}P labelled cDNA was electrophoresed through a 6% polyacrylamide gel under reducing conditions.


DAY 1
DAY 2
DAY 3
DAY 4
DAY 5
WELLS

Confirmation of intron-exon boundaries with a cDNA library

To confirm intron-exon boundaries, a 13-day-embryo cDNA library made from brain mRNA, was screened using the 294bp PCR fragment. Fifteen clones were isolated and 4 clones completely sequenced. All intron-exon boundaries were confirmed (Figure 3.4). However, in cDNA prepared from a library lysate, the boundary between exons 4 and 5 was found to vary: in some bands, the boundary was at position 5703bp and in the other the boundary had slid nine bases downstream to position 5712 bp.

Figure 3.4: Schematic for alternative splicing of the embryonic chicken *grf/pacap* transcripts to produce only two different mRNAs. The gene is shown on top with the two different mRNAs shown below. The first mRNA was the most abundant and encodes GRF_{1-46}. The second mRNA has the splice site 9bp downstream from the first site and encodes GRF_{1-43}.


62bp
EXON 1
EXON 2
EXON 3
EXON 4
EXON 5
CAAT
TATAAA
GRF1-46
GRF1-43

Discussion

Expression of the *grf/pacap* gene occurs early in development

To gain insight into *pacap* and *grf* gene expression in the developing chicken embryo, I analysed the number and structure of transcripts for this gene in the developing chick embryo. I extracted mRNA from individual embryos every 24hr until hatching. Using poly A^+ mRNA in a cDNA extension reaction and mRNA from each day with primers specific for each exon, I determined that a full length cDNA containing exons 1-5 encoding chicken *grf/pacap* is produced. Expression of the full length cDNA begins on day 5 of embryonic development and persists till hatching on day 21 and into adulthood (Figure 3.1). Early *pacap* mRNA synthesis also was observed in rainbow trout embryos (*Oncorhynchus mykiss*) where full length mRNA transcripts were detected in 4-day-old embryos (S. Krueckl, personal communication). However, within the chicken, we detected an incomplete *grf/pacap* mRNA in 1-day-old embryos as shown by the short extension reaction band of approximately 275 bases (Figure 3.3) and the lack of a PCR band except with primers C and D directed to exons 4 and 5 (Figures 3.1 and 3.2). This mRNA would encode only the GRF and PACAP, but not the signal or cryptic peptides. The truncated mRNA is probably not translated because of the lack of nucleotides encoding known translation initiation sites (i.e. nucleotides CAGATG) at the

beginning of the signal peptide. The shortened chicken *pacap* mRNA was present until day 4 of development when exon 3 encoding the cryptic peptide was detected with primers B and D. Finally at day 5, the nucleotides that encode the signal peptide and 5'UTR are present for translation into a prepropeptide. The function of the loose transcription is not known. During early development the embryo may lack the proper factors for correct transcription or RNA processing of this transcript until day 5. Only then is mRNA of the correct composition available for translation.

The production of the full length *grf/pacap* mRNA coincides with differentiation and development of the brain. By day 5, the chick cerebral hemispheres of the telencephalon have developed and the lateral walls of the diencephalon and mesencephalon have begun differentiation (Patten 1964). The production of the full length *grf/pacap* mRNA at the same time as development of the central nervous system and pituitary suggests that PACAP and GRF are potential growth factors for developing neuroblasts and somatotrophs.

Exon sliding produces 2 different mRNAs

The organization of the chicken *grf/pacap* gene is similar to both the human *grf* and *pacap* genes and to other members of the glucagon superfamily. As described in Chapter 2, the chicken *grf/pacap* gene is composed of 5 exons and as sequenced, encompassed 6525bp. All 5 exon locations and intron/exon

boundaries were confirmed by isolating cDNA clones from a 13-day-embryo library and with 5' and 3' RACE reactions. However, in sequences of cDNA library clones and PCR fragments, we observed that the intron/exon boundary between exons 4 and 5 varied (Figures 3.4). The dominate boundary occurs at position 5703bp (Figure 2.2) and the second boundary slides 9bp downstream to position 5712. At both splice sites nine bases apart, proper consensus splice sites exist. Therefore, the chicken *grf/pacap* mRNA transcript has splice donor sites that encode a 43-amino-acid GRF. The acceptor site was also shown to slide 9bp upstream to encode a GRF of 46 residues. The intron nucleotides at the 5' splice site of intron 4, GG:GT(A) and the last 4 nucleotides of intron 4 at both 3' splice sites (NCAG:C), closely match the splice site consensus sequences found in vertebrates (Padgett *et al.* 1986; Green 1991). Exon sliding has not been previously reported for this family of peptides. The function of exon sliding is not known other than to encode two GRFs of different length that potentially have two different functions. Recent evidence shows that human GRF_{1-44} in the chicken affects somatotroph differentiation in the embryonic chicken pituitary (Porter *et al.* 1995) and the development of chick neuroblasts and their neurotransmitters (Kentori and Vernadakis 1990). These effects on early brain development and GH-releasing somatotrophs in the pituitary imply an early role of GRF_{1-43} and/or GRF_{1-46} in birds.

Transcription of the embryonic chick *grf/pacap* gene includes exon 4

Within Chapter 2, the processing of the chicken *grf/pacap* transcript was shown to result in three different mRNAs. Screening the embryonic cDNA library or amplifying cDNA with the PCR, I did not detect any clones that lacked exon four. Therefore, the chick embryo encodes only GRF_{1-43} and GRF_{1-46} in addition to PACAP (Figure 3.4). The skipping of exon four was seen to occur only in mRNA from the juvenile and adult chicken brain. However, sometime between hatching and day 25 of juvenile growth, exon four begins to be skipped. The skipping of exon four is a common occurrence in members of this gene family. Exon four is skipped in adult mRNA in salmon PACAP (Parker *et al.* 1993), turkey VIP (Seungkwon *et al.* 1995) chicken VIP (Talbot *et al.* 1995), human secretin (Kopin *et al.* 1991) and human GIP (Takeda *et al.* 1987). However, this is the first report of an age-specific skipping event and the function is unclear.

References

DiCicco-Bloom EM 1992 Pituitary adenylate cyclase activating polypeptide (PACAP) potently stimulates mitosis, neurogenesis and survival in cultured rat sympathetic neuroblasts. Regulatory Peptides 37: 319.

Green MR 1991 Biochemical mechanisms of constituitive and regulated pre-mRNA splicing. Annual Review of Cell Biology 7: 559-599.

Kentori S, Vernadakis A 1990 Growth hormone-releasing hormone and somatostatin influence neuronal expression in developing chick brain. II Cholinergic neurons. Brain Research 512: 297-303.

Kopin AS, Wheeler MB, Nishitani J, McBride EW, Chang T, Chey W, Leiter AB. 1991 The secretin gene: evolutionary history, alternative slicing and developmental regulation. Proceedings of the National Academy of Science (USA) 88: 5335-5339.

Masuo Y, Tokito F, Matumoto Y, Shimamooto N, Fujino M 1994 Ontogeny of pituitary adenylate cyclase activating polypeptide (PACAP) and its binding sites in the rat brain. Neuroscience letters 170: 43-46.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA 1986 Splicing of messenger RNA precursors. Annual Reviews in Biochemistry 55: 1119-1150.

Parker DB, Coe IR, Dixon GH, Sherwood NM 1993 Two salmon neuropeptides encoded by one brain cDNA are structurally

related to members of the glucagon superfamily. European Journal of Biochemistry 215: 439-448.

Patten BM 1964 Development of the chick during the third and fourth days of incubation. In: Foundations of Embryology. pp225-264. McGraw-Hill, Toronto.

Pesce M, Canipari R, Ferri G-L, Siracusa G, De Felici M 1996 Pituitary adenylate cyclase-activating polypeptide (PACAP) stimulates adenylate cyclase and promotes proliferation of mouse primordial germ cells. Development 122: 215-221.

Porter TE, Couger GS, Dean CE, Hargis BM 1995 Ontogeny of growth hormone (GH)-secreting cells during chicken embryonic development: initial somatotrophs are responsive to GH-releasing hormone. Endocrinology 136: 1850-1856.

Robberecht P, Woussen-Colle MC, Vertongen P, De Neef P, Hou X, Salmon I, and Brotchi J 1994 Expression of pituitary adenylate cyclase-activating polypeptide (PACAP) receptors in human glial cell tumors. Peptides 15: 661-665.

Seungkwon Y, Silsby JL, Farris J, Foster DN, El Halawani ME 1995 Tissue-specific alernative splicing of turkey preprovasoactive intestinal peptide messenger ribonucleic acid, its regulation and correlation with prolactin secretion. Endocrinology 136: 2602-2610.

Takeda J, Seino Y, Tanaka K-I, Fukumoto H, Kurono M, Suzuki T, Tobe T, Imura H 1987 Sequence of an intestinal cDNA encoding human gastric inhibitory polypeptide precursor.

Proceedings of the National Academy of Science (USA) 84: 7005-7008.

Talbot RT, Dunn IC, Wilson PW, Sang HM, Sharp P 1995 Evidence for alternative splicing of the chicken vasoactive intestinal peptide gene transcript. Journal of Molecular Endocrinology 15: 81-91.

CHAPTER 4

Evolution of growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase activating-polypeptide (PACAP) in fish.

A version of this chapter has been published: McRory JE, Parker DB, Ngamvongchon S, Sherwood NM 1995 Sequence and expression of cDNA for pituitary adenylate cyclase activating polypeptide (PACAP) and growth hormone-releasing hormone (GHRH)-like peptide in catfish. Molecular and Cellular Endocrinology 108: 169-177.

Summary

In this chapter I consider the molecular evolution of GRF and PACAP by isolating and sequencing the cDNA that encodes these two peptides from catfish. Growth hormone-releasing hormone (GRF) and pituitary adenylate cyclase-activating polypeptide (PACAP) are two neuropeptides that are associated with the release of pituitary growth hormone. Here, a cDNA of 2501 base pairs encoding both a GRF-like peptide and a PACAP was isolated from a brain cDNA library made from Thai catfish (*Clarias macrocephalus*). The organization is similar to that of chicken, but is unlike that of the mammalian gene where PACAP and PACAP-related peptide (PRP) are encoded in one gene, and the GRF peptide is on a separate gene.

Northern blot analysis of catfish brain mRNA indicated *grf/pacap* mRNA of three sizes. These bands of 6000, 2500, and 1000 bases suggest alternative splicing of the gene. Reverse transcriptase/PCR detected *grf/pacap* mRNA in tissue from the brain, testis, ovary, and stomach, but not from the pancreas, pituitary, muscle, and liver. Our hypothesis that the two mammalian genes encoding GRF or PACAP originated from a gene duplication in ancestral mammals is supported by the present findings of similar mRNA organization and pattern of expression for the single *grf/pacap* gene in fish and birds and the separate *grf* and *pacap* mammalian genes.

Introduction

The location and actions of PACAP suggest that one function is to release pituitary hormones. PACAP immunoreactive fibers have been detected in both the external and internal zones of the median eminence. *In vitro* studies showed that PACAP caused an increase in the release of not only GH, but also of several other pituitary hormones (Hart *et al.* 1992); this effect may result from PACAP's action in stimulating the accumulation of intracellular cyclic AMP from dispersed rat anterior pituitary cells (Propato-Mussafiri *et al.* 1992). Also, PACAP stimulated GH and PRL release from GH3 tumour cells (Propato-Mussafiri *et al.* 1992) and enhanced the effect of gonadotropin-releasing hormone on LH and FSH release (Culler and Paschall, 1991). Finally, PACAP is assumed to have functions outside of the pituitary because it is detected by antisera in many locations: brain, testis, ovary, adrenal, pancreas, gastrointestinal tract, and respiratory tract (Koves *et al.* 1990; Gottschall *et al.* 1990; Uddman *et al.* 1991; Arimura, 1992)

The primary role of GRF is to release GH from the pituitary. This is unlike PACAP, which appears to release several pituitary hormones. However, GRF has functions in addition to the release of GH because GRF is expressed not only in the hypothalamus, but also in the testis, ovary, and placenta (Berry and Pescovitz, 1988, 1990; Suhr *et al.* 1989; Moretti *et al.* 1990; Margioris *et al.* 1990; Bagnato *et al.* 1992). In nonmammalian vertebrates, an undisputed GRF has not been isolated, but the

human GRF has been shown to stimulate the release of GH from the pituitaries of turtles (Denver and Licht, 1989, 1991) and frogs (Malagon *et al.* 1991). Recently, Vaughan *et al.* (1992) demonstrated that a carp GRF-like molecule stimulated the release of GH from perfused goldfish pituitaries in a dose dependent manner.

In mammals, PACAP and PRP are encoded in one gene (Ohkubo *et al.* 1992; Hosoya *et al.* 1992), whereas GRF is on a separate gene (Mayo *et al.* 1985). However, in other vertebrates a gene encoding only GRF has not been identified (Mayo *et al.* 1985). Rather, GRF along with PACAP appears to be encoded in one gene , at least in salmon (Parker *et al.* 1993). Among nonmammalian vertebrates, peptide sequences have been determined only for frog $PACAP_{1-38}$ (Chartrel *et al.* 1991) and for common carp (*Cyprinus carpio*) GRF (Vaughan *et al.* 1992).

The evolutionary and functional relationships of PACAP, PRP, and GRF are not clear. A function has not been ascribed to mammalian PRP. If gene duplication occurred within the vertebrates to produce the *prp/pacap* and *grf* genes, the timing is obscure. As a first step to study the evolutionary relationships of PACAP, PRP, and GRF, we have isolated the catfish cDNA that encodes both PACAP and a GRF-like peptide. To study the functional relationships of these growth-related peptides, we used a reverse transcriptase/PCR assay to show *grf/pacap* mRNA is expressed not only in the brain, but in gonadal and gut tissues as well.

Materials and Methods

Construction of cDNA library

Thai catfish (*Clarias macrocephalus*) brains were taken from anesthetized fish and placed immediately in liquid nitrogen and stored at -80°C. Total RNA was extracted with an acidic guanidinium thiocyanate method (Chomczynski and Sacchi, 1987) and poly A+-rich mRNA was purified on two consecutive oligo dT_{12-18} columns. Double stranded cDNA was synthesized using poly T(20) and reverse transcriptase, (Pharmacia) ligated to EcoR1/Not1 adaptors and Lambda ZapII vector arms (Stratagene). Gigapack gold (Stratagene) was used to package the lambda vectors with inserts.

Amplification of *grf/pacap* cDNA by PCR

DNA was amplified using a cDNA library lysate and nondegenerate primers directed against the salmon *grf/pacap* cDNA sequence (Parker *et al.* 1993). The SP1 primer (5'-ggaatc ataatgcactacagtgtc) was based on amino acids 13-20 of the salmon signal peptide. The NMS3 primer (5'-tcggtagcggctgtagc tatctg) was designed for amino acids 8-14 of the salmon PACAP peptide. First round amplification was done in a 50µl volume (0.2µg cDNA, 5U Taq, 1x Taq buffer (Promega), 200µM dNTP's, 2mM $MgCl_2$, 20 pmol of primers Sp1 and NMS 3) with 35 cycles at 94°C for 1.5 min, 45°C for 2 min, 72°C for 2.5 min and a 5.3 min extension at 72°C. Reamplified bands were cloned into pBluescript KS+ (Stratagene), electroporated into XL-1

competent cells, and prepared for sequencing with an alkaline hydrolysis method (Birnboim, 1983).

Assay by tissue mRNA reverse transcription/PCR

Messenger RNA was isolated from the following tissues: brain, pituitary, testis, ovary, stomach, liver, pancreas, and muscle. Complementary DNA was synthesized from 0.5µg mRNA for each tissue using 200U avian reverse transcriptase (H^- Superscript RT, BRL), 10mM dithiothreitol, 1mM each dNTP, 160U RNA guard, 2mM Oligo dT_{12-18}, and 1X H^- RT buffer in a total reaction volume of 20µl. The reaction proceeded for 90 min at 41°C followed by 5 min at 90°C. DNA amplifications were done in a 50µl volume that contained 0.5µl cDNA, 1x TAQ buffer (BRL), 200µM dNTP's, 5U Taq, 1.5mM $MgCl_2$, and 20pmol of each primer. Primer 3 was made against the 5'UTR, position 294-314bp of the catfish clone, and primer 4 corresponded to bases 901-925. The sequence for primer 3 was 5'cctcgctctctgaccaaactg and for primer 4 was 5'cctgggtaactccaattaatacct. All reagents, except the Taq enzyme, were mixed, layered with mineral oil, and heated to 95°C for 5 min. The mixture was frozen rapidly in a dry ice/ethanol bath and 5U Taq DNA polymerase were added. The tubes were replaced in the preheated 95° thermal cycler where 35 cycles at 95°C for 1 min, 55°C for 2 min, and 74°C for 2.5 min were completed. Tubulin PCR amplifications were done in a 50µl reaction volume that contained 0.5µg cDNA from each

tissue, 5U Taq DNA polymerase, 1x Taq buffer (Promega), 200μM each dNTP, 3mM $MgCl_2$, and 20pmol of each primer. The reactions were carried out for 35 cycles of 94° for 1 min, 55° for 1.5 min, and 72° for 1.5 min. Primers 5 and 6 correspond to bases 523-545 and 719-740 of the salmon tubulin cDNA clone respectively (Coe *et al.* 1992). The sequence for primer 5 was 5'caggtgtccacgg ctgtggtg and for primer 6 was 5'agggctccatcgaaacgcag.

Sequencing of clone inserts

Both strands of the plasmid were sequenced with [α-^{3}S]dATP using the chain termination method (Sanger *et al.* 1977) and Sequenase 2.0 (US Biochemical Corp.). All sequencing gels were 6% polyacrylamide/7M urea wedge gels, dried under vacuum at 80°C and exposed to Kodak XAR-5 film for 12-24h.

Screening of cDNA library

A total of 50 000 pfu from the catfish library ($2.3x10^{9}$ pfu/μl) were screened using the 346bp PCR probe. Duplicate nylon membrane (Bio-Rad) lifts were prehybridized at 50°C in 6x SSC, 5x Denhardt's solution, 0.5% SDS and 30μg/ml blocking DNA for 4 h. The hybridization solution, consisting of 6x SSC, 0.5% SDS, and 30ug/ml blocking DNA was added to the [α-^{32}P]dCTP (Dupont) labeled probe ($2.4x10^{7}$ cpm/ml), and incubated at 50°C for 14 h. The membranes were washed under high stringency (0.1xSSC/0.1%SDS) for 50 min at 50°C, then

exposed to Kodak XAR-5 film for 7 days at -80°C. Isolated single positive clones were cored and inserts rescued with *in vitro* excision as to manufacturer's (Stratagene) instructions.

Determination of mRNA size and copy number using northern analysis

Brains were obtained from Thai catfish (*Clarius macrocephalus*) and African catfish (*C. gariepinus*). A gel consisting of 1.2% agarose, 20mM MOPS, and 10.5ml formaldehyde was prepared. Thai and African catfish brain poly A+ mRNA (20μg each) were loaded onto the gel with Thai catfish total RNA (20μg) and RNA markers. After the samples were electrophoresed in 1X MOPS buffer (20mM MOPS, pH 7.0, 5mM $NaC_2H_3O_2$, 2mM EDTA), the RNA was transferred to a nylon membrane, dried at 80°C for 2h, and prehybridized in 25mM KPO_4, 5x SSC, 10X Denhardt's solution, 50% formamide and 30μg/ml blocking DNA at 40°C for 3h. Hybridization (prehybridization solution plus 5% dextran sulfate) was done overnight at 40°C in a 10ml volume with the PCR generated 346bp probe labeled with [α-^{32}P]dCTP. The membrane was washed at 55°C successively in 2XSSC/0.1%SDS, 1XSSC/0.1%SDS, 0.5XSSC/0.1%SDS and two times in 0.1XSSC/0.1%SDS. The membrane was air dried and exposed to XAR-5 film (Kodak) for 12h at -80°C.

Results

Isolation and sequence of catfish *grf/pacap* cDNA

Four bands resulted from first round DNA amplification of the Thai catfish cDNA library lysate with the two primers. Band 2 contained 346bp of the catfish *grf/pacap* clone. In three other clones of band 2, sequences were found to be identical to the first band. All other bands contained sequences without similarity to GRF or PACAP.

The 346bp *grf/pacap* fragment was used to screen a Thai catfish cDNA library. After three rounds of screening, two positive clones (clones 3.51 and 1.44) were isolated from the library (Figure 4.1). The catfish *grf/pacap* cDNA clone 3.51 was 2501bp and contained 6 distinct regions (Figure 4.2). The cDNA consisted of the 5' untranslated (5'UTR) region, signal peptide, cryptic peptide, GRF-like region, PACAP and the 3' untranslated region. The 5'UTR, which is 321bp long, contains (T)n, $(CT)_n$, and $(CA)_n$ repeat sequences and has two possible ATG translation start sites in positions 322bp and 490bp (Figure 4.2). Immediately downstream of the catfish GRF-like peptide is the portion of cDNA which encodes both PACAP1-27 and the extended PACAP1-38. Cleavage at the first processing site (Gly-Arg-Arg) of the preproGRF/PACAP precursor, at amino acids 157-159, would result in a $PACAP_{1\text{-}27NH_2}$ form. Cleavage at residues 168-170 (Gly-Arg-Arg) would produce a complete $PACAP_{1\text{-}38NH_2}$ peptide.

Figure 4.1: Comparison of clones 1.44 and 3.51 coding for *grf/pacap* cDNA. The addition of an extra adenosine in the PACAP coding region of clone 1.44 is illustrated by an arrow on the top line of nucleotide bases. The shift in reading frame for one of the two clones is shown by the difference in amino acids after the arrow. A dibasic processing site is present in both clones at position 29/30 implying PACAP 1-27 could be processed from both clones although the sequence differs from amino acids 22 to 27. The full length clones for *grf/pacap* are shown by box diagrams below the nucleotide sequence. The shaded boxes represent in order, from the left: 1) 5'UTR, 2) signal peptide, 3) 5' cryptic peptide, 4) GRF-like peptide, 5) PACAP, 6) 3'cryptic peptide, and 7) 3'UTR. In clone 1.44, the two open boxes show the poly A+ binding protein and an unknown mRNA connected to the truncated *grf/pacap* cDNA.

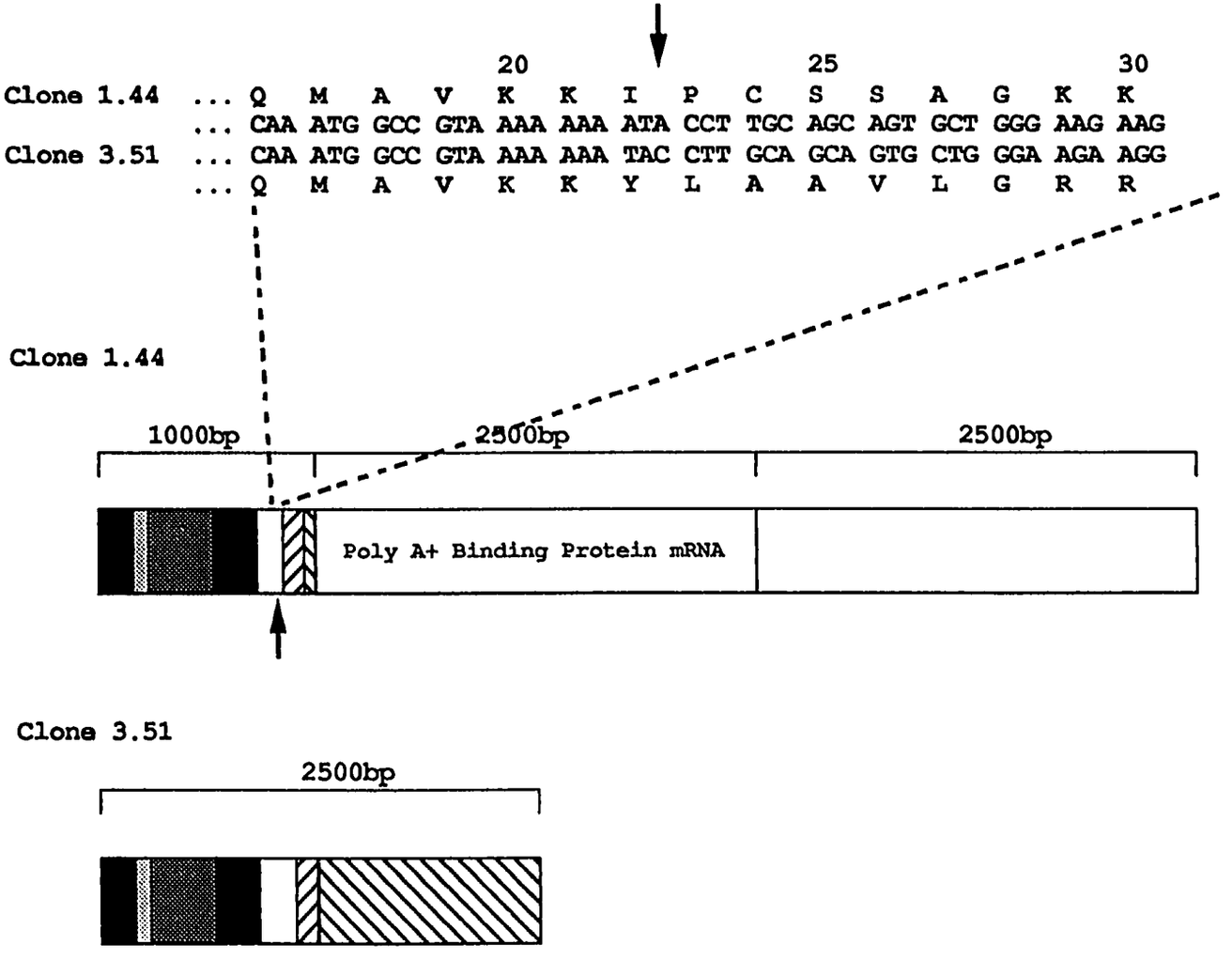
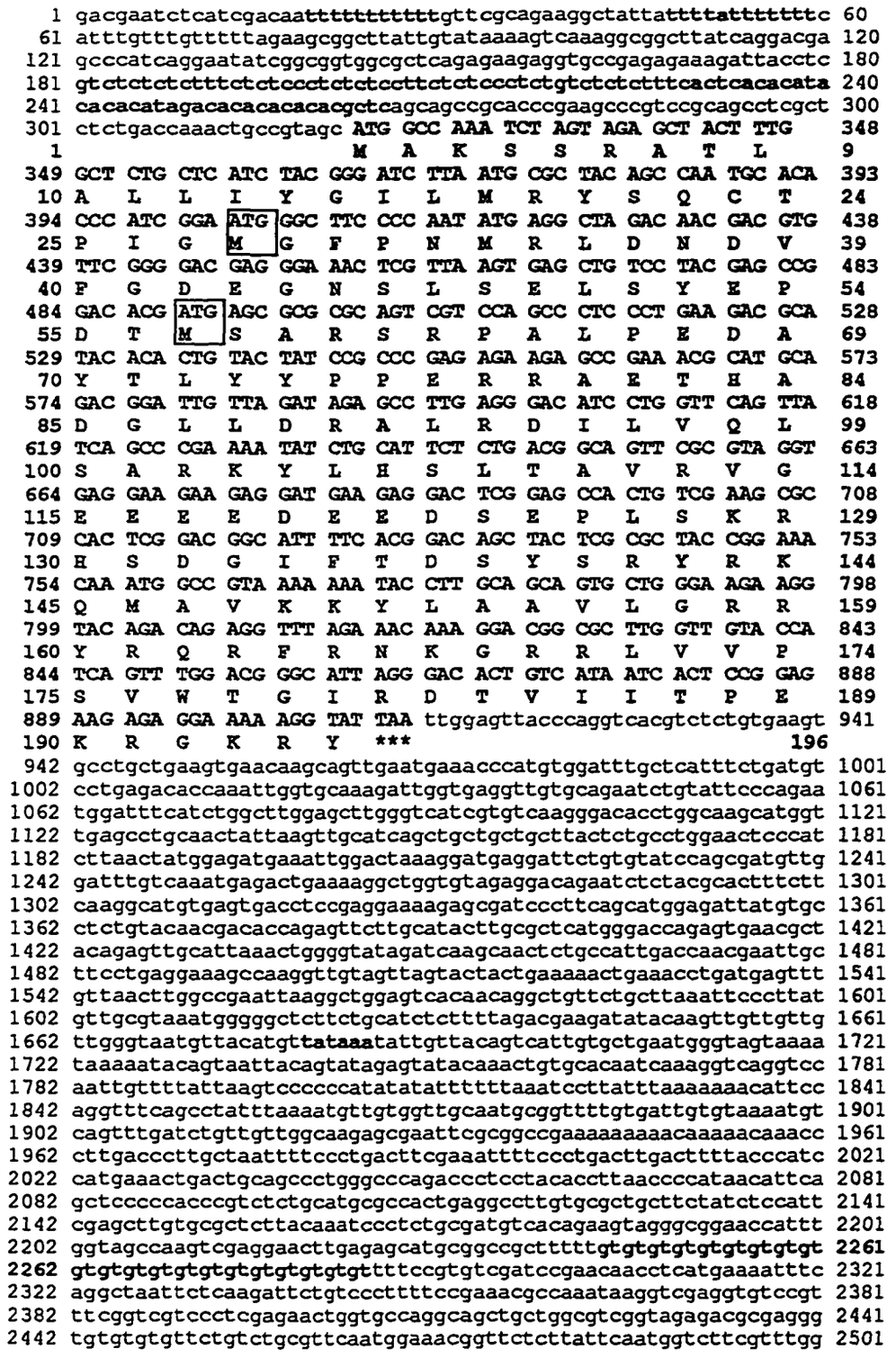

20
25
30
Clone 1.44 ... Q M A V K K I P C S S A G K K
... CAA ATG GCC GTA AAA AAA ATA CCT TGC AGC AGT GCT GGG AAG AAG
Clone 3.51 ... CAA ATG GCC GTA AAA AAA TAC CTT GCA GCA GTG CTG GGA AGA AGG
... Q M A V K K Y L A A V L G R R
Clone 1.44
1000bp
2500bp
2500bp
Poly A+ Binding Protein mRNA
Clone 3.51
2500bp

Figure 4.2: Nucleotide sequence and deduced amino acid sequence of Thai catfish *grf/pacap* cDNA. The boxed ATGs at nucleotide positions 403-405 and 490-492 correspond to the alternative translation initiation start sites. The nucleotides in the UTRs are in small letters; the repeats are in bold. The translated region is in bold capital letters.

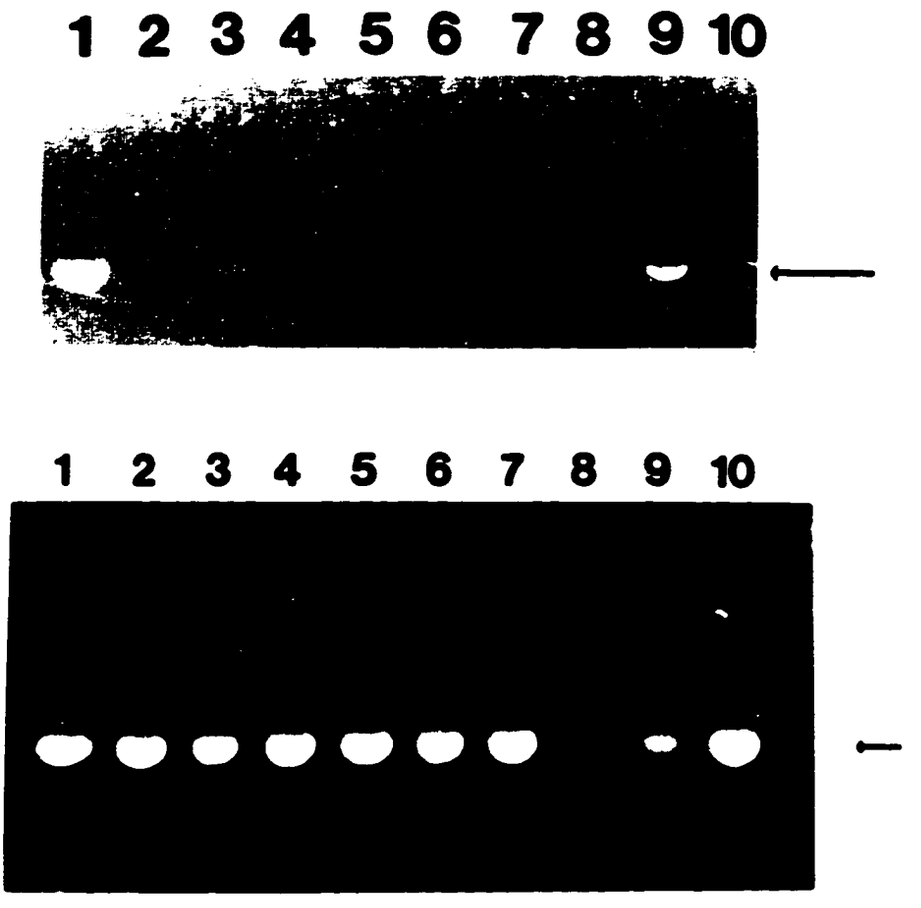

1 gacgaatctcatcgacaattttttttttttgttcgcagaaggctattatttttattttttttc 60
61 atttgtttgtttttagaagcggcttattgtataaaagtcaaaggcggcttatcaggacga 120
121 gcccatcaggaatatcggcggtggcgctcagagaagaggtgccgagagaaagattacctc 180
181 gtctctctctttctctccctctctccttctctccctctgtctctctttcactcacacata 240
241 cacacatagacacacacacacgctcagcagccgcacccgaagcccgtccgcagcctcgct 300
301 ctctgaccaaactgccgtagc ATG GCC AAA TCT AGT AGA GCT ACT TTG 348
1 M A K S S R A T L 9
349 GCT CTG CTC ATC TAC GGG ATC TTA ATG CGC TAC AGC CAA TGC ACA 393
10 A L L I Y G I L M R Y S Q C T 24
394 CCC ATC GGA ATG GGC TTC CCC AAT ATG AGG CTA GAC AAC GAC GTG 438
25 P I G M G F P N M R L D N D V 39
439 TTC GGG GAC GAG GGA AAC TCG TTA AGT GAG CTG TCC TAC GAG CCG 483
40 F G D E G N S L S E L S Y E P 54
484 GAC ACG ATG AGC GCG CGC AGT CGT CCA GCC CTC CCT GAA GAC GCA 528
55 D T M S A R S R P A L P E D A 69
529 TAC ACA CTG TAC TAT CCG CCC GAG AGA AGA GCC GAA ACG CAT GCA 573
70 Y T L Y Y P P E R R A E T H A 84
574 GAC GGA TTG TTA GAT AGA GCC TTG AGG GAC ATC CTG GTT CAG TTA 618
85 D G L L D R A L R D I L V Q L 99
619 TCA GCC CGA AAA TAT CTG CAT TCT CTG ACG GCA GTT CGC GTA GGT 663
100 S A R K Y L H S L T A V R V G 114
664 GAG GAA GAA GAG GAT GAA GAG GAC TCG GAG CCA CTG TCG AAG CGC 708
115 E E E E D E E D S E P L S K R 129
709 CAC TCG GAC GGC ATT TTC ACG GAC AGC TAC TCG CGC TAC CGG AAA 753
130 H S D G I F T D S Y S R Y R K 144
754 CAA ATG GCC GTA AAA AAA TAC CTT GCA GCA GTG CTG GGA AGA AGG 798
145 Q M A V K K Y L A A V L G R R 159
799 TAC AGA CAG AGG TTT AGA AAC AAA GGA CGG CGC TTG GTT GTA CCA 843
160 Y R Q R F R N K G R R L V V P 174
844 TCA GTT TGG ACG GGC ATT AGG GAC ACT GTC ATA ATC ACT CCG GAG 888
175 S V W T G I R D T V I I T P E 189
889 AAG AGA GGA AAA AGG TAT TAA ttggagttacccaggtcacgtctctgtgaagt 941
190 K R G K R Y *** 196
942 gcctgctgaagtgaacaagcagttgaatgaaacccatgtggatttgctcatttctgatgt 1001
1002 cctgagacaccaaattggtgcaaagattggtgaggttgtgcagaatctgtattcccagaa 1061
1062 tggatttcatctggcttggagcttgggtcatcgtgtcaagggacacctggcaagcatggt 1121
1122 tgagcctgcaactattaagttgcatcagctgctgctgcttactctgcctggaactcccat 1181
1182 cttaactatggagatgaaattggactaaaggatgaggattctgtgtatccagcgatgttg 1241
1242 gatttgtcaaatgagactgaaaaggctggtgtagaggacagaatctctacgcactttctt 1301
1302 caaggcatgtgagtgacctccgaggaaaagagcgatcccttcagcatggagattatgtgc 1361
1362 ctctgtacaacgacaccagagttcttgcatacttgcgctcatgggaccagagtgaacgct 1421
1422 acagagttgcattaaactggggtatagatcaagcaactctgccattgaccaacgaattgc 1481
1482 ttcctgaggaaagccaaggttgtagttagtactactgaaaaactgaaacctgatgagttt 1541
1542 gttaacttggccgaattaaggctggagtcacaacaggctgttctgcttaaattcccttat 1601
1602 gttgcgtaaatgggggctcttctgcatctcttttagacgaagatatacaagttgttgttg 1661
1662 ttgggtaatgttacatgttataaatattgttacagtcattgtgctgaatgggtagtaaaa 1721
1722 taaaaatacagtaattacagtatagagtatacaaactgtgcacaatcaaaggtcaggtcc 1781
1782 aattgttttattaagtccccccatatatatttttttaaatccttatttaaaaaaacattcc 1841
1842 aggtttcagcctatttaaaatgttgtggttgcaatgcggttttgtgattgtgtaaaatgt 1901
1902 cagtttgatctgttgttggcaagagcgaattcgcggccgaaaaaaaaacaaaaacaaacc 1961
1962 cttgacccttgctaattttccctgacttcgaaattttccctgacttgacttttacccatc 2021
2022 catgaaactgactgcagccctgggcccagaccctcctacaccttaaccccataacattca 2081
2082 gctcccccacccgtctctgcatgcgccactgaggccttgtgcgctgcttctatctccatt 2141
2142 cgagcttgtgcgctcttacaaatccctctgcgatgtcacagaagtagggcggaaccattt 2201
2202 ggtagccaagtcgaggaacttgagagcatgcggccgctttttgtgtgtgtgtgtgtgtgt 2261
2262 gtgtgtgtgtgtgtgtgtgtttttccgtgtcgatccgaacaacctcatgaaaatttc 2321
2322 aggctaattctcaagattctgtccctttttccgaaacgccaaataaggtcgaggtgtccgt 2381
2382 ttcggtcgtccctcgagaactggtgccaggcagctgctggcgtcggtagagacgcgaggg 2441
2442 tgtgtgtgttctgtctgcgttcaatggaaacggttctcttattcaatggtcttcgtttgg 2501

The second clone 1.44 was 6000bp and appeared to contain 3 mRNAs (Figure 4.1). The 5' sequence was identical to the *grf/pacap* like cDNA clone 3.51 except that 1) only 1000bp were present, 2) an additional adenosine was present within the PACAP coding region, 3) the stop codon occurred prematurely in the PACAP region, and 4) the 3'UTR was about 1500bp shorter than clone 3.51. In clone 1.44 the second mRNA in the middle portion of the clone was about 2500bp and coded for a poly A+ binding protein (Figure 4.1). The third mRNA within clone 1.44 was at the end of the clone. This portion of the clone was about 3450bp. A region of 753bp was sequenced, but did not match sequences in the Blastn database.

Expression of *grf/pacap* mRNA as detected by PCR

The *grf/pacap* mRNA was detected not only within the brain, but also in tissues external to the brain using a sensitive PCR method to amplify the cDNA. *grf/pacap* mRNA was detected only in the brain, testis, ovary, and stomach of the catfish (Figure 4.3A). The *grf/pacap* cDNA was not detected in the pituitary, muscle, liver or pancreas. The RT/PCR products from the brain, testis, ovary and stomach were sequenced to verify the presence of an authentic *grf/pacap* sequence. The cDNA appeared to be of good quality as judged by the PCR products obtained with the tubulin primers 5 and 6 (Figure 4.3B).

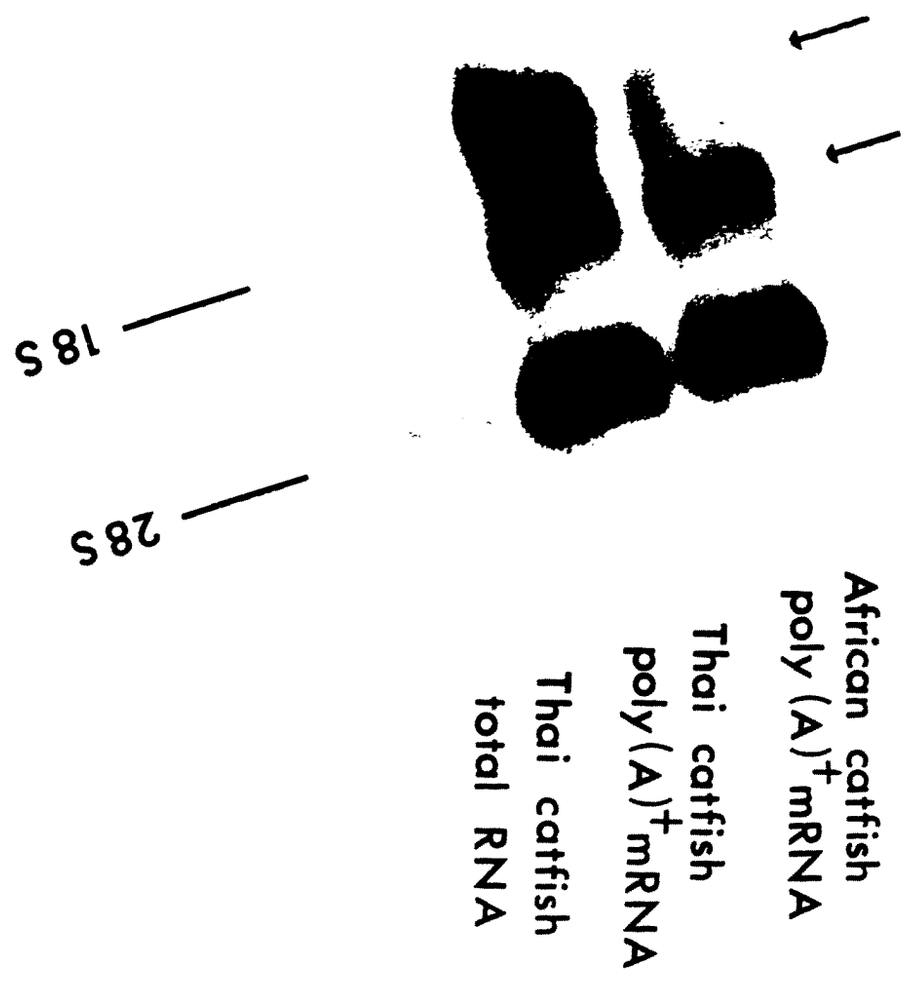
Figure 4.3: Reverse transcriptase assay to detect the *grf/pacap* mRNA in various tissues of Thai catfish. **(A)** Tissue cDNA used for each reaction is as follows: lane 1, brain; lane 2, pituitary; lane 3, stomach; lane 4, muscle; lane 5, testis; lane 6, ovary; lane 7, liver; lane 8, pancreas; lane 9, positive control (2501bp *grf/pacap* clone); lane 10, negative control. **(B)** Tubulin detected by a reverse transcriptase/PCR assay to determine the quality of mRNA/cDNA for the assay in Fig. 4A. The lanes contained the cDNA/mRNA as above except lanes 9 and 10, which contained the herring brain cDNA and tubulin clone (positive controls), respectively.


1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10

mRNA size as determined with northern blot analysis

A northern blot of Thai catfish and African catfish poly A+ rich mRNA and Thai catfish total RNA was screened using the 346bp probe. After several high stringency washes, the northern blot had three bands in the Thai catfish lane, two bands in the African catfish lane, and a single band in the lane containing the Thai catfish total RNA (Figure 4.4). The largest band in each lane is approximately 6000 bases and probably represents co-migration of the *grf/pacap* mRNA with 28S ribosomal mRNA. The second band in each lane is approximately 2500 bases followed by a single band of 1000 bases in the Thai catfish poly A+ mRNA lane. The 18S ribosomal RNA migrated at 2000 bases and did not align with any of these other bands.

Figure 4.4: Northern blot analysis of African and Thai catfish poly A+ rich mRNA. Migration of the 18S and 28S rRNA is indicated on the right. The top arrow on the left is 2500bp and the lower arrow indicates 1000bp.


18S
28S
African catfish
poly (A)+mRNA
Thai catfish
poly(A)+mRNA
Thai catfish
total RNA

Discussion

Catfish cDNA shows PACAP is highly conserved

I have isolated from Thai catfish brains the cDNA encoding a single precursor containing both GRF-like and PACAP peptides. The deduced amino acid sequence of the catfish $PACAP_{1-38}$ is 86% identical to the human sequence, whereas the catfish GRF-like $peptide_{1-45}$ is only 31% identical to human GRF. The evolutionary aspects of these observations are important as the conservation of the *pacap* nucleotide sequence suggests an important role for this molecule. The origin of the mammalian gene that encodes PACAP is clearly rooted in fish, but the origin of the second gene in mammals encoding only GRF is more elusive. If the present fish study is coupled with the chicken study, it is clear the duplication event of the *pacap* gene occurred after the reptilian-bird stem line separated from the mammalian line. The implication of both studies is that the *grf* coding region had a higher nucleotide substitution rate than the *pacap* gene and evidence in mammals supports this idea (Sherwood *et al.* 1994).

Catfish *grf/pacap* mRNA is expressed in brain, gut and gonads

The catfish and chicken data are the first, for vertebrates other than mammals, that the mRNA encoding both the GRF and PACAP peptides is expressed not only in the brain, but also in

the ovary and testis. The expression within the gastrointestinal tract is peculiar to catfish. Tissue-specific expression was not present in pituitary, pancreas, muscle or liver. In evolutionary terms, this expression pattern is similar to the expression of the mammalian *grf* and *pacap* genes. Mammalian *grf* mRNA expression is reported for the hypothalamus, testis, ovary and placenta. To date, studies of mammalian *pacap* mRNA have been limited to the brain and testis; expression was found in both tissues. This supports the idea that the separate mammalian *grf* and *pacap* genes resulted from a duplication of an ancestral *grf/pacap* gene in ancestral mammals. Cross-reactivity studies have detected both mammalian peptides in gastrointestinal tissue (and PACAP in several other tissues) (Arimura 1992), but immunological detection is less specific than mRNA expression, provided the mRNA (cDNA) is sequenced. The importance of the vertebrate PACAP and GRF peptides in gonads and gastrointestinal tract is not yet clear, but GRF is known to act directly on the ovary to promote follicular maturation (Moretti *et al.* 1990). It remains for another study to systematically measure mRNA expression for GRF and PACAP in a variety of mammalian tissues to compare the pattern of tissue expression for *grf/pacap* mRNA in fish and chicken.

Catfish GRF is not well conserved in the GRF family

The functional importance of the GRF peptide present in the catfish precursor is more tenuous than that of PACAP based on reduced structural conservation. The GRF peptide has sequence similarity to both growth hormone-releasing hormone and PACAP-related peptide (PRP). The catfish GRF peptide has 64% identity to the recently identified salmon GRF-like peptide, 58% identity to the carp GRF peptide, 38% to human PRP, and 31% identity to human GRF. This divergence among species is not surprising because the GRF sequence is not highly conserved, even between the human, rat and mouse. The identity between the catfish and salmon GRF peptides is 64%, not unlike the 66% sequence identity between human and rat. The critical comparison, however, may be the biologically active part of the molecule, amino acids 1-29. The sequence of catfish GRF_{1-29} is 38% identical to human GRF_{1-29} and 56% identical to human PRP_{1-29}.

GRF cleavage at N-terminal is important for function

The catfish GRF/PACAP precursor has two potential cleavage sites for the amino terminal of the GRF-like peptide (Figure 4.2). Cleavage at a threonine, which is part of a 4-amino-acid processing site (Nagamatsu *et al.* 1991) would result in a peptide like the carp and salmon GRF-like peptides of 45 amino acids with a free hydroxyl carboxy terminus. Cleavage at the Arg-Arg site found 4 amino acids upstream (Arg-Arg-Ala-Glu-

Thr) is a second possibility. If the threonine cleavage site is used, the catfish GRF-like peptide would have histidine at position 1. The His_1 is identical to GRFs in rat, mouse, carp and salmon, but is unlike other mammalian GRF peptides that have a tyrosine at position 1. If the Arg-Arg site is a processing site, the first amino acid would be alanine, a non-aromatic amino acid. The conservation of an aromatic amino acid such as histidine at position 1 retains the biological activity of GRF; removal diminishes the potency (Ling *et al.* 1984).

Catfish *grf/pacap* cDNA encodes three possible translation start sites

Conservation of sequence is not evident in the 5' untranslated region (UTR) among 3 mammalian, 1 avian and 2 picsine cDNAs for the *grf/pacap* precursors. However, the catfish 5' UTR of 321bp is similar to the human and chicken cDNA in having two possible translation start sites (positions 322 and 490) (Figure 4.2). Catfish and chicken 5'UTRs also contain T, $(CA)_n$, and $(CT)_n$ repeats: the latter repeat sequences have been shown to have a distinct role in chromatin structure and transcriptional activation (Lu *et al.* 1993). The $(CT)_n$ repeats appear to contribute to the formation of a wild-type chromatin structural array that leaves the gene elements in a nucleosome-free, DNAse I-hypersensitive state. This nucleosome-free structure allows regulatory sequences and the promoter/transcriptional start site of the gene to be in an accessible conformation.

Figure 4.5: Amino acid comparison of two teleost, an avian and a mammalian preproPACAPs. Identical residues are shown by blank spaces and "---" is used to frame shift for better amino acid alignment of the sequences.

```
THAI CATFISH    met ala lys ser ser arg ala thr leu ala leu leu ile tyr gly ile leu met
SALMON          --- --- met         lys                                     ile
CHICKEN             ser gly asn val tyr lys thr     leu thr     leu val tyr leu ile
HUMAN PACAP         thr mer cys     gly     atg                 val             ile

                arg tyr ser gln cys thr pro --- ile gly met gly phe pro asn met arg leu asp
                his         val tyr ser ser pro leu     leu asn tyr         leu         glu
                his cys asn val tyr cys ser pro     arg trp thr pro val pro gly ala lys leu
                his ser     val tyr ser ser pro ala ala ala     leu arg phe pro gly ile arg

                asn asp val --- phe gly asp glu --- gly asn ser leu ser glu leu ser tyr glu
                    glu     --- --- tyr         asp                 pro ala     ala phe asp
                leu glu glu glu val tyr         asp         thr     gln asp phe ala leu arg
                pro glu glu glu ala tyr gly     asp         pro     pro asp phe gly --- gly

                pro asp thr met ser ala arg ser arg pro ala leu pro glu asp ala tyr thr leu
                ser     gln ile ala ile         pro     ser val ala asp     leu
                ala gly ala pro gly gly gly gly pro arg pro arg trp gly arg cys thr ala leu
                ser glu pro pro gly     gly     pro ala ser ala     arg ala     ala ala trp

                                                <-------------------------------------------

                tyr tyr pro pro glu arg arg ala glu thr his --- ala asp gly leu leu asp lys
                                    lys gly thr     arg     ---             met phe asn
                                gly lys         --- --- ---     ---             ile phe ser
                            ala gly --- --- --- arg arg asp val     his     ile     asn glu


                ------------------- GHRH-LIKE PEPTIDE------------------->
                ala leu arg asp ile leu val gln leu ser ala arg lys tyr leu his ser leu thr
                    tyr     lys ala     gly                                         met
                    tyr     lys leu     gly                     asn                 met
                    tyr     lys val     asp                         his     gln     val

                ------------------------------------------------------->
                ala val arg val gly --- --- glu glu glu glu asp glu glu asp ser glu
                    lys             --- --- gly gly ser thr met     asp     thr
                    lys             --- --- ser gly leu gly         ala glu pro leu
                    arg gly         gly ser leu gly gly gly ala gly asp     ala

                ------->        <-------------------------------------------
                pro leu ser lys arg his ser asp gly ile phe thr asp ser tyr ser arg tyr arg

                                        ile

                ------------------------ PACAP------------------------------>
                lys gln met ala val lys lys tyr leu ala ala val leu gly arg arg tyr arg gln
                                                                    lys
                                                                    lys         lys
                                                                    lys         lys

                ------------->
                arg phe arg asn lys gly arg arg leu val val pro ser val trp gly ile arg asp
                    tyr                             gly tyr leu ***
                    val lys                     val ala tyr leu ***
                    val lys                     ile ala tyr leu ***


                trp val ile trp pro glu lys arg gly lys arg tyr ***
```

As with other known *grf/pacap* cDNA sequences, translation is believed to be initiated at the first ATG. If the precursor is cleaved between amino acid positions 23 and 24, the catfish signal peptide would be one amino acid shorter than the human, chicken and ovine and one residue longer than the salmon signal peptides. Amino acid similarity of the catfish signal peptide is quite high compared with salmon (87%), ovine (58%), and human (57%) PACAP signal peptide sequences (Figure 4.5). However, another initiation start site which has higher identity to Kozak's (1991a; 1991b) consensus sequence (GCC$^{A}/_{G}$CCAUGG) was found at position 490bp (Figure 1). Only the first start site (position 322) possessed a guanine in the +4 position; this nucleotide is found in 47% of the reported mRNA transcripts. If translation were initiated at either site, the resulting signal peptides would fulfill the criteria of a signal peptide: 1) proper length (23 and 26 residues), 2) amino terminal region of the signal peptide with at least one positively charged residue, 3) hydrophobic core, 4) cleavage site preceded by a sequence of about five residues that are more polar than the hydrophobic core and 5) cleavage site with small neutral side chain residues at the amino-terminal side. These rules of translation suggest initiation is equally likely to occur at either position along the mRNA sequence.

Tetraploid catfish express two different *grf/pacap* cDNAs

One of the two positive clones recovered from the cDNA library, clone 1.44 of 6000bp was unusual compared to clone 3.51 of 2501bp (Figure 4.1). The 6000bp clone contained coding regions not only for a *grf/pacap* protein (1000bp), but also for a poly A+ binding protein (≈2500), and for an unknown sequence (≈2500bp). The significance of the poly A+ binding protein and the unknown mRNA region is unclear and may be due to a cloning artifact. The 1000bp segment of clone 1.44 coding for *pacap/grf* peptide was identical to clone 3.51 in the coding region for the GRF peptide, but differs in the PACAP coding region by 1 nucleotide and was shorter due to the lack of an extended 3'UTR (Figure 4.6). This shortening of the 3'UTR may explain the presence of the 1000 base band in the northern analysis even though a clone for only 1000bp was not isolated from the catfish library. An extra adenosine in the catfish PACAP coding region of clone 1.44 resulted in a frame shift in which the coding for the processing site of $PACAP_{1-38}$ was missing. However, like the cDNA of other PACAP peptides, a dibasic processing site (Gly-Lys-Lys) was present at residues 28-30 of PACAP and cleavage here would result in a modified $PACAP_{1-27NH2}$ (Figure 4.1). One explanation for the two distinct mRNA transcripts is that they are products of two

genes due to the tetraploid nature of catfish. One gene produces only an altered $PACAP_{1-27}$ of unknown significance, whereas the other gene produces conventional $PACAP_{1-27}$ and $PACAP_{1-38}$.

REFERENCES

Arimura A 1992. Pituitary adenylate cyclase-activating polypeptide (PACAP): discovery and current status of research. Regulatory Peptides 37: 287-304.

Bagnato A, Moretti C, Ohnishi J, Frajese G, Catt KJ 1992 Expression of the growth hormone-releasing hormone gene and its peptide product in the rat ovary. Endocrinology 130: 1097-1102.

Bell GI 1986 The glucagon superfamily: Precursor structure and gene organization. Peptides 7: 27-36.

Berry SA, Pescovitz OH 1988 Identification of a rat GRF-like substance and its messenger RNA in rat testis. Endocrinology 123: 661-663.

Berry SA, Pescovitz OH 1990 Ontogeny and pituitary regulation of testicular growth hormone-releasing hormone-messenger ribonucleic acid. Endocrinology 127: 1404-1411.

Birnboim M 1983 A rapid alkaline extraction method for the isolation of plasmid DNA. Methods in Enzymology 100: 243-255.

Campbell RM, Scanes CG 1992 Evolution of the growth hormone-releasing factor (GRF) family of peptides. Growth Regulation 2: 175-191.

Chartrel N, Tonon MC, Vaudry H, Conlon JM 1991 Primary structure of frog pituitary adenylate cyclase-activating polypeptide (PACAP) and effects of ovine PACAP on frog pituitary. Endocrinology 129: 3367-3371.

Chomczynski P, Sacchi N 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Annals in Biochemistry 162: 156-159.

Coe IR, Munroe R, Sherwood NM 1992 Isolation of different brain-specific isotypes of α-tubulins from chum salmon. DNA sequence 3: 257-262.

Culler MD, Paschall CS 1991 Pituitary adenylate cyclase activating polypeptide (PACAP) potentiates the gonadotropin-releasing activity of luteinizing hormone-releasing hormone. Endocrinology 129: 2260-2262.

Denver RJ, Licht P 1989 Neuropeptides influencing in vitro pituitary hormone secretion in hatchling turtles. Journal of Experimental Zoology 251: 306-315.

Denver RJ, Licht P 1991 Several hypothalamic peptides stimulate thyrotropin and growth hormone secretion by adult turtle pituitary glands. Comparative Biochemistry and Physiology 100: 603-606.

Giladi E, Shani Y, Gozes I 1990 The complete structure of the rat VIP gene. Molecular Brain Research 7: 261-267.

Gottschall PE, Tatsuno I, Miyata A, Arimura A 1990 Characterization and distribution of binding sites for the hypothalamic peptide, pituitary adenylate cyclase-activating polypeptide. Endocrinology 127: 272-277.

Hart GR, Gowing H, Burrin JM 1992 Effects of a novel hypothalamic peptide, pituitary adenylate cyclase activating

polypeptide, on pituitary hormone release in rats. Journal of Endocrinology 134: 33-41.

Hosoya M, Kimura C, Ogi K, Ohkubo S, Miyamoto Y, Kugoh H, Shimizu M, Onda H, Oshimura M, Arimura A, Fujino, M 1992 Structure of the human pituitary adenylate cyclase-activating polypeptide gene. Biochemica Biophysica Acta 1129: 199-206.

Koves K, Arimura A, Somogyvari-Vigh A, Vigh S, Miller J 1990 Immunohistochemical demonstration of a novel hypothalamic peptide, pituitary adenylate cyclase activating polypeptide, in the ovine hypothalamus. Endocrinology 127: 264-271.

Kozak, M. (1991a). Structural features in eukaryotic mRNAs that modulate the initiation of translation. Journal of Biological Chemistry. 266: 19867-19870.

Kozak, M. (1991b). An analysis of vertebrate mRNA sequences: Intimidations of translational control. Journal of Biological Chemistry 115: 887-903.

Ling N, Baird A, Wehrenberg WB, Ueno W, Munegun T, Brazeau P 1984 Synthesis and in vitro bioactivity of C-terminal deleted analogs of human growth hormone-releasing factor. Biochemical Biophysical Research Communications. 123: 854-861.

Lu, Q., Wallrath, L.L., Granok, H., and Elgin, S.C.R. (1993). Molecular and Cellular Biology 13: 2802-2814.

Malagon NM, Castaño JP, Dobado-Berrios PM, Garcia-Navarro S, Garcia-Navarro F 1991 Human pancreatic growth hormone-

releasing factor (1-44) stimulates GH cells in an anuran amphibian (*Rana perezi*). General and Comparative Endocrinology 84: 461-469.

Margioris AM, Brockman G, Bohler Jr. HCL, Grino M, Vanvakopoulos N, Chrousos GP 1990 Expression and localization of growth homone-releasing hormone messenger ribonucleic acid in rat placenta: in vitro secretion and regulation of its peptide product. Endocrinology 126: 151-158.

Mayo KE, Cerelli GM, Rosenfeld MG, Evans RM 1985 Characterization of cDNA and genomic clones encoding the precursor to rat hypothalamic growth hormone-releasing factor. Nature 314: 464-467.

Moretti C, Fabbri A, Gnessi L 1990 Immunohistochemical localization of growth hormone releasing hormone in human gonads. Journal of Endocrine Investigations 13: 301-305.

Nagamatsu S, Chan SJ, Falkmer S, Steiner DF 1991 Evolution of the insulin superfamily. Journal of Biological Chemistry 266: 2397-2402.

Ohkubo S, Kimura C, Ogi K, Okazaki K, Hosoya M, Onda H, Miyata A, Arimura A, Fujino M 1992 Primary structure and characterization of the precursor to human pituitary adenylate cyclase activating polypeptide. DNA and Cell Biology 11: 21-30.

Parker DB, Coe IR, Dixon GH, Sherwood NM 1993 Two salmon neuropeptides encoded by one brain cDNA are structurally

related to members of the glucagon superfamily. European Journal of Biochemistry 215: 439-448.

Propato-Mussafiri R, Kanse SM, Ghatei MA, Bloom SR 1992 Pituitary adenylate cyclase activating polypeptide releases 7B2, adrenocorticotrophin, growth hormone and prolactin from the mouse and rat clonal pituitary cell lines AtT-20 and GH3. Journal of Endocrinology 132: 107-113.

Rivier J, Spiess J, Thorner M, Vale W 1982 Characterization of growth hormone releasing factor from a human pancreatic islet tumor. Nature (London) 300: 276-278.

Sanger F, Nicklen S, Coulson AR 1977 DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science (USA) 74: 5463-5467.

Sherwood NM, Parker DB, McRory JE and Lescheid DW 1994 Molecular evolution of growth hormone-releasing hormone and gonadotropin-releasing hormone. In: Fish Physiology, N.M. Sherwood and C.L. Hew (eds). 13: 3-66 Academic Press, Orlando.

Suhr ST, Rahal OJ, Mayo KE 1989 Mouse growth hormone-releasing hormone: precursor structure and expression in brain and placenta. Molecular Endocrinology 3: 1693-1700.

Uddman R, Luts A, Arimura A, Sundler F 1991 Pituitary adenylate cyclase-activating polpeptide (PACAP), a new member of the vasoactive intestinal peptide (VIP)-like peptide in the respiratory tract. Cell and Tissue Research 265: 197-201.

Vaughan JM, Rivier J, Speiss J, Peng C, Chang JP, Peter RE, Vale W 1992 Isolation and characterization of hypothalamic growth-hormone releasing factor from common carp, *Cyprinus carpio.* Neuroendocrinology 56: 539-549.

Chapter 5

Origin of the glucagon superfamily as determined from two protochordate genes encoding pituitary adenylate cyclase-activating polypeptide (PACAP) and related family members.

Summary

To address the origin of the glucagon superfamily, I isolated and sequenced the cDNA and partial gene that encode pituitary adenylate cyclase-activating polypeptide (PACAP) from a protochordate (tunicate), a sister group of the amphioxus and vertebrates, but one that evolved before the amphioxus. This is the first report of any superfamily member sequenced from an invertebrate. Transcription of the tunicate *pacap* gene results in a mRNA that is 507bp. The gene contains 3 exons that encode a signal peptide, GRF-like peptide$_{1-27}$ and PACAP$_{1-27}$; the tunicate GRF-like peptide has 59% identity with human GRF. The deduced amino acids of tunicate PACAP$_{1-27}$ have 96% identity with the ovine, human and salmon PACAP$_{1-27}$ forms. Another cDNA clone was isolated and encodes a signal peptide, a cryptic peptide, a glucagon-like peptide (59% identity with human glucagon) and a vasoactive intestinal polypeptide (VIP)-like molecule with 67% identity to human VIP. A comparison of the two clones shows high inter- and intra-exon nucleotide sequence identity. Sequence analysis suggests that an exon duplication followed by a gene duplication was responsible for the origin of the two genes. It is argued that these two genes are derived from the protochordate ancestral genes that led to the vertebrate forms of GRF, PACAP, VIP, PHM and glucagon.

Introduction

Over the past century, studies on protochordates (tunicates/sea squirts) have increased our understanding of chordate phylogeny (Garstang 1928; Berril 1955). Protochordate maturation consists of a two step process involving a motile juvenile stage followed by morphogenesis to a settled stage when the organism matures. The Garstang theory postulates that vertebrates arose when an ancestral relative forewent settling and became reproductive within the larval stage (Satoh and Jeffery 1995). Thereafter, this neotenous motile organism became reproductive and possibly gave rise to the ancestral cephalochordates (amphioxus) and vertebrates. This theory helps us to understand physiological features of vertebrate development and evolution. However, the evolution of the nervous system, in the transition from invertebrates to vertebrates, is not clearly understood. The origin of such features as the forebrain and pituitary in vertebrates still remain unknown because organisms that contain the primordial structures are lacking.

In comparison to present day vertebrates, the protochordate nervous system is simple, consisting of a neural gland, neural ganglion (brain) with nerve roots innervating the branchial sac and viscera and a well developed nerve plexus associated with the dorsal strand. The processes of the dorsal strand plexus go to the dorsal blood sinus, the brain and various internal organs

(Bullock and Horridge 1965; Ruppert EE 1990; Gorbman A 1995; Mackie GO 1995).

To understand the evolution of the nervous system, one approach is to compare the structure and location of neuropeptides in protochordates and mammals. However, limited structural data exists in general for genes, cDNAs and proteins with homology to their mammalian counterparts. Studies have shown that protochordates produce neuropeptides that cross-react with antiserum raised against mammalian peptides (Thorndyke and Goldsworthy 1988) including LHRH, somatostatin (Fritsch *et al.* 1982), and neurotensin-like peptides (Fritsch *et al.* 1982). Also, glucagon (Pestarino 1990) and αMSH (Pestarino and Facchinetti 1995) immunoreactivity have been detected within the protochordate *Ciona intestinalis.* In the tunicate *Ciona intestinalis*, cross-reactivity is detected in the nervous system and alimentary tract with antiserum raised against mammalian insulin (O'Neil *et al.* 1986; Ebberink *et al.* 1989; Galloway and Cutfield 1988), neuropeptide Y (Pestarino 1992; Fritsch *et al.* 1982), PHI (Pestarino 1993), VIP, secretin and pancreatic polypeptide (Fritsch *et al.* 1982). Evidence relating to immunocytochemical localization, tissue expression and nucleotide and amino acid sequences of protochordate genes and peptides, should clarify aspects of the evolution of the vertebrate nervous system. To date, the sequence of 2 protochordate neuropeptides with identity to their mammalian counterparts have been isolated.

Immunocytochemical evidence initially showed a gonadotropin-releasing hormone (GnRH)-like peptide was associated with tunicate neural structures (Georges and Dubois 1980; Kelsall et al. 1990). Mackie (1995) showed that immunoreactive GnRH cells form a neural plexus surrounding the dorsal strand. Determination of the peptide sequence of two forms of GnRH made it possible to prepare synthetic forms of GnRH (Powell *et al.* 1996). Injections of these two tunicate GnRH forms resulted in an increase in the content of estradiol in the gonads (Sherwood, Rivier and Mackie, unpublished observation). Hence, a separation of approximately 650 million years (Doolittle *et al.* 1996) did not obscure the origin of GnRH in that the tunicate GnRH has sequence identity (60%) and a conserved function to that found for mammalian GnRH.

As mentioned in Chapter 1, the human glucagon superfamily is composed of GRF (Rivier *et al.* 1982; Guillemin *et al.* 1982), glucagon (Thompsen *et al.* 1972), secretin (Carlquist *et al.* 1985), VIP (Itoh *et al.* 1983), GIP (Brown 1971) and PACAP (Miyata *et al.* 1989). PACAP, the newest family member, was of particular interest to me because 1) the nucleotide and amino acid sequence is highly conserved among mammals, birds and fish and 2) in addition to its role as a releaser of pituitary hormones, it appears to have a role as a growth factor in developing nervous systems (DiCicco-Bloom 1992; Robberecht *et al.* 1994; Pesce *et al.* 1996; McRory Chapter 3). It is assumed that these family members share a common ancestor based on

similar amino acid sequences and intron/exon structure (Campbell and Scanes 1992; Sherwood *et al.* 1994). The hypothesis that extant superfamily members originated from a common ancestor was speculative because structural evidence was not available. To investigate the origin of the glucagon superfamily, I used molecular techniques to determine PACAP's nucleotide sequence and tissue expression within the protochordate *Chelyosoma productum*. This chapter reports the structure of two protochordate genes and cDNAs in which gene 1 encodes a GRF_{1-27}-like peptide and $PACAP_{1-27}$ and gene 2 encodes two peptides with amino acid sequences similar to glucagon and VIP.

Materials and Methods

Construction of cDNA library

Tunicates (*Chelyosoma productum*) were dissected from their tunic. The dorsal internal organs, mainly the neural ganglion, neural gland and dorsal strand were separated from the gonads, gut, and branchial basket. These dorsal organs were placed immediately in liquid nitrogen and stored at -80°C. Total RNA was extracted with an acidic guanidinium thiocyanate method (Chomczynski and Sacchi, 1987) and poly A+-rich mRNA was purified with the Poly Attract System (Invitrogen). The cDNA library was constructed with the UNI-Zap-cDNA synthesis kit (Stratagene) and Gigapack packaging mix.

Amplification of tunicate *grf/pacap* cDNA by PCR

DNA was amplified using a cDNA library lysate and the degenerate primers PA-1 (5'-cattcggatgggatcttcacggatag) and 3'-PA (5'-catgtttggacagaacacaacgtgagcg). First round amplification was done in a 50µl volume reaction that contained 0.2µg cDNA, 5U Taq, 1x Taq buffer (Pharmacia), 200mM dNTP's, 2mM $MgCl_2$, 20 pmol of primers PA-1 and 3'-PA with 40 cycles at 94°C for 1.5 min, 45°C for 2 min, 72°C for 2.5 min. PCR reactions were electrophoresed on a 2% agarose gel. The cDNA in the bands from the gel was cloned into pBluescript KS+ (Stratagene) and the resulting plasmids were electroporated into XL-1 Blue (MRF') competent cells, and

prepared for sequencing with an alkaline hydrolysis method (Birnboim, 1983). Both strands of the plasmid were sequenced with [α-^{35}S] dATP using the chain termination method (Sanger *et al.* 1977) with Sequenase 2.0 (US Biochemical Corp.) and Vent (exo-) (New England Biolabs). All sequencing gels were 6% polyacrylamide/7M urea wedge gels, dried under vacuum at 80°C and exposed to Kodak XAR-5 film for 12-24h.

Tissue assay by reverse transcription/PCR

Messenger RNA was isolated from the following tissues: neural gland, dorsal strand/neural ganglion, gonad, gonad/digestive gland, intestine, heart, tunic, and branchial basket. Single stranded cDNA was synthesized from 0.5µg mRNA for each tissue using 200U avian reverse transcriptase (H$^-$ Superscript reverse transcriptase, BRL), 10mM dithiothreitol, 1mM each dNTP, 160U RNA guard, 2mM oligo dT_{20}, and 1X H$^-$ RT buffer to a total reaction volume of 20µl. The reaction proceeded for 90 min at 41°C followed by 5 min at 90°C. DNA amplifications were done in a 50µl volume that contained 0.5µg cDNA, 1x Taq buffer (Pharmacia), 200µM dNTP's, 1.5mM $MgCl_2$ and 20pmol of each primer. The *pacap*-specific primers were Tun 3, (5'-tacactgg attgtcttggccgcc) and Tun 4 (5'-cgctcag catgagttctgtc) and the *vip*-specific primers were Tun 5 (5'-gacggtaacgattcttatgc) and Tun 6 (5'-gcctaacagatagcctagtc). All reagents, except the Taq enzyme, were mixed, layered with mineral oil, and heated to 95°C for 5 min. The mixture was

frozen rapidly in a dry ice/ethanol bath and 5U Taq DNA polymerase (Promega) was added. The tubes were replaced in the preheated 95° thermal cycler where 40 cycles at 95°C for 1 min, 55°C for 2 min, and 74°C for 1.5 min were completed. Tubulin PCR amplifications were in a 50µl reaction that contained 0.5µl cDNA from each tissue, 5U Taq DNA polymerase, 1x Taq buffer (Pharmacia), 200µM each dNTP, 3mM $MgCl_2$, and 20pmol of each tubulin primer (Chapter 4). The reactions were carried out for 35 cycles of 94° for 1 min, 55° for 1.5 min, and 72° for 1.5 min.

Screening of cDNA library

A total of $5x10^4$ pfu from the tunicate library was screened. Duplicate nylon membrane (Bio-Rad) lifts were prehybridized at 50°C in 6x SSC, 5x Denhardt's solution and 0.5% SDS for 4 h. The hybridization solution consisted of 6x SSC and 0.5% SDS to which the probe was added. The probe was a 163bp product amplified by PCR with primers PA-1 and 3'-PA. The probe was labeled with [α-^{32}P]dCTP ($7.8x10^6$ cpm/ml), and the membranes and the probe were incubated at 50°C overnight. The membranes were washed under high stringency (0.1xSSC/0.1%SDS) for 50 min at 65°C, then exposed to Kodak XAR-5 film for 5 days at -80°C. Isolated single positive clones were cored and inserts rescued with *in vitro* excision.

Genomic DNA amplification

High molecular weight tunicate genomic DNA was extracted after Proteinase K (Sigma) treatment by successive phenol:chloroform:isoamyl alcohol (24:24:1) washes, and dialysis against 0.01M Tris-HCl/EDTA (TE). Primers to the 5' untranslated region (UTR) and 3'UTR of the tunicate *pacap* (Tun 3 and 4) and *vip* (Tun 5 and 6) cDNAs were used in the amplification of the tunicate genes. Amplifications were done in a 50µl volume (1.0µl DNA, 5U Taq, 1x Taq buffer (Pharmacia), 200µM dNTP's, 2mM $MgCl_2$, 20 pmol of each primer) with 40 cycles at 94°C for 1.5 min, 45°C for 2 min, 72°C for 2.5 min. The reaction was electrophoresed through a 1.5% agarose gel. A band was cloned into pBluescript KS+ (Stratagene); the plasmid and insert were electroporated into XL-1 (MRF') competent cells; and the extracted DNA was prepared for sequencing with an alkaline hydrolysis method (Birnboim, 1983).

Zoo blot and Southern analysis

For the zoo blot, DNA (10ug) was prepared from rat (*Rattus norviegus*), starling (*Sturnus vulgarus*), chicken (*Gallus domesticus*), alligator (*Alligator mississippiensis*), salmon (*Oncorhynchus nerka*), catfish (*Clarias macrocephalus*), reedfish (*Calamoichthys calabaricus*), tunicate (*Chelyosoma productum*), urchin (*Strongylocentrotus purpuratus*), *Drosophila* (*Drosophila melanogaster*), yeast (*Saccharomyces cerevisiae*), and bacteria

(*Escherichia coli*). DNA for the zoo blot and Southern blot (tunicate DNA only) were digested to completion with Eco RI and electrophoresed in a 0.8% agarose gel. The DNA was transferred as to manufacturer's (Bio-Rad) specification for the alkaline Zeta-Probe GT membrane. Prehybridization was in 7% SDS, 0.5M $NaHPO_4$, and 1M EDTA at 65°C for 15min. Hybridization was overnight (14hr) at 65°C in fresh prehybridization solution plus the 163bp probe labeled with [α-^{32}P]dCTP. The hybridized membranes were rinsed with 5%SDS, 40mM $NaHPO_4$, 1mM EDTA and then washed for 45min at 45°C with fresh solution. The wash solution was then changed to 1% SDS, 40mM $NaHPO_4$, 1mM EDTA and washed twice for 45min at 65°C with fresh solution. After washing, the membrane was sealed in plastic and exposed at -80°C for 8 days to Kodak BIOMAX film.

In situ hybridization of tunicate *grf/pacap* and *glucagon/vip* mRNA

Localization of the tunicate *grf/pacap* or *glucagon/vip* mRNA in sections from the neural ganglion of *Cheylosoma productum* was done by *in situ* hybridization using a digoxigenin (DIG)-labeled tunicate insulin RNA probe. All RNA probes were synthesized, purified, and tested as to the manufacturer's (Boehringer Mannheim) instructions. The changes in protocol for fixation, prehybridization and hybridization are listed below. The tunicate neural gland and ganglia were dissected and pinned on Sylgard coated dishes and fixed for 3 h in 4%

paraformaldehyde in phosphate buffered saline (PBS) (pH 7.4). The fixed tissues were washed in PBS and soaked overnight in 30% sucrose. Tissue was embedded in O.T.C compound, then sectioned (10μm) and allowed to dry on poly-L-lysine coated slides. Sections were fixed again in 4% paraformaldehyde for 5 min, washed 3 X in PBS (5 min each) and placed for 10 min in 2 X SSC. Prehybridization was in 2 X SSC for 2 h at room temperature. This solution was exchanged for the hybridization solution that consisted of 2 X SSC with a DIG-labeled RNA probe diluted 1:200. The hybridization solution was incubated overnight at 42°C. The sections were washed with SSC (0.5 X SSC) followed with 2% normal goat serum in TBS buffer for 30 min at room temperature. The remaining steps, involving the anti-digoxigenin antibody and the substrate detection, were performed as to the manufacturer's instructions.

Results

Isolation of the tunicate *grf/pacap* and *glucagon/vip* mRNAs

I have isolated a cDNA that encodes for $PACAP_{1-27}$ from a tunicate cDNA library. Encoded within the *pacap* cDNA is another peptide that could be a GRF-like peptide (59% identity to human GRF) or glucagon-like peptide (63% identity to human glucagon). However, with GRF as the 5' adjacent peptide in preproPACAP in birds and fish, the peptide in the same position tunicates is most likely a GRF-like peptide and not glucagon. Another clone, distinct from the *pacap* clone was isolated from the tunicate cDNA library. This clone, which encodes a VIP-like peptide (67% identity to human VIP) and glucagon-like peptide (59% identity to human glucagon), was not the same as the *pacap* cDNA clone because of nucleotide changes in the exons, the presence of a region encoding a cryptic peptide, and differences in the introns of the two isolated genes.

For isolation of the tunicate *pacap* mRNA, two primers, PA-1 and 3'-PA, were used with the PCR to amplify a 163bp fragment that encoded only tunicate PACAP and 57bp of the 3'untranslated region. This clone was used to screen a tunicate cDNA library. After screening $5x10^4$ pfu from the library, I found 15 plaques that hybridized to the radioactive probe. Of the 15 clones that hybridized to the probe, 10 were purified. These clones were removed from the phage with *in vitro* excision, purified and sequenced. It was determined that all

clones were identical and all coded for tunicate GRF_{1-27}/$PACAP_{1-27}$ (Figure 5.1A) The remaining 5 phage DNA clones that hybridized to the probe were purified and one encoded a different cDNA. One exon has identity to human glucagon (59% identity) and the other exon to human VIP (67% identity) (Figure 5.1B). Therefore, tunicates contain two different mRNAs; one encodes tunicate $PACAP_{1-27}$ named because of the 96% identity to the mammalian $PACAP_{1-27}$ protein. The other clone was referred to as a tunicate *glucagon/vip* cDNA because it had sequence characteristics similar to human glucagon and VIP.

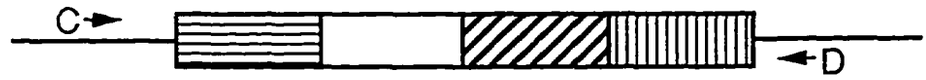
The tunicate *grf/pacap* cDNA clone was 507bp long and encodes a signal peptide, a GRF-like $peptide_{1-27}$ and $PACAP_{1-27}$; no other proteins are encoded. The tunicate cDNA does not encode the longer version of $PACAP_{1-38}$ or the GRF-like peptide of 43-46 amino acids.

Figure 5.1: Nucleotide and deduced amino acid sequences of the tunicate *grf/pacap* and *glucagon/vip* cDNA clones. **(A)** shows the 507bp tunicate *grf/pacap* cDNA and **(B)** shows the tunicate 883bp *glucagon /vip* cDNA. The primers used for the expression study are underlined or overlined in the sequences and shown on the box diagrams. Coding regions are indicated by boxes and 5' and 3' untranslated regions by thin lines. The box with horizontal lines encodes the signal peptide and the white box encodes a cryptic peptide. In box diagram A, the diagonally striped box encode GRF-like peptide and the vertical lined box encodes PACAP. In box diagram B, the diagonally striped box encodes a glucagon-like peptide and the box with vertical lines encodes a VIP-like peptide.

A


```
tttgacgatcgttttacggcgcggcggagcacgcgcgcgcgcgcgcgggagctagggatttttacactggatgtcttggcc 80
gcc ATG GCC AAG TCA AGT GGG GCG ACC CTC GCC CTC CTA ATT TAC GGA ATC ATT ATG AGA  140
    M   A   K   S   S   G   A   T   L   A   L   L   I   Y   G   I   I   M   R    19
TAC TCC CAC TCC GAT GGG ATA TTC ACA AAA GAT TAT CGG AAG TAC CTC GGG CAA CTG CGA  200
Y   S   H   S   D   G   I   F   T   K   D   Y   R   K   Y   L   G   Q   L   R    39
GCT CAA AAA TTC CTG CAA TGG CTT ATG AAG CGC CAC TCG GAT GGG ATC TTC ACG GAC AGC  260
A   Q   K   F   L   Q   W   L   M   K   R   H   S   D   G   I   F   T   D   S    59
TAT AGC CGC TAC CGG AAT CAA ATG GCT GTT AAG AAA TAC CTG GCG GCA GTG CTT GGG AAA  320
Y   S   R   Y   R   N   Q   M   A   V   K   K   Y   L   A   A   V   L   G   K    79
AGG TAT AAA CAA AGG TAA agaaacaatggacgcaattttaccaatatacatgttacgacagaactcatgctgag 394
R   Y   K   Q   R   *                                                           85
cgaaattttaacgcaccaccgactaggctatctgttaataaagtatttacatttaagcattatcatttacgttaaaaaca 474
gttatatattttaaaaaaaaaaaaaaaaaaaa                                               507
```

B




```
aagtcaagtggggcgaccctcgccctcctaatttacggaatcattatgagatactcccttatttcgatcgacgtttgcga 80
taccttgacaacgctgtttatgacggtaacgattcttatgctaattaccacaaaacttggcttttgtgaaggacatcaaa 160

atcgaataacaattccccat ATG CTC GCT ACG ACG GGG AGT ACT TTA GCA CTT CTT CCC GGC GTT 225
                     M   L   A   T   T   G   S   T   L   A   L   L   P   G   V    15
GTG TCA AAT ACA AGA ATC TCG ATC GAC GTT TGC GAT ACC TGG ACA ACG CTT CTT ATG CTA  285
V   S   N   T   R   I   S   I   D   V   C   D   T   W   T   T   L   L   M   L    35
ATT ACC ACA AAA CTT GGC TTT GTT GAA GGA CAT CAA AAT ACG ATA ACA ATT CCC CAT ATG  345
I   T   T   K   L   G   F   V   E   G   H   Q   N   T   I   T   I   P   H   M    55
AGG CGG ACG ACG GGG GCA ACT TTA TAC TAC CCC CCC GGC GAT GTG TAC AAT AAG AGA CAC  405
R   R   T   T   G   A   T   L   Y   Y   P   P   G   D   V   Y   N   K   R   H    75
TCC GAT GGG ATA TTC ACA AGT GAT TAT CGG AGG TAC CTC GGG CAA CTG AGT GCT CAA AAA  465
S   D   G   I   F   T   S   D   Y   R   R   Y   L   G   Q   L   S   A   Q   K    95
TTC CTG CAA TGG CTT ATG AAG CGC CAC TCG GAT GGG ATC TTC ACG GAC AGC TAT AGC CGC  525
F   L   Q   W   L   M   K   R   H   S   D   G   I   F   T   D   S   Y   S   R   115
TAC CGG AAT CAA ATG GCT GTT AAG AAA TAC ATA AAC GCA CTG CTT GGG AAA AGG TAT AAA  585
Y   R   N   Q   M   A   V   K   K   Y   I   N   A   L   L   G   K   R   Y   K   135
CAA AGG TAA agaaacaatggacgcaattttaccaaaggcttacatataaaattttaacgcaccaccgactaggctat 662
Q   R   *                                                                       138
ctgttaggcgggagagataaagtatttacctcgacatttaattcatctatctacatgtaacgagagaacacttcatgagc 742
gcgttataatcgttacgtggaaatacattataaacaattacggcgcttataaacaggggcttgccggagcccccgtatta 822
ttatttataaaatttcggatataaaagtgtttacgttaaaaacagttatatatttaaaaa                    883
```

The tunicate *glucagon/vip* mRNA isolated was 883bp encoding a signal peptide, cryptic peptide, glucagon-like$_{1-27}$ peptide, and VIP-like$_{1-27}$ peptide. Encoded within the tunicate *glucagon/vip* cDNA sequence, but not found in the tunicate *grf/pacap* cDNA are 159 nucleotides that encode a 53-amino-acid cryptic peptide, if the first ATG start site is assumed to be the correct one. However, if the third ATG start site is correct then a cryptic peptide does not exist, but the 5'UTR is longer. The tunicate glucagon-like peptide has amino acid identity (30-59%) to all superfamily members in humans (Table 5.1).

The two cDNAs are very similar in sequence. The nucleotides in the coding regions for the tunicate PACAP$_{1-27}$ and VIP$_{1-27}$-like proteins are 93% identical, that is only 6 bases are different. Identity of the tunicate PACAP amino acids to other members of the glucagon superfamily ranges from 96% with human PACAP to 22% for human GIP (Table 5.1).

Isolation of the tunicate *grf/pacap* and *glucagon/vip* genes

The partial tunicate *grf/pacap* gene isolated was 1590bp long and encoded three exons, as deduced from the cDNA (Figure 5.2). Located on the first exon is the 5'UTR and the signal peptide. The second exon encodes GRF-like peptide and the third exon encodes PACAP$_{1-27}$. No exon encoding a cryptic peptide was found between the signal and the bioactive peptides in the 14 cDNA clones isolated. This is a similar intron/exon

Figure 5.2: Nucleotide sequence of the partial (1590bp) tunicate *grf/pacap* gene. PCR using primers directed to the 5' and 3' untranslated regions were used to amplify the gene. The exons are in bold print and have the encoded amino acid listed below. The region of intron 1 that has high sequence identity to exon 2 of the *glucagon/vip* cDNA clone is underlined.

```
TACACTGGATCTCTTGGCCGCC ATG GCC AAG TCA AGT GGG GCG ACC CTC GCC CTC CTA ATT TAC   64
                       M   A   K   S   S   G   A   T   L   A   L   L   I   Y     14

GGA ATC ATT ATG AGA TAC TC gtggcaaaaacgctttatatatttcgcgcgatcgtacgtacgtacatcgcgta 137
G   I   I   M   R   Y   S                                                       21

cccggggcatcgatcgtttcaactacgtttactactattctttatttcgctattttcgggctatttcggctttgcgattt 217
atattatttattatttactattcggtttccgggtatatactttctattcggttggcccccttttacatctcttactagctt 297
atttcgatcgacgtttacgatacttgacaacgctatttatgacggtaacgattcttatgctaattaccacaaaaacttgg 377
cttttgacggacatcaaaatcgaataacaattccccatatgaggcggacgacgggggacactttatactacccccccggc 457
gttagtcttgggtatatctatcacttacttccggtctctctctatcacaacacgggggcttattacacccggctttgcact 537
cgcgcgtatgcttcgatcggcgaccgtcactgattgattgatcgatcaatttggatcgacgtgtatttttttctccaactt 617
tttatataatataacggcacccgctgatcgcttatgcgatgcggcgatctataacccccccaaaattcgggcgcatcttta 697
cacacacgggaaacgatcgggcgcgcgcgggcgggacctatactaccctaccgaaaccaaatgcccttaccaacgactta 777
ccttattcgggatcgaactcttctctattccctcttctctctctcgcgcgcgatcgattataaaaaacgaatcgatgcta 857
cctttacgatccgatatttttttcgatcgatcgggggaaacgatcgatcggatattgggctattacgatcgtacgatgct 937
accatggcatattattatcacacgggtttctaactaccgcatcgcgatggttgtgggctatttggggtatcttaattata 1017

cgtatcgcgatgcatttttttgccag C CAC TCC GAT GGG ATA TTC ACA AAA GAT TAT CGG AAG TAC  1082
                           H   S   D   G   I   F   T   K   D   Y   R   K   Y     34

CTC GGG CAA CTG CGA GCT CAA AAA TTC CTG CAA TGG CTT AT gtaatttaccatttcgcggcgcgat 1148
L   G   Q   L   R   A   Q   K   F   L   Q   W   L   M                           48

ttatatcggctatcgtacctatctagcgcgctacgattatatatataaaaacggggtttttcggcgcgcggctctatcgc 1228
tcttcttcgtgatatgcggacttcctggttaaattttattattcgggctgatgcgcgctattatattatattttcgatcg 1308
ggcgatttaaaacactttcgtttctccctatcgtacgtcgtattcgggagcatcgcttatcgtcggggcgcgcgcaaata 1388

tcgttgggctttaatggcttattttcttag G AAG CGC CAC TCG GAT GGG ATC TTC ACG GAC AGC TAT 1455
                               K   R   H   S   D   G   I   F   T   D   S   Y    60

AGC CGC TAC CGG AAT CAA ATG GCT GTT AAG AAA TAC CTG GCG GCA GTG CTT GGG AAA AGG   1515
S   R   Y   R   N   Q   M   A   V   K   K   Y   L   A   A   V   L   G   K   R    80

TAT AAA CAA AGG TAA AGAAACAATCGACGCAATTTTACCAATATAAAATTTTAACGCACCACCGACTAGGCTATC 1590
Y   K   Q   R   *                                                               85
```

TABLE 5.1: Percent identity of the four tunicate peptides in comparison to the human members of the glucagon superfamily.

	TUNICATE			
	PACAP	GRF-like	VIP	glucagon
HUMAN				
PACAP	96	48	85	48
VIP	67	37	67	37
PHM	70	33	74	37
GRF	33	59	37	59
glucagon	30	63	30	59
secretin	37	41	41	44
GIP	22	26	22	30
PRP	22	44	26	48

structure as found in the genes encoding mammalian glucagon, GIP, secretin and GRF in that a cryptic peptide exon is not 5' to the hormone-encoding region

However, within the first intron of the *grf/pacap* gene, nucleotides exist (position 295-457) that have a high identity (93%) with the encoded *glucagon/vip* cryptic segment (Figure 5.3). In addition, a partial tunicate *glucagon/vip* gene of 1435bp was isolated and as deduced from its cDNA clones contains four exons (Figure 5.3). The four exons contain nucleotides encoding a signal peptide, a cryptic peptide, glucagon-like peptide, VIP-like peptide and a 3'UTR. Unlike the *grf/pacap* mRNA, the tunicate *glucagon/vip* gene encodes a cryptic peptide. The nucleotides encoding the cryptic peptide have high sequence identity to nucleotides within intron 1 of the *grf/pacap* gene. These base pairs in the *glucagon/vip* cryptic peptide region are 93% identical to the nucleotides of intron 1 of the *grf/pacap* gene, but corresponding intron/exon splice sites are not found in the *grf/pacap* gene.

Three putative start sites for the glucagon/vip prepropeptide

There are 3 ATG codons that are potential start sites for the glucagon/VIP prepropeptide (Figure 5.3). If the first ATG is the correct start site, the propropeptide is as described above and a cryptic peptide is encoded. The second ATG site would mean that the 5'UTR is longer and the cryptic peptide is shorter than

the first prepropeptide. The third ATG site would result in the longest 5'UTR and loss of a cryptic peptide compared to the other possible preprohormones.

Tissue expression of tunicate *grf/pacap* and *glucagon/vip* mRNA by a PCR method

A sensitive PCR detection method for the presence of tunicate *grf/pacap* and *glucagon/vip* in various tissues was developed. For each cDNA clone, primers Tun 3 & 4 and Tun 5 & 6 (Figure 5.1A and B) were designed for the 5' and 3' untranslated regions. These regions were distinct and allowed the specific detection of either *grf/pacap* or *glucagon/vip* mRNA/cDNA. Reverse transcribed cDNA of various tunicate tissues was amplified with clone-specific primers and the results are shown in Figure 5.4A and 5.4B. Tunicate *grf/pacap* mRNA was detected specifically in the neural ganglion and the reaction containing the tunicate gonad, gonad/digestive gland, intestine, heart, branchial basket, negative control or the reaction containing the tunicate *glucagon/vip* clone. Tunicate *glucagon/vip* mRNA was detected (Figure 4b) in the neural ganglion, dorsal strand, the intestine, and the reaction containing the tunicate *glucagon/vip* cDNA (positive control). No bands were detected in the lanes containing the mRNA/cDNA from the neural gland, gonad, gonad/digestive gland, heart, branchial basket, negative control, or the lane containing the tunicate *grf/pacap* clone.

Figure 5.3: Nucleotide sequence of the partial (1105bp) tunicate *glucagon/vip* gene. The nucleotides for introns are shown in lower case and for exons in upper case. The translated amino acids are shown below their respective nucleotides. Both the amino acids and nucleotides are numbered on the right. The three possible start site ATG codons are boxed. The GRF-like peptide includes amino acids 75-101 and PACAP is amino acids 104-130. the cleavage sites (KR) between GRF and PACAP and at the end of PACAP are shown.

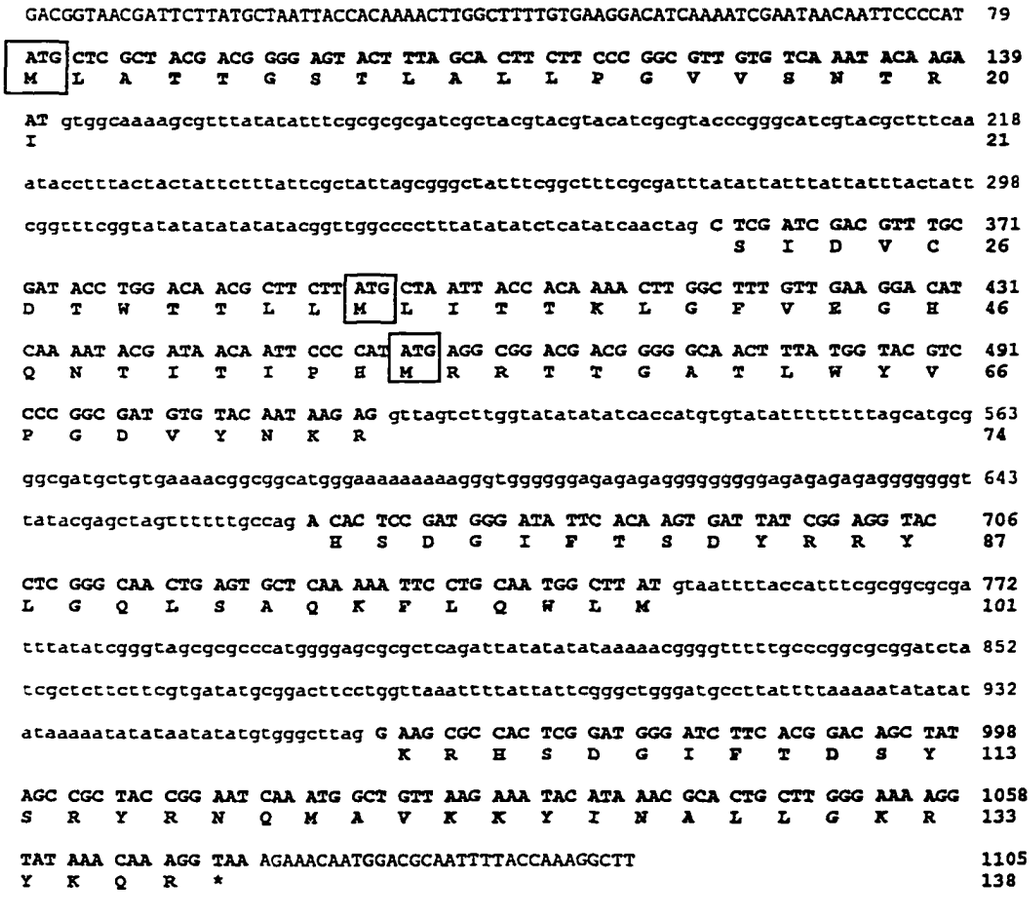
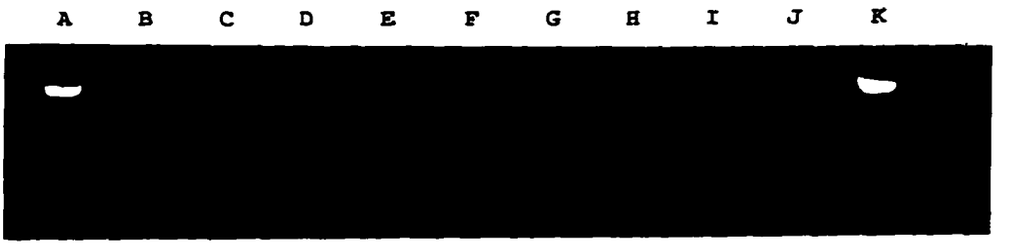
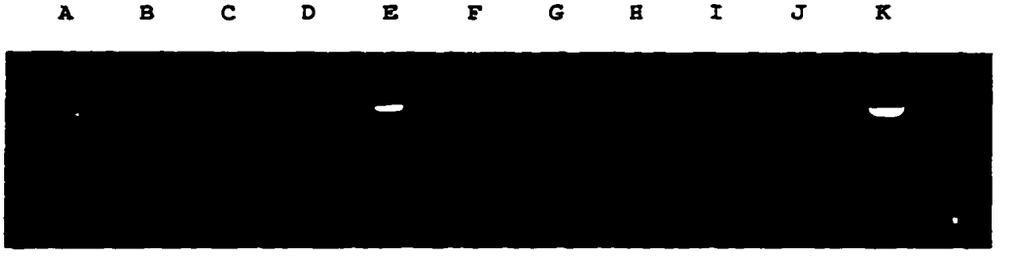

GACGGTAACGATTCTTATGCTAATTACCACAAAACTTGGCTTTTGTGAAGGACATCAAAATCGAATAACAATTCCCCAT 79
ATG CTC GCT ACG ACG GGG AGT ACT TTA GCA CTT CTT CCC GGC GTT GTG TCA AAT ACA AGA 139
M L A T T G S T L A L L P G V V S N T R 20
AT gtggcaaaagcgtttatatatttcgcgcgcgatcgctacgtacgtacatcgcgtacccgggcatcgtacgctttcaa 218
I 21
atacctttactactattctttattcgctattagcgggctatttcggctttcgcgatttatattatttattatttactatt 298
cggtttcggtatatatatatatacggttggccccttttatatatctcatatcaactag C TCG ATC GAC GTT TGC 371
S I D V C 26
GAT ACC TGG ACA ACG CTT CTT ATG CTA ATT ACC ACA AAA CTT GGC TTT GTT GAA GGA CAT 431
D T W T T L L M L I T T K L G F V E G H 46
CAA AAT ACG ATA ACA ATT CCC CAT ATG AGG CGG ACG ACG GGG GCA ACT TTA TGG TAC GTC 491
Q N T I T I P H M R R T T G A T L W Y V 66
CCC GGC GAT GTG TAC AAT AAG AG gttagtcttggtatatatatcaccatgtgtatatttttttagcatgcg 563
P G D V Y N K R 74
ggcgatgctgtgaaaacggcggcatgggaaaaaaaaagggtgggggggagagagagggggggggagagagagggggggt 643
tatacgagctagttttttgccag A CAC TCC GAT GGG ATA TTC ACA AGT GAT TAT CGG AGG TAC 706
H S D G I F T S D Y R R Y 87
CTC GGG CAA CTG AGT GCT CAA AAA TTC CTG CAA TGG CTT AT gtaattttaccatttcgcggcgcga 772
L G Q L S A Q K F L Q W L M 101
tttatatcgggtagcgcgcccatggggagcgcgctcagattatatatataaaaacggggtttttgcccggcgcggatcta 852
tcgctcttcttcgtgatatgcggacttcctggttaaattttattattcgggctgggatgccttatttttaaaaatatatat 932
ataaaaatatataatatatgtgggcttag G AAG CGC CAC TCG GAT GGG ATC TTC ACG GAC AGC TAT 998
K R H S D G I F T D S Y 113
AGC CGC TAC CGG AAT CAA ATG GCT GTT AAG AAA TAC ATA AAC GCA CTG CTT GGG AAA AGG 1058
S R Y R N Q M A V K K Y I N A L L G K R 133
TAT AAA CAA AGG TAA AGAAACAATGGACGCAATTTTACCAAAGGCTT 1105
Y K Q R * 138

Figure 5.4: Tissue expression of tunicate *grf/pacap* and *glucagon/vip* mRNA. **(A)** Tunicate *grf/pacap* mRNA detected by a reverse transcriptase/PCR assay. PCR reactions contained tissue cDNA as follows: neural ganglion (lane A), neural gland (lane B), gonad (lane C), gonad/digestive gland (lane D), intestine (lane E), heart (lane F), tunic (lane G), branchial basket (lane H), tunicate ILP cDNA clone (lane I), negative control (lane J), and *grf/pacap* cDNA clone (positive control; lane K). **(B)** Tunicate *vip/glucagon*-like mRNA detected by a reverse transcriptase/ PCR assay. PCR reactions contained tissue cDNA as follows: neural ganglion (lane A), neural gland (lane B), gonad (lane C), gonad/digestive gland (lane D), intestine (lane E), heart (lane F), tunic (lane G), branchial basket (lane H), tunicate insulin cDNA clone (lane I), negative control (lane J), and *glucagon/vip* cDNA clone (positive control; lane K).

A)



B)



grf/pacap cDNA (positive control). No bands were detected in the lanes containing the mRNA/cDNA from the neural gland,

Localization of tunicate *grf/pacap* and *glucagon/vip* mRNA with *in situ* techniques

To investigate the presence of *grf/pacap* and *glucagon/vip* mRNA expression within the tunicate neural ganglion, we used *grf/pacap* and *glucagon/vip* RNA probes. Both mRNAs were localized in cortical cells of the neural ganglion (Figures 5.6 and 5.7). The *in situ* hybridization evidence confirms the presence of both tunicate mRNAs in the cells of the neural ganglion. The *in situ* hybridization sections were compared to sections stained with hematoxylin and eosin, which stains all cells in the neural ganglion (Figure 5.8).

Figure 5.5: Sections (11μM) of tunicate (*Cheylosoma productum*) neural gland and ganglion stained with a hemotoxylin and eosin stain. The neural ganglion can be seen in the bottom half of the section. The ring of large cells around the periphery of the ganglion are overlaid with the blood sinus which contains a few blood cells. The neural gland (faint rust color) is shown by the strip of tissue at the top of the section (magnification is X 40).

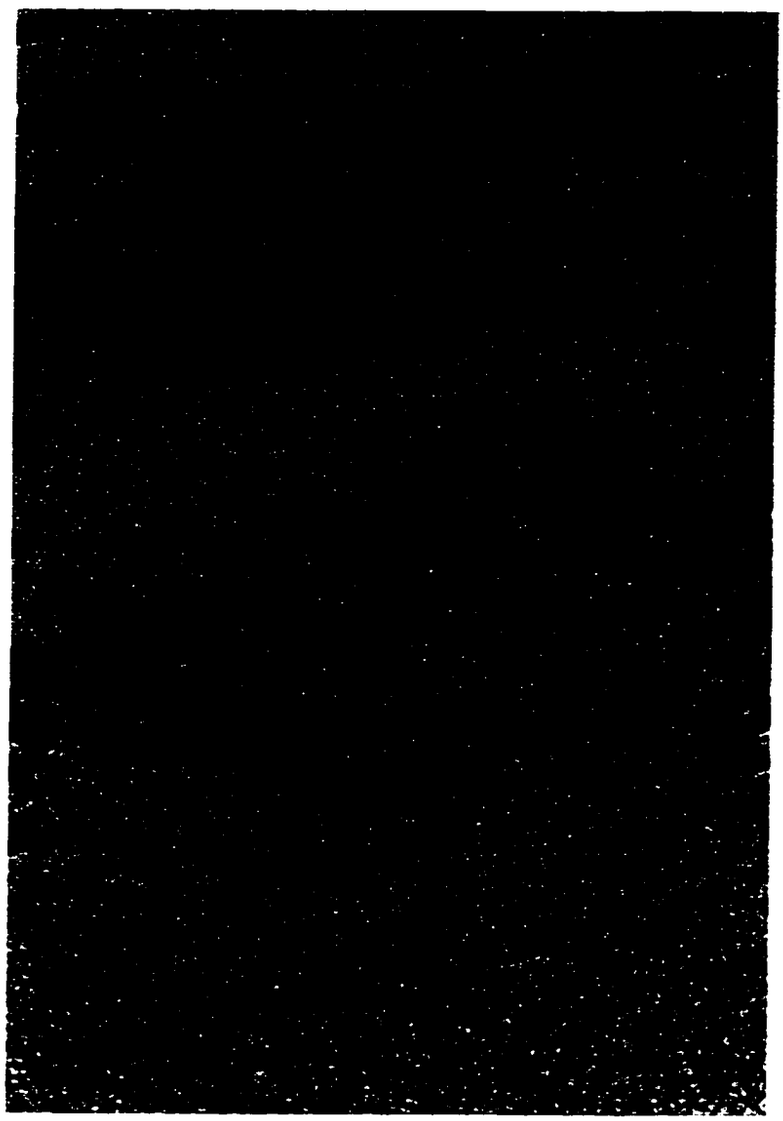
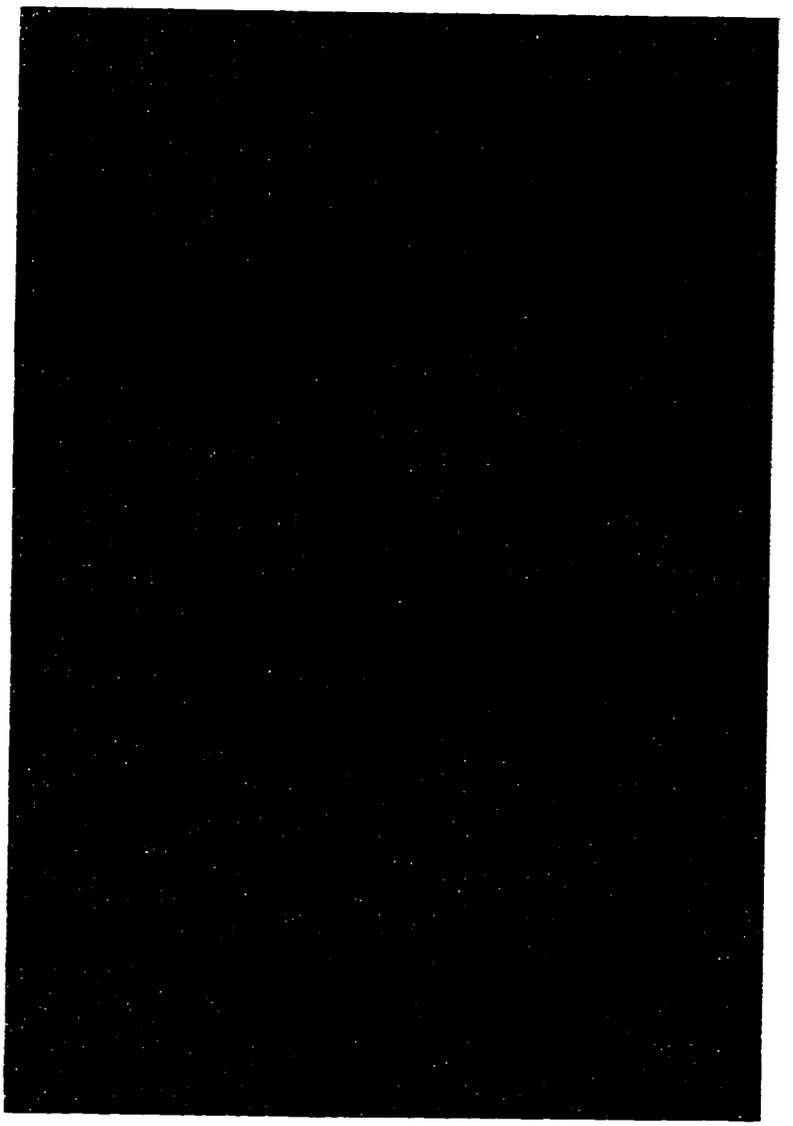


Figure 5.6: Localization of tunicate *grf/pacap* mRNA by *in situ* hybridization. **(A)** *grf/pacap* anti-sense mRNA (magnification: x40) and **(B)** *grf/pacap* sense negative control mRNA (magnification: x16) are shown in sections of the neural ganglion of *Cheylosoma productum* by *in situ* hybridization using a DIG-labelled RNA probe.

A



B

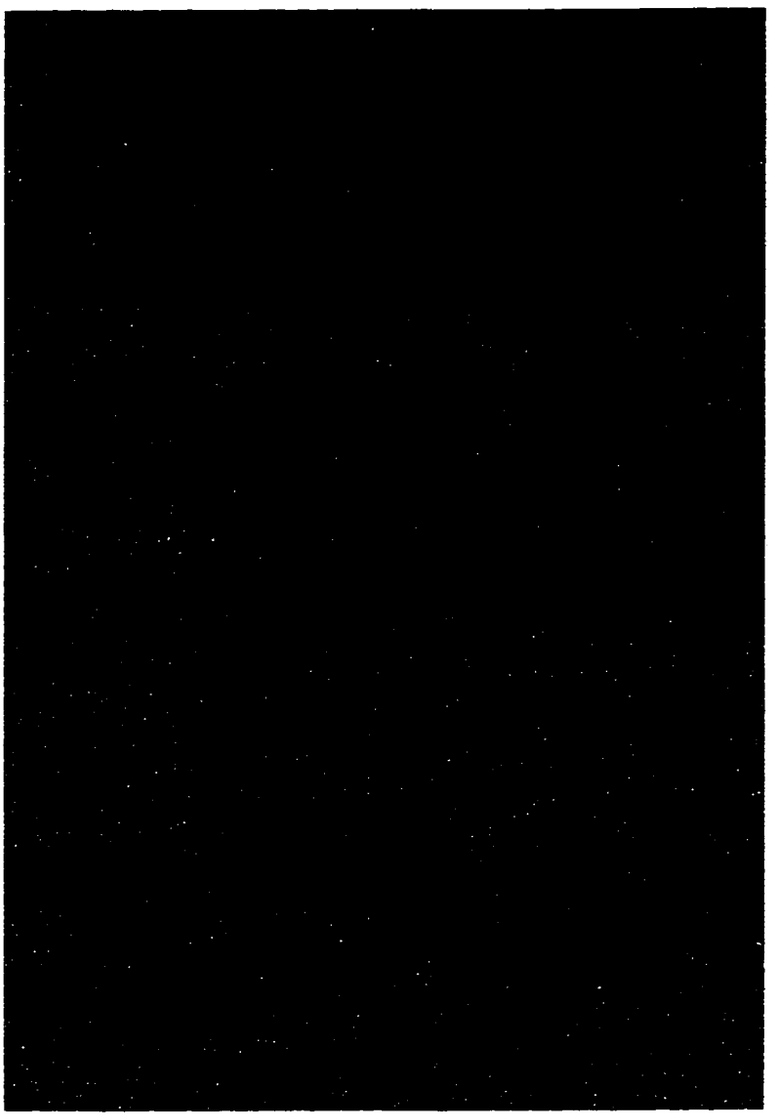
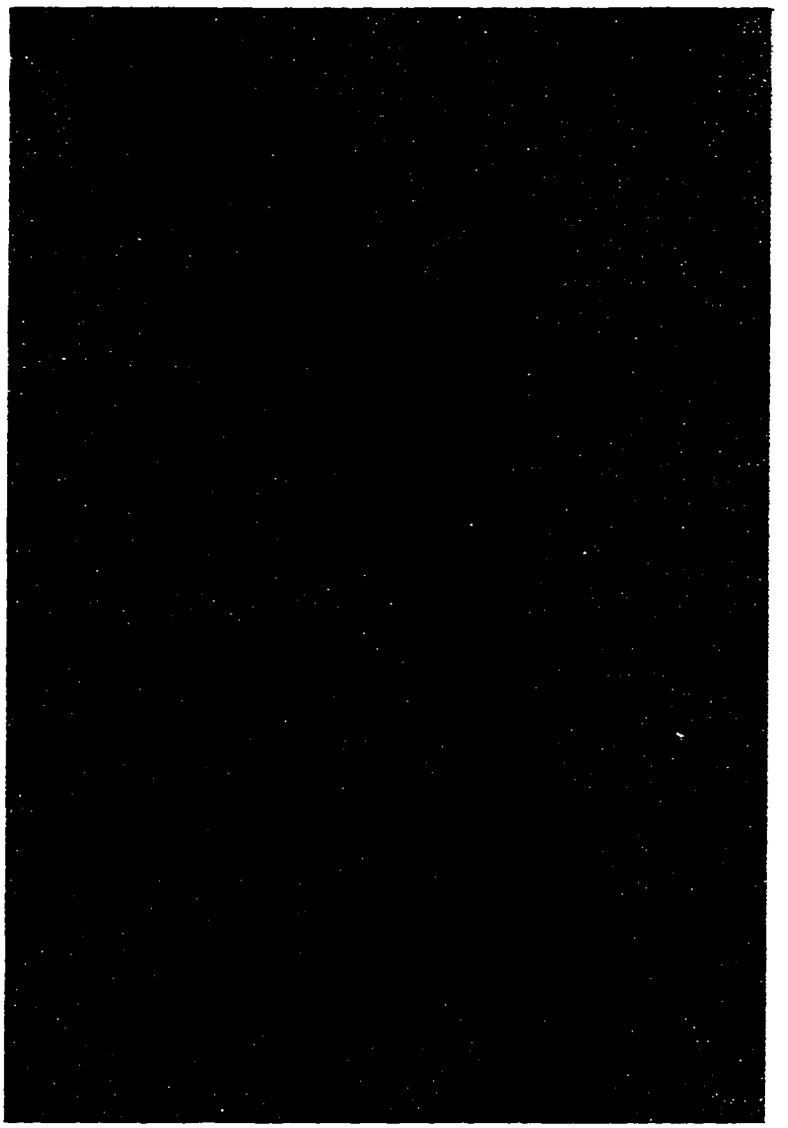
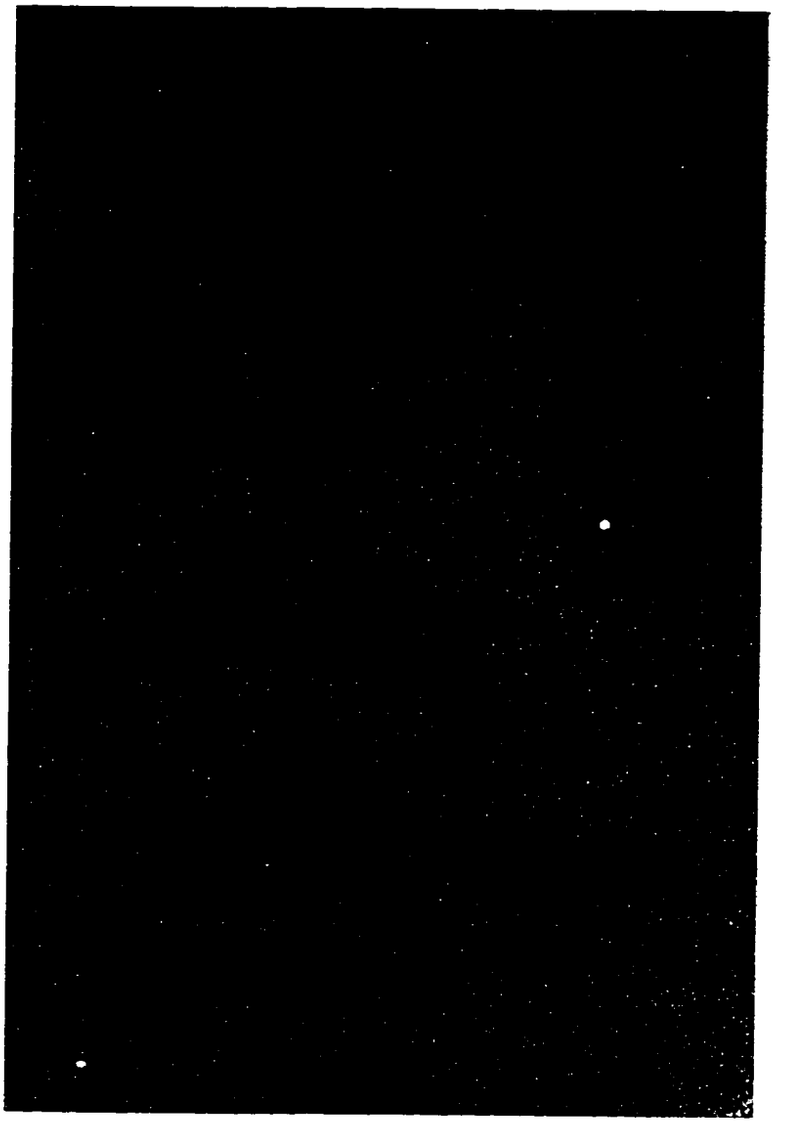


Figure 5.7: Localization of tunicate *glucagon/vip* mRNA by *in situ* hybridization. **(A)** *glucagon/vip* anti-sense mRNA (mag. X40) and **(B)** *glucagon/vip* sense (negative control) mRNA (mag. X16) are shown in sections of the neural ganglion of *Cheylosoma productum* by *in situ* hybridization using a DIG-labelled RNA probe (magnification: x16).

A



B



Conservation of the *grf/pacap* gene

A zoo blot containing genomic DNA from rat, starling, chicken, alligator, salmon, catfish, reedfish, tunicate, urchin, *Drosophila,* yeast, and *E. coli* was probed with the PCR-generated 163bp tunicate PACAP cDNA (Figure 5.8). This probe has high sequence identity to all known *pacap* cDNAs and detected two bands in all lanes except the *Drosophila,* yeast, and *E. coli.* The two bands produced in the zoo-blot may be due to the conservation of the *pacap* and *vip* nucleotides and therefore the probe may have hybridized to the *grf/pacap* and *glucagon/vip* genes. In Figure 5.9, tunicate DNA was digested to completion with EcoRI and probed with tunicate *grf/pacap* cDNA. Lane A was probed with the 163bp *pacap* PCR cDNA fragment and due to a high identity between the two clones is thought to have hybridized to both the *glucagon/vip* and *grf/pacap* genomic sequences. Lane B was probed with a tunicate *grf/pacap*-specific probe (bases 387-531) that hybridized only to *grf/pacap* genomic sequences. Specific hybridization to the tunicate *glucagon/vip* gene was obtained with a *glucagon/vip* specific probe (bases 713-907). In lane D a clone encoding tunicate *glucagon/vip* cDNA was used, and therefore hybridization to both tunicate *grf/pacap* and *glucagon/vip* occurred. Therefore, the two bands can be identified as either *vip* (7.5 Kb band) or *pacap* (5Kb band) when digested with Eco RI.

Figure 5.8: Zoo blot of DNA from various organisms probed with the 163bp *pacap* PCR fragment. Zoo blot using rat (lane A), alligator (lane B), starling (lane C), chicken (lane D), salmon (lane E), catfish (lane F), tunicate (lane G), reedfish (lane H), tunicate (lane I), urchin (lane J), *Drosophila* (lane K), yeast (lane L), and *E. coli* (lane M) DNA which was hybridized with the 163bp *pacap* PCR fragment. The two band may be due to the hybridization of the probe to both *pacap* and *vip* .

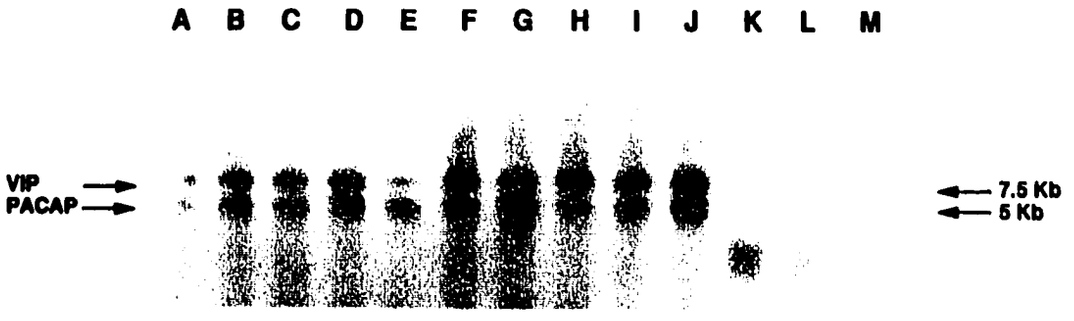
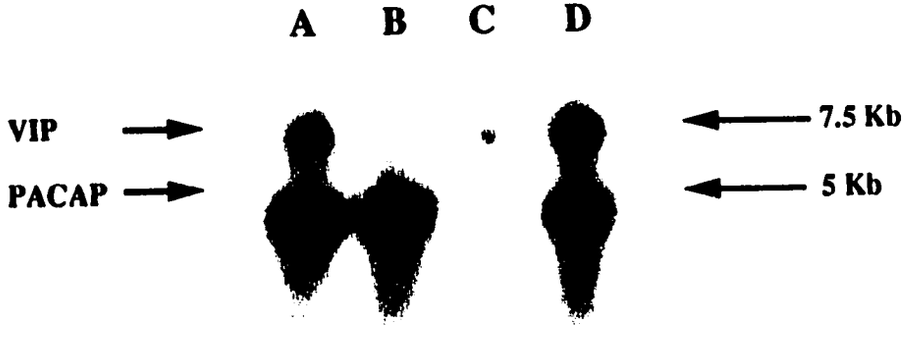

A B C D E F G H I J K L M
VIP
PACAP
7.5 Kb
5 Kb

Figure 5.9: Southern blot analysis of tunicate DNA. The probes were the 163bp *pacap* cDNA PCR probe (lane A), *grf/pacap*-specific probe (lane B), *vip*-specific probe (lane C) and a *glucagon/vip* probe (lane D). The tunicate digested with Eco RI results in two bands that can be identified as either *vip* (7.5 Kb band) or *pacap*, the 5Kb band. No EcoRI sites are present in the exons and introns and therefore no cuts would happen within the areas probed.


A
B
C
D
VIP
PACAP
7.5 Kb
5 Kb

Discussion

Tunicates contain two mRNAs encoding GRF/PACAP and glucagon/VIP-like peptides

Two tunicate genes and their respective cDNAs appear to be members of the glucagon superfamily. The partial tunicate *glucagon/vip* gene was 1435bp and, as deduced from the cDNA, two of the exons encode proteins with identity to mammalian glucagon and VIP. The other cDNA encodes tunicate GRF_{1-27}-like peptide and $PACAP_{1-27}$. This cDNA is transcribed from a gene that contains only three exons. This is in contrast to mammalian members of the superfamily, which have at least four exons, such as human *secretin* gene, and up to 7 exons as found in the human *vip* gene. The first tunicate exon contains the nucleotides that encode the 5'UTR and signal peptide. The tunicate *grf/pacap* cDNA does not have an exon which encodes a cryptic or N-terminal peptide, a feature also found in other superfamily genes, except the mammalian *pacap* and *vip* genes. However, the nucleotides that encode a tunicate PACAP cryptic peptide are present within intron 1 (Figure 5.2), but appear to lack splice sites on either side for processing as a separate exon. These nucleotides are not transcribed and therefore the tunicate PACAP precursor does not have a cryptic peptide.

With the exception of the human *grf* gene where a recent report shows the cryptic peptide has a role in stimulating sertoli cell activity (Breyer *et al.* 1996), the function of the

cryptic peptides is not known for any member of the superfamily. However, one possible explanation of the origin of the cryptic peptide in tunicates is that the nucleotide coding is present in the intron and requires only a few base changes to create an intron/exon splice site on each side resulting in a new exon. This appears to have occurred after the divergence of the *glucagon/vip* and *grf/pacap* genes. Alternatively, the exon encoding the cryptic peptide may have been present before the protochordates evolved, but within the tunicate *grf/pacap* gene, splice-site nucleotides were altered. However, within both tunicate genes isolated, consensus intron/exon splice sites are present at all intron/exon boundaries for proper intron removal.

This lack of an exon within the *grf/pacap* gene is interesting because it demonstrates the evolution of an exon in relation to an intronic sequence. The ancestral gene is assumed to have contained a nucleotide sequence that is either being incorporated or lost as an exon. Depending on the start site, the exon encodes a cryptic peptide or 5'UTR. It was speculated by Campbell and Scanes (1992) that the ancestral gene giving rise to the present day family contained only three exons. One possibility is that the tunicate *pacap* gene is derived from such an ancestral gene because it contains only three exons.

Conservation of PACAP may extend to taxons predating tunicates

A feature of the many *pacap* genes is the conservation of the nucleotides encoding PACAP and the amino acids specifying the protein. The nucleotides that encode tunicate $PACAP_{1-27}$ are very conserved, 90% (73/81), in comparison to human $PACAP_{1-27}$ (Figure 5.10). In addition, the deduced amino acids for tunicate $PACAP_{1-27}$ are 96% identical to the human, sheep, rat, frog, catfish and salmon $PACAP_{1-27}$ forms. Chicken $PACAP_{1-27}$ has two amino acid substitutions in cPACAP1-27 compared to tunicate $PACAP_{1-27}$; the additional substitution is an alanine at position 2 (Figure 5.11). Also, evidence of *pacap*'s nucleotide conservation is shown by hybridization of the tunicate *pacap* probe within the zoo-blot. The conservation of the chordate *pacap* gene sequence is extended to the nucleotides of the 5' and 3' UTRs as shown by the conservation of the restriction enzyme sites (Eco RI) used for the zoo-blot creating a migration of all PACAP genes at approximately the same position. Within these family members, conservation of the untranslated regions, both 3'UTR (Parker et al. 1993) and 5'UTR (Sena et al. 1994), has been reported. However, the extent of PACAP's sequence conservation between species that are separated by 650 million years of evolution was unexpected. Using the tunicate *pacap* PCR fragment as a probe, I was able to detect a *pacap* gene in rat, starling, chicken, alligator, salmon, catfish, tunicate, reedfish and sea urchin.

Figure 5.10: A comparison of the nucleotides that encode tunicate and human $PACAP_{1-27}$. Identical nucleotides are shown by a connecting line (|).

NUCLEOTIDES ENCODING HUMAN PACAP1-27

K

```
cactcggacgggatcttcacggacagctacagccgctaccggaaacaaatggctgtcaagaaatacttggcggccgtccta
|||||||| |||||||||||||||||||| |||||||||||||| |||||||||| ||||||||| ||||||| || ||
cactcggatgggatcttcacggacagctatagccgctaccggaatcaaatggctgttaagaaatacctggcggcagtgctt
```

N

NUCLEOTIDES ENCODING TUNICATE PACAP1-27

No other known hormone has such sequence conservation separated by at least 600 million years of evolution. This conservation is even more intriguing because the basic function of PACAP is still speculative.

The other band that hybridized to the probe in the lanes containing the vertebrate DNA was probably the *vip* gene since the tunicate *pacap* probe has a high degree of identity to vertebrate VIP. However, since the lane containing the sea urchin DNA (lane I) also had two bands it would be expected that sea urchins would have a *grf/pacap* -like and a *glucagon/vip*-like gene similar to the tunicate. Echinoderms branched from a stem line that led to vertebrate evolution about 100 million years before tunicates. It is possible that a sea urchin *grf/pacap* and *glucagon/vip* genes have a higher degree of sequence identity than found in tunicates provided there has not been a high number of substitutions in these genes between the present day and ancestral sea urchins. One can speculate that organisms evolving before the sea urchins do not contain two genes, but rather a single parent *pacap* gene. Conclusions about evolution in regard to gene organization, number of exons, length and conservation of the bioactive core of the peptide are clear. More difficult are deductions about the amino acids or nucleotides that might have been present within the ancestral genes.

Figure 5.11: Comparison of PACAP proteins from different species. $PACAP_{1-27}$ amino acid conservation as shown by protein sequences from tunicate, human, mouse, chicken, frog, catfish and salmon. Amino acids that are different are indicated.

```
TUNICATE   H S D G I F T D S Y S R Y R N Q M A V K K Y L A A V L
HUMAN                                  K
MOUSE                                  K
CHICKEN      A                         K
FROG                                   K
CATFISH                                K
SALMON                                 K
```

Exon and gene duplication produced two tunicate cDNAs

A feature of the tunicate *grf/pacap* and *glucagon/vip* genes is the high sequence identity found between the two genes. The strong identity suggests that the two genes arose from a relatively recent gene duplication. The importance of a gene duplication in evolution was first noted by Haldane (1932) and Muller (1935), coincidentally even before the structure of DNA was known. In 1970, Ohno stated that the most effective means by which a new gene can evolve is by a gene duplication. It is apparent that the two tunicate genes arose from a complete gene duplication. However, the mechanism of the duplication could be due to a single gene duplication or to aneuploidy or polyploidy of the genome. The tunicate *glucagon/vip* mRNA, encodes amino acids that have strong sequence identity to the tunicate PACAP peptide (89%) and, in addition, to the human PACAP (85%) peptide. This similarity is indicative of a gene duplication of the parent *pacap* gene resulting in two tunicate genes. One can argue that the duplication allowed one of the two genes, presumably *glucagon/vip* to evolve into a distinct gene encoding related, but different peptides.

The identity (89%) of the exons between tunicate PACAP and VIP and identity between tunicate GRF and glucagon support the idea of a relatively recent gene duplication (Figure 5.12). However, the identity shared between exons on the same gene suggests exon duplication occurred prior to the gene duplication. Within the *glucagon/vip* gene, the identity of exons 3 and 4

suggest an exon duplication. The nucleotides encoding exon 3 and exon 4 are similar in length, the nucleotides are 57% identical, and the encoded amino acids have 48% identity. This data from the *glucagon/vip* gene is the best proof to date to confirm the speculation that the origin of the two exons is by an exon duplication. In addition to the 57% nucleotide identity between exons 3 and 4 within the tunicate *glucagon/vip* gene, exon 4 of tunicate VIP has 89% identity with exon 3 of the nucleotides of the tunicate *pacap* gene. The 89% nucleotide identity between exons of two different genes supports the idea that these two genes are the result of a gene duplication followed by nucleotide substitutions resulting into two distinct sequences.

Tunicate *grf/pacap* and *glucagon/vip* genes are related to other family members

The two tunicate genes appear to be the result of an exon and gene duplication. The origin of the superfamily may be closely related to ancestral forms of these two tunicate genes. It is apparent from the high conservation that the gene encoding $PACAP_{1-27}$ is the most likely ancestral molecule to have undergone an exon and a gene duplication resulting in the tunicate *glucagon/vip* gene. The tunicate *grf/pacap* gene sequence also has a high degree of identity to other known glucagon sequences. The encoded amino acids of the glucagon-

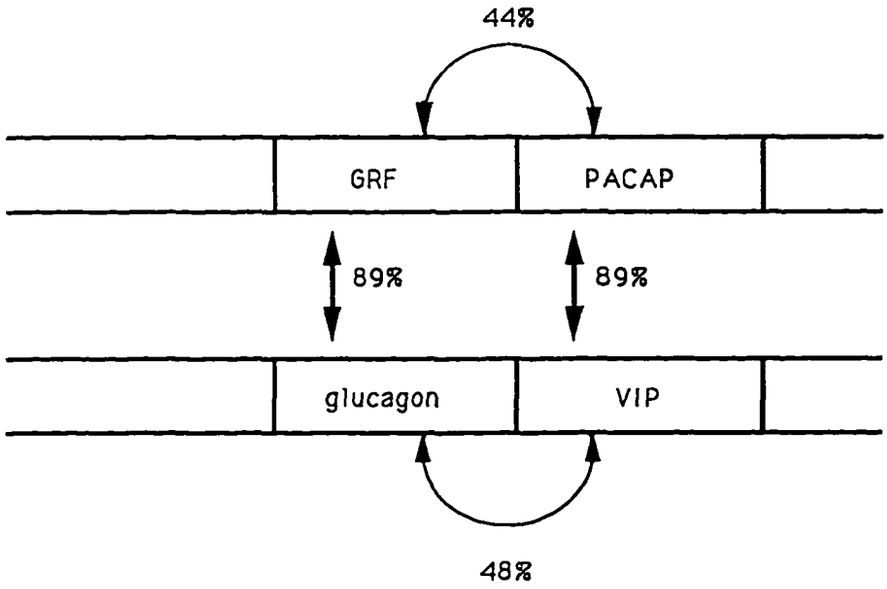
Figure 5.12: Comparison of the amino acid sequences for the four tunicate (t) glucagon superfamily peptides. The amino acid sequence of each peptide is shown on top and the percent identity among the four deduced peptides is shown in the schematic diagram.

```
tPACAP      H S D G I F T D S Y S R Y R N Q M A V K K Y L A A V L
tVIP        H S D G I F T D S Y S R Y R N Q M A V K K Y I N A L L

tGRF        H S D G I F T K D Y R K Y L G Q L R A Q K F L Q W L M
tglucagon   H S D G I F T S D Y R R Y L G Q L S A Q K F L Q W L M

tVIP        H S D G I F T D S Y S R Y R N Q M A V K K Y I N A L L
tglucagon   H S D G I F T S D Y R R Y L G Q L S A Q K F L Q W L M

tGRF        H S D G I F T K D Y R K Y L G Q L R A Q K F L Q W L M
tPACAP      H S D G I F T D S Y S R Y R N Q M A V K K Y L A A V L
```



like segment of the tunicate *glucagon/vip* gene has 59% identity to human glucagon, and in addition, the glucagon-like sequence is located between the signal and C-terminal peptides. Therefore, glucagon may have arisen from a gene duplication of the *glucagon/vip* gene in which one product of the duplication underwent nucleotide changes to encode a distinct glucagon gene. These data suggest the ancestral gene that gave rise to the tunicate glucagon-like exon also gave rise to vertebrate glucagon.

Therefore, the tunicate has two genes, the ancestors of which may be the progenitors for mammalian PACAP, GRF, glucagon and VIP. Based on sequence data available for the two remaining peptides, (secretin and GIP), they may have evolved as separate genes recently from gene duplications of the *vip* and *pacap* genes for they are only detected within mammals. Also, based on sequence data from chickens and fish, distinct *grf* and *pacap* genes probably originated from a *pacap* gene duplication. This is because the GRF is encoded on the same gene as PACAP, at least in chicken and fish. Therefore, these two genes within the tunicates appear to be the progenitors from which the present day vertebrate glucagon superfamily originated.

References

Berril NJ 1955 The Origin of Vertebrates, Oxford University Press.

Birnboim M 1983 A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymology 100: 243-255.

Breyer PR, Rothrock JK, Beaudry N and Pescovitz OH 1996 A novel peptide from the growth hormone releasing hormone gene stimulates sertoli cell activity. Endocrinology 137: 2159-2162.

Brown JC and Dryburgh JR 1971 A gastric inhibitory polypeptide. II. The complete amino acid sequence. Canadian Journal of Biochemistry 49: 867-872.

Bullock TH and Horridge GA 1965 Structure and Function in the Nervous Systems of Invertebrates. W.H. Freeman and Company, San Francisco. pp-1577-1592.

Carlquist M, Jörvall H, Forssmann W-G, Thulin L, Johansson C, and Mutt V 1985 Human secretin is not identical to the porcine/bovine hormone. IRCS Medical Science 13: 217-218.

Campbell RM and Scanes CG 1992 Evolution of the growth hormone-releasing factor (GRF) family of peptides. Growth Regulation 2: 175-191.

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

DiCicco-Bloom EM 1992 Pituitary adenylate cyclase activating polypeptide (PACAP) potently stimulates mitosis,

neurogenesis and survival in cultured rat sympathetic neuroblasts. Regulatory Peptides 37: 319

Doolittle RF, Feng D-F, Tsang S., Cho G, Little E 1996 Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271: 470-477.

Ebberink RHM, Smit AB and Van Minnen J 1989 The insulin family: evolution of structure and function in vertebrates and invertebrates. Biological Bulletin 177: 176-182.

Fritsch HAR, Van Noorden S, and Pearse AGE 1982 Gastro-intestinal and neurohormonal peptides in the alimentary tract and cerebral complex of *Ciona intestinalis* (Ascidiaceae). Cell and Tissue Research 223: 369-402.

Galloway SM and Cutfield JF 1988 Insulin-like material from the digestive tract of the tunicate *Pyura pachydermatina* (sea tulip). General and Compartive Endocrinology 69: 106-113.

Garstang W 1928 Q. J. Microsc. Sci. 72: 51-187.

Georges D and Dubois MP 1980 Mise en evidence par les techniques d'immunofluorescence d'un antigene de type LH-RH dans le systeme nerveux de *Ciona intestinalis* (Tunicier ascidiace) Comptes rendus de l'Academie des Science, Paris 290: 29-31

Gorbman A 1995 Olfactory origins and evolution of the brain-pituitary endocrine system: fact and speculation. General and Comparative Endocrinology. 97: 171-178.

Guillemin R, Brazeau P, Bohlen P, Esch F, Ling N, and Wehrenberg W (1982). Growth hormone-releasing factor from a human pancreatic tumor that caused acromegaly. Science 218: 585-587.

Haldane JBS 1932 The Causes of Evolution. Longmans and Green, London.

Itoh N, Obata K, Yanaihara N, and Okamoto H 1983 Human preprovasoactive interstinal polypeptide contains a novel PHI-27-like peptide, PHM-27. Nature 304: 547-549.

Kelsall R, Coe IR and Sherwood NM 1990 Phylogeny and ontogeny of gonadotropin-releasing hormone: comparison of guinea pig, rat and a protochordate. General and Compartive Endocrinology. 78: 479-494.

Mackie GO 1995 On the "visceral nervous system" of *Ciona.* J. Mar. Biol. Assoc. U.K. 75: 141-151.

Miyata A, Arimura A, Dahl RR, Minamino N, Uehara A, Jian L, Culler MD, Coy DH 1989 Isolation of a novel 38 residue-hypothalamic polypeptide which stimulates adenylate cyclase in pituitary cells. Biochemical and Biophysical Research Communication 164: 567-574.

Muller HJ 1935 The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. Genetics 17: 237-252.

Ohno S 1970 Evolution by gene duplication. Springer-Verlag, Berlin, pp 156.

O'Neil, Falkmer GS, and Thorndyke, M.C. 1986. Insulin-like immuoreactivity in the neural ganglion of the ascidian *Ciona intestinalis*. Acta Zool. (Stockholm) 67, 147-153.

Parker DB, Coe IR, Dixon GH, Sherwood NM 1993 Two salmon neuropeptides encoded by one brain cDNA are structurally related to members of the glucagon superfamily. European Journal of Biochemistry 215: 439-448.

Pesce M, Canipari R, Ferri G-L, Siracusa G, and De Felici M 1996 Pituitary adenylate cyclase-activating polypeptide (PACAP) stimulates adenylate cyclase and promotes proliferation of mouse primordial germ cells. *Development* 122: 215-221.

Pestarino M 1990 Immunocytochemical localization of glucagon-related peptides in the cerebral ganglion or the protochordate ascidia, *Styela plicata*. Journal of Neuroendocrinology 2: 641-645.

Pestarino M 1992 The distribution of Neuropeptide Y-like immunoreactivity in the ascidian *Styela plicata*. Cellular and Molecular Biology 38: 255-262.

Pestarino M 1993 Localization of peptide histidine isoleucine-like immunoreactivity in the gut of the ascidian, Stylea plicata. Animal Biology 2: 91-95.

Pestarino M and Facchinetti F 1995 Immunocytochemical localization and biochemical characterization of melatropin-like peptides in the gonads of a protochordate. Peptides 16: 1269-1272.

Powell JFF, Reska-Skinner SM, Om Prakash M, Fischer WH, Park M, Rivier JE, Craig AG, Mackie GO and Sherwood NM 1996 Gonadotropin-releasing hormone in a tunicate and evidence regarding its ancestral mode of action. Proceedings of the National Academy of Science (USA) in press.

Rivier J, Spiess J, Thorner MO and Vale W 1982 Characterization of growth hormone-releasing factor from a human pancreatic islet tumour. Nature 300: 276-278.

Robberecht P, Woussen-Colle MC, Vertongen P, De Neef P, Hou X, Salmon I and Brotchi J 1994 Expression of pituitary adenylate cyclase-activating polypeptide (PACAP) receptors in human glial cell tumors. Peptides 15: 661-665.

Ruppert EE 1990 Structure, ultrastructure and function of the neural gland complex of Ascidia interrupta (Chordata, Ascidiacea): clarification of hypotheses regarding the evolution of the vertebrate pituitary. Acta Zoology 71: 135-149.

Sanger F, Nicklen S, Coulson AR 1977 DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science (USA) 74: 5463-5467.

Satoh N and Jeffery WR 1995 Chasing tails in ascidians: developmental insights into the origin and evolution of chordates. Trends in Genetics. 11: 354-359.

Sena M, Bravo DT, Agoston D, Waschek JA 1994 High conservation of upstream regulatory sequences on the human

and mouse vasoactive intestinal (VIP) genes. DNA sequence 5: 25-29.

Sherwood NM, Parker DB, McRory JE and Lescheid DW 1994 Molecular evolution of growth hormone-releasing hormone and gonadotropin-releasing hormone. In: Fish Physiology, N.M. Sherwood and C.L. Hew (eds). 13: 3-66 Academic Press, Orlando.

Thomsen J, Kristiasen K, Brunfeldt K and Sundby F 1972 The amino acid sequence of human glucagon. FEBS Letters 21: 315-319.

Thorndyke MC and Goldsworthy GJ 1988 Functional aspects of peptide neurohormones in protochordates. in: Neurohormones in Invertebrates. (eds.) Thorndyke M.C. and G.J. Goldsworthy. Cambridge University Press. pp 235-257.

CHAPTER 6

A brain-specific insulin-like growth factor-I (IGF-I)

A version of this chapter has been published: McRory JM, Sherwood NM 1994 Catfish express two forms of insulin-like growth factor-I (IGF-I) in the brain. Journal of Biological Chemistry. 269: 18588-18592

Summary

This chapter examines the evolution of the insulin superfamily. In this research I found that the insulin-like growth factor family had not only a "ubiquitous IGF-I", but also a "brain-specific IGF-I" with a distinct structure. At the beginning of the research it was known that IGF-I is expressed not only in the liver, but also in brain and other tissues. This ubiquitous IGF-I has a complex pattern of expression due to multiple transcription start sites, polyadenylation sites and exon skipping. In this chapter I describe the isolation of a cDNA encoding a brain-specific IGF-I from a catfish brain cDNA library. Also, a fragment encoding ubiquitous IGF-I was amplified from brain and liver mRNA and the deduced protein was shown to be distinct (66% sequence identity) from brain-specific IGF-I. Consistent with other IGF-I prepropeptides, the brain-specific IGF-I has a 43 residue signal peptide followed by B, C, A, D, and E domains. Retained in the catfish brain-specific IGF-I peptide are residues predicted to be involved with the correct tertiary folding, disulfide linkages, and receptor binding. Northern blot analysis of poly A+ rich mRNA from brain indicated a single 1600bp transcript; a band was not detected from mRNA of liver, stomach, pancreas, pituitary, blood, herring brain or brain poly A- RNA. A sensitive reverse transcriptase/ PCR assay also showed that brain-specific IGF-I mRNA was expressed solely in the Thai catfish brain and not in liver, stomach, pancreas, pituitary, ovary, or African catfish brain.

Introduction

Insulin-like growth factor I (IGF-I) is important for proper animal growth, tissue development, and differentiation (Daughaday and Rotwein 1989; Froesch *et al.* 1985). IGF-I belongs to a family which includes insulin, IGF-II, relaxin, molluscan insulin-related hormone, and insect prothoracicotropic hormone. The release of IGF-I from the liver is under the control of pituitary growth hormone (GH), as evidenced by increased numbers of *igf-I* transcripts in the liver and other tissues after administration of GH (Bichell *et al.* 1992; Foyt *et al.* 1992). It is IGF-I rather than GH that is thought to promote cell division and differentiation in extrahepatic tissues (Romagnolo *et al.* 1992).

The IGF-I cDNA sequence has been reported from a variety of animals including the human (Jansen *et al.* 1983), rat (Casella et al, 1987; Murphy *et al.* 1987), pig (Tavakkol *et al.* 1988), cow (Wong *et al.* 1989; Francis *et al.* 1988), chicken (Kajimoto and Rotwein 1989), frog (Kajimoto and Rotwein 1990), trout (Shamblott and Chen 1992), coho salmon (Cao *et al.* 1989; Duguay *et al.* 1992), chinook salmon (Wallis and Devlin 1993) and hagfish (Nagamatsu *et al.* 1991). Also, amphioxus has a cDNA sequence similar to both insulin and IGF-I, and thus may be ancestral to both peptides (Chan *et al.* 1990). The sequence identity of IGF-I among all species examined is very high; the identity between the human IGF-I protein sequence and that of salmon (Cao *et al.* 1989; Duguay *et al.* 1992), frog (Kajimoto and

Rotwein 1990), and chicken (Shamblott and Chen 1992) is 80%, 81% and 86%, respectively. Within mammals and lower vertebrates, IGF-I is produced predominately in the liver, but lower concentrations are found in muscle, spleen, fat, heart, testis, and kidney. To date, only hagfish has IGF-I expression that is limited to the liver (Nagamatsu *et al.* 1991).

In addition to the role of IGF-I in the linear growth of tissue, IGF-I is believed to be important in the growth and development of the central nervous system (CNS). It is not known if IGF-I acts on the brain in an autocrine and/or paracrine manner or crosses the blood-brain barrier to act as an endocrine agent. Roles such as regulating neuronal and glial function (Sara *et al.* 1986), acting as a neurotrophic factor (Schwartz *et al.* 1992), and modulating synaptic transmission (Schwartz *et al.* 1992) have been reported.

Both complete and truncated forms of IGF-I have been purified from human fetal brain tissue (Sara *et al.* 1986). The truncated form of human fetal brain IGF-I, which lacks the first three amino acids, also has been isolated from porcine uterus (Ogasawara *et al.* 1989) and bovine colostrum (Francis et al 1988); all have an amino acid sequence identical with the human liver form. The tripeptide Gly-Pro-Glu, which originates from the intact IGF-I peptide, is a product of post-translational processing and functions within the CNS (Sara *et al.* 1986). Compared to the complete molecule, the truncated IGF-I lacking the tripeptide is 5-10 times more potent in its ability to

stimulate DNA and protein synthesis and in competing for binding sites on the brain membrane (Sara *et al.* 1989). It has been suggested that the tripeptide may itself act as a neuroactive peptide because it has a potent stimulatory action on the release of acetylcholine (Sara *et al.* 1989).

In the present study a brain-specific form of IGF-I is identified for the first time from any species. The sequence of the brain-specific IGF-I is compared to a second distinct catfish IGF-I isolated from both the liver and brain.

Methods and Materials

Screening of the catfish cDNA library

A cDNA library was constructed from cDNA isolated from a single Thai catfish (*Clarias macrocephalus*) brain, Lambda Zap II, and Gigapack Gold packaging mix (Stratagene). A brain specific IGF-I clone was isolated using low density phage plating, high stringency washes (0.1X SSC, 0.1% SDS, 65°C), and a probe that encoded catfish growth hormone-releasing hormone and pituitary adenylate cyclase activating polypeptide (see Fig. 6.2, nucleotides 244-560).

Exonuclease III digestion of positive clones

IGF-I cDNA, in pBluescript KS- (5μg), was purified and hydrated in 50μl sterile-filtered water. Digestion with Sac1 and Xba1 (Pharmacia) were used to linearize the clone. The DNA was precipitated, washed with 75% ethanol and hydrated in 40μl of Exonuclease III buffer; Exonuclease III (300U, Boehringer Mannheim) was added. The DNA solution was heated to 30°C. Aliquots were removed after 30s and placed in tubes on ice containing 7.5μl of S1 nuclease mix and 60U S1 nuclease (Boehringer Mannheim). The tubes were left at room temperature for 45min.

Linear plasmids were circularized with 4U T4 DNA ligase (Stratagene) and electroporated into XL-1 competent cells. White colonies were selected as recombinant plasmids and

prepared for sequencing with an alkaline hydrolysis method (Birnboim 1983). The opposite strand of the clone was sequenced as described in the above procedure. However, Kpn I and Xho I (Pharmacia) were used to linearize the clone prior to digestion.

Sequencing of cDNA clones

Both strands of the plasmid were sequenced with Sequenase 2.0 and [α-^{35}S]ATP using the chain termination method (Sanger *et al.* 1977). All sequencing gels were 6% polyacrylamide/7M urea wedge gels that were dried under vacuum at 80°C and exposed to Kodak XAR-5 film for 12-24h.

mRNA size as determined with northern blot analysis

Total RNA was isolated from tissues of a single female catfish using the method of Chomczyski and Sacchi (1987). The exception was pituitary tissue, which was obtained from several females. Poly A+ rich mRNA was isolated using a Fast-Track mRNA isolation kit (Invitrogen). In separate lanes brain (1µg), liver (1µg), liver (10µg), stomach (1µg), pancreas (1µg), pituitary (1µg), ovary (1µg), blood (1µg), brain poly A- mRNA (10µg) and herring poly A+ mRNA (1µg) were electrophoresed in a gel of 1.2% agarose, 3% formaldehyde, and 1X MOPS (3[N-morpholino] propanesulfonic acid). The gel was blotted onto a nylon membrane (Amersham) and hybridized in a solution of 25mM KPO_4, 5XSSC, 10X Denhardt's solution, 50% formamide and

30mg/ml sonicated sea urchin DNA. The probe was an IGF-I [α-^{32}P]ATP randomly-primed cDNA fragment that corresponded to nucleotides 339-762. The filter was washed under high stringency conditions (0.1X SSC, 0.1% SDS, 65°C) and exposed to X-ray film (Kodak) with intensifying screens at -80°C for 7 days. The quality of mRNA was assessed for each tissue by probing with a randomly-primed salmon tubulin cDNA fragment as defined below. The membrane was washed under high stringency conditions (0.1X SSC, 0.1% SDS, 60°C) and exposed to X-ray film with intensifying screens for 1h.

Reverse transcriptase/PCR assay for brain-specific IGF-I expression

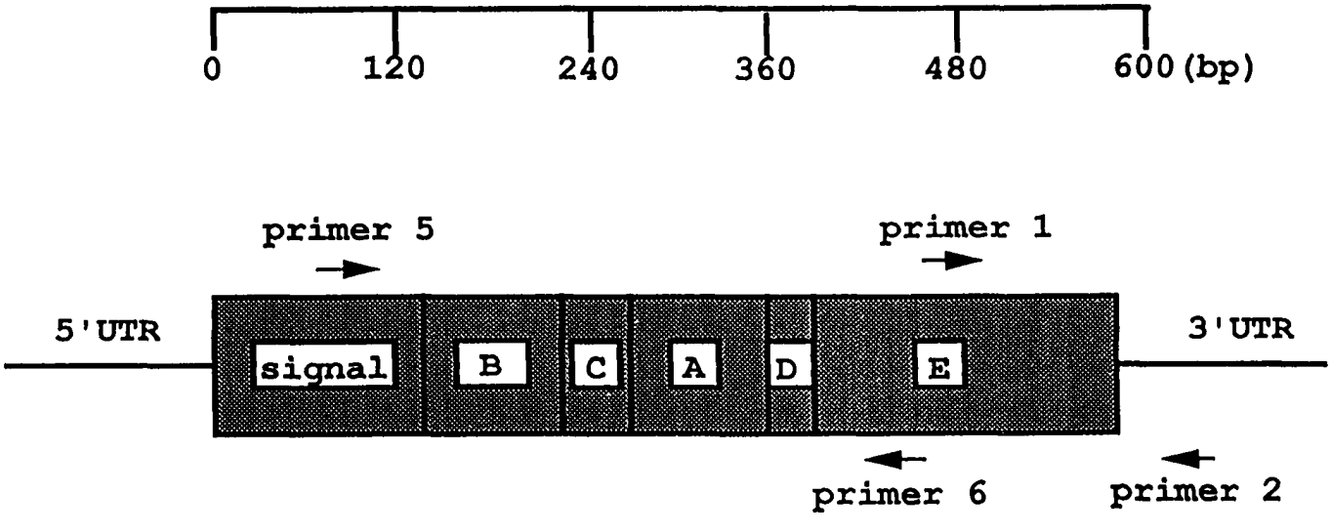
Complementary DNA was synthesized from 1µg of mRNA using 200U avian reverse transcriptase (H^- RT Superscript, BRL), 10mM dithiothreitol, 1mM each dNTP, 160U RNA guard, 2mM Oligo dT_{12-18}, and 1X H^- RT buffer for a total reaction volume of 20µl. The reaction proceeded for 90min at 41°C followed by 5min at 90°C. IGF-I and tubulin PCR amplifications were done separately with 0.5µl of newly transcribed cDNA from each tissue, 5U Taq DNA polymerase, 1x Taq buffer (Promega), 0.2mM each dNTP, 0.4mM of each primer (1-4), and 3mM MgCl in a 50µl reaction for 35 cycles (94°-55°-72°C). Primers 1 and 2 correspond to bases 612-636bp and 829-852bp of the catfish brain IGF-I clone, an area not conserved between brain-specific IGF-I and salmon IGF-I (Figure 6.1). Primers 3

and 4 correspond to bases 523-545 and 719-740 of the salmon tubulin cDNA clone (Coe *et al.* 1992).

RT/PCR assay for Liver (ubiquitous) IGF-I expression

Polymerase chain reactions included 1.0μg of brain and liver transcribed cDNA, 5U Taq polymerase, 1x Taq buffer (BRL), 0.2mM each dNTP, 0.4mM of each primer (5 and 6, Figure 6.1), and 1.5mM MgCl in a 50μl reaction for 35 cycles (94°-50°-72°C). A negative control containing water in place of cDNA was included. Primer 5 corresponded to base pairs 122-148 (sense strand) and primer 6 matched base pairs 439-459 on the antisense strand of salmon IGF-I (Wallis and Devlin 1993). The DNA was recovered with the Geneclean system (Geneclean) and cloned into dephosphorylated pBluescript KS+. Plasmids were electroporated into DH5-α competent cells and white colonies were prepared for sequencing with an alkaline hydrolysis method (Birnboim 1983).

Figure 6.1: Schematic map of the catfish brain-specific *igf-I* cDNA clone. The hybridization sites for the four primers used in the reverse transcriptase/PCR assay are shown. Primers 1 and 2 were used to detect brain-specific *igf-I* mRNA, whereas primers 5 and 6 were used to detect ubiquitous IGF-I.


0
120
240
360
480
600 (bp)
primer 5
primer 1
5'UTR
signal
B
C
A
D
E
3'UTR
primer 6
primer 2

Results

Sequence of catfish brain-specific IGF-I

In the process of screening a Thai catfish brain cDNA library with a cDNA probe to growth hormone-releasing hormone (GRF), a positive clone was isolated. This clone was partially sequenced and the 5' untranslated region was shown to have a high sequence identity with salmon IGF-I. The full length brain-specific IGF-I cDNA clone was then sequenced. This clone, referred to hereafter as catfish brain-specific IGF-I, had a cDNA nucleotide sequence that contained a 204 base pair (bp) 5' untranslated region and a 582bp open reading frame, followed by an 850bp 3'untranslated region for an entire length of 1633bp (Figure 6.2).

Distinct structures for brain-specific IGF-I and ubiquitous IGF-I

A single band was amplified from catfish brain and liver cDNA using primers 5 and 6 that matched regions in salmon liver IGF-I. A band was not detected in the negative control. The sequence of the PCR product was identical for both catfish brain and liver. A comparison of the catfish liver/brain (ubiquitous) form with salmon liver IGF-I showed 88% sequence identity for the nucleotides, and 100% identity for amino acids. The catfish ubiquitous form, however, had only 66% amino acid similarity with catfish brain-specific IGF-I.

Figure 6.2: Nucleotide and deduced amino acid sequence of a 1633bp catfish brain-specific *igf-I* cDNA. Domains B, A, C, and D have been underlined. The 5'UTR and 3'UTR are in small letters without spaces between the letters. It is not known if the signal peptide begins at the MET at amino acid position 1 or 19. Numbering of nucleotides is on the right side and for amino acids is at the left and center of the figure.

```
ttaaagacgtctgctgttcgataaatgtgattttcccaaaattcgccgagcaatggttctttaaaggcttattagagag      79
tgatgccatgtgcactccggccactttctcactgtttttaatgagttctaacgcgtgtatttttggcgggcattaccgt     158
gtccagacttgtgcatatctcctttgtcgtgcggagacccgtgggg                                       204
1                                   11
atg tct agc ggt cat tta ttc cag tgg cat tta tgt gat gtc ttc aag agt gcg atg tgc  264
Met ser ser gly his leu phe gln trp his leu cys asp val phe lys ser ala met cys
21                                  31
tgt att tct tgt aca cat act ctg tct ctg ctg ctt acc gta ttg tgg ttg acg ggg gag  324
cys ile ser cys thr his thr leu ser leu leu leu thr val leu trp leu thr gly glu
41                                  51
ctt ggg gcg ggt ccg tat acc ttg tgc cgg gcg gaa ctc gtt gac tcg ctg caa ttc gtg  384
leu gly ala gly pro tyr thr leu cys arg ala glu leu val asp ser leu gln phe val
61                                  71
tgt ggt cag cgc ggg ttc tac ttt agt aga cct atg ggg tat ggt tcc agc aca tgc tgg  444
cys gly gln arg gly phe tyr phe ser arg pro met gly tyr gly ser ser thr cys trp
81                                  91
aca cac aat cgt ggt atc ctg gat aag tgt tgt ttt caa tct tgc cag ctt cga tgg atc  504
thr his asn arg gly ile leu asp lys cys cys phe gln ser cys gln leu arg trp ile
101                                 111
gag atg tat tgc gca gct ttc gaa cct agc tta gta act cgc tcc gta cgt gca caa cgt  564
glu met tyr cys ala ala phe glu pro ser leu val thr arg ser val arg ala gln arg
121                                 131
gtg acg gat atg tgc agg aca tta aaa aaa agt gcc gtt cag aac gta gaa aga cgt act  624
val thr asp met cys arg thr leu lys lys ser ala val gln asn val glu arg arg thr
141                                 151
gaa ata tgg act gca cag cat cct gac aaa act aag ccg aaa aag gaa gtg tat cag cag  684
glu ile trp thr ala gln his pro asp lys thr lys pro lys lys glu val tyr gln gln
161                                 171
aat aca ctt cgt ggg aat aca tgg ggc ggg tac tac gga atg tac aag ttg aca cta gct  744
asn thr leu arg gly asn thr trp gly gly tyr tyr gly met tyr lys leu thr leu ala
181                                 191         194
aag gga tat gaa gac tcg gga tat cag acc acg ggc cca tag ctagtgcctgtgtgcaatc       805
lys gly tyr glu asp ser gly tyr gln thr thr gly pro AMB

gatcgggtaacgtaactaaaaggtgaattaagctactaggattataataataagcttgtgtctattctcttcccatacg  884
catcatgttcatcgatttcacacagtcgatcaccgatcgacatatatacatcgatgggacatagctacgtacgtacgaa  963
cacaaaaaagtgtacaccgggggacacatcatcgggatcagggatttaatcacacgtacgtagtgctagctgatgctaa 1042
cacatgcatgtgacatgtgtagtccccccacatatatatcacacacactcgatgctacacgtgtacgtagcacccccca 1121
ctcgtacgtagctacgcctgacccaaacgtatttttacacgttagtagatcacgtagtggctcgactgctagcattcag 1200
ctgatttctgcacaagagatgtggatttcgattgctgatgccttcgcatcgtgctgggaaaagtgatgcacatgtgtgc 1279
aactccaaatacgtcgtgagggaaacacactttaaagttttggacacacttacacccatcgatcacccacaccacaccg 1358
tagtcgtacggggttaaaaaagcctagtcaccgcgcgcgatcgatgcacacacagtttgcatgctagtgggctatcgga 1437
attcgcggccgcatgtgtgcacacgggatcgatgcgactattatatcctttaacatcgatcatgctagtgcatcgtaca 1516
cacatcgtttcgatcgtaatttattgatgtaggggcatcgttcgttcgctatgtccgcttactgttataaagaaggat 1595
tatcctttatatacgtcgatttacaaccactcacagta                                           1633
```

Northern blot analysis of brain-specific IGF-I

The level of expression and size of the mRNA transcript was determined by northern analysis with poly A+ rich mRNA from brain, liver, stomach, pancreas, pituitary, ovary, blood, poly A- RNA, and herring poly A+ mRNA. Only a single band at approximately 1700 bases was detected in the lane containing the brain poly A+ rich mRNA (Figure 6.3). Further exposure of the X-ray film to the blot with intensifying screens at -80°C for 10 days did not reveal any bands in other lanes. As an internal control for the quality of the mRNA, a sample of the poly A+ rich mRNA from each tissue was screened for tubulin and a band appeared in all lanes at the correct size for a tubulin mRNA (1500-1600 bases). Also, the RNA preparation did not appear to be degraded as judged by a consistent smear on an ethidium bromide-stained gel.

Expression of brain specific IGF-I detected with a PCR method

A sensitive PCR detection method for the presence of brain IGF-I in other tissues was developed. Primers were designed for the E-domain and 3' untranslated region of the catfish brain-specific IGF-I because these regions were distinct from the ubiquitous IGF-I form. Reverse transcribed cDNA of various catfish tissues were PCR amplified with these primers and the results are shown in Fig. 3A. A single band was detected in the lanes containing the brain cDNA and the positive control, but

other bands were not detected in lanes containing the cDNA from 5 other catfish tissues including liver, the negative control, or the cDNA from a chinook salmon liver IGF-I clone. Again, the cDNA appeared to be of good quality as judged by the PCR products obtained with tubulin primers 3 and 4 (Figure 6.4.b).

Figure 6.3: Northern blot analysis of mRNAs isolated from different catfish tissues. The migration of 18S and 28S ribosomal RNA is are shown on the right and the migration of brain-specific IGF-I mRNA is shown by the arrow on the left.

BRAIN

LIVER (10 ug)

STOMACH

PANCREAS

PITUITARY

OVARY

BLOOD

LIVER (1ug)

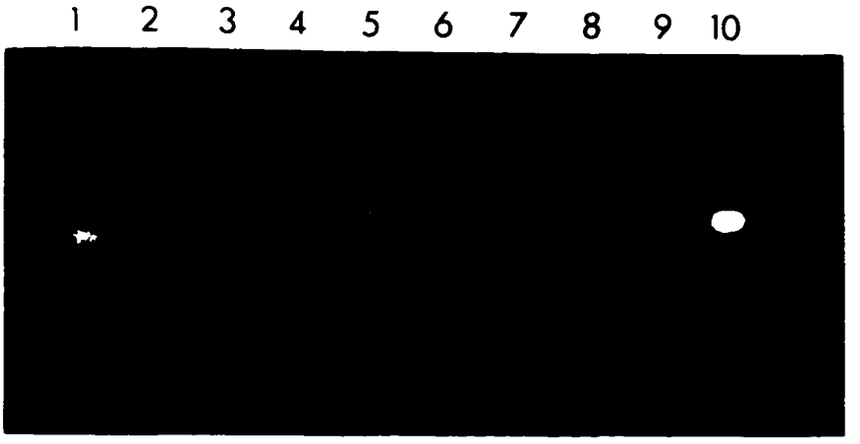
HERRING BRAIN

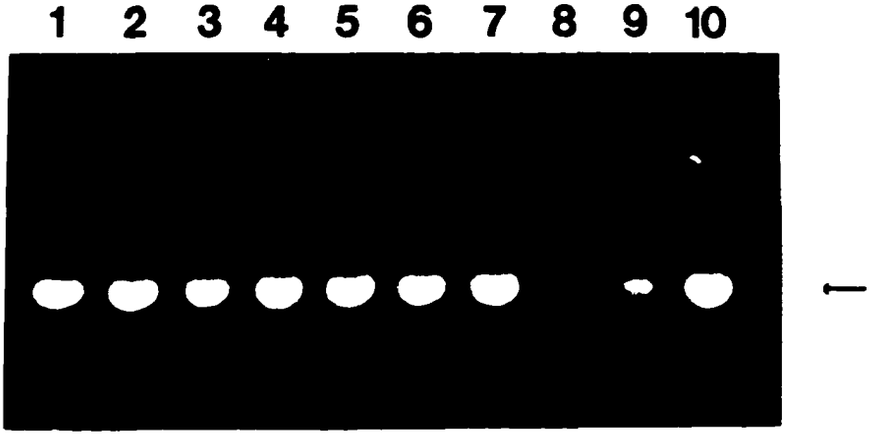
POLY A$^-$

28 S

18 S

Figure 6.4: (A) Brain specific *igf-I* cDNA of catfish detected by a reverse transcriptase/PCR assay. PCR reactions contained tissue cDNA/mRNA as follows: brain (lane 1), liver (lane 2), stomach (lane 3), pancreas (lane 4), pituitary (lane 5), ova (lane 6), African catfish brain (lane 7), negative control (lane 8), salmon liver IGF-I clone (lane 9), and brain specific IGF-I clone (positive control; lane 10). **(B)** Tubulin detected by a reverse transcriptase/PCR assay to determine the quality of mRNA/cDNA for the assay in Fig. 6.4A. The lanes contained the cDNA/mRNA as above except lanes 9 and 10, which contained the herring brain cDNA and tubulin clone (positive control), respectively.


1 2 3 4 5 6 7 8 9 10


1 2 3 4 5 6 7 8 9 10

Discussion

Isolation of brain-specific IGF-I is based on a heterologous probe

In this paper we report the structure and tissue expression of a brain-specific IGF-I that has high sequence identity compared with other forms of IGF-I, but is produced only within the brain. More appropriately, this peptide should be called a brain-specific IGF-I peptide because it may have different biological and receptor binding properties compared with ubiquitous IGF-I. In the process of screening a catfish brain cDNA library with a catfish growth hormone-releasing hormone (*grf*) cDNA probe, a positive clone, initially thought to be a *grf* clone, was isolated, sequenced and identified as an IGF-like molecule. The rational for hybridization of the *grf* probe to the *IGF-I* clone is not clear because an area of identity is not obvious between the clones.

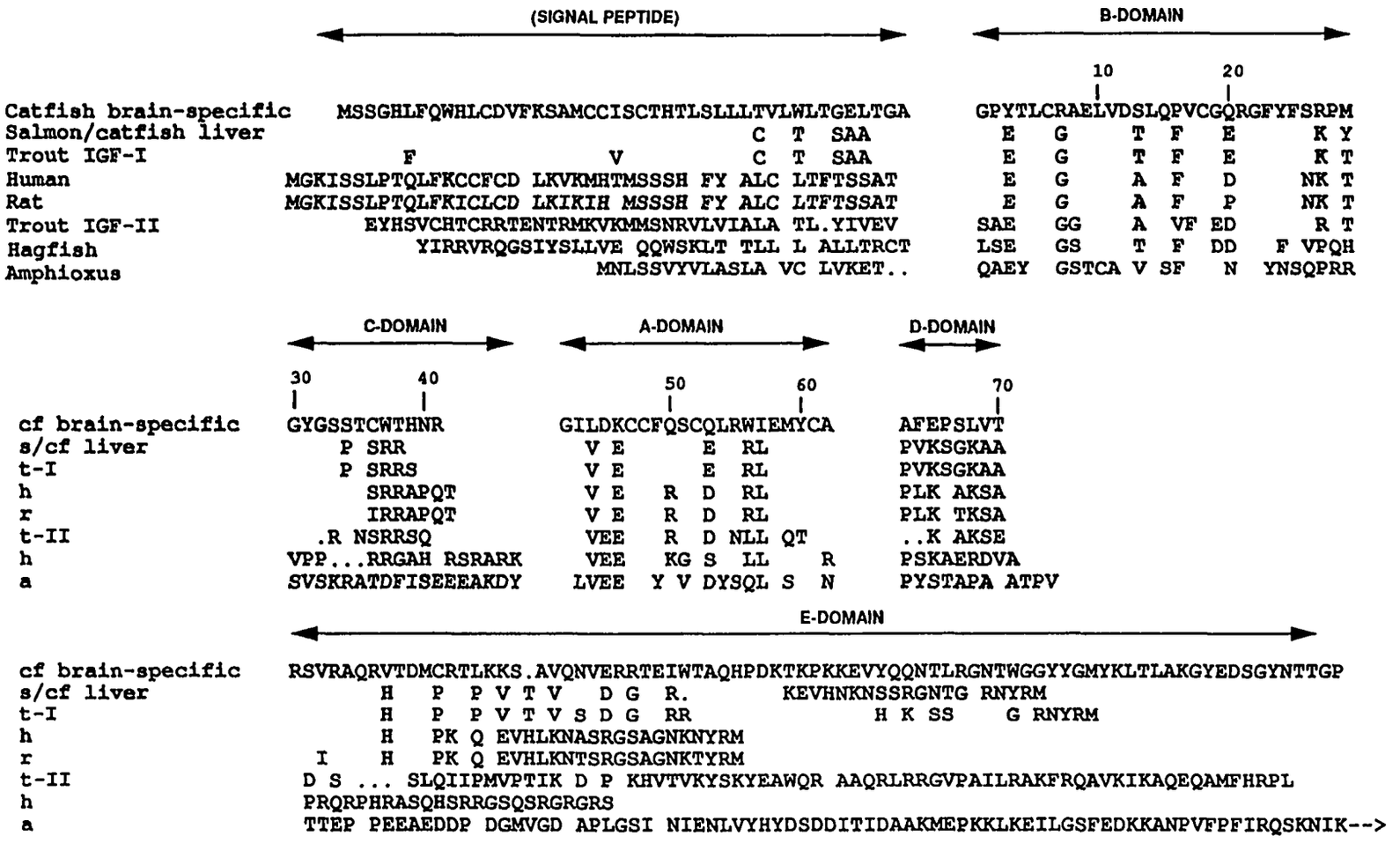
Brain-specific IGF-I is a member of the IGF family

The brain-specific IGF-I amino acid sequence deduced from the cDNA contains many of the residues that are essential for IGF-I structure and bioactivity (Nagamatsu 1991). The 6 cysteine amino acids responsible for the intra- and inter-domain disulfide linkages, located at positions 6, 18, 47, 48, 52, 61 in the catfish brain-specific IGF-I (Fig. 6.2), are

conserved in all IGF-I peptides determined to date. Also conserved are the hydrophobic residues, positions 5, 10, 11, 14, 17, 43, 44, 54, 57, and 60 that create the hydrophobic core of the IGF-I peptide, thought to be responsible for secondary structure. Conserved in all teleost and mammalian IGF-I and insulin sequences are three aromatic amino acids that appear to be important for binding to the IGF type-I receptor. The aromatic residues Phe-Tyr-Phe, (residues 23-24-25, Fig. 6.2) at the carboxy end of the B-domain are conserved in all IGF-I peptides except those from the hagfish and amphioxus, which possess Phe-Phe-Phe and Tyr-Asn-Ser, respectively (Nagamatsu *et al.* 1991; Chan *et al.* 1990). Therefore, the deduced protein sequence from catfish brain IGF-I cDNA is predicted to be bioactive because it contains all the required cysteine amino acids for disulfide linkages, the hydrophobic amino acids for the conservation of a hydrophobic core, and the Phe-Tyr-Phe sequence required for receptor binding. However, the sequence identity of the brain-specific IGF-I deduced peptide is not as high as that normally seen among IGF-I molecules (Figure 6.5).

The identity between the deduced catfish brain-specific IGF-I peptide B-C-A-D domains is 66% with salmon, 66% with trout, 61% with human, 61% with rat, 40% with hagfish, and 27% with amphioxus. The catfish brain-specific IGF-I does not appear to be an IGF-II molecule because its amino acid identity with trout IGF-II is only 44%. Other amino acid changes in brain-specific IGF-I may be required for binding to a neuron-specific IGF-I

Figure 6.5: Alignment of the catfish brain-specific preproIGF-I with several other ubiquitous IGF-I prepropeptides. IGF amino acids of catfish brain-specific IGF-I are aligned with salmon (s), catfish (cf), human (h), rat (r) IGF-I, trout IGF-I and IGF-II, and hagfish and amphioxus IGF prepropeptides. The catfish liver IGF-I sequence is based on the PCR fragment shown in Fig. 6.6. Only the coding region of the mature peptide is compared. Identity between the catfish sequence and the other sequences are indicated by a space. The amino acids of the bioactive IGF-I peptide are numbered starting at the beginning of the indicated B-domain and terminating at the end of the E-domain. A period was used to frame shift the peptide for maximal alignment between sequences.


(SIGNAL PEPTIDE)
B-DOMAIN
10
20
Catfish brain-specific MSSGHLFQWHLCDVFKSAMCCISCTHTLSLLLTVLWLTGELTGA GPYTLCRAELVDSLQPVCGQRGFYFSRPM
Salmon/catfish liver C T SAA E G T F E K Y
Trout IGF-I F V C T SAA E G T F E K T
Human MGKISSLPTQLFKCCFCD LKVKMHTMSSSH FY ALC LTFTSSAT E G A F D NK T
Rat MGKISSLPTQLFKICLCD LKIKIH MSSSH FY ALC LTFTSSAT E G A F P NK T
Trout IGF-II EYHSVCHTCRRTENTRMKVKMMSNRVLVIALA TL.YIVEV SAE GG A VF ED R T
Hagfish YIRRVRQGSIYSLLVE QQWSKLT TLL L ALLTRCT LSE GS T F DD F VPQH
Amphioxus MNLSSVYVLASLA VC LVKET.. QAEY GSTCA V SF N YNSQPRR
C-DOMAIN
A-DOMAIN
D-DOMAIN
30
40
50
60
70
cf brain-specific GYGSSTCWTHNR GILDKCCFQSCQLRWIEMYCA AFEPSLVT
s/cf liver P SRR V E E RL PVKSGKAA
t-I P SRRS V E E RL PVKSGKAA
h SRRAPQT V E R D RL PLK AKSA
r IRRAPQT V E R D RL PLK TKSA
t-II .R NSRRSQ VEE R D NLL QT ..K AKSE
h VPP...RRGAH RSRARK VEE KG S LL R PSKAERDVA
a SVSKRATDFISEEEAKDY LVEE Y V DYSQL S N PYSTAPA ATPV
E-DOMAIN
cf brain-specific RSVRAQRVTDMCRTLKKS.AVQNVERRTEIWTAQHPDKTKPKKEVYQQNTLRGNTWGGYYGMYKLTLAKGYEDSGYNTTGP
s/cf liver H P P V T V D G R. KEVHNKNSSRGNTG RNYRM
t-I H P P V T V S D G RR H K SS G RNYRM
h H PK Q EVHLKNASRGSAGNKNYRM
r I H PK Q EVHLKNTSRGSAGNKTYRM
t-II D S ... SLQIIPMVPTIK D P KHVTVKYSKYEAWQR AAQRLRRGVPAILRAKFRQAVKIKAQEQAMFHRPL
h PRQRPHRASQHSRRGSQSRGRGRS
a TTEP PEEAEDDP DGMVGD APLGSI NIENLVYHYDSDDITIDAAKMEPKKLKEILGSFEDKKANPVFPFIRQSKNIK-->

receptor or may occur in areas not known to be essential for IGF-I function.

Within the 5' untranslated region are five potential ATG translation initiation sites that create open reading frames terminating before the authentic translation initiation site as judged by the length of other IGF-I cDNA molecules. At the beginning of the open reading frame, nucleotides 205-207 (Fig. 6.2) encode an initiator methionine residue similar to those found in other IGF-I molecules except hagfish IGF-I, which begins with a tyrosine. The signal peptide is 129bp (43 amino acids). This is unusually long for most signal peptides, but IGF-I signal peptides in chicken, human and salmon are 48, 48, and 44 amino acids, respectively. As with most other IGF-I sequences, the B, C, and A domains are 29, 12, and 21 amino acids, respectively, and have higher similarity than D and E domains to other IGF-I molecules. The brain-specific IGF-I D domain has little or no identity to other forms. However, the 171bp E domain has 53%, 70%, and 40% identity to the salmon, trout (IGF-I) and human forms, respectively. The function of the E-domain in liver IGF-I peptides is not known, but the conservation of its sequence and the alternative processing of the E-domain (Wallis and Devlin 1993; Duguay *et al.* 1992) suggest an important biological function.

Brain has a second form of IGF-I that is identical with liver IGF-I

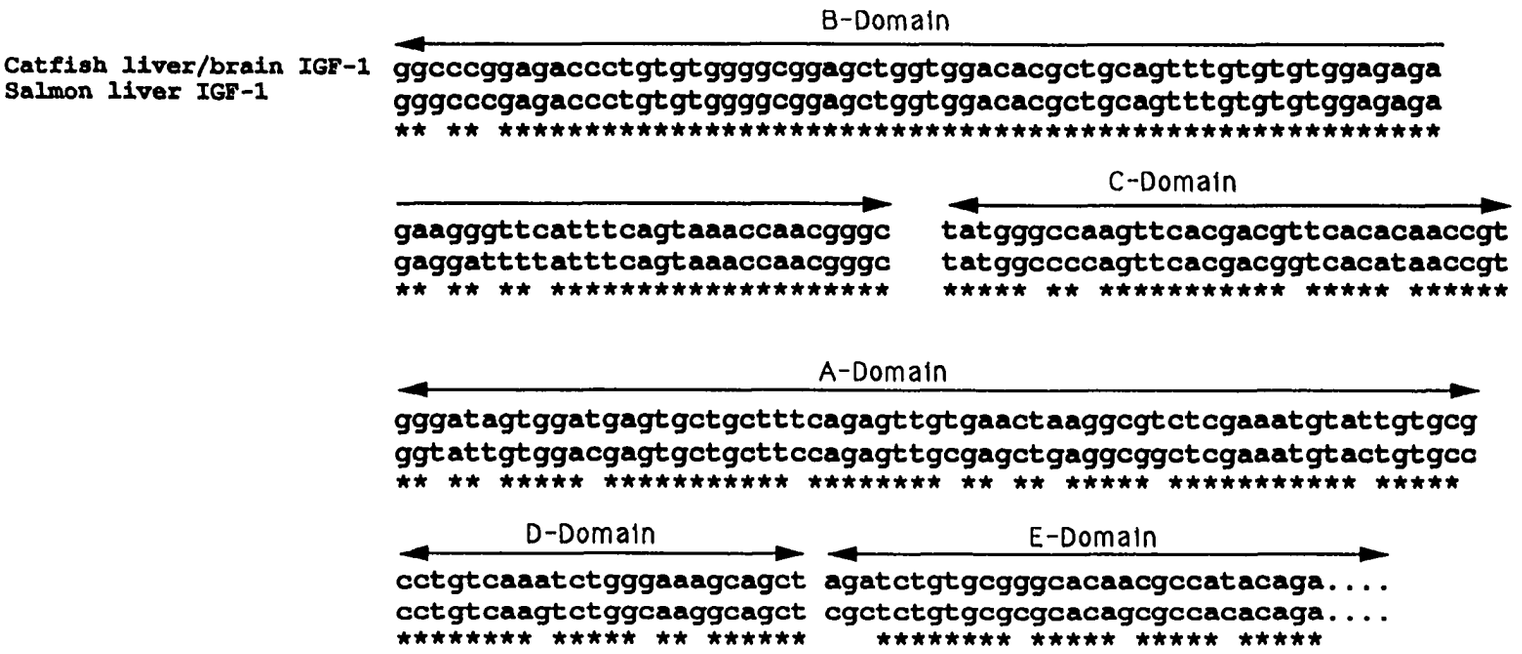
To ensure the brain specific IGF-I was indeed unique and not the product of liver IGF-I expression in the brain, PCR was used to determine whether a ubiquitous form of IGF-I is present in catfish liver and brain. Two primers, corresponding to salmon IGF-I (Wallis and Devlin 1993), amplified an IGF-I sequence from both the brain and liver of catfish. This cDNA sequence was distinct from brain-specific IGF-I. This established the presence of two forms of IGF-I in the brain: one form is brain-specific and the other form is identical to liver IGF-I. The cDNA sequence of this mature catfish IGF-I has 88% nucleotide identity to the salmon form, (Figure 6.6) but amino acid changes do not result from these base substitutions. A comparison of the catfish ubiquitous and brain-specific IGF-I hormones showed 66% sequence identity for the mature peptides.

Brain-specific IGF-I has several potential roles

The occurrence of a brain specific IGF-I clone helps to explain the neurotrophic and neuromodulatory role previously suggested for IGF-I. Regulating neuronal and glial function, acting as a neurotrophic factor, and modulating synaptic transmission have been suggested for IGF-I, but previously it was hypothesized that liver IGF-I, rather than a distinct brain IGF-I, was expressed within the brain. The function of brain

IGF-I is not known. The recent identification and localization of the insulin-related receptor (IRR) is intriguing because it demonstrates the existence of neuron-specific receptors that have high sequence similarity to the receptors for insulin and IGF-I (Reinhardt *et al.* 1993; Shier and Watt 1989). Insulin-related receptor mRNA has been shown to be present within sensory neurons of the trigeminal and dorsal root ganglia during early embryonic development. Therefore, expression of the IRR appears to be restricted, as is the brain IGF-I peptide, to the nervous system. The presence of both a brain specific IGF-I and a ubiquitous (liver) form of IGF-I in the brain emphasizes the importance of these growth factors for the nervous system.

Figure 6.6: Nucleotide alignment of the ubiquitous IGF-I found in catfish liver and brain with the salmon IGF-I. Identity of nucleotides between the catfish and salmon forms of IGF-I is indicated by a * under the sequences.


B-Domain
Catfish liver/brain IGF-1
Salmon liver IGF-1
ggcccggagaccctgtgtggggcggagctggtggacacgctgcagtttgtgtgtggagaga
gggcccgagaccctgtgtggggcggagctggtggacacgctgcagtttgtgtgtggagaga
C-Domain
gaagggttcatttcagtaaaccaacgggc
gaggattttatttcagtaaaccaacgggc
tatgggccaagttcacgacgttcacacaaccgt
tatggccccagttcacgacggtcacataaccgt
A-Domain
gggatagtggatgagtgctgctttcagagttgtgaactaaggcgtctcgaaatgtattgtgcg
ggtattgtggacgagtgctgcttccagagttgcgagctgaggcggctcgaaatgtactgtgcc
D-Domain
E-Domain
cctgtcaaatctgggaaagcagct
cctgtcaagtctggcaaggcagct
agatctgtgcgggcacaacgccatacaga....
cgctctgtgcgcgcacagcgccacacaga....

References

Bichell DP, Kikuchi K, Rotwein P 1992 Growth hormone rapidly activates inulin-like growth factor-I gene transcription in vivo. Molecular Endocrinology 6: 1899-1908.

Birnboim M. 1983 A rapid alkaline extraction method for the isolation of plasmid DNA. Methods in Enzymology 100: 243-255

Cao Q-P, Duguay SJ, Plisetskaya E, Steiner DF, Chan SJ 1989 Nucleotide sequence and growth hormone-regulated expression of salmon insulin-like growth factor-I mRNA. Molecular Endocrinology 3: 2005-2010.

Casella SJ, Smith EP, Van Wykm JJ, Joseph DR, Hynes MA, Hoyt EC, Lund PK 1987 Isolation of rat testis cDNAs encoding an insulin-like growth factor-I precursor. DNA 6: 325-330.

Chan SJ, Cao Q, Steiner DF 1990 Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. Proceedings of the National Academy of Science (USA) 87: 9319-9323.

Chomczyski P, Sacchi N 1987 Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162: 156-159.

Coe IR, Munroe R, Sherwood NM 1992 Isolation of different brain-specific isotypes of α-tubulins from chum salmon. DNA sequence 3: 257-262.

Daughaday W H, and Rotwein P 1989 Insulin-like growth factors I and II. Peptide, messenger ribonucleic acid and

gene structures, serum and tissue concentrations. Endocrine Reviews 10: 68-91.

Duguay SJ, Park LK, Samadpour M, Dickhoff WW 1992 Nucleotide sequence and tissue distribution of three insulin-like growth factor I prohormones in salmon. Molecular Endocrinology 6: 1202-1210.

Foyt HL, Lanau F, Woloschak M, Leroith D, Roberts CT 1992 Effects of growth hormone on levels of differentially processed insulin-like growth factor I mRNAs in total and polysomal mRNA populations. Molecular Endocrinology 6: 1881-1888.

Francis GL, Upton FM, Ballard FJ, McNeil KA, Wallace JC. 1988 Insulin-like growth factors I and 2 in bovine colostrum: sequence and biological activities compared to a potent truncated form. Journal of Biochemistry 251: 95-103.

Froesch ER, Schmidt C, Schwander J, Zapf J 1985 Actions of insulin-like growth factors. Annual Reviews in Physiology 47: 443-467.

Jansen M, van Schaik FMA, Ricker AT, Bullock B, Woods DE, Gabbay KH, Nussbaum AL, Sussenbach JS, Van den Brande JL 1983 Sequence of cDNA encoding human insulin-like growth factor I precursor. Nature 306: 609-611.

Kajimoto Y, Rotwein P 1989 Structure and expression of a chicken insulin-like growth factor I precursor. Molecular Endocrinology 3: 1907-1913.

Kajimoto Y, Rotwein P 1990 Evolution of insulin-like growth factor I (IGF-I): structure and expression of an IGF-I precursor from *Xenopus laevis.* Molecular Endocrinology 4: 217-226.

Murphy LJ, Bell GI, Duckworth ML, Friesen HG 1987 Identification, characterization, and regulation of a rat complementary deoxyribonucleic acid which encodes insulin-like growth factor-I. Endocrinology 121: 684-691.

Nagamatsu S, Chan SJ, Falkmer S, Steiner DF 1991 Evolution of the insulin superfamily. Journal of Biological Chemistry 266: 2397-2402.

Ogasawara M, Karey KP, Marquardt H, Sirbasku DA 1989 identification and purification of truncated insulin-like growth factor I from porcine uterus. Evidence for high biological potency. Biochemistry 28: 2710-2721.

Reinhardt RR, Chin E, Zhang B, Roth RA, Bondy CA 1993 Insulin receptor-related receptor messenger ribonucleic acid is focally expressed in sympathetic and sensory neurons and renal distal tubule cells. Endocrinology 133: 3-10.

Romagnolo D, Akers RM, Wong EA, Boyle PL, McFadden TB, Turner JD 1992 Overexpression of ovine insulin-like growth factor-I stimulates autonomous autocrine or paracrine growth in bovine mammary-derived epithelial cells. Molecular Endocrinology 6: 1774-1780.

Sanger F, Nicklen S, Coulson AR 1977 DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science (USA). 74: 5463-5467.

Sara VR, Carlsson-Skwirut C, Andersson C, Hall E, Sjogren B, Holmgren A, Jornvall H 1986 Characterization of somatomedins from human fetal brain: Identification of a variant form of insulin-like growth factor I. Proceedings of the National Academy of Science (USA). 83: 4904-4907.

Sara VR, Carlsson-Skwirut C, Bergman T, Jornvall H, Roberts PJ, Crawford M, Hakansson LN, Civalero I, Nordberg A 1989 Identification of GLY-PRO-GLU (GPE), the aminoterminal tripeptide of insulin-like growth factor I which is truncated in brain, as a novel neuroactive peptide. Biochemical and Biophysical Research Communications 165: 766-771.

Schwartz MW, Figlewicz DP, Baskin DG, Woods SC, Porte D 1992 Insulin in the brain: hormonal regulation of energy balance. Endocrine Reviews 13: 387-414.

Shamblott MJ, Chen TT 1992 Identification of a second insulin-like growth factor in a fish species. Proceedings of the National Academy of Science (USA) 89: 8913-8917.

Shier P, Watt VM 1989 Primary structure of a putative receptor for a ligand of the insulin family. Journal of Biological Chemistry 264: 14605-14608.

Tavakkol A, Simmen FA, Simmen RCM 1988 Porcine insulin-like growth factor-I (pIGF-I): complementary deoxyribonucleic acid cloning and uterine expression of

messenger ribonucleic acid encoding evolutionarily conserved IGF-I peptides. Molecular Endocrinology 2: 674-681.

Wallis AE, Devlin RH 1993 Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. Molecular Endocrinology 7: 409-422

Wong EA, Ohlsen SM, Godfredson JA, Dean DM, Wheaton JE 1989 Cloning of ovine insulin-like growth factor-I cDNAs: heterogeneity in the mRNA population. DNA 8: 649-657.

CHAPTER 7

Ancient Divergence of Insulin and Insulin-Like Growth Factor (IGF)

Summary

In this chapter I examined the molecular evolution of insulin and IGF, peptides that are members of the insulin superfamily as discussed in chapter 6. Studies on the evolutionary pathway of insulin and IGF suggested they became distinct molecules only after the vertebrates arose. For amphioxus, a hybrid insulin/IGF molecule was reported. In contrast I found that tunicates, another sister group to the vertebrates, possess two cDNAs of which one encodes a distinct preproinsulin with B, C and A domains, and the other encodes a distinct preproinsulin-like growth factor (IGF) with five domains in their proper sequence. Both clones are expressed in the nervous system, digestive tract, gonad, and heart, but not in branchial basket or tunic. *In situ* hybridization shows tunicate *insulin* and *igf* mRNA in the cortical cells of the neural ganglion. Hence, tunicate insulin and IGF have similar expression patterns in tunicates, but separate gene lineages in both protochordate and vertebrate evolution. Insulin and IGF have a distinct evolutionary history of more than 600 million years.

Introduction

The insulin superfamily is comprised of insulin, insulin-like growth factors-I and -II (IGF-I and -II), relaxin and several invertebrate insulin-related peptides. The superfamily represents a group of structurally related peptides that are essential in growth, development and metabolism (Froesch *et al.* 1985: Daughaday and Rotwein, 1989). Many different vertebrate insulin and IGF sequences have been isolated and sequenced (Daughaday and Rotwein, 1989). Even within the same species, several different forms of insulin and IGF-I have been identified (Chan *et al.* 1979; Shamblot and Chen, 1992; Adham *et al.* 1993; McRory and Sherwood, 1994; Perfetti *et al.* 1994; Kavson *et al.* 1994). Sequencing of peptides and cloning of different insulin and IGF-related cDNAs reveal a high degree of identity in the sequence of the functional domains and in gene organization. The similarity between family members suggests insulin and IGF may have evolved from a common ancestral gene (Steiner *et al.* 1984; Ebberink *et al.* 1989; Chan *et al.* 1990). However, it is not known when the divergence of insulin, IGF, and insulin-related peptides took place.

A unique group within the insulin superfamily of peptides includes silkworm bombyxin A1, B1, and IV, molluscan insulin-related peptide-I and -II (MIP-I and -II) and locust insulin-related peptide (LIRP). These peptides have sequence and structural identity with insulin, although the sequence for

receptor binding is distinct from that of vertebrate insulin (Adachi *et al.* 1989; Smit *et al.* 1991). For invertebrates there are radioimmunoassay (RIA) data and immunocytochemical evidence for the presence of insulin and related molecules, but the structures are not known (Bevis and Thorndyke, 1978; Plisetskaya *et al.* 1978; O'Neil *et al.* 1986; Galloway and Cuttfield, 1988; Thorndyke and Goldsworthy, 1988). Immunocytochemical evidence suggested that an insulin-like molecule was synthesized within the tunicate *Pyura pachydermatina* (Galloway and Cuttfield, 1988). High pressure liquid chromatography (HPLC) and RIA evidence supported this observation in that two fractions contained immunoreactive insulin from *P. pachydermatina's* digestive tract. Sequence data was not obtained. In addition, O'Neil *et al.* (1986) have reported the presence of insulin-like immunoreactive cells in the neural ganglion of the ascidian *Ciona intestinalis*. However, an authentic insulin molecule has yet to be identified and sequenced in an organism ancestral to the vertebrates.

The developmental pattern and evolutionary position of the tunicate suggest that a neotenous relative of the tunicate was the progenitor from which vertebrates evolved (Ebberink *et al.* 1989). To investigate the evolutionary pathway of the insulin gene family, we have cloned, sequenced and determined tissue expression of an invertebrate insulin and insulin-like growth factor from a tunicate *Chelyosoma productum*.

Materials and Methods

Preparation of tunicate cDNA by construction of library

Tunicates (*Chelyosoma productum*) were removed from their tunic and the top half, consisting predominately of the neural ganglion, neural gland and dorsal strand plexus, was removed from the gonad, gut and branchial basket. The tissue was placed in liquid nitrogen and stored at -80°C. Total RNA was extracted with an acidic guanidinium thiocyanate method (Chomczyski and Sacchi, 1987) and poly A^+-rich mRNA was purified using the Poly A+ Attract System (Invitrogen). A cDNA library was constructed with the uniZap-cDNA synthesis kit and Gigapack packaging mix (Stratagene) according to the manufacturer's instructions.

Amplification of tunicate *igf* by PCR

DNA was amplified by PCR using an aliquot from the cDNA library lysate and primers corresponding to conserved amino acid sequences in salmon IGF-I. Primer A (5'-ggccccgagaccctgtgtg gg-3') corresponded to the nucleotides encoding the first seven amino acids of the B domain and primer B (5'cacaagtacattcgagccgc-3') to the nucleotides encoding the last 7 residues of the A chain. Amplification was done in a 50µl volume (0.2µg cDNA, 5U Taq, 1x Taq buffer (Pharmacia), 200mM dNTP's, 2.5mM $MgCl_2$, 20 pmol primer A and B) with 40 cycles at 94°C for 1.5 min, 46°C for 2 min, 72°C for 2.5 min and

a 5.3 min extension at 72°C. PCR reactions were purified by electrophoresis through a 3% agarose gel. The resulting bands were cloned into pBluescript KS+ (Stratagene), electroporated into XL-1 Blue (MRF') competent cells, and prepared for sequencing with an alkaline hydrolysis method (Birnboim, 1983). Both strands of the plasmid were sequenced with [α-^{35}S] dATP using the chain termination method (Sanger *et al.* 1977) with Sequenase 2.0 (US Biochemical Corp.) and Vent (exo-) (New England Biolabs). All sequencing gels were 6% polyacrylamide/7M urea wedge gels, dried under vacuum at 80°C and exposed to Kodak XAR-5 film for 12-24h.

Assay of mRNA by tissue reverse transcription and PCR

Messenger RNA was isolated using the poly A+ Attract System (Invitrogen) from the following tissues: neural gland, dorsal strand/neural ganglion, gonad, gonad/digestive gland, intestine, heart, tunic and branchial basket. Single stranded cDNA was synthesized from 0.5µg poly A+ mRNA for each tissue using 200U avian reverse transcriptase (H$^-$ Superscript, BRL), 10mM dithiothreitol, 10mM each dNTP, 160U RNA guard, 2mM Oligo dT_{20}, and 1X H$^-$ RT buffer in a total reaction volume of 20µl. The reaction proceeded for 90 min at 42°C followed by 5 min at 90°C. For DNA amplifications primers were prepared to match the tunicate sequence. Primers 1 (5'-ccgtaccttactcga gtaa) and 2 (5'-ctgcatcctatcgtaacg) were designed to hybridize with the tunicate insulin 5'UTR and 3'UTR and primers 3 (5'-

tacgtagcttacccattgc) and 4 (5'-gatcgtacgatgctgaatcgtcg) were designed to hybridize to the IGF 5'UTR and 3'UTR. DNA amplifications were done in a 50µl volume that contained 0.5µl cDNA/RNA mixture diluted 1:5, 1x Taq buffer (Pharmacia), 200µM dNTP's, 2.5mM $MgCl_2$, and 20 pmol each of primers for tunicate insulin (primers 1 and 2) and tunicate IGF (primers 3 and 4) (Figure 6.1). All reagents, except the DNA polymerase, were mixed, layered with mineral oil, and heated to 95°C for 5 min. The mixture was frozen rapidly in a dry ice/ethanol bath and 5U Taq DNA polymerase was added on top of the frozen oil. The tubes were replaced in the preheated 95°C thermal cycler where 40 cycles at 95°C for 1 min, 56°C for 2 min, and 74°C for 1.5 min were completed.

Detection of positive clones with library screening

A total of 50,000 plaque forming units from a tunicate cDNA library were screened. Duplicate nylon membrane (Bio-Rad) lifts were prehybridized at 50°C in 6x SSC, 5x Denhardt's solution, 0.5% SDS for 4 h. The hybridization solution, consisting of 6x SSC and 0.5% SDS, was added to the [α-^{32}P]dCTP (Dupont) labelled probe, and incubated at 50°C overnight. The probe was a 165bp fragment obtained from the PCR reaction above. The membranes were washed under high stringency (0.1xSSC/0.1%SDS) for 50 min at 65°C, then exposed to Kodak XAR-5 film for 5 days at -80°C. Isolated single positive clones had inserts rescued with *in vitro* excision.

Detection of *insulin* mRNA using *in situ* hybridization

Localization of the tunicate preproinsulin mRNA in whole mounts of the neural ganglion of *Chelyosoma productum* was done by *in situ* hybridization using a digoxigenin (DIG)-labeled (Boehringer Mannheim) tunicate *insulin* (bp 351-495) and *igf* (bp507-671) RNA probes. From the *insulin* and *igf* cDNA clones, the regions to be used as the RNA probes were removed with restriction enzymes and subcloned into pBluescript. All RNA probes were synthesized, purified, and tested as to the manufacturer's (Boehringer Mannheim) instructions. The changes in protocol for fixation, prehybidization and hydridization are listed below. The tunicate neural gland and ganglia were dissected and pinned on Sylgard coated dishes and fixed for 3 h in 4% paraformaldehyde in phosphate buffered saline (PBS) (pH 7.4). The fixed tissues were washed in PBS and soaked overnight in 30% sucrose. Tissue was embedded in O.T.C compound and 10μm sections were taken, then allowed to dry on poly-L-lysine coated slides. Sections were fixed again in 4% paraformaldehyde for 5 min, washed 3 times in PBS (5 min each) and placed for 10 min in 2 X SSC. Prehybridization was in 2 X SSC for 2 h at room temperature. This solution was exchanged for the hybridization solution that consisted of 2 X SSC with a DIG-labeled RNA probe diluted 1:200. The hybridization solution was incubated overnight at 42°C. The sections were washed with SSC (0.5 X SSC) followed with 2% normal goat serum in

TBS buffer for 30 min at room temperature. The remaining steps involving the anti-digoxigenin antibody and the substrate detection were done as to the manufacturer's instructions.

Results

cDNAs encoding tunicate insulin-like and IGF-like peptides.

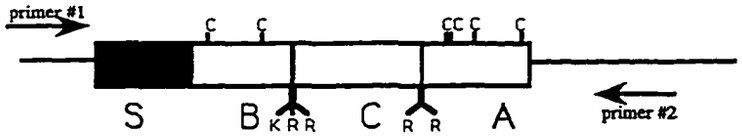
To generate a probe from tunicate *insulin*-like and *igf*-like cDNAs, the PCR method was used with primers corresponding to salmon IGF-I A and B domains (Wallis and Devlin, 1993). The resulting fragment, which encoded tunicate IGF, was used to screen the cDNA library. Screening of 50,000 clones resulted in 10 positive clones that were purified and sequenced. Six of the clones encoded insulin, one clone encoded tunicate IGF (Figure 7.1a), and the remaining three clones could not be identified. The cDNA and deduced amino acid sequences of both clones have a high degree of identity. The deduced amino acid sequence of tunicate preproinsulin (signal peptide and B-C-A domains) has 87% identity with the tunicate preproIGF, although tunicate preproinsulin is considerably shorter than preproIGF.

The tunicate *insulin* cDNA clone is 495bp and the deduced amino acids encode a signal peptide, and domains for B, C, and A chains (Figure 7.1A). Like other vertebrate insulins, but unlike amphioxus insulin-like peptide (ILP) (Chan *et al.* 1990), D and E domains do not exist due to the presence of a stop codon at the end of the A domain. The deduced peptide of the tunicate insulin C-domain is 4 amino acids shorter than that found in human insulin, contains little sequence homology, and is flanked by dibasic cleavage

Figure 7.1: Nucleotide and deduced amino acid sequence of **(A)** a tunicate 495bp *insulin* cDNA and **(B)** a tunicate 671bp *igf* cDNA clone. The primers used for the expression study are underlined or overlined. Coding regions are indicated by the B, C, A, D, and E labelled boxes and the 5' and 3' untranslated regions by the thin lines. The black box encodes the signal peptide and the box with the horizontal lines indicates the extra base pairs in the B-domain of the *igf* cDNA clone compared to the tunicate *insulin* clone. Dibasic cleavage sites for the C domains are indicated by lysine and arginine. The cysteine residues are illustrated above each clone with C, and the clone-specific primers (1-4) are shown.

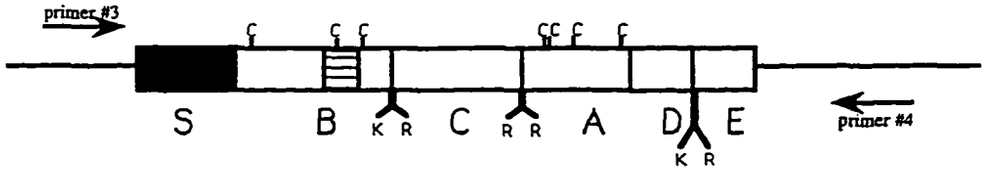
A

```
    1 tttcctaaccccgtaccttactcgagtaaagagactactatagcatcgggc ATG AAT TCC TCC AGC TAC AGC  72
  1                                                       M   N   S   S   S   Y   S    7
 73 GTG GTG CTC AAT AAG GAG GTG ATG GCC AAG GAG CAG AAC GAG TAC CTG TGC GGG TCC CAC  132
  8 V   V   L   N   K   E   V   M   A   K   E   Q   N   E   Y   L   C   G   S   H   27
133 CTG GCC GAC GTG CTT TCC TTC GTC TGC GGG AAC AGA GGT TAT TTC TTT GAG CCC CGG ACG  192
 28 L   A   D   V   L   S   F   V   C   G   N   R   G   Y   F   F   E   P   R   T   47
193 AGG CGG TCC GTG GAC GAC AGA GCA GTG GAC TTT ATC TCG GAA CAA CAG GCG AAA GAC TAT  252
 48 R   R   S   V   D   D   R   A   V   D   F   I   S   E   Q   Q   A   K   D   Y   67
253 ATG GGG GCC ATG CCC CAC ATC CGG AAG CGC CGG GGT ATT GTG GAG CAG TGC TGC TAT CGG  312
 68 M   G   A   M   P   H   I   R   K   R   R   G   I   V   E   Q   C   C   Y   R   87
313 GTG TGC AGC CTT CGG CAA CTC GAG ACG TAT TGC AAC TAG ccagtcggctatattaaatagaagccgt 379
 88 V   C   S   L   R   Q   L   E   T   Y   C   N   *                               100
380 tacgataggatgcagatgcttttaaaaataaacccacgtgggaaaaatggacacacaggtataatgttatttatttaaa 459
460 accaggattttaaatatttaacaaaaaaaaaaaaaa                                           495
```



B

```
   1 tgggtatcctagagggcatcacagttaagctagtccggtacgtagcttacccattgcacacagcgtacccatctcgacta  80
  81 tagcatcgggc ATG AAT TCC TCC AGC TAC AGC GTG GTG CTC AAT AAG GAG GTG GCC AAG GAG  142
   1            M   N   S   S   S   Y   S   V   V   L   N   K   E   V   A   K   E   17
 143 CAG GCG GAG TAC CTG TGC GGG TCC ACC CTG GCC GAC GTG CTT TCC TTC GTC GCC GAG CAC  202
  18 Q   A   E   Y   L   C   G   S   T   L   A   D   V   L   S   F   V   A   E   H   37
 203 CTC TGC GGA TCC TAC GCC GAA ATC CTT AGC GTG TGT GGA AAC AGA GGC TAC TAC TTT CAA  262
  38 L   C   G   S   Y   A   E   I   L   S   V   C   G   N   R   G   Y   Y   F   Q   57
 263 CCT CGG ACG TCC GTG GGC AAA AGA GCA ATC GAC TTT ATC TCG GAA AAG CAG GCG AAA GAC  322
  58 P   R   T   S   V   G   K   R   A   I   D   F   I   S   E   K   Q   A   K   D   77
 323 TAT ATG GGG GCC ATG CCC CAC ATC CGG CGG CGC CGG GGT TTG GTG GAG GAG TGC TGC TAT  382
  78 Y   M   G   A   M   P   H   I   R   R   R   R   G   L   V   E   E   C   C   Y   97
 383 CGG GTG TGC GAC TTC CGG CAA CTC GAG ACG TAT TGC AAC CCC TAC TCG ACG GCC TCC AGC  442
  98 R   V   C   D   F   R   Q   L   E   T   Y   C   N   P   Y   S   T   A   S   S  117
 443 CAC CCC CAC CCC CTC AAA AGA CAC CGA GCC GCA CCC AGT GGT GGA CGA CGA TCT CCT CTC  502
 118 H   P   H   P   L   K   R   H   R   A   A   P   S   G   G   R   R   S   P   L  137
 503 TGA gctattattacggggcatttgcgataggttacgatgtggcatttccacacatttgagatatacagggggggcgcgat 581
 138 *                                                                              138
 582 ctattatctatctacgacgattcagcatcgtacgatcattttatattatttaaattaaccgggtaataaagcggcgatcg 661
 662 agcgcacaaa                                                                     671
```

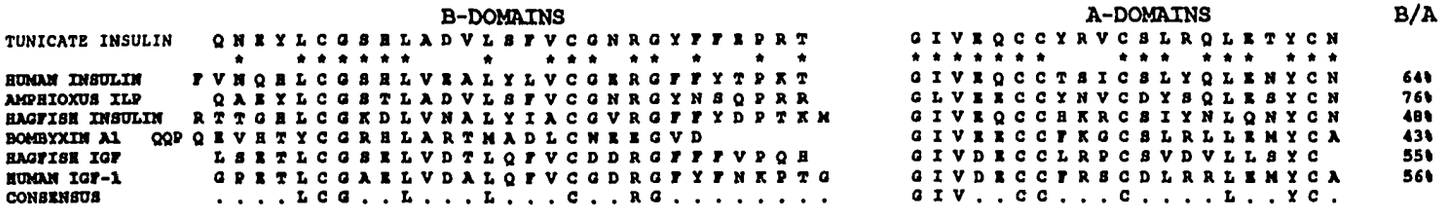


sites in the precursor protein. However, the residues of the tunicate insulin B and A domains have a high degree of identity to human insulin, hagfish insulin, and the amphioxus ILP (Figure 7.2A). A comparison of the deduced amino acids of mature tunicate and human insulin show 64% identity.

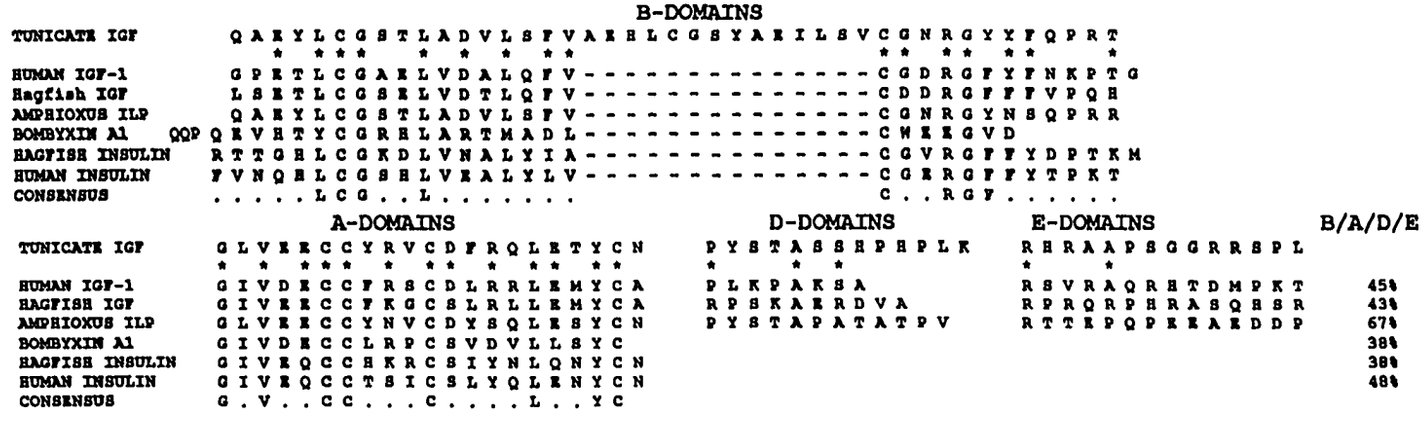
The other clone that was sequenced has structural and amino acid sequence characteristics (Figure 7.1B, 7.2B) consistent with that of vertebrate IGF-I and -II (Rinderknecht and Humber, 1978a, b; Jansen *et al.* 1983; Bell *et al.* 1984), and amphioxus ILP (Chan *et al.* 1990). The tunicate *igf* cDNA is larger than the *insulin* clone at 671bp, has dibasic processing sites flanking the C domain, but does not contain a stop codon at the end of the A domain, and therefore, codes for additional D and E domains. Like the tunicate *insulin* cDNA, the tunicate *igf* cDNA contains coding regions for B, C, and A domains. However, an unusual feature within the B domain is that tunicate *igf* has an additional region of 42bp. This sequence, which appears to be a repeat within the B domain, was not found in the tunicate *insulin* cDNA sequence. The number of deduced amino acids for the tunicate IGF D and E domains are 13 and 14 amino acids, respectively.

Figure 7.2: Alignment of tunicate insulin andIGF with several other family members. Tunicate insulin **(A)** and IGF **(B)** are aligned with amphioxus ILP, hagfish insulin and IGF, silkworm bombyxin A1, human IGF-I and human insulin. Identical amino acids between human and tunicate insulin and IGF are indicated by a *, and for better alignment of the two sequences a dotted line (---) was used to frame shift the clones.

A)


B-DOMAINS
A-DOMAINS
B/A
TUNICATE INSULIN
HUMAN INSULIN
AMPHIOXUS ILP
HAGFISH INSULIN
BOMBYXIN A1
HAGFISH IGF
HUMAN IGF-1
CONSENSUS
64%
76%
48%
43%
55%
56%

B)


B-DOMAINS
TUNICATE IGF
HUMAN IGF-1
Hagfish IGF
AMPHIOXUS ILP
BOMBYXIN A1
HAGFISH INSULIN
HUMAN INSULIN
CONSENSUS
A-DOMAINS
D-DOMAINS
E-DOMAINS
B/A/D/E
45%
43%
67%
38%
38%
48%

Tissue expression of tunicate *insulin* and *igf* mRNA

A sensitive PCR detection method for the presence of tunicate *insulin* and *igf* mRNAs in various tissues was developed. For each cDNA clone, primers 1 & 2 and 3 & 4 (Figure 7.1) were designed for the 5' and 3' untranslated regions. These regions were distinct and allowed the specific detection of either *insulin* or *igf* mRNA/cDNA. Reverse transcribed cDNA of various tunicate tissues was amplified with clone-specific primers and the results are shown in Figure 7.3. In Figure 7.3A tunicate *insulin* was amplified in the lanes containing the cDNA from neural gland, neural ganglion/dorsal strand, intestine, gonad/digestive gland, gonad only, heart, and positive control. No bands were detected in the lanes containing the cDNA of branchial basket, tunic, negative control, or the tunicate *igf* cDNA clone. In Figure 7.3B tunicate *igf* was amplified in the lanes containing the cDNA from the same tissues as above. Bands were not detected in the lanes containing the cDNA of branchial basket, tunic, negative control, or the tunicate insulin cDNA clone. The RNA preparation used in all DNA amplifications did not appear to be degraded as determined by a consistent smear on an ethidium bromide stained gel.

In addition, to investigate the presence of insulin and IGF mRNA expression within the tunicate neural ganglion I used insulin and IGF RNA probes to find insulin expression localized in cortical cells of the neural ganglion (Figure 7.5 and 7.6). The *in situ* evidence confirms the presence of tunicate insulin

mRNA in the cortical cells of the neural ganglion. Therefore, tunicate insulin mRNA expression in nervous tissue, digestive gland, and intestine confirms a more widespread expression in animals that evolved earlier than vertebrates.

Figure 7.3: Tunicate *insulin* and *igf* cDNA detected by a reverse transcriptase/PCR assay. **(A)** PCR reactions with tunicate insulin primers contained tissue cDNA as follows: neural gland (lane A), dorsal strand/neural ganglia (lane B), gonad (lane C), gonad/digestive gland (lane D), intestine (lane E), heart (lane F), tunic (lane G), branchial basket (lane H), tunicate *igf* cDNA clone (lane I), negative control (lane J), and *insulin* cDNA clone (positive control; lane K). **(B)** PCR reactions with tunicate *igf* primers contained tissue cDNA as follows: neural gland (lane A), dorsal strand/neural ganglia (lane B), gonad (lane C), gonad/digestive gland (lane D), intestine (lane E), heart (lane F), tunic (lane G), branchial basket (lane H), tunicate *insulin* cDNA clone (lane I), negative control (lane J), and *igf* cDNA clone (positive control; lane K).

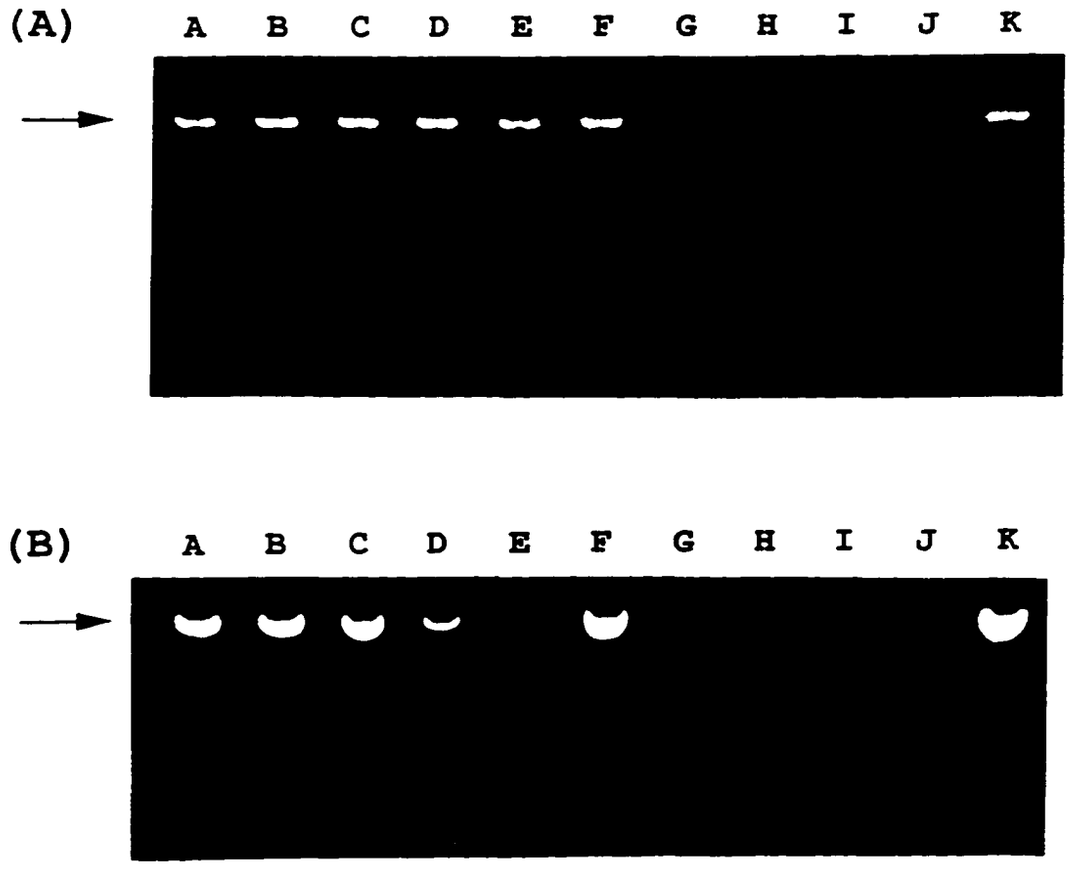

(A)
A B C D E F G H I J K
(B)
A B C D E F G H I J K

Figure 7.4: Sections (11μM) of tunicate (*Chelyosoma productum*) neural gland and ganglion stained with a hemotoxylin and eosin stain (magnification=40 times). The neural ganglion can be seen in the bottom half of the section. The ring of large cells around the periphery of the ganglion are overlaid with the blood sinus which contains a few blood cells. The neural gland (faint rust color) is shown by the strip of tissue at the top of the section.

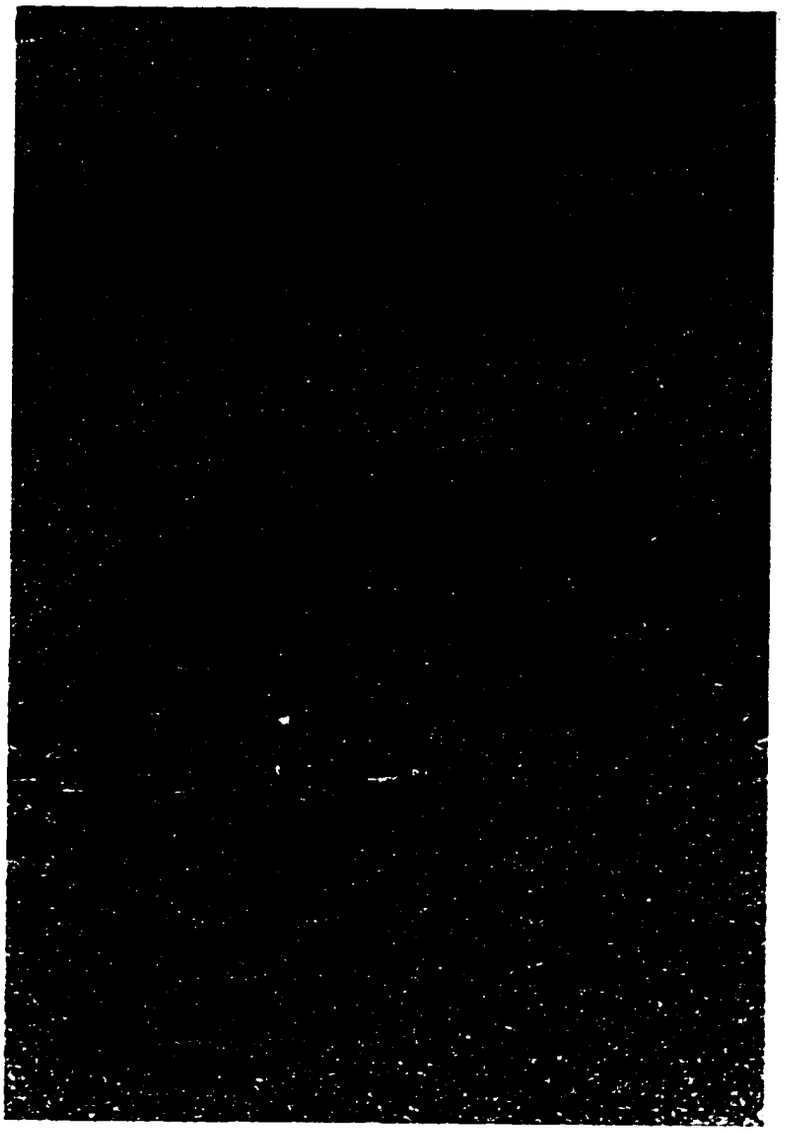
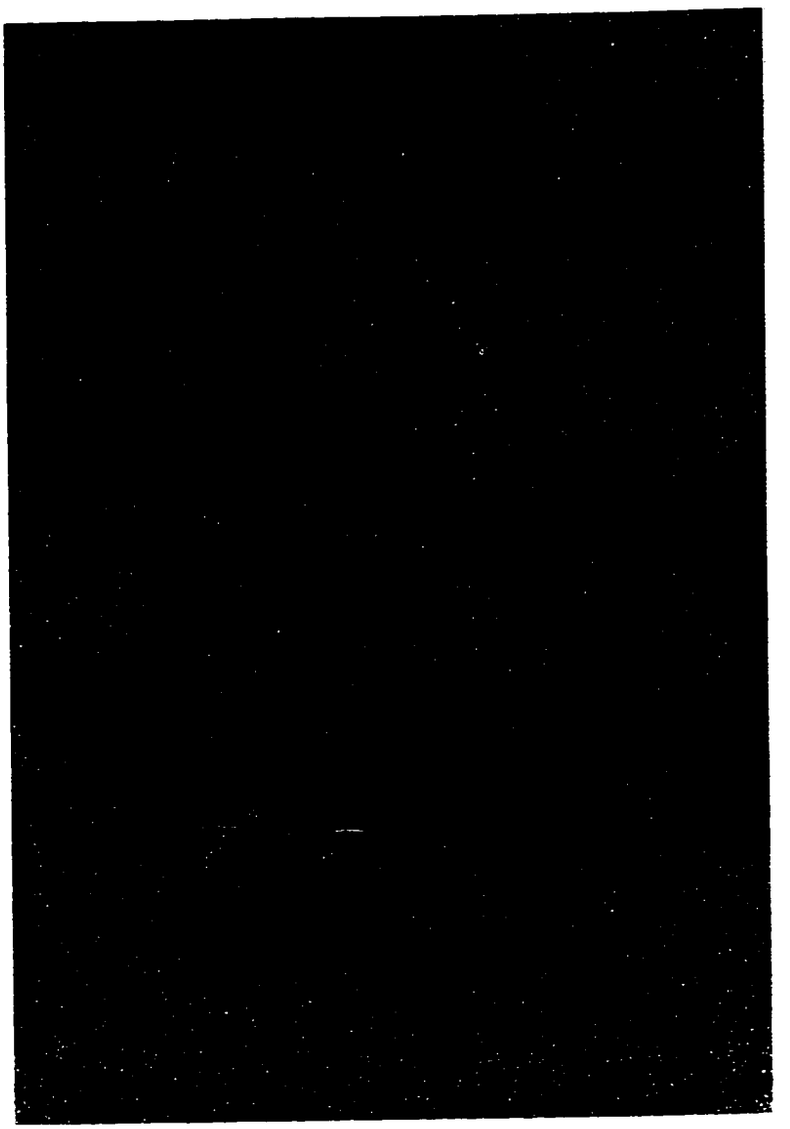


Figure 7.5: Localization of tunicate insulin mRNA by *in situ* hybridization. **(A)** *insulin* anti-sense mRNA (magnification=40 times) and **(B)** *insulin* sense mRNA (magnification=16 times) are shown in sections of the neural ganglion of *Chelyosoma productum* by *in situ* hybridization using a DIG-labelled RNA probe.

A



B

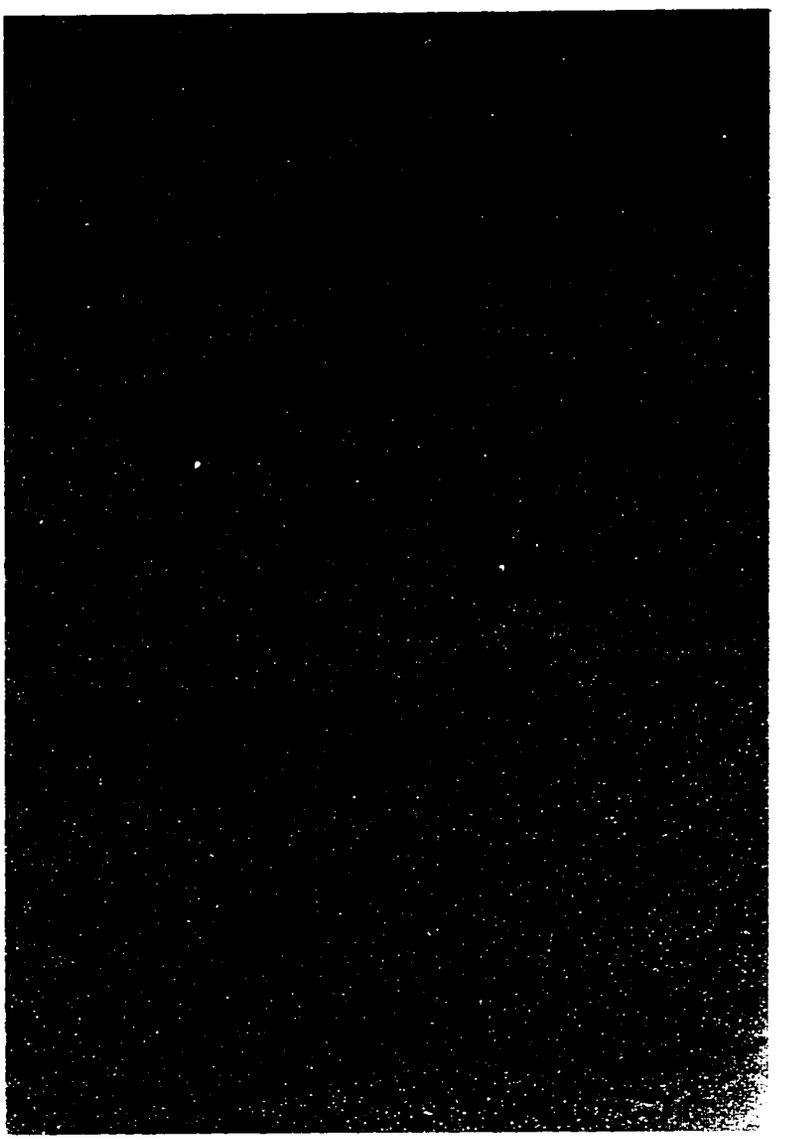
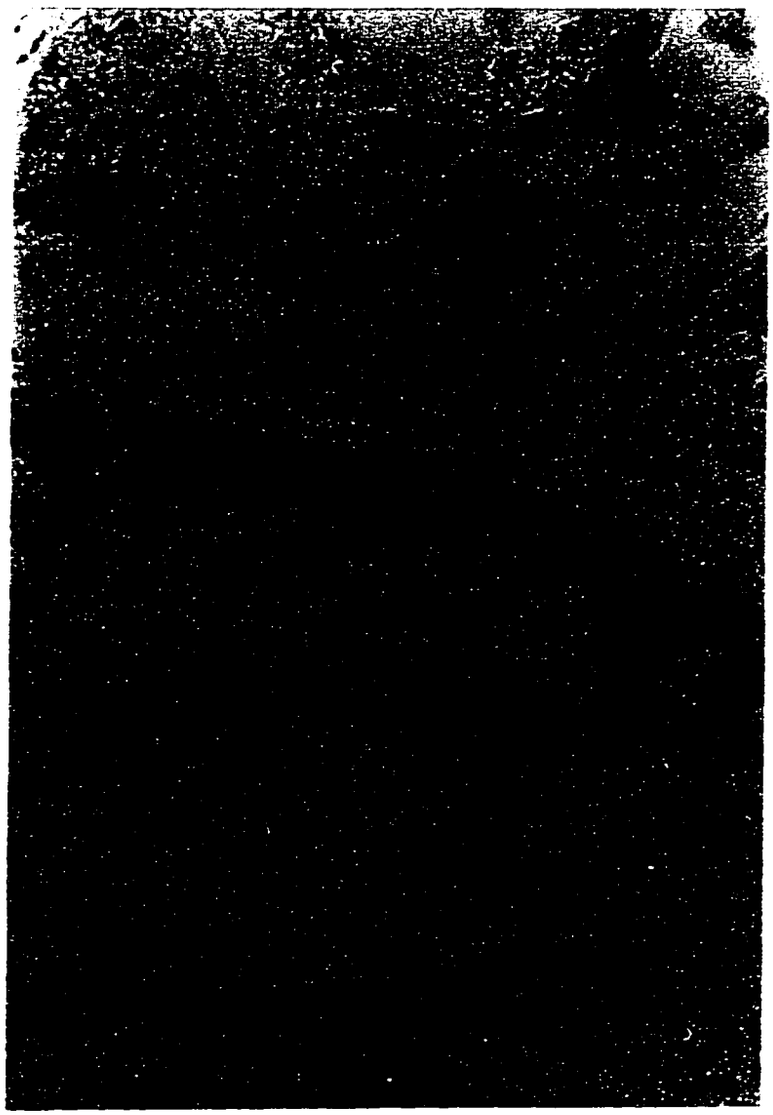
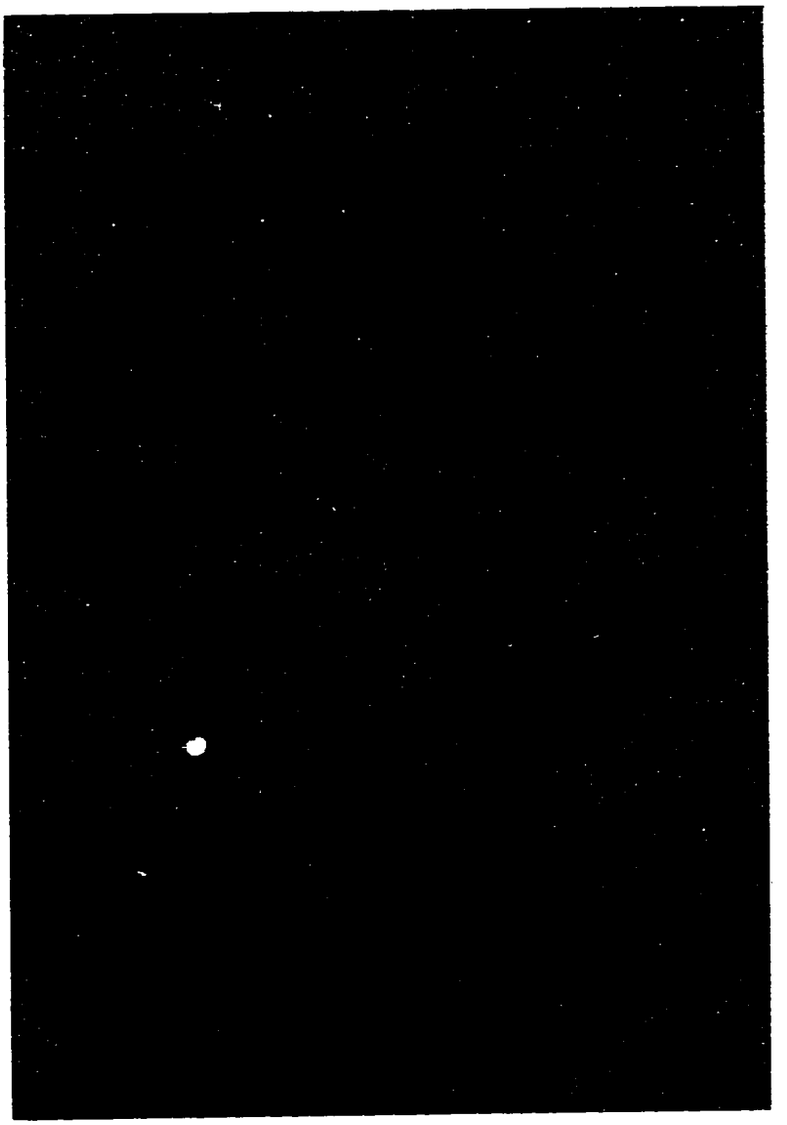


Figure 7.6: Localization of tunicate *igf* mRNA by *in situ* hybridization. **(A)** *igf* anti-sense mRNA (magnification=40 times) and **(B)** *igf* sense (magnification=40 times) mRNA are shown in sections of the neural ganglion of *Chelyosoma productum* by *in situ* hybridization.

A



B



Discussion

I report here the structure and tissue expression of two distinct tunicate cDNAs that have high sequence and structural identity to mammalian insulin and IGF-I and -II. In addition, the tunicate cDNAs encode precursors with the characteristic organization of other members of the insulin superfamily in that both tunicate cDNAs encode B, C, and A domains, and tunicate IGF encodes additional D and E domains. These common factors demonstrate that the two tunicate cDNAs are indeed members of this superfamily.

Tunicate insulin and IGF are structurally related to other family members.

The tunicate insulin amino acid sequence as deduced from the cDNA contains many of the residues that are essential for structure and function. The hydrophobic residues at positions B5, B10, B11, B14, B17, A60, A61, A71, A74 and A77 (Figure 7.1) are responsible for the hydrophobic core necessary for proper tertiary structure (Nagamatsu *et al.* 1991). Also conserved are the six cysteine residues at positions B6, B18, A64, A65, A69, and A78 essential for the disulfide linkages of the B and A domains in the mature peptide (Figure 7.1) (Nagamatsu *et al.* 1991). The 6 cysteine residues and the hydrophobic core are characteristics of the insulin superfamily. These characteristics are seen in other related peptides

including human and hagfish insulin, insect bombyxin, locust insulin-related peptide (LIRP) and the amphioxus ILP. In addition to the cysteine and hydrophobic residues, the tunicate insulin and IGF have, like other vertebrate insulins and IGFs, but not like the insect family members, residues thought to be responsible for receptor binding (Nakagawa and Tager, 1986, 1987; Cascieri *et al.* 1988; De Meyts *et al.* 1990). In vertebrates and tunicates, the aromatic residues in the B domain at positions 23, 24, 25 are always Tyr or Phe, although the order varies. Conservation of aromatic residues at positions 23-25 is seen for all vertebrate insulins and IGF peptides but not that of the amphioxus ILP (Chan *et al.* 1990), which has Tyr-Asn-Ser.

The tunicate *insulin* cDNA has all the amino acids necessary for proper secondary structure, receptor binding and disulfide links. In addition, tunicate insulin has a very high sequence identity to other known superfamily members. The deduced amino acids of the mature tunicate insulin have 74% identity with the B and A domains of the amphioxus ILP, 48% identity to mature hagfish insulin, and 44% to the insect peptide Bombyxin A1 (Figure 7.2) (Adachi *et al.* 1989). However, the mature tunicate insulin peptide has a high degree of identity to human insulin and IGF-I with 64% and 58% identity, respectively. Therefore, the tunicate insulin has higher identity with human insulin than hagfish insulin has with human insulin (51%). The high sequence identity of tunicate and mammalian insulin was predicted by Galloway and Cutfield (1988) who showed

immunoreactive insulin from the digestive tract of the tunicate *P. pachydermatina* using an antisera directed against porcine insulin. The lack of success in showing cross-reactivity using a hagfish insulin antiserum led them to predict correctly that the amino acid sequence identity of tunicate insulin would be closer to mammalian than hagfish insulin.

The tunicate *igf* cDNA encodes the correct amino acids for proper IGF-I and -II tertiary structure, receptor binding and disulfide links. Also, characteristic of other IGFs, the tunicate IGF contains D and E domains. However, an extra 14 amino acids are encoded within the B domain. The nucleotides and deduced 14 amino acids have a high degree of identity to 16 residues that immediately precede them (Figure 7.2). Positions 2-17 of the tunicate IGF B-domain match closely the repeat segment (residues 18-31), and probably arose by a duplication event.

```
2                 10              17
A E Y L C G S T L A D V L S F V
* *   * * * *     *     * *   *
A E H L C G S Y - A E I L S - V
18                              31
```

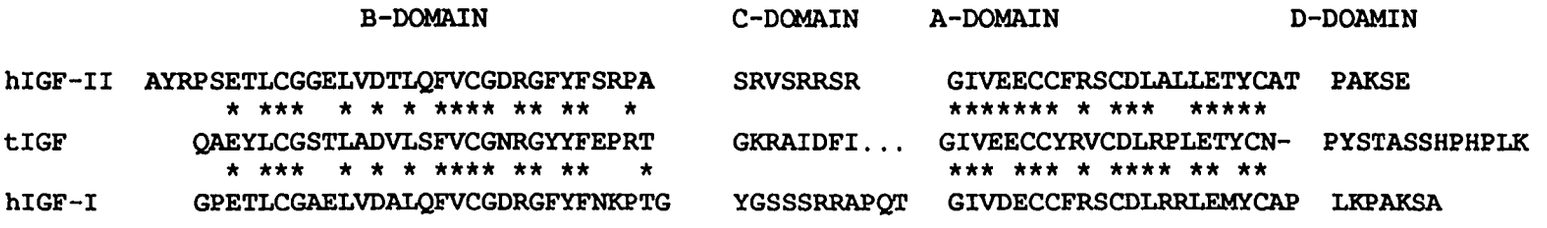
The addition of extra bases is not unique among this family. For example, one of the molluscan insulin related peptides (MIP-1) has additional base pairs within its C domain (Smit *et al.* 1991) and among family members it is not uncommon for the C domain to vary in length. However, this is the first report of extra bases resulting in an extended B domain; neither the

function of the extra coding region nor the effect of an extra cysteine within the B domain is known. The D and E domains of the tunicate IGF have a structure similar to that of mammalian IGF-I sequences and amphioxus ILP. The tunicate D domain of 13 amino acids is 1 residue longer than the amphioxus ILP D domain and 3 residues longer than the human IGF-I D domain. The deduced sequence of the tunicate E domain is shortened to only 14 amino acids and is 21 and 176 residues shorter than the respective human IGF-I and amphioxus ILP E domains. The E-domain's 14 amino acids are the least conserved region of the tunicate IGF precursor, which may reflect the removal of the E domain in the mature peptide. It is not known if this shortened E domain is normal for the tunicate or is the result of alternative intron-exon splicing as seen in vertebrate *igf-I* genes (Baskin *et al.* 1987; Lowe *et al.* 1988; Wallis and Devlin, 1993).

Due to the high degree of identity between the two tunicate cDNAs, the deduced tunicate IGF sequence also has a high degree of identity to other family members (Figures 7.2 and 7.7). A comparison of the deduced amino acids shows the tunicate IGF has the highest identity to amphioxus ILP with 73%. Some degree of identity to tunicate IGF is seen with other structurally related IGFs such as human and hagfish IGF, which are 55% and 48% identical, respectively.

Figure 7.7: Comparison of tunicate insulin and IGF to related molecules. **(A)** The amino acids identity is given as the number of identical amino acids to the total number of amino acids and as a percent. **(B)** Alignment of the deduced amino acids for human IGF-I and IGF-II in comparison to the tunicate (t) IGF. The tunicate sequence is located in the middle and the two human sequences on the top and bottom. Identical amino acids are indicated by a *.

	Tunicate insulin	Tunicate IGF
Amphioxus ILP	36/50 (72%)	45/62 (73%)
Hagfish insulin	24/50 (48%)	22/50 (44%)
Bombyxin A1	18/44 (43%)	13/44 (30%)
Hagfish IGF	30/50 (60%)	29/62 (48%)
Human IGF-I	28/50 (56%)	32/58 (55%)
Human IGF-II	29/50 (58%)	32/58 (55%)
Human insulin	31/50 (62%)	29/50 (58%)



Tunicate *insulin* and *igf* mRNA are expressed in neural ganglion and other tissues

To gain insight into the role of each of the two tunicate mRNAs, we prepared primers specific for each clone and studied tissue distribution of each mRNA. Tunicate *insulin* and *igf* mRNA expression was not restricted to a single organ as hagfish IGF expression was restricted to the liver. Rather, both mRNAs were expressed in all tissues studied, except the tunic and branchial basket. In separate reports, it was shown that an insulin-like peptide was expressed solely in the digestive tract (Galloway and Cuttfield, 1988) or in the neural ganglion (O'Neil *et al.* 1986) of two different tunicates using antisera directed to porcine and salmon insulins, respectively. However, these differences may be due to the species of tunicate, antibody, or form of insulin-like peptide in the studies. Also, reverse transcriptase-polymerase chain reaction (RT-PCR) is considerably more sensitive than immunological detection because the cDNA is amplified many times. Here for this thesis, *in situ* techniques were used with the same species of tunicate from which insulin cDNA was isolated. However, only the neural ganglion and gland have been studied to date.

The DIG-labeled RNA probes were produced in our lab and the sectioning and staining were performed by Lijuan Sun. Using the RNA probe specific to tunicate (*C. productum*) insulin and IGF, we found expression of both genes localized in cortical cells of the neural ganglion (Figure 7.5 and 7.6). However, neither

insulin nor IGF containing cells were clearly detected in the neural gland using the RNA probe. This lack of staining within the neural gland may have been due to low levels of mRNA not detectable with *in situ* techniques. However, the positive results with the PCR may be due the sensitivity of the PCR or conversely due to a piece of neural ganglion contaminating the neural gland. Therefore, the *in situ* evidence confirms the presence of insulin in the neural ganglion. The widespread detection of insulin-like peptides in previous studies is confirmed with our evidence that tunicate insulin and IGF are expressed in the tunicate neural ganglion, neural gland, intestine, gonad/digestive gland, gonad only and heart. Previous reports identified insulin in pancreas and brain of vertebrates (Baskin *et al.* 1987), intestine of hagfish (Peterson *et al.* 1975) and the brain of insects (Adachi *et al.* 1989) and molluscs (Smit *et al.* 1988). Therefore, tunicate *insulin* mRNA expression in nervous tissue, digestive gland, and intestine confirms a more widespread expression in animals that evolved earlier than vertebrates. The expression in gonad and heart is interesting and may reflect the common origin of insulin and IGF because the latter is known to be expressed in all the above tissues including the gonad and heart (Duguay *et al.* 1992).

The origin of separate insulin and IGF molecules precedes tunicates

The expression of distinct forms of tunicate insulin and IGF confirms that insulin is an ancient peptide whose origin is still unknown. It is clear that the tunicate contains a molecule that is a true insulin in terms of structure and, therefore, the most ancient form known to date. The tunicate IGF, which is also very similar in nucleotide sequence to the amphioxus ILP, fits Chan's proposed evolutionary pathway for IGF-I and II (Chan *et al.* 1990). However, the presence of a true insulin within the tunicate shows that the amphioxus ILP gene is not the progenitor to both insulin and IGF. Rather, the high sequence and structural identity of insulin and IGF suggest that an ancestral parent molecule existed. However, in tunicates the two cDNA's are likely to have arisen from a gene duplication of an IGF-insulin progenitor preceding the protochordates.

References

Adachi T, Takiya S, Suzuki Y, Iwami M, Kawakamis A, Takahashi SY, Ishizaki H, Nagasawa H, Suzuki A 1989 cDNA structure and expression of bombyxin, an insulin-like brain secretory peptide of the silkworm *Bombyx mori* Journal of Biological Chemistry 264: 7681-7685.

Adham IM, Burkhardt E, Benahmed M, Engel W 1993 Cloning of a cDNA for a novel insulin-like peptide of the testicular Leydig cell. Journal of Biological Chemistry. 268: 26668-26672

Baskin DG, Figlewicz DP, Woods SC, Porte D, Dorsa D 1987 Insulin in the brain. Annual Reviews of Physiology 49: 335-347

Bell G I, Merryweather J P, Sanchez-Pescador R, Stempien M M, Priestly L, Scott J, Rall L B 1984 Sequence of a cDNA clone encoding human preproinsulin-like growth factor II. Nature 310: 775-777.

Bevis PJR, Thorndyke MC 1978 Endocrine cells in the oesophagus of the Ascidian *Styela clava*, a cytochemical and immunofluorescence study. Cell and Tissue Research 187: 153-158.

Birnboim M. 1983 A rapid alkaline extraction method for the isolation of plasmid DNA. Methods in Enzymology 100: 243-255.

Cascieri MA, Chicchi GG, Applebaum J, Green BG, Bayne ML 1988 Mutants of human insulin-like growth factor-I with reduced

affinity for the type I insulin-like growth factor receptor. Biochemistry 27: 3229-3233.

Chan SJ, Cao Q, Steiner DF 1990 Evolution of the insulin superfamily: Cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. Proceedings of the National Academy of Science (USA) 87: 9319-9323,

Chan SJ, Noyes BE, AgarwaL KL, Steiner DF 1979 Construction and selection of recombinant plasmids containing full length cDNAs corresponding to rat insulins I and II. Proceedings of the National Academy of Science (USA) 76: 5036-5040.

Chomczyski P, Sacchi N 1987 Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162: 156-159.

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

De Meyts P, Gu J-L, Shymko RM, Kaplan BE, Bell GI, Whittaker 1990 Identification of a ligand-binding region of the human insulin receptor encoded by the second exon of the gene. Molecular Endocrinology 4: 409-416.

Duguay SJ, Park LK, Samadpour M, Dickhoff WW 1992 Nucleotide sequence and tissue distribution of three insulin-like growth factor I prohormones in salmon. Molecular Endocrinology 6: 1202-1210

Ebberink RHM, Smit AB, Van Mimmen J 1989 The insulin family: evolution of structure and function in vertebrates and invertebrates. Biological Bulletin 177: 176-182.

Froesch ER, Schmidt C, Schwander J, Zapf J 1985 Actions of insulin-like growth factors. Annual Reviews in Physiology 47: 443-467.

Galloway SM, Cuttfield JF 1988 Insulin-like material from the digestive tract of the tunicate *Pyura pachydermatina* (sea tulip). General and Comparative Endocrinology 6: 106-113.

Jansen M, van Schaik FMA, Ricker AT, Bullock B, Woods DE, Gabbay KH, Nussbaum AL, Sussenbach JS, Van den Brande JL 1983 Sequence of cDNA encoding human insulin-like growth factor I precursor. Nature 306: 609-611.

Kavson VM, Grebenjuk VA, Koval AP, Skorokhod AS, Roberts Jr. CT, LeRoith D 1994 Structure of the chum salmon insulin-like growth factor I gene. DNA and Cell Biology 13: 555-559.

Lowe WL, Lasky SR, LeRoith D, Roberts CT 1988 Distribution and regulation of rat insulin-like growth factor I messenger ribonucleic acids encoding alternative carboxyterminal E-peptide: evidence for differential processing and regulation in liver. Molecular Endocrinology 2: 528-535.

McRory JM, Sherwood NM 1994 Catfish express two forms of insulin-like growth factors-I (IGF-I) in the brain. Journal of Biological Chemistry 269: 18588-18592.

Nagamatsu S, Chan SJ, Falkmer S, Steiner DF 1991 Evolution of the insulin superfamily. Journal of Biological Chemistry 266: 2397-2402.

Nakagawa SW, Tager HS 1986 Role of the phenylalanine B25 side chain in directing insulin interaction with its receptor: steric and conformational effects. Journal of Biological Chemistry 261: 7332-7341.

Nakagawa SW, Tager HS 1987 Role of the COOH-terminal B-chain domain in insulin-receptor interactions: indentification of pertubations involving the insulin main chain. Journal of Biological Chemistry 262: 12054-12508.

O'Neil GS, Falkmer S, Thorndyke MC 1986 Insulin-like immunoreactivity in the neural ganglion of the ascidian *Ciona intestinalis.* Acta Zoology (Stockholm) 67: 147-153.

Perfetti R, Scott LA, Shuldiner AR 1994 The two nonallelic insulin-like growth factor-I genes in *Xenopus laevis* are differentially regulated during development. Endocrinology 135: 2037-2044.

Peterson JD, Steiner DF, Emdin SO, Falkmer S 1975 The amino acid sequence of the insulin from a vertebrate, the Atlantic hagfish (*Myxine glutinosa*). Journal of Biological Chemistry 250: 5183-5191.

Plisetskaya E, Kazakov VK, Soltitskaya L and Leibson LG 1978 Insulin-producing cells in the gut of freshwater mollusks

Andota cygnea and *Unio pictorum* and the role of insulin in the regulation of their carbohydrate metabolism. General and Comparative Endocrinology 35: 133-145.

Rinderknecht E, Humber RE 1978a The amino acid sequence of human insulin-like growth factor I and its structural homology with proinsulin. Journal of Biological Chemistry 253: 2769-2874

Rinderknecht E, Humber RE 1978b Primary structure of human insulin-like growth factor II. FEBS Letters 89: 283-287.

Sanger F, Nicklen S, Coulson AR 1977 DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science (USA). 74: 5463-5467.

Shamblott MJ, Chen TT 1992 Identification of a second insulin-like growth factor in a fish species. Proceedings of the National Academy of Science (USA). 89: 8913-8917.

Smit AB, Geraerts WPM, Meester I, van Heerikhuizen H, Joosse J 1991 Characterization of a cDNA clone encoding molluscan insulin-related II of *Lymnaea stagnalis*. European Journal of Biochemistry 199: 699-703.

Smit AB, Vreugdenhil E, Ebberink RHM, Geraerts WPM, Klootwijk J, Joosse J. 1988 Growth-controlling mulluscan neurons produce the precursor of an insulin-related peptide. Nature 331: 535-538.

Steiner DF, Chan SJ, Docherty K, Emdin SO, Dodson GG, Falkmer S 1984 Evolution of polypeptide hormones and their precursor processing mechanisms. In: Evolution and

tumour pathology of the neuroendocrine system. Elsevier, New York, pp 203-223

Thorndyke MC, Goldsworthy GJ 1988 Functional aspects of peptide neurohormones in protochordates. In: Neurohormones in Invertebrates. Cambridge University Press pp. 236-258

Wallis AE, Devlin RH Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. 1993 Molecular Endocrinology 7: 409-422.

Chapter 8
General discussion

The theory of molecular evolution indicates that in the majority of cases, a domain duplication at the protein level implies an exon duplication at the DNA level. It has been suggested that exon duplication is one of the most important types of internal duplication. Eukaryotic genes generally consist of many exons and introns, and the neighboring exons may be at the very least, similar in nucleotide sequence to one another. These facts suggest that related genes in extant organisms have evolved by internal duplication and subsequent modifications of primordial genes that contained one or few exons. In addition to the increase in the number of exons within a specific gene, the evolution of a gene family is initiated by the duplication of a gene or genome to produce two different genes. Different mechanisms exist which may result in the production of two new genes. The different mechanisms are usually classified according to the extent of the genomic region involved. The following types of duplication are recognized: (1) partial or internal gene duplication, (2) complete gene duplication, (3) partial chromosomal duplication, (4) aneuploidy (chromosomal duplication), and (5) polyploidy (genome duplication).

DNA duplication events have long been recognized as an important means in the evolution of increasing the genomic

size. In particular, the duplication of the entire genome or a major portion of it, such as a chromosome, may result in a sudden substantial increase in the size of the genome. Genomic duplication events have been reported throughout the evolution of different groups of organisms. Duplication events (polyploidy) have occurred in notable groups such as some plants, bony fish and amphibians in which varying degrees of polyploidy have occurred to increase the genomic size many fold. After the duplication of exons or genes, three evolutionary paths exist (1) both genes encode the same protein and do not change, (2) one gene maintains its original coding for a protein, however, the other one mutates and evolves into a new but related peptide or (3) both mutate and become two new genes.

Evolution of the glucagon superfamily of peptides

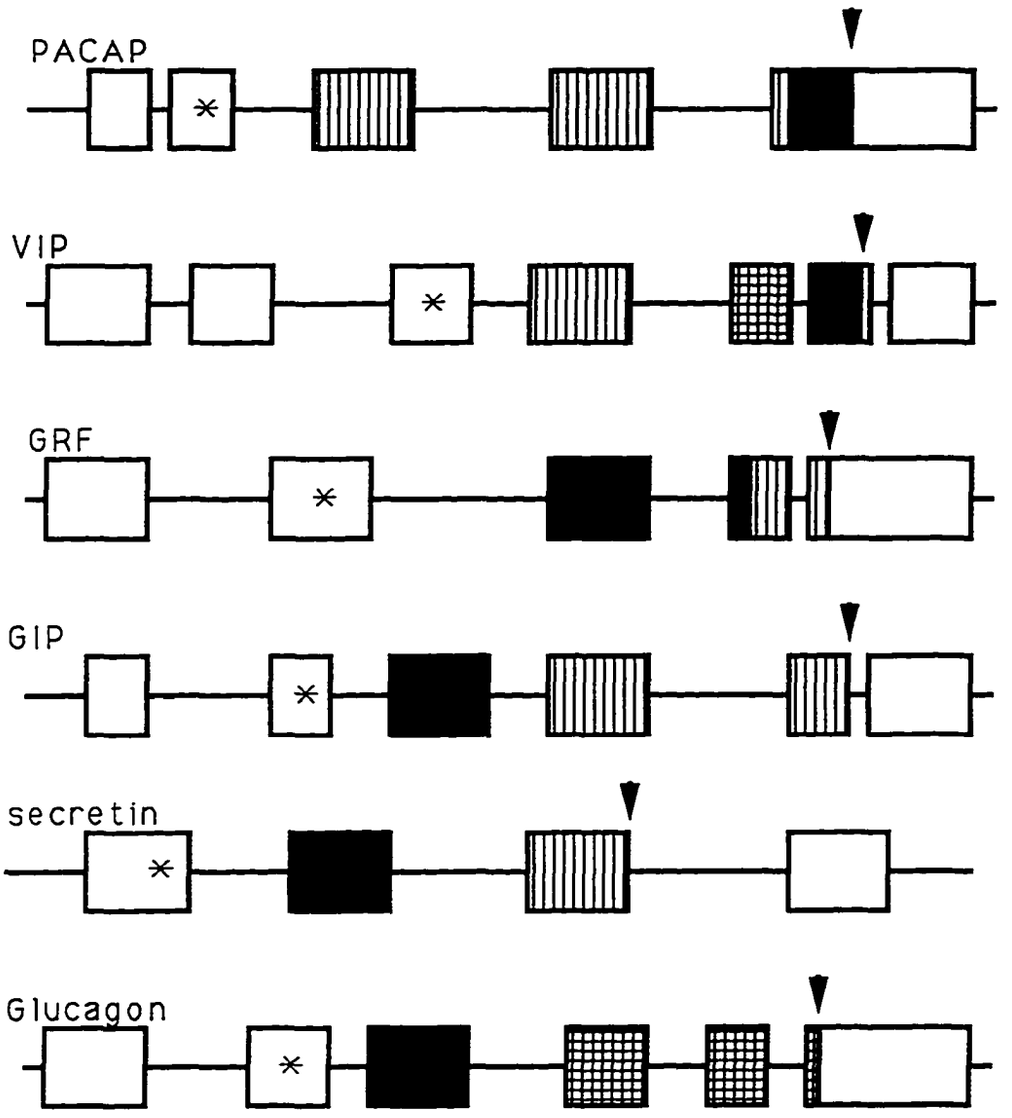
In the first portion of this thesis, *grf/pacap* cDNAs and genes were isolated and sequenced from a bird (chicken), fish (catfish) and a protochordate (tunicate). It is commonly thought that the glucagon superfamily arose from a single gene. Exon and gene duplications of this gene followed by nucleotide substitutions resulted in multiple distinct genes that encode mammalian forms of VIP, PHM, secretin, GIP, GRF, glucagon, glucagon-like peptide-I, glucagon-like peptide-II and PACAP. The *grf/pacap* gene from the tunicate, catfish and chicken span over 600 million years of vertebrate evolution, and hence,

provide an excellent opportunity to demonstrate the evolution of the glucagon superfamily.

The glucagon superfamily, as it is observed in mammals, is composed of nine members that are encoded by genes containing a similar intron/exon structure, sequence and length (Figure 8.1). Within all genes of the superfamily, the first exon encodes a 5'-untranslated region and the second exon encodes the signal peptide with the exception of the VIP gene, which has 2 exons for the 5'UTR (Fig. 8.1). As dictated by function, the length of the signal peptide is usualyy limited to about 20 amino acids (60bp), although for several genes in the glucagon family the signal peptide may be about 40 amino acids.

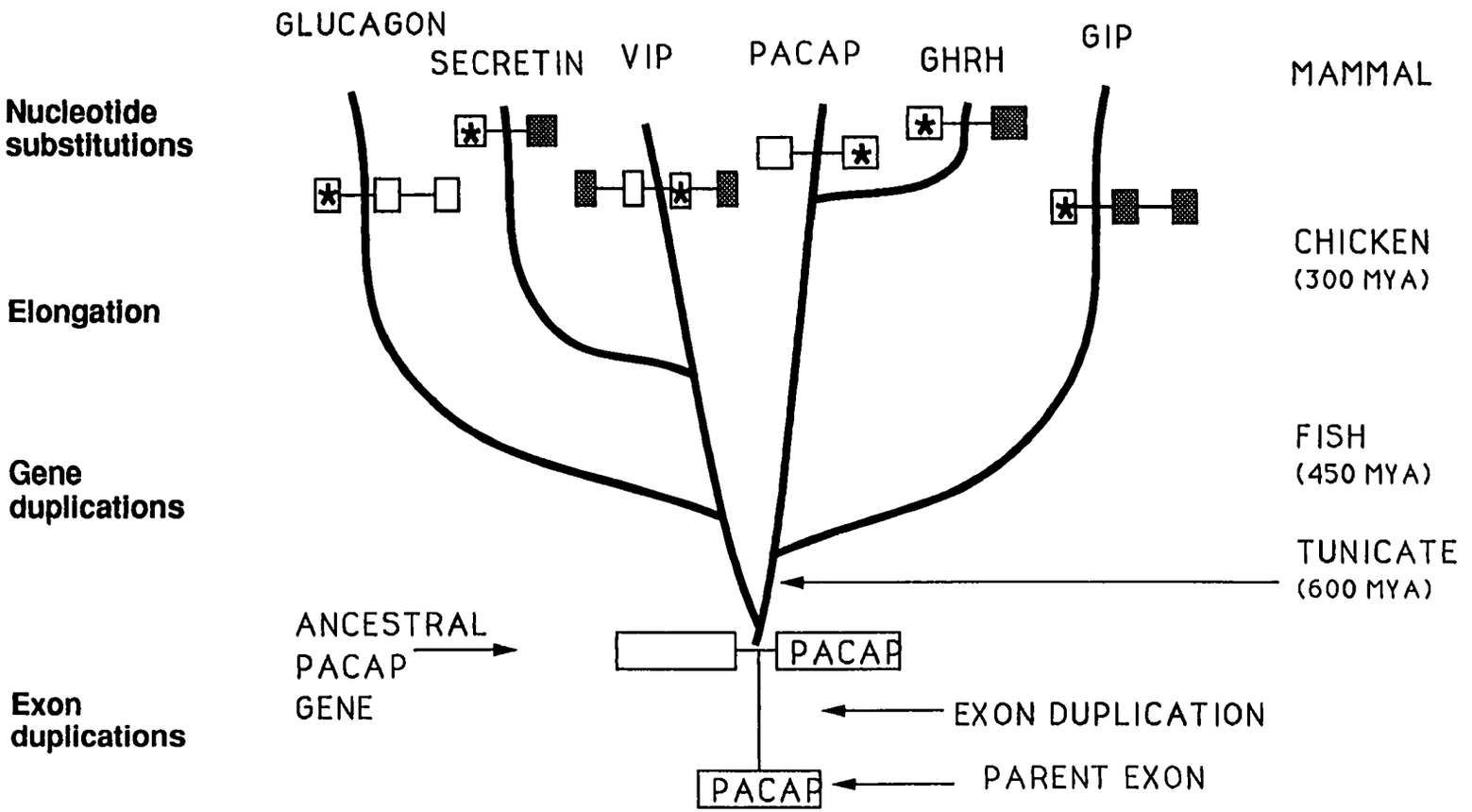
The third exon represents the most noticeable difference between family members (Figure 8.1). Only the PACAP and VIP genes contain a complete exon that encodes a cryptic peptide although the GRF and GIP genes encode cryptic peptides after the bioactive regions. It is interesting only the VIP and PACAP genes contains an exon encoding a cryptic peptide that is upstram to the exon encoding the bioactive molecule because the tunicate data suggests these two ancestral genes were the genes from which the extant glucagon superfamily evolved. Gene duplications of the ancestral gene appear to have resulted in new members that lack exon 3 encoding the cryptic peptide.

Figure 8.1: Schematic diagram of seven human glucagon superfamily genes. The exons are shown by boxes and the introns are shown by lines. The bioactive peptide products of each gene are shown by a black box or by a hatched box if a second active product is encoded. The cryptic peptide exons of the VIP and PACAP gene are shown by the box with the vertical lines. The 5' and 3'UTR exons are white boxes and the signal peptide exons have a star in the box.


PACAP
VIP
GRF
GIP
secretin
Glucagon

Not all family members have been isolated from nonmammalian vertebrates. For example, a *secretin* gene and/or protein has been isolated only from birds and mammals and a GIP has been found exclusively in mammals. Also, a distinct gene that specifically encodes GRF has only been found in mammals. As shown in Chapters 2-5, GRF is encoded on the same gene as PACAP in nonmammalian species. We hypothesize that the *grf/pacap* gene duplicated and gave rise to the two separate genes after the split between the avian/reptilian and mammalian lineages. Therefore, the largest number of members for the glucagon family is within the mammals. The number of superfamily members is thought to be decreased in tunicates because the duplication events that created the new members did not occur in the tunicate lineage (Figure 8.2). A note of caution, however, is that a tunicate secretin, GIP or other novel members related to the glucagon superfamily may not have been detected by my probe. In tunicates, only two genes in the superfamily were detected; the high degree of identity between them suggests that gene duplication was more recent than exon duplication events. These data suggest the ancestral gene that gave rise to the glucagon superfamily preceded the protochordates to a point in time much earlier than previously thought.

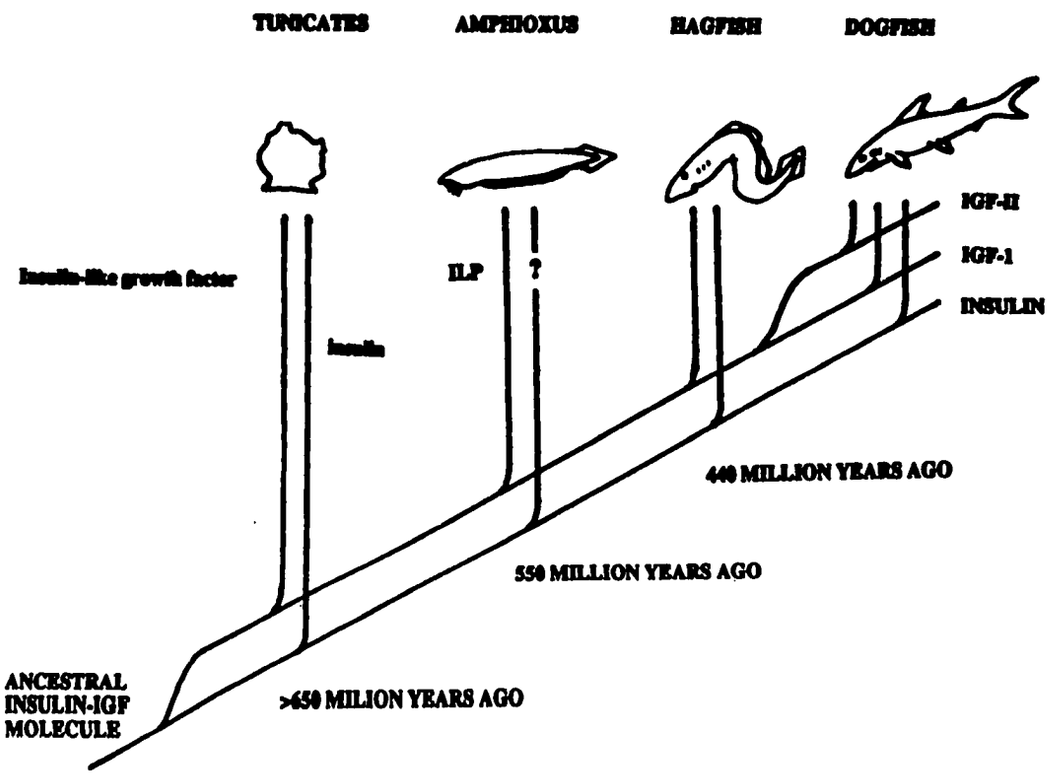
Figure 8.2: Proposed evolutionary pathway for members of the glucagon superfamily. The diagram is based on the presence of genes and proteins isolated from different vertebrates and invertebrates (tunicates). The presence of the highly conserved $PACAP_{1-27}$ within the tunicate suggests that it was the gene from which all other members of the family originated. A box with a * inside indicates the exon encoding the biologically active peptide, a white box indicates a second or third biologically active peptide in the gene and the shaded box indicates a cryptic peptide


GLUCAGON
SECRETIN
VIP
PACAP
GHRH
GIP
MAMMAL
Nucleotide substitutions
CHICKEN
(300 MYA)
Elongation
FISH
(450 MYA)
Gene duplications
TUNICATE
(600 MYA)
ANCESTRAL PACAP GENE
PACAP
Exon duplications
EXON DUPLICATION
PACAP
PARENT EXON

Evolution of the insulin superfamily of peptides

The insulin and insulin-like growth factor-I (IGF-I) evolutionary story shows a similar pattern to that of the glucagon superfamily although the gene duplication occurred at a different time. It was observed in Chapter 7 that the tunicates encode for two members of the insulin superfamily. These two members are 87% identical in nucleotide sequence and have identical expression patterns. It is apparent that the two tunicate insulin family members evolved from the same progenitor from which insulin, IGF-I and IGF-II originated. With the isolation of a true insulin and IGF from the tunicate, the amphioxus insulin-like peptide is most likely an IGF and not the progenitor of the insulin family as previously thought (Chan *et al.* 1990). A more complete picture is now available to demonstrate how the insulin/IGF family has evolved from protochordates to mammals. In Figure 8.3, the tunicate is shown to contain two peptides (insulin and IGF) which are the result of a recent gene duplication. However, it is not known if either the insulin or the IGF was the original gene that gave rise to the tunicate members.

Figure 8.3: Proposed evolutionary pathway for insulin and insulin-like growth factors. Distinct tunicate insulin and IGF clones suggest that the insulin/IGF ancestor precedes the tunicates. The question mark in the amphioxus insulin line indicates that insulin has not been found to date.


TUNICATES
AMPHIOXUS
HAGFISH
DOGFISH
IGF-II
IGF-1
INSULIN
Insulin-like growth factor
ILP
?
Insulin
440 MILLION YEARS AGO
550 MILLION YEARS AGO
ANCESTRAL
INSULIN-IGF
MOLECULE
>650 MILION YEARS AGO

Therefore, the two different tunicate cDNAs are products from a point in evolution in which a gene duplication event produced a new gene. Due to the restriction that one gene product is needed for its classical function, a slower evolutionary rate for this gene is predicted. However, the other gene is speculated to have the opportunity to undergo base pair changes and evolve into a new but related gene.

The insulin family has undergone several changes to encompass what we now refer to as the insulin superfamily. It is apparent a gene duplication of the ancestral gene gave rise to insulin and IGF. The IGF lineage has undergone several duplications to encode IGF-I and IGF-II. In addition, in Chapter 6, two different IGF-Is were presented to indicate, at least in catfish, another gene duplication had occurred to encode brain-specific and ubiquitous IGF-I. These events that led to the development of the insulin superfamily are in contrast to those of the glucagon superfamily. The development of the insulin superfamily appears to have occurred solely by gene duplication and not exon duplication or elongation of exons. Unlike the exon and gene duplications that gave rise to the nine different members of the glucagon superfamily, the insulin superfamily members originated from many fewer events.

Future directions

With the isolation and sequencing of cDNAs for two members of two superfamilies from an invertebrate, it is obvious that the origin of both families precedes the vertebrates. This idea of such an early origin of peptide superfamilies is very exciting. Utilizing the tunicate cDNAs as probes, we can examine organisms that evolved before the tunicates to determine whether both families can be detected. The most logical first step is to investigate the presence of PACAP within organisms that are phylogenetically older than tunicates. In the zoo blot of Chapter 5, a hybridization band was observed in the lane that contained the sea urchin DNA suggesting the presence of *pacap* gene in the urchin. However, as in the tunicate, there are two bands within the urchin DNA suggesting the urchins also contain a *grf/pacap* and *glucagon/vip* gene. A logical second step would be to investigate the presence of PACAP and/or VIP in organisms that are more ancient than the echinoderms. Using the argument of exon and gene duplication, I think that phylogenetically older organism will be found with only one gene encoding a PACAP-like molecule. This idea is supported by the high identity between the PACAP and VIP exons and genes in tunicates. The data imply that the order of duplication was first the exon followed by the gene. Therefore, I speculate that at point earlier than protochordates in evolution, a single gene encoding a PACAP-like exon will be found.

The evolutionary position of the tunicate glucagon superfamily is shown in Figure 8.2. However, the events which occur before the tunicate appeared in evolution are strictly hypothetical. The events that are shown in the vertebrate lineage are based on available sequence data. To find an organism that contains a single PACAP exon without exons encoding related family members, a probe of the tunicate PACAP cDNA could be used. The main problem of isolating sequences from these phylogenetically ancient organisms is the degeneracy of the nucleotides and hence the inability of primers or probes to bind to the target DNA.

PACAP and GRF as growth factors

The other direction in which this research may proceed is to determine the effects of GRF and PACAP together or alone. It is well established that GRF is an effective releaser of GH in mammalian systems. However, the function of PACAP has been shown tentatively to be 1) a releaser of anterior pituitary hormones, 2) a putative posterior pituitary hormone, 3) a factor that may have a role in the development and maintenance of various cells within embryos. In addition to PACAP's potential role as a releaser of pituitary hormones, preliminary evidence shows that PACAP promotes neuroblast proliferation and neurite outgrowth in the cerebral cortex, cerebellar granule cells and sympathetic ganglia of fetal mice. Further evidence of PACAP's role as a potential growth factor is based on the

following: PACAP promotes proliferation of mouse primordial germ cells; PACAP is produced in specific tumour cells; PACAP receptors are detected on human glial cell tumours; and PACAP is expressed early in rat brain development with levels peaking at birth. Meanwhile, mammalian GRF has also been shown to affect somatotroph differentiation in the embryonic chicken pituitary and to accelerate the development of chick neuroblasts and synthesis of neurotransmitters. In addition, the evidence presented in Chapter 4 shows that the chicken *grf/pacap* gene is expressed very early in development (day 5), a time when neither the pituitary nor the portal system is developed. This early expression is also seen in the mammal in which the two independent genes are expressed prior to the development of their classical target, the pituitary. Within the rodents, PACAP mRNA has been shown to be expressed starting at embryonic day 9.5 and protein is detected on day E14 with protein levels peaking just prior to birth. Within the mammalian system only GRF has a defined role in the release of growth hormone. In contrast to this, PACAP's role as a releaser of pituitary hormones may be only one function in a network of functions. In addition, GRF has been shown to enhance the differentiation of pituitary somatotrophs and PACAP has been shown to promote the proliferation of primordial germ cells. Therefore, another role of these neuropeptides may be as a growth or differentiating factor within the nervous system.

In the human fetus a developmental disorder has been identified as holocephaly. The symptoms are defined as the lack of a forebrain, and nonfunctional eyes and ears resulting in death within hours after birth. In this disorder, the defect is that the short arm of chromosome 18 is missing. Further investigations have localized the chromosome defect to a finite region of the short arm of chromosome 18-PT1100, a region that also encodes PACAP, at least in humans. It is extremely speculative but the lack of PACAP by itself or in combination with another protein also encoded on the short arm of chromosome 28 may result in this disorder. This proposed additional role in brain development for the two peptides is exciting in contrast to the original idea that the only role of these neuropeptides is to release pituitary hormones.

A technique that could provide evidence of PACAP's role is the preparation of a transgenic mouse in which the native *pacap* gene is replaced with a nonfunctional gene. The "knock-out" mouse is produced using a gene encoding a mutated form of PACAP. The altered gene, injected into the dividing cells, results in the exchange of the native PACAP gene with the mutated one. Any physical abnormalities in the mouse can then be attributed to the lack of PACAP. Another method to determine the role of PACAP is to destroy the cells that express PACAP. A construct that contains a truncated diphtheria toxin cDNA driven by the PACAP gene promoter will express the toxin only in cells that normally transcribe the PACAP gene. The

truncated toxin is about 30 times less potent than the normal form, but is sufficient to cause the death of the PACAP transcribing cell. This cell-specific ablation using the truncated diphtheria toxin would ablate all cells that normally produce PACAP and, in turn, prevent all native PACAP expression. The physiological effects can be attributed to the lack of PACAP. These defects created by the ablation of the PACAP synthesizing cells would create a physiological scenario similar to the knockout mouse.

Finally, the physiological effects of PACAP and GRF can be studied by overexpressing the peptide in a transgenic animal, or testing the peptide in vivo or in vitro in cell culture. The underlying question is whether the structural changes in evolution have resulted in functional changes.