Expression, Regulation and Evolution of Proglucagon Genes in Vertebrates

by

Ellen Rain Bnsby

B.Sc., University of Victoria, 1995

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

Supervisory Committee:

Dr. T.P. Mommsen, Supervisor (Department of Biochemistry and Microbiology)

Dr. R.W. Olafson, Co-supervisor (Department of Biochemistry and Microbiology)

Dr. W.W. Kay, Departmental Member (Department of Biochemistry and Microbiology)

Dr. T.W. Pearson, Departmental Member (Department of Biochemistry and Microbiology)

Dr. N.M. Sherwood, Outside Member (Department of Biology)

Dr. J.T. Silverstein, External Examiner (Genetics, National Center for Cool and Cold Water Aquaculture, United States Department of Agriculture – Agriculture Research Services)

© Ellen Rain Busby, 2002
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Expression, Regulation and Evolution of Proglucagon Genes in Vertebrates

by

Ellen Rain Busby
B.Sc., University of Victoria, 1995

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

We accept this dissertation as conforming to the required standard

Dr. T.P. Mommsen, Supervisor (Department of Biochemistry and Microbiology)

Dr. R.W. Olafson, Co-supervisor (Department of Biochemistry and Microbiology)

Dr. W.W. Kay, Departmental Member (Department of Biochemistry and Microbiology)

Dr. T.W. Pearson, Departmental Member (Department of Biochemistry and Microbiology)

Dr. N.M. Sherwood, Outside Member (Department of Biology)

Dr. J.T. Silverstein, External Examiner (Genetics, National Center for Cool and Cold Water Aquaculture, United States Department of Agriculture - Agriculture Research Services)

© Ellen Rain Busby, 2002
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Supervisors: Dr. Thomas P. Mommsen and Dr. Robert W. Olafson

ABSTRACT

Expression, Regulation and Evolution of Proglucagon Genes in Vertebrates

Three biologically active peptide hormones, glucagon, glucagon-like peptide (GLP)-1, and GLP-2, are co-encoded by the precursor proglucagon. As all three peptides have distinct functions, regulation of proglucagon expression, translation, and processing is necessary. In all vertebrates, glucagon release from the pancreatic α-cells, leads to an increase in circulating glucose levels through liberation of glucose from the liver. Unlike glucagon, a drastic change in GLP-1 function occurs between mammals and bony fishes. In mammals, GLP-1 acts as an incretin hormone, stimulating glucose-dependent production and release of insulin from pancreatic β-cells. In fish, like glucagon, GLP-1 stimulates the hepatic release of glucose. GLP-1 may also play a role in regulation of food intake in mammals and teleosts. To date, GLP-2 function has only been determined in mammals where it acts as an intestinal growth factor.

I studied two teleost species, copper rockfish (*Sebastes caurinus*) and channel catfish (*Ictalurus punctatus*), and detected multiple forms of proglucagon regulation. First, I identified two distinct proglucagon genes, one of which does not encode GLP-2. Second, I found differential expression of the two proglucagon genes in three copper rockfish tissues, the endocrine pancreas, brain and intestine. This, along with analysis of the putative proglucagon-derived peptide sequences, suggests that the two genes encode functionally distinct proglucagon-derived peptides. Third, teleost proglucagon transcripts exhibit a form of alternative splicing whereby the GLP-2 sequence is removed from the subsequent message. Fourth, analysis of peptide production using mass spectrometry, identified the presence of some but not all peptides predicted from the mRNA, suggesting differential peptide degradation. Thus, analysis of teleost proglucagon transcripts and peptides identifies multiple levels of regulation that allow tissue specific expression and production of select proglucagon-derived peptides, demonstrating unexpected complexity of vertebrate proglucagon regulation.

Identification and phylogenetic evolutionary analysis of proglucagons from teleosts, and other vertebrate groups, including elasmosbranchs, lungfish, amphibians, reptiles and mammals demonstrates the dynamic evolutionary history of proglucagon. A plethora of

evolutionary processes have molded the proglucagon gene, including gene duplication, exon duplication, whole or partial exon loss, and increased sequence diversification due to duplication possibly leading to new peptide function. Thus, proglucagon is an ideal model for studying evolutionary processes and methods of functional adaptation.

Examiners:

Dr. T.P. Mommsen, Supervisor (Department of Biochemistry and Microbiology)

Dr. R.W. Olafson, Co-supervisor (Department of Biochemistry and Microbiology)

Dr. W.W. Kay, Departmental Member (Department of Biochemistry and Microbiology)

Dr. T.W. Pearson, Departmental Member (Department of Biochemistry and Microbiology)

Dr. N.M. Sherwood, Outside Member (Department of Biology)

Dr. J.T. Silverstein, External Examiner (Genetics, National Center for Cool and Cold Water Aquaculture, United States Department of Agriculture - Agriculture Research Services)

Table of Contents

List of Figures viii

List of Tables x

List of Abbreviations xi

Proglucagon: An Introduction 1

Introduction 1
Glucagon superfamily and message transduction 2
Structure 2
Glucagon 4
GLP-1 5
GLP-2 6
Other Proglucagon-derived Peptides 7
Proglucagon Expression, Tissue Distribution, and Peptide Function 7
Pancreas 9
Mammals 9
Non-mammalian Vertebrates 10
Intestine 11
Mammals 11
Non-mammalian Vertebrates 13
Brain 14
Mammals 14
Non-mammalian Vertebrates 18
Regulation 19
At the peptide level 19
N-terminal extension 19
Differential Processing of Prohormone Precursor 20
Peptide Degradation 22
At the transcript level 23
At the gene level 24
Receptors 25
Molecular Evolution 27

Research Objectives 30

Methodological Approach 32

Chapter 1 – Regulation of expression and translation of proglucagon genes in Copper Rockfish: variations on a theme. 33

Abstract 33
Introduction 33
Methods 36
Fish 36
Total RNA isolation 36
cDNA synthesis 37
PCR 37

PCR programs 37
Primers 39
mRNA Isolation 40
5' and 3' RACE PCR 40
Cloning 41
Plasmid Preparations 41
Sequencing 41
Genomic DNA isolation 42
Southern Blotting 42
PCR Walking 44
Mass Spectrometry 46
Peptide Extraction 46
C-18 SepPak 46
Sephadex G-50 46
C-18 Reverse Phase High Performance Liquid Chromatography 47
Mass Spectrometry Analysis 47
Results 48
Proglucagon mRNA Sequences 48
Genomic DNA Sequences and Southern Blot for Proglucagon Genes 52
Tissue Distribution of Proglucagon Messenger RNAs 57
Brockmann body (endocrine pancreas) 61
Brain 61
Gastrointestinal tract 62
Mass Spectrometry Data 63
Discussion 70
Comparison of the two rockfish proglucagon sequences to anglerfish 70
Different putative peptide sequences 71
At least two proglucagon genes identified in rockfish 74
Truncated proglucagon II gene lacking GLP-2 74
Alternative splicing in proglucagon I 75
Differential tissue expression of the two proglucagon genes 76
Brockmann Body - Proglucagon I transcript encoding GLP-2 identified 76
Brain - Absence of messenger RNA for GLP-2 78
Gastrointestinal tract - Zonation of proglucagon II expression in intestine 79
Expression of two forms of the same gene in the same tissue 80
Identification of proglucagon derived peptides in pancreas 81
Summary – It all comes back to regulation 82

Chapter 2 – Proglucagon Expression and Distribution in Channel Catfish and Identification of a Glucagon-like Receptor in Teleosts 84

Abstract 84
Introduction 84
Methods 87
Fish 87
Total RNA isolation 87
cDNA synthesis 88
PCR 88

PCR programs ... 89
Primers ... 89
mRNA Isolation ... 90
5' and 3' RACE PCR ... 90
Cloning ... 91
Plasmid Preparations ... 91
Sequencing ... 92
Peptide Extraction ... 92
C-18 SepPak ... 92
Sephadex G-50 ... 93
C-18 Reverse Phase High Performance Liquid Chromatography ... 93
Mass Spectrometry Analysis ... 93
Results ... 94
Sequence Data ... 94
Tissue Distribution ... 97
Brockmann Body ... 98
Brain ... 99
Gastrointestinal tract ... 101
Stomach ... 101
Intestine ... 102
Anus ... 102
Starved and Refed Experiments ... 102
Peptide Identification by Mass Spectrometry ... 102
Receptor sequences ... 104
Discussion ... 107
Two catfish proglucagon sequences encode distinct peptides ... 107
Tissue Distribution ... 108
Brockmann Body ... 108
Brain ... 108
Gastrointestinal tract ... 109
Starved and Refed Catfish ... 110
Proglucagon-derived peptide receptors ... 111
Summary ... 112

Chapter 3 – Molecular Evolution of Proglucagon in Non-Mammalian Vertebrates ... 113

Abstract ... 113
Introduction ... 113
Methods ... 115
Fish ... 115
Total RNA isolation ... 116
cDNA synthesis ... 117
PCR ... 117
PCR programs ... 117
Primers ... 118
3'RACE ... 119
Cloning ... 119

Plasmid Preparations 119
Sequencing 120
Databases 120
Phylogenetic Analysis 121
Results 122
cDNA sequences 122
Fugu database search. 128
Phylogenetic Analysis 131
Glucagon phylogenetic tree 131
GLP-1 phylogenetic tree 134
GLP-2 phylogenetic tree 137
Teleost phylogenetic tree 139
A Phylogenetic tree including all three peptides 141
Discussion 143
Lungfish 143
Elasmobranchs 143
Teleosts 144
Exon duplication 148
Duplication leads to diversification of function 149
Gene duplication 150
Summary 151
Conclusions and Future Work 153
Acknowledgements 155
References 156

List of Figures

FIGURE 1. SCHEMATIC DEPICTION OF A MAMMALIAN PROGLUCAGON TRANSCRIPT..3

FIGURE 2. THE GLUCAGON AMINO ACID SEQUENCE FOR REPRESENTATIVE VERTEBRATES. 5

FIGURE 3. THE GLP-1 AMINO ACID SEQUENCE FOR REPRESENTATIVE VERTEBRATES. 6

FIGURE 4. THE GLP-2 AMINO ACID SEQUENCE FOR REPRESENTATIVE VERTEBRATE SPECIES. 7

FIGURE 5. SCHEMATIC REPRESENTATION OF ALTERNATIVE SPLICING IN TELEOSTS. 24

FIGURE 6. COMPLETE MRNA SEQUENCE AND PUTATIVE TRANSLATED AMINO ACIDS FOR COPPER ROCKFISH PROGLUCAGON I INCLUDING THE ALTERNATIVELY SPLICED 3'UTR. 50

FIGURE 7. COMPLETE MRNA SEQUENCE AND PUTATIVE TRANSLATED AMINO ACIDS FOR COPPER ROCKFISH PROGLUCAGON II. 51

FIGURE 8. A SCHEMATIC REPRESENTATION OF THE ROCKFISH PROGLUCAGON GENES. 52

FIGURE 9. GENOMIC DNA SEQUENCE FOR PROGLUCAGON I, INCLUDING AN ALIGNMENT WITH THE 3'UTR OF THE ALTERNATIVELY SPLICED PROGLUCAGON I CDNA. 54

FIGURE 10. GENOMIC DNA SEQUENCE FOR PROGLUCAGON II. 55

FIGURE 11. SOUTHERN BLOT FOR THE PROGLUCAGON GENE IN COPPER ROCKFISH DNA DIGESTED BY ECORI AND BGLII. 57

FIGURE 12. SCHEMATIC REPRESENTATION OF THE PROGLUCAGON TRANSCRIPTS AMPLIFIED BY RT-PCR. ...59

FIGURE 13. TISSUE DISTRIBUTION OF PROGLUCAGON TRANSCRIPTS IN COPPER ROCKFISH TISSUES, SCREENED BY RT-PCR OF MRNA. 60

FIGURE 14. DIFFERENTIAL EXPRESSION OF PROGLUCAGON II ALONG THE INTESTINE OF COPPER ROCKFISH. 63

FIGURE 15. PARTIAL PURIFICATION OF PANCREATIC PEPTIDES. ABSORBANCE PROFILE AT 230 NM OF SEPHADEX BB PEAK 1 SAMPLE SEPARATED BY C-18 REVERSE PHASE HPLC. 65

FIGURE 16. PARTIAL PURIFICATION OF PANCREATIC PEPTIDES. ABSORBANCE PROFILE AT 230 NM OF SEPHADEX BB PEAK 2 SAMPLE SEPARATED BY C-18 REVERSE PHASE HPLC. 66

FIGURE 17. MASS SPECTROMETRY DATA FOR HPLC PEAK 10 IDENTIFYING GLP1 I 31AA. 67

FIGURE 18. MASS SPECTROMETRY DATA FOR HPLC PEAK 9 IDENTIFYING GLP1 II 34AA. 68

FIGURE 19. MASS SPECTROMETRY DATA FOR HPLC PEAK 7 IDENTIFYING GLUCAGON II AND GLP1 I 34AA. 68

FIGURE 20. MASS SPECTROMETRY DATA FOR HPLC PEAK 11 IDENTIFYING GLP1 II 31AA. 69

FIGURE 21. MASS SPECTROMETRY DATA FOR SEPHADEX PEAK 2 WHOLE IDENTIFYING GLUCAGON I. 69

FIGURE 22. AMINO ACID ALIGNMENT OF SOME GLUCAGON SEQUENCES, INCLUDING THE TWO PUTATIVE ROCKFISH SEQUENCES. 72

FIGURE 23. AMINO ACID ALIGNMENT OF SOME GLP-1 SEQUENCES, INCLUDING THE TWO PUTATIVE ROCKFISH SEQUENCES. 73

FIGURE 24. MRNA SEQUENCE AND PUTATIVE AMINO ACID SEQUENCE FOR CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) PROGLUCAGON I. 95

FIGURE 25. MRNA SEQUENCE AND PUTATIVE AMINO ACID SEQUENCE FOR CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) PROGLUCAGON II. 96

FIGURE 26. A SCHEMATIC DIAGRAM OF THE CATFISH PROGLUCAGON TRANSCRIPTS, INDICATING THE LOCATION OF THE GENE SPECIFIC PRIMERS USED IN RT-PCR REACTIONS. 97

FIGURE 27. TISSUE DISTRIBUTION OF PROGLUCGON TRANSCRIPTS IN CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) BROCKMANN BODY AND BRAIN AS IDENTIFIED BY RT-PCR. 100

FIGURE 28. TISSUE DISTRIBUTION OF PROGLUCGON TRANSCRIPTS IN CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) GASTROINTESTINAL TRACT AS IDENTIFIED BY RT-PCR. 101

FIGURE 29. AMINO ACID ALIGNMENT OF MAMMALIAN AND AMPHIBIAN GLUCAGON AND MAMMALIAN GLP-1 RECEPTORS WITH PUTATIVE ROCKFISH AND RAINBOW TROUT GLP-1 RECEPTOR SEQUENCES 105

FIGURE 30. TRANSMEMBRANE DOMAIN PREDICTION FOR THE PARTIAL GLUCAGON-LIKE RECEPTOR AMINO ACID SEQUENCE FROM COPPER ROCKFISH (*S. CAURINUS*). 106

FIGURE 31. CLUSTALW PHYLOGENETIC TREE FOR TELEOST, MAMMALIAN, AND AMPHIBIAN GLUCAGON AND GLP-1 RECEPTORS. 106

FIGURE 32. ALIGNMENT OF THE AMINO ACID SEQUENCES FOR THE CHANNEL CATFISH GLP-1S 108

FIGURE 33. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM AFRICAN LUNGFISH (*PROTOPTERUS DOLLOI*) 123

FIGURE 34. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM AUSTRALIAN LUNGFISH (*NEOCERATODUS FORSTERI*). 124

FIGURE 35. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM SPINY DOGFISH (*SQUALUS ACANTHIAS*). 125

FIGURE 36. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM SPOTTED RATFISH (*HYDROLAGUS COLLIEI*). 126

FIGURE 37. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM CANE TOAD (*BUFO MARINUS*). 127

FIGURE 38. PARTIAL PROGLUCAGON CDNA SEQUENCE FROM COTTONMOUTH SNAKE (*AGKISTRODON PISCIVORUS*). 128

FIGURE 39. PROGLUCAGON I CDNA SEQUENCE FOR PUFFERFISH, *TAKIFUGU RUBRIPES*, DETERMINED FROM THE GENOMIC DNA SEQUENCE DATABASE. 130

FIGURE 40. PROGLUCAGON II CDNA SEQUENCE FOR PUFFERFISH, *TAKIFUGU RUBRIPES*, DETERMINED FROM GENOMIC DNA SEQUENCE 130

FIGURE 41. GLUCAGON III CDNA SEQUENCE FOR PUFFERFISH, *TAKIFUGU RUBRIPES*, DETERMINED FROM GENOMIC DNA SEQUENCE 131

FIGURE 42. PHYLOGENETIC TREE FOR GLUCAGON 133

FIGURE 43. PHYLOGENETIC TREE FOR GLP-1 136

FIGURE 44. PHYLOGENETIC TREE FOR GLP-2 138

FIGURE 45. A PHYLOGENETIC TREE FOR TELEOSTS FROM GLUCAGON THROUGH TO THE END OF GLP-1. 140

FIGURE 46. A PHYLOGENETIC TREE FOR GLUCAGON, GLP-1 AND GLP-2. 142

FIGURE 47. AMINO ACID ALIGNMENT OF TELEOST GLUCAGON AND GLP-1 SEQUENCES. 146

FIGURE 48. A SCHEMATIC CLADOGRAM FOR THE EVOLUTION OF PROGLUCAGON GENES IN VERTEBRATES. 152

List of Tables

TABLE 1. NAMES AND SEQUENCES OF PRIMERS USED IN THE IDENTIFICATION OF COPPER ROCKFISH PROGLUCAGON SEQUENCES. 39

TABLE 2. NON-SPECIFIC PRIMERS USED IN 5' AND 3' RACE AND PCR WALKING REACTIONS. 39

TABLE 3. THEORETICAL AND EXPERIMENTAL MASSES FOR COPPER ROCKFISH PROGLUCAGON DERIVED PEPTIDES. 66

TABLE 4. NAMES AND SEQUENCES OF PRIMERS USED IN IDENTIFICATION OF CHANNEL CATFISH PROGLUCAGON SEQUENCES. 89

TABLE 5. NAMES AND SEQUENCES OF PRIMERS USED IN IDENTIFICATION OF GLUCAGON-LIKE RECEPTOR SEQUENCES. 90

TABLE 6. CALCULATED AND EXPERIMENTAL MASSES FOR CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) PROGLUCAGON-DERIVED PEPTIDES ISOLATED FROM THE BROCKMANN BODY AND ANALYZED BY MALDI-TOF MASS SPECTROMETRY. 103

TABLE 7. PERCENT SIMILARITIES OF PARTIAL REGIONS OF RECEPTOR cDNA SEQUENCES BETWEEN SEVERAL TELEOST GLUCAGON AND GLP-1 RECEPTOR SEQUENCES. 105

TABLE 8. NAMES AND SEQUENCES OF PRIMERS USED IN THE IDENTIFICATION OF VARIOUS NON-MAMMALIAN VERTEBRATE PROGLUCAGON SEQUENCES 118

List of Abbreviations

ACN	acetonitrile
ACTH	adrenocorticotropin hormone
AP	area postrema
BB	Brockmann body
BSA	bovine servum albumin
cAMP	cyclic adenosine monophosphate
CNS	central nervous system
CC	channel catfish
CPE	carboxy-peptidase E
CRF	copper rockfish
CTA	conditioned taste aversion
DEPC	diethyl pyrocarbonate
DPP-IV	dipeptylpeptidase-IV
DTT	dithiothreitol
GIP	gastric inhibitory peptide or glucose-dependent insulinotropic peptide
GHRH	growth hormone releasing hormone
GLI	glucagon-like immunoreactivity
GLP-1	glucagon-like peptide-1
GLP-2	glucagon-like peptide-2
GRPP	glicentin related polypeptide
GSP	gene specific primer
ICV	intracerebroventricular
IP_3	inositol 3-phosphate
MALDI-TOF	matrix assisted laser desorption/ionization - time of flight
MCT	microcentrifuge tube
MPGF	major proglucagon fragment
NPY	neuropeptide Y
NTS	nucleus of the solitary tract
OXM	oxyntomodulin
RACE	rapid amplification of cDNA ends
RP-HPLC	reverse phase - high performance liquid chromatography
RT-PCR	reverse transcription - polymerase chain reaction
PACAP	pituitary adenylyl cyclase-activating peptide
PBN	parabranchial nulceus
PC	prohormone convertase
PHM	peptide histidine methionine
POMC	pro-opiomelanocortin
PVN	periventricular nucleus
SDS	sodium dodecyl sulphate
SPC	subtilisin-like pro-protein convertase
SSC	sodium chloride sodium citrate
TFA	trifluoroacetic acid
UTR	untranslated region
VIP	vasoactive intestinal peptide

Proglucagon: An Introduction

Introduction

The 29 amino acid residue glucagon is one of the primary hormones involved in the regulation of glucose metabolism. Glucagon is mainly produced in the alpha cells of the endocrine pancreas, released into the blood stream, and travels to the liver, where it stimulates the release of glucose through either liberation from glycogen (glycogenolysis), or de novo synthesis (gluconeogenesis). In mammals, precise regulation of circulating glucose and glucose metabolism is essential for healthy, productive, efficient maintenance of all organs and the animal as a whole. While glucagon leads to the production and release of glucose into the bloodstream, insulin counters glucagon in regulation of glucose levels by stimulating uptake of glucose. Disease states of glucose regulation include diabetes mellitus, which can lead to debilitating effects on health. Understanding production, regulation, and functions of glucagon and related peptides is essential to fight this disease, which is considered primarily an insulin related condition, yet abnormal circulating levels of glucagon compound the ill effects of diabetes (Unger, R.H. 1976).

In this body of work, I will start by introducing current work and thoughts on the proglucagon genes and their expression, function and regulation in vertebrates. The first chapter presents the work done on our model animal, the copper rockfish (*Sebastes caurinus*), where two proglucagon genes are identified and intricacies of the regulation of these genes are investigated. Proglucagon sequences, tissue distribution, and some insights on GLP-1 receptor from channel catfish (*Ictalurus punctatus*) and other teleosts comprise the second chapter. The third and final chapter, on molecular evolution of new and some previously reported non-mammalian vertebrate proglucagon sequences provides new insights into the evolution of this gene. This is followed by a summary of my research and a general discussion of the present and future of proglucagons.

Glucagon superfamily and message transduction

Glucagon belongs to a family of peptide hormones that, like many biologically active peptides, are encoded in prohormone structures that require proteolytic processing in the endoplasmic reticulum before secretion. This family includes secretin, vasoactive intestinal peptide (VIP), gastric inhibitory peptide aka glucose-dependent insulinotropic peptide (GIP), growth hormone releasing hormone (GHRH), peptide histidine methionine (PHM), pituitary adenylyl cyclase-activating polypeptide (PACAP), PACAP-related peptide, helospectin, helodermin, exendin, and of course glucagon, glucagon-like peptides 1 and 2 (GLP-1 and GLP-2) (Plisetskaya, E.M. et al. 1996). Many of these peptide hormones bind and activate receptors that belong to the glucagon receptor superfamily. These receptors have seven transmembrane domains coupled to G-proteins that activate downstream signal transduction pathways. Many of these receptors, including the glucagon receptor, are primarily coupled to G_s-proteins involved in activation of adenylyl cyclase and the production of cAMP. In the case of the glucagon receptor in a liver cell, increased production of cAMP activates protein kinases such as protein kinase A, which in turn phosphorylates numerous target proteins including glycogen phosphorylase (Moon, T.W. et al. 1999). Phosphorylation of glycogen phosphorylase activates this enzyme to break glycogen down into glucose for release into the blood.

Structure

Mammalian glucagon is encoded and translated as a prohormone, similar to other members of the glucagon superfamily. Often these precursors contain more than one biologically active peptide as in the case of that containing GHRH and PACAP (Parker, D.B. et al. 1993). Proglucagon contains glucagon, glucagon-like peptide-1 (GLP-1) and glucagon-like peptide-2 (GLP-2). Other peptides and fragments, whose biological significance is less clear, such as glicentin, oxyntomodulin, and mini-glucagon can also be released from the proglucagon precursor. As depicted in Figure 1, the N-terminal end of the prohormone starts with a secretory signal peptide, followed by glicentin-related pancreatic peptide (GRPP), in the position of the cryptic peptide of many glucagon

superfamily members. Next are glucagon and a short intervening peptide (IP-1). Glucagon-like peptide 1 (GLP-1), a second short intervening peptide (IP-2), and glucagon-like peptide 2 (GLP-2) follow glucagon. The prohormone can be proteolytically processed in the endoplasmic reticulum into an assortment of peptides and fragments before secretion from the cell. The combination of peptides depends on the presence and activity of prohormone convertases that cleave at specific single and double dibasic amino acid sites, as indicated in Figure 1. Differential processing at these sites can produce the peptides oxyntomodulin, glicentin, or larger fragments that include more than one peptide, such as GLP-1, IP-2, and GLP-2 released together as an extended precursor fragment. As well, mature glucagon, GLP-1, and GLP-2 are produced. Glicentin consists of GRPP, glucagon, and IP-1. Oxyntomodulin is a C-terminally extended version of glucagon comprised of glucagon and IP-1. All of the peptides synthesized and secreted from proglucagon are shown in Figure 1. Not all of these peptides have a clearly defined role in mammalian metabolism.

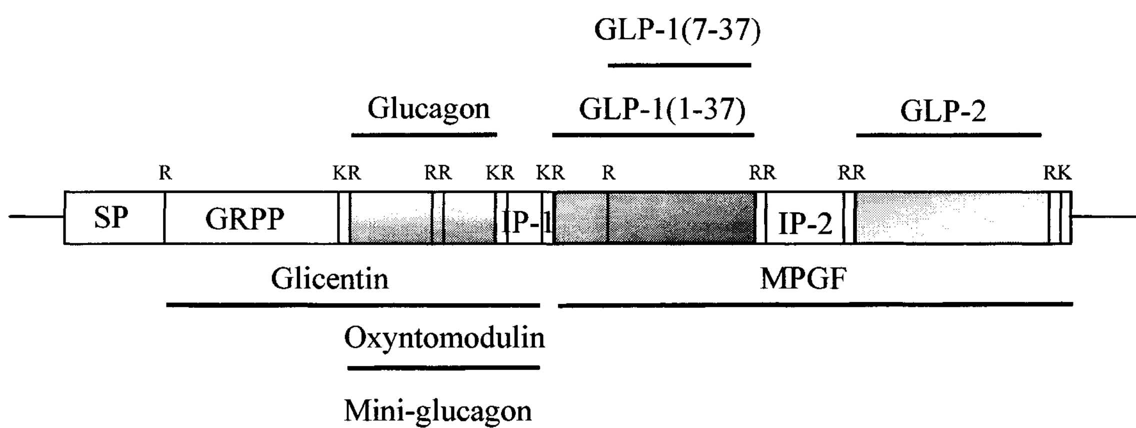

Figure 1. Schematic depiction of a mammalian proglucagon transcript.

The proglucagon precursor structure is not consistent throughout all vertebrates. For instance, in amphibians the proglucagon structure is altered from the depicted mammalian structure containing an exon duplication of the exon encoding GLP-1. Thus, in all frog proglucagons at least two GLP-1 sequences occur consecutively within the prohormone. In *Xenopus laevis* an additional exon duplication has occurred to reveal three consecutive GLP-1 sequences (Irwin, D.M. et al. 2000; Irwin, D.M. et al. 1997).

Glucagon

Throughout vertebrate evolution, the glucagon peptide sequence is highly conserved. The most variation occurs between the parasitic sea lamprey (*Petromyzon marinus*) and human (*Homo sapiens*) glucagons which have 72.4% identity at the amino acid level. Excluding a group of New World rodents (guinea pig, degu, and chinchilla), that possess unique biological activities (Seino, S. et al. 1988), the peptide sequence of glucagon is invariant across mammals. Even including the above New World rodents in the mammalian group, glucagon is still highly conserved with at least 82.8% amino acid identity. Mammalian, avian, reptilian, amphibian, and piscine glucagons are 29 amino acids in length. The only exception is the North American paddlefish (*Polyodon spathula*), whose glucagons are 31 amino acids long (Nguyen, T.M. et al. 1994). When comparing glucagon peptide sequence across the vertebrates (Figure 2), eight amino acid positions are found to be invariant, disregarding a few allegedly non-functional glucagons and the glucagons from the New World rodents. These positions are His^{1}, Gly^{4}, Phe^{6}, Asp^{9}, Tyr^{10}, Lys^{12}, Trp^{25}, and Leu^{26} that surely must play a significant role in the specific function of glucagon (Irwin, D.M. 2001). In fact, residues His ^{1}, Asp ^{9}, and Ser ^{16} have been proposed to play a role in the activation of the mammalian glucagon receptor, but are not essential for receptor binding (Unson, C.G. et al. 1998; Unson, C.G. et al. 1994). Asp ^{15}, has been shown to be essential for receptor binding while Ser ^{8} contributes to a structure needed for proper binding (Unson, C.G. et al. 1994). Also in mammalian glucagon, substitutions of Lys ^{12}, Arg ^{17}, and Arg ^{18} indicated these three residues are necessary for optimum ligand binding and potency (Unson, C.G. et al. 1998). Considering that some of these positions, such as Ser ^{8}, are not conserved throughout all vertebrates, the amino acid requirements of each position will vary slightly with the specificity and ligand interactions of the glucagon receptor in each vertebrate species. This is especially true of Ser^{16}, as mammalian systems show a direct role in receptor activation, but throughout the vertebrates, it is one of the most highly variable positions of glucagon, suggesting that its requirement for glucagon function may be restricted to mammals (Irwin, D.M. 2001) with repercussions on receptor structure. Also, in some cases, such as Arg ^{17}, arginine itself is not conserved, but amino acids of similar biochemical nature are found in this position, most likely maintaining the same peptide

conformation and thus displaying similar ligand receptor interactions, although processing into the biologically active miniglucagon may be compromised (Dalle, S. et al. 2002).

```
Lamprey           HSEGTFTSDY SKYLENKQAK DFVRWLMNA
Electric ray      HSEGTFTSDY SKYLDNRRAL DFVQWLMNT
Anglerfish I      HSEGTFSNDY SKYLEDRKAQ EFVRWLMNN
Anglerfish II     HSEGTFSNDY SKYLETRRAQ DFVQWLKNS
Rainbow trout I   HSEGTFSNDY SKYQEERMAQ DFVQWLMNS
Leopard frog      HSQGTFTSDY SKYLDSRRAQ DFVQWLMNS
Gila monster      HSQGTFTSDY SKYLDTRRAQ DFVQWLMNT
Chicken           HSQGTFTSDY SKYLDSRRAQ DFVQWLMST
Human             HSQGTFTSDY SKYLDSRRAQ DFVQWLMNT
                  *  * *  **  *             **
```

Figure 2. The glucagon amino acid sequence for representative vertebrates. Asterisks indicate invariant residues throughout vertebrates.

GLP-1

Biologically active GLP-1 is generally 31 amino acids in length, but can also occur as a C-terminally amidated 30 amino acid peptide with equal potency to the non amidated version (Orskov, C. et al. 1994). With only five invariant amino acid positions, GLP-1 varies in sequence across vertebrates more than glucagon (Irwin, D.M. 2001). Mammalian, avian, and reptilian proglucagons encode a GLP-1 that is N-terminally extended by six amino acids (cf. Figure 1). The 37 amino acid precursor is not biologically functional. Yet, during maturation of the peptide this six amino acid extension is cleaved to produce the functional 31 amino acid peptide. The reptilian GLP-1 and first GLP-1 sequences of amphibians also reveal the N-terminal extension, whereas the remaining one or two GLP-1(s) in amphibians, as well as the GLP-1 sequences in fish do not have this extension (Chen, Y.E. et al. 1997; Irwin, D.M. et al. 2000; Irwin, D.M. et al. 1997; Irwin, D.M. et al. 1995). Therefore, after translation, these GLP-1 peptides require one less processing step before biological activity.

Several experiments used individual alanine substitutions of residues in mammalian GLP-1 to identify which residues play a role in receptor binding and activation. Generally, the essential residues were determined to be His^{1}, Gly^{4}, Phe^{6}, Thr^{7}, Asp^{9},

Phe^{22}, and Ile^{23}, but Trp^{25} and Leu^{26} also assisted in potent receptor activation (Adelhorst, K. et al. 1994; Gallwitz, B. et al. 1994; Parker, J.C. et al. 1998; DaCambra, M.P. et al. 2000). Considering some of these residues are among the five invariant residues in all known vertebrate sequences, His^{1}, Ala^{2}, Gly^{4}, Thr^{7}, and Phe^{22}, and others are only substituted with biochemically similar residues (Figure 3), the functional interaction between receptor and ligand is probably also conserved throughout vertebrates. This becomes evident as both glucagon and GLP-1 are conserved enough that mammalian peptide will activate receptors in fish cells, such as copper rockfish, to the same degree as other closely related teleostean peptides, e.g. salmon GLP-1 (Plisetskaya, E.M. et al. 1996).

```
Lamprey               HADGTFTNDM TSYLDAKAAR DFVSWLARSD K
Dogfish               HAEGTYTSDV DSLSDYFKAK RFVDSLKSY
Anglerfish I          HADGTFTSDV SSYLKDQAIK DFVDRLKAGQ V
Anglerfish II         HADGTYTSDV SSYLQDQAAK DFVSWLKAGR G
Rainbow trout I       HADGTYTSDV STYLQDQAAK DFVSWLKSGR A
Leopard frog 1a       HAEGTYTNDV TQFLEEKAAK EFIDWLIKGK P
Leopard frog 1b       HADGTFTSDM SSYLEEKAAK EFVDWLIKGR Q
Gila monster          HADGRYTSDI SSYLEGQAAK EFIAWLVNGR G
Chicken               HAEGTYTSDI TSYLEGQAAK EFIAWLVNGR G
Rat                   HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR G
Human                 HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR G
                      ** *  *                *
```

Figure 3. The GLP-1 amino acid sequence for representative vertebrates. Asterisks indicate invariant residues throughout vertebrates.

GLP-2

GLP-2, in non-mammalian vertebrates, has not been studied as extensively as the other peptides of proglucagon, thus it is difficult to determine evolutionary variability in this 33 amino acid peptide. Not surprising, considering the relatively short evolutionary history, GLP-2 sequence is highly conserved within mammals. The peptide is less highly conserved in the few non-mammalian vertebrates for which sequence is known (Irwin, D.M. 2001). Because of the lack of sequences available and the lack of conservation across non-mammalian vertebrates, it is difficult to infer which residues are important for functionality. Individual alanine substitutions in mammalian GLP-2 did indicate residues

5, 6, 17, 20, 22, 23, 25, 26, 30, and 31 are important for receptor binding, while residues 1, 3, 12, and 21 are important for receptor activation, but not binding (DaCambra, M.P. et al. 2000). Since some of these residues are important for peptide function in all three proglucagon derived peptides, these residues are most likely essential for interaction with the peptide receptors, which also share similarity as members of the same superfamily.

```
Lamprey            HSDGSFTNDM NVMLDRMSAK NFLEWLKQQG RG-
Rainbow trout      HVDGSFTSDV NKVLDSLAAK EYLLWVMTSK TSG
Leopard frog       HADGSFTSDF NKALDIKAAQ EFLDWIINTP VKE
Gila monster       HADGTFTSDY NQLLDDIATQ EFLKWLINQK VTQ
Rat                HADGSFSDEM NTILDNLATR DFINWLIQTK ITD
Human              HADGTFTSDY NQLLDDIATQ EFLKWLINQK VTQ
```

Figure 4. The GLP-2 amino acid sequence for representative vertebrate species.

Other Proglucagon-derived Peptides

Mammalian oxyntomodulin is the 37 amino acid C-terminally extended form of glucagon including IP-1, whose length varies among the vertebrates as the length of the intervening peptide-1 varies. Glicentin, in mammals, is 69 amino acids, starting at the beginning of GRPP, including glucagon and continuing to IP-1 (Collie, N.L. et al. 1994). Its length also varies in non-mammalian vertebrates, as both GRPP and IP-1 vary in length. GRPP is the least conserved portion of proglucagon, excluding IP-1 and IP-2.

Two basic amino acids in positions 17 and 18 of glucagon create an internal cleavage site in mammalian and most non-mammalian vertebrate sequences. Proteolytic processing after this position results in a glucagon fragment consisting of positions 19-29 known as mini-glucagon.

Proglucagon Expression, Tissue Distribution, and Peptide Function

Mammals and other vertebrates share the same prohormone layout for proglucagon, but posttranslational processing varies greatly with the organ of production and processing and with the animal species.

In mammals, proglucagon mRNA, proglucagon-derived peptides, or glucagon-like immunoreactivity (GLI) have been found in three tissues, pancreas, intestine, and brain (CNS). Consequently, these tissues have been the focus for identifying and characterizing possible functions for the biologically active peptides from proglucagon. The function of proglucagon-derived peptides in these organs is inter-related, as these peptides and organs are involved in glucose uptake and metabolism and establish an endocrine bridge between the three tissues. The pancreas, brain and intestine work together with the liver to provide a highly regulated metabolic system. Because of this, some proglucagon-derived peptides are produced in more than one of these organs to achieve the same effect. In the same vein, the intricacies of metabolic regulation are not fully understood, hence, some of the patterns of proglucagon-derived peptide tissue distribution may not be, as yet, fully recognized.

Although proglucagon has not been studied as extensively in non-mammalian vertebrates as in mammals, proglucagon derived peptides or mRNA have been detected in pancreas and intestine of many non-mammalian vertebrates, including birds (chicken), reptiles (gila monster, python), amphibians (claw-toed frog, bullfrogs, and toads) and several groups of fish (agnathan, elasmobranch, and teleost). So far, due to a lack of analysis, glucagon immunoreactivity or mRNA sequences have been located in brain for only one fish species (*Clarias batrachus*) (Sarkar, S. et al. 2001) and one frog species (*Rana tigrina rugulosa*) (Yeung, C.M. et al. 2001).

The function of GLP-1 as well as some derivatives of glucagon has not been consistent throughout all species studied.

The production of each proglucagon derived peptide (glucagon, GLP-1, GLP-2, oxyntomodulin, glicentin) varies between tissues and tissue distribution is not consistent between species of animals. Since the proglucagon precursor relies on proteolytic cleavage to release the biologically active peptides, the combination of peptides found in each tissue depends on the presence of processing enzymes, which can lead to a different complement of functional peptides in each tissue. To add to this, not all peptides are

expressed in all three tissues as alternative splicing occurs in some species, leading to truncated mRNA sequences missing portions of the precursor.

Such variation in the availability of the proglucagon-derived peptides within an animal also suggests the potential for multiple functions for these hormones. Besides the presence of the peptide and the biological response of the target tissue, the presence and activation of specific receptors have assisted in determining and characterizing the functions of proglucagon derived peptides in all three tissues.

Pancreas

Mammals

In the α-cells of the mammalian pancreas, the primary protein product is glucagon. Small amounts of oxyntomodulin are also found, possibly representing a precursor yet to be processed. GLP-1 and GLP-2 are released together with IP-2 in the α-cells, in an uncleaved nonfunctional precursor fragment known as the "major proglucagon fragment" (MPGF) (Patzelt, C. et al. 1984; Orskov, C. et al. 1986), together with very small amounts of the biologically inactive N-terminally extended form of GLP-1 (Holst, J.J. et al. 1994). Miniglucagon (19-29), processed from existing glucagon, has also been localized to α-cells actively secreting glucagon and consists of a small percentage of the glucagon concentrations (Dalle, S. et al. 2002).

After release from the pancreas, the function of glucagon in all vertebrates is the increase of blood glucose levels by activation of its receptor on the liver and stimulation of gluconeogenesis and glycogenolysis. As a general, but not unequivocal antagonist of insulin, glucagon is also involved in the regulation of lipid, amino acid and urea metabolism (Plisetskaya, E.M. et al. 1996). While glucagon receptors are the most abundant in mammalian liver, pancreas, cerebral cortex and lung also show glucagon receptor transcripts (Dunphy, J.L. et al. 1998). Another action of glucagon and the more potent miniglucagon is inhibition of the release of insulin from β-cells. Although miniglucagon shares this function with glucagon, it does not stimulate other functions of

glucagon, including regulation of glucose release (Mallat, A. et al. 1987). Considering the central importance of glucagon in metabolism, it is not surprising that the function of glucagon has not changed with time throughout vertebrates.

Although no functional truncated GLP-1 (7-37) is released from the pancreas in mammals, it is produced in the intestine in response to a meal (Holst, J.J. et al. 1987) and released into the plasma (Orskov, C. et al. 1987). It then travels to the pancreas and in the presence of glucose, stimulates the release of insulin (Wang, Z. et al. 1995). This is known as an incretin effect. Other aspects of this effect include stimulation of expression and biosynthesis of insulin, and inhibition of the release of glucagon which both lead to a subsequent decrease in blood glucose levels. GLP-1 is important for normal postprandial glucose homeostasis in humans (Edwards, C.M. et al. 1999), thus its incretin effect is the primary action of GLP-1 on the pancreas. Yet, there is also recent evidence for a potential role of GLP-1 in pancreatic islet differentiation and growth (Ling, Z. et al. 2001; Abraham, E.J. et al. 2002).

Non-mammalian Vertebrates

All non-mammalian vertebrates produce and release functional glucagon and GLP-1 in the endocrine pancreas. Teleost fish (Lund, P.K. et al. 1983), chicken (Irwin, D.M. et al. 1995), a reptile (Chen, Y.E. et al. 1997) and *X. laevis* (Irwin, D.M. et al. 1997) produce an alternatively spliced pancreatic proglucagon mRNA that does not encode GLP-2. Other frogs (Irwin, D.M. et al. 2000; Yeung, C.M. et al. 2001) and an agnathan fish (lamprey) (Irwin, D.M. et al. 1999) produce a mammalian-like full length mRNA including GLP-2. Yet, unlike glucagon and GLP-1, that have been isolated from pancreas of many non-mammalian vertebrates, GLP-2 has only been isolated from an amphibian (*Amphiuma tridactylum*) pancreas (Cavanaugh, E.S. et al. 1996).

However, not all vertebrates are 'insulin' driven, like the mammals. For instance, in birds, glucagon appears to be the main player in the regulation of blood glucose and, incidentally fatty acid metabolism, with insulin playing a subordinate role only. This is reflected in the bird pancreas where glucagon-producing α-cells predominate and an

entire half is completely devoid of β-cells. Unfortunately, in birds, the roles of glucagon, that may provide a glimpse of the multifaceted potentials of this hormone in all vertebrates, remain yet to be defined.

Although the function of glucagon is conserved throughout mammals, GLP-1 does not appear to act as an incretin hormone in some non-mammalian vertebrates. In many teleostean species, GLP-1 binds to a receptor on the liver and stimulates the release of glucose. While in some frogs, and perhaps all amphibians, reptiles, and birds, GLP-1 does not activate the liver receptors (Mommsen, T.P. et al. 1994), and does stimulate the release of insulin in mammalian cells, suggesting a possible insulinotropic role in amphibians (Irwin, D.M. et al. 1997). Thus, in teleosts, GLP-1 functions in the same manner as glucagon, and in fact, has been considered a 'better glucagon' as GLP-1 is actually more potent at stimulating glucose release than glucagon (Plisetskaya, E.M. et al. 1996). There seems to be no overlap in GLP-1 function between teleosts and mammals as GLP-1 has does not stimulate glucose release in mammalian liver and no insulinotrophic effects of GLP-1 are found in teleosts (Murayama, Y. et al. 1990; Plisetskaya, E.M. et al. 1996).

Intestine

Mammals

Glicentin, oxyntomodulin, GLP-1 and GLP-2, but not glucagon, are all produced and released by the enteroendocrine intestinal L-cells (Kervran, A. et al. 1987; Orskov, C. et al. 1987). The role and fate of glicentin and oxyntomodulin in the mammalian intestine are under debate. Although glicentin is produced in the intestine (Tager, H.S. et al. 1979) and some effects on gastric acid secretion have been reported, no clear role for glicentin has been demonstrated. In fact, some researchers have demonstrated no effects of glicentin on pancreas or intestine (Holst, J.J. 1997; Ghatei, M.A. et al. 2001; McGregor, G.P. et al. 1998). Also, oxyntomodulin has been detected in intestine without glicentin suggesting glicentin may merely act as a precursor to oxyntomodulin (Collie, N.L. et al. 1994). For oxyntomodulin, on the other hand, inhibition of gastric acid secretion,

inhibition of gastric motility, stimulation of production of cAMP in oxyntic cells, and stimulation of insulin release have all been described (Kervran, A. et al. 1987; Dakin, C.L. et al. 2001; Jarrousse, C. et al. 1984). All of these functions are also common to GLP-1. Not surprisingly, oxyntomodulin has been shown to act as a weak agonist to the GLP-1 receptor (Holst, J.J. 1997; Gros, L. et al. 1993), although it is difficult to confirm that oxyntomodulin stimulates these effects *in vivo*.

Besides the incretin effect stimulated by the intestinal release of GLP-1 postprandially, GLP-1 inhibits gastric acid secretion and gastric emptying. This action is known as the 'ileal break' as GLP-1 is released upon the detection of nutrients in ileum, causing the inhibition of the rate of gastric emptying to facilitate uptake of nutrients (Eissele, R. et al. 1992). The actions of GLP-1 in the ileal break are signalled to the central nervous system via the intestinal innervations of the vagus nerve (Wettergren, A. et al. 1997; Imeryuz, N. et al. 1997). GLP-1 release from the L-cells is triggered by fatty acids or carbohydrates, but not protein, in the ileum (Layer, P. et al. 1995).

Unlike GLP-2, glicentin and oxyntomodulin have no stimulatory effect on intestinal weight or cell proliferation in rats, while GLP-1 has a slight stimulatory effect on stomach and small intestine weights as well as cell proliferation in small and large intestine (Ghatei, M.A. et al. 2001).

In intestine, GLP-2 acts as a growth factor by stimulating epithelial cell differentiation and proliferation in the small and large intestine (Drucker, D.J. et al. 1996; Ghatei, M.A. et al. 2001) and inhibiting epithelial apoptosis (Burrin, D.G. et al. 2001). The specific intestinal epithelial structures that respond to GLP-2 are the crypts and villi (Bjerknes, M. et al. 2001). The crypt structure contains stem, progenitor and Paneth cells. Stem cells differentiate into absorptive columnar enterocytes, enteroendocrine cells (including L-cells), mucous producing goblet cells, and secretory Paneth cells (Mills, J.C. et al. 2001). All except Paneth cells migrate from crypts to villi during differentiation. Previously, it was thought that enteroendocrine cells display the GLP-2 receptors that signal stimulation of crypts and villi. However, recently Bjerknes and Cheng (2001)

demonstrated molecular evidence of GLP-2 receptors and specific response of these receptors, in enteric neurons of mice. Activation of GLP-2 receptors on enteric neurons leads to increased proliferation of columnar enterocytes. Thus, a regulatory feedback loop exists from L-cells that produce GLP-2 to the enteric neurons that bind GLP-2 and stimulate the intestinal epithelium, including enterocytes, involved in nutrient uptake. Although the signalling of GLP-2 appears to be contained within the intestine, plasma GLP-2 levels in humans and rats increases 1.5 to 3.6 fold after feeding (Brubaker, P.L. et al. 1997), and the intestinotrophic effect is stimulated by infusions of GLP-2 (Tsai, C.H. et al. 1997), suggesting that there may also be a peripheral aspect to the function of GLP-2.

GLP-2 also suppresses gastric motility and gastric acid secretion and increases hexose and amino acid transport into intestine (Burrin, D.G. et al. 2001). These functions of GLP-2 are also most likely mediated through the enteric nervous system, as it regulates motility, blood flow, and secretion in the intestine (Mills, J.C. et al. 2001). Therefore, similar to GLP-1, the role of GLP-2 in the intestine is to slow motility and ingestion of food, while increasing nutrient absorption in the small intestine. All these properties make GLP-2 a useful choice for treatment of people suffering with intestinal dysfunction, injury or insufficiency, such as patients of small bowel resection, short bowel syndrome, inflammatory bowel syndrome, and chemotherapeutic injury.

Non-mammalian Vertebrates

In non-mammalian vertebrates, the proglucagon picture in intestine is somewhat less clear, as no roles have been identified for proglucagon-derived peptides. Proglucagon message has been isolated and sequenced from intestine of many non-mammalian vertebrates, including several fish, frogs and a reptile and a bird. Like mammals, the proglucagon message from intestine does include GLP-2, but in some frogs alternative splicing, to remove either the GLP-1 sequence or the GLP-2 sequence does occur in intestine (Yeung, C.M. et al. 2001; Irwin, D.M. et al. 1997). Unfortunately, at this time, no proglucagon-derived peptides have been isolated from intestine of species that don't have endocrine pancreatic tissue (islets) incorporated into the intestine, as occurs in

several fish species (agnathan and some teleostean). Obviously, in peptide isolation from these animals, it is impossible to determine what component of the proglucagon-derived peptides is from an endocrine pancreas source and which are from the intestinal L-cells. Yet, a clear function for proglucagon-derived peptides in intestine is indicated in non-mammalian vertebrates as glucagon and glucagon-like immunoreactivity has been demonstrated in intestinal cells of a turtle (Ku, S.K. et al. 2001), a lungfish (Tagliafierro, G. et al. 1996), a teleost (Gomez-visus, I. et al. 1998), and two ancient teleosts (Groff, K.E. et al. 1997; Al Mahrouki, A.A. et al. 1998). Also, increased cAMP production in response to GLP-1 in rockfish enterocytes and an increase in intestinal glucose uptake with GLP-1 treatment in catfish (*Ictalurus melas*) supports a role for this peptide in intestine (Mommsen, T.P. et al. 1998; Soengas, J.L. et al. 1998).

Brain

Mammals

GLP-1 has been determined to play a role in food intake in mammals. Evidence leading to this conclusion includes the presence of GLP-1 and its receptor in the mammalian brain, as well as inhibition of food intake by GLP-1 treatment.

Proglucagon-derived peptides are produced in only one part of the brain, the nucleus of the solitary tract (NTS) in the brain stem, as identified by the presence of proglucagon mRNA (Larsen, P.J. et al. 1997a; Han, V.K. et al. 1986). Yet, proglucagon-derived peptide immunoreactivity, including oxyntomodulin (Tager, H. et al. 1980), glucagon (Jin, S.L. et al. 1988), GLP-1 (Larsen, P.J. et al. 1997a; Shimizu, I. et al. 1987) and GLP-2 (Tang-Christensen, M. et al. 2000) is found throughout the rodent brain with highest concentrations found in the periventricular nucleus (PVN) of the hypothalamus. GLP-1 receptors have also been identified in several parts of the rat brain, by specific GLP-1 binding and receptor immunoreactivity (Uttenthal, L.O. et al. 1992; Goke, R. et al. 1995; Larsen, P.J. et al. 1997b). Some of these areas include the PVN and arcuate nucleus of the hypothalamus, area postrema and subfornical organ and the NTS of the brain stem (Goke, R. et al. 1995; Uttenthal, L.O. et al. 1992). As the PVN region of the brain is

involved in food intake regulation and appetite, and both GLP-1 and its receptor are found there, GLP-1 may be involved in regulation of food intake.

In rats, intracerebroventricular (ICV) injections of GLP-1 and GLP-1 injections directly in the periventricular nucleus (PVN) of the hypothalamus inhibit food intake in rats (Turton, M.D. et al. 1996; Tang-Christensen, M. et al. 1996). These ICV injections, but not the PVN injections, also stimulate taste aversion, measured using conditioned taste aversion (CTA) (McMahon, L.R. et al. 1998; Thiele, T.E. et al. 1997), which may be responsible for the inhibition of food intake. This is supported by the similarity seen in c-fos induction patterns of rats injected ICV with GLP-1 and rats treated with toxins considered to simulate interoceptive stress or visceral illness, such as lithium chloride (Thiele, T.E. et al. 1998).
Peripheral treatment with GLP-1 in rats is not effective at inducing inhibition of food intake (Turton, M.D. et al. 1996; Tang-Christensen, M. et al. 1996), while several experiments with intravenous infusions of GLP-1 did stimulate an increased feeling of satiation, and fullness, with a decrease in hunger and food intake in type 2 diabetic (Gutzwiller, J.P. et al. 1999a; Toft-Nielsen, M.B. et al. 1999), obese (Naslund, E. et al. 1999; Flint, A. et al. 2001), and normal weight humans (Flint, A. et al. 1998; Gutzwiller, J.P. et al. 1999b). Regions exterior to the blood brain barrier, area postrema and subfornical organ have been shown to specifically bind GLP-1 in rats (Goke, R. et al. 1995) and area postrema has been identified as involved in signalling the conditioned taste aversion, caused by lithium chloride treatment in rats (Curtis, K.S. et al. 1994). Reports of GLP-1 peripheral treatments in humans have reported no side effects except if higher concentration of GLP-1 are used, in which case nausea and vomiting were observed (Flint, A. et al. 1998). It is possible that the higher effective concentrations used in ICV injections in rodents stimulates inhibition of food intake through receptors in the area postrema and other regions of the hind brain, thus stimulating taste aversion and indicating a signal of malaise or visceral illness. Peripheral treatments done in rats are usually single bolus injections, whereas in humans infusion usually occurs for several hours, plus circulating GLP-1 has been shown to be degraded exceptionally quickly in rats (Kieffer, T.J. et al. 1995), possibly explaining the discrepancy of effectiveness in

peripheral treatments between rats and humans. This hypothesis may suggest, taken with the fact that PVN injection of GLP-1 in rats does not induce CTA (McMahon, L.R. et al. 1998), that one region of the brain, the hypothalamus, may be involved in signalling satiety and hunger, while another part of the brain, the hind brain, may be involved in signalling malaise, which in turn leads to decreased food intake (van Dijk, G. et al. 1999). Also, if inhibition of food intake were solely due to activation through taste aversion, blockage of the GLP-1 receptor in the hypothalamus should not stimulate food intake. Yet, injection with the GLP-1 specific antagonist, exendin (9-39), increases food intake in rats (Turton, M.D. et al. 1996).

A general hypothesis for the role of GLP-1 in regulation of food intake would also most likely involve the intestinal signal of satiation in which GLP-1 may be one of the mediators of this effect (Read, N. et al. 1994). In regular food consumption the intake of food into the stomach and small intestine leads to the release of GLP-1 (Layer, P. et al. 1995). GLP-1 then binds to receptors in the intestine and stimulates activation of the vagus nerve, which carries the signal to the brain stem of the CNS. Here the signal to inhibit gastric acid secretion and motility is processed (Imeryuz, N. et al. 1997; Rocca, A.S. et al. 1999), but the NTS in the brain stem could also be stimulated to produce and release GLP-1 as a neurotransmitter in neurons that project to the PVN of the hypothalamus (Larsen, P.J. et al. 1997a). GLP-1 then stimulates the inhibition of food intake as the regular signal that the animal is full. If the animal eats a contaminated meal or perhaps eats too much, the hypothesis is that potentially higher concentrations of GLP-1, or at least a different pathway, now signal the CNS in multiple brain stem locations, including the NTS, area postrema (AP) and lateral parabrachial nucleus (PBN), signalling malaise, which may induce vomiting and at the very least reduces food intake. This hypothesis indicates that the pathways involved in regular food intake inhibition due to satiety and food intake inhibition stimulated by malaise are independent pathways (van Dijk, G. et al. 1999), and may primarily rely on the initial stimulus from the intestine for any GLP-1-induced food intake inhibition.

It is possible that not all GLP-1 brain signals come directly from the intestine. It has recently been shown that circulating GLP-1 can cross the blood brain barrier by passive

diffusion (Kastin, A.J. et al. 2002) and enters into several parts of the brain. Some regions of the brain, such as the arcuate nucleus contain GLP-1 receptors, but not many GLP-1 immunoreactive fibers, suggesting a role for circulating GLP-1 in activating these receptors (Kastin, A.J. et al. 2002). Also, the area postrema and subfornical organ have GLP-1 receptors and are outside the blood brain barrier, thus are also available for activation by circulating GLP-1 (Orskov, C. et al. 1996). At this point the extent to which GLP-1 from each source, central or peripheral, plays in the regulation of food intake is unclear, but it is possible they represent parallel pathways both involved in regulation of feeding behavior (Havel, P.J. 2001).

Surprisingly, removal of the GLP-1 receptor in GLP-1R -/- knockout mice does not lead to obesity, suggesting that GLP-1 does not affect long-term food or caloric intake or body mass (Scrocchi, L.A. et al. 1996; Havel, P.J. 2001) and indicates that additional peptides are involved in regulation of food intake, compensating for the loss of GLP-1 signalling. These hormones include CCK, also involved in short-term inhibition of food intake, insulin and leptin, involved in long-term regulation of food and caloric intake related to body weight (Havel, P.J. 2001).

The production, presence and some evidence for inhibition of food intake exists for other proglucagon-derived peptides, namely oxyntomodulin, glucagon, and GLP-2, but these peptides have not been studied in as much detail as GLP-1. ICV or PVN injections of oxyntomodulin result in inhibition of food intake, but addition of the GLP-1-specific antagonist, exendin (9-39), negates inhibitory effect of oxyntomodulin, suggesting oxyntomodulin may have been acting through the GLP-1 receptor (Dakin, C.L. et al. 2001). Peripheral infusion of rats with glucagon also stimulates inhibition of food intake, but if rats are hepatically vagatomized, which breaks the nerve connection between the liver and central nervous system, the inhibitory effects are eliminated (Geary, N. et al. 1993), suggesting glucagon inhibits food intake by binding to liver receptors and stimulating changes in circulating glucose concentrations, which also affects food intake (Havel, P.J. 2001). Rat brain membranes do specifically bind glucagon and stimulate

adenylyl cyclase activity, but at this point the function is unclear (Hoosein, N.M. et al. 1984b).

GLP-2 has been added to the list of candidates for regulation of food intake as it has been found in high concentrations in the hypothalamus (Tang-Christensen, M. et al. 2000), the GLP-2 receptor mRNA and immunoreactivity have also been identified in the hypothalamus (Tang-Christensen, M. et al. 2000; Yusta, B. et al. 2000), intracerebroventricular injection of GLP-2 in rat decreases food intake (Tang-Christensen, M. et al. 2000; Lovshin, J. et al. 2001), and unlike GLP-1 activation of the GLP-2 receptor by central administration of GLP-2 induces c-fos expression only in the hypothalamus (Tang-Christensen, M. et al. 2001). Thus the effects of GLP-2 on the brain may be less complex than the multiple pathway actions of GLP-1. At this point the signals stimulating GLP-2 release in the brain have not been determined, but it is possible, like GLP-1, to originate from the gastrointestinal tract.

Obviously the distinct roles of GLP-1 and GLP-2 in food intake regulation are not yet clear, but there seems to be crosstalk between the two peptides and their receptors. Considering that both peptides are produced together and both receptors have been found in the same brain tissues and perhaps even the same neural cells, this system may be difficult to clarify.

Non-mammalian Vertebrates

A role of GLP-1 in regulation of food intake is also observed in some non-mammalian vertebrates, as central treatment of chickens (Furuse, M. et al. 1997) and channel catfish (Silverstein, J.T. et al. 2001) with GLP-1 leads to decreased food intake. GLP-1 immunoreactivity is demonstrated in several parts of the channel catfish brain, including the PVN of the hypothalamus (Sarkar, S. et al. 2001), and proglucagon mRNA has been isolated from tiger frog brain (Yeung, C.M. et al. 2001), indicating presence and production of GLP-1 in non-mammalian brain. Treatment of isolated copper rockfish brain membranes with GLP-1 leads to increased production of cAMP (Mommsen, T.P. et al. 1998), demonstrating receptors and a transduction pathway for GLP-1 in non-mammalian brain. Although other proglucagon-derived peptides have not yet been

studied in non-mammalian vertebrates and the entire role of GLP-1 is still uncertain, at least one aspect of GLP-1 function is conserved between mammals and other vertebrates, suggesting similar functions of the proglucagon-derived peptides may also exist in non-mammalian vertebrates.

Regulation

Regardless of whether, mammalian or other systems are considered, three distinct hormones, with different functions, are being produced by the same prohormone, potentially at the same time. Considerable regulation of this system must occur in order for it to function efficiently and effectively, especially considering the varying functions and importance of their roles in carbohydrate metabolism. Thus it is essential that regulation of these hormones is successful, and in turn, not surprising that proglucagon and the peptide hormones derived from it are regulated on many levels.

At the peptide level

N-terminal extension

In mammals, and possibly birds, reptiles and amphibians, the incretin function of GLP-1 opposes the glucose stimulatory function of glucagon. Considering they are produced from the same precursor, a regulatory mechanism exists to prevent the role of one from negating the role of the other. GLP-1 is processed and initially released in an N-terminally extended form, consisting of an additional six amino acids. This form of GLP-1 is biologically inactive, and additional convertase processing is necessary to release the truncated, biologically active GLP-1 (7-37). This adds another level of regulation to the activity of GLP-1 allowing versatility necessary for these two opposing peptides.

In fish, where the function of glucagon and GLP-1 are similar, with both stimulating the release of glucose and opposing insulin, the role of the N-terminal extension become unnecessary, and it is not found in the proglucagon sequence (Irwin, D.M. et al. 1995; Lund, P.K. et al. 1983).

Differential Processing of Prohormone Precursor

Processing of proglucagon into mature active peptides relies on action of prohormone convertases (PCs), a family of enzymes responsible for maturation of most prohormones, including proinsulin, POMC, prosomatostatin, progastrin, proneurotensin, and proneuropeptide Y (Dhanvantari, S. et al. 1996). In mammals, this family consists of seven subtilisin-like pro-protein convertases (SPC), furin, PC1 (also known as PC3), PC2, PC4, PACE4, PC5/6 (PC5/6a), and PC7/8 (PC5/6b), which are related to kexin serine proteases. These enzymes cleave peptides primarily at dibasic cut sites, as seen in proglucagon flanking glucagon, GLP-1 and GLP-2, but some can also cut at single basic amino acids, such as the arginine in position 6 of GLP-1. When discussing proglucagon, PC1/3 and PC2 are of interest as they are primarily expressed in neural and endocrine tissue (Brakch, N. et al. 2000).

Differential prohormone convertase processing of proglucagon is seen in mammalian tissues, as different proglucagon derived peptide profiles are found in pancreas and intestine. Pancreatic α-cells secrete primarily glucagon and the major proglucagon fragment, with very small amounts of oxyntomodulin and N-terminally extended GLP-1 (Patzelt, C. et al. 1984; Holst, J.J. et al. 1994), while intestinal L-cells secrete oxyntomodulin, glicentin, GLP-1 and GLP-2, but not glucagon (Tager, H.S. et al. 1979). The production of different profiles is mostly due to expression of different PCs in each tissue. Pancreatic α-cells contain PC2, while intestinal L-cells express primarily PC1/3 (Rouille, Y. et al. 1997a).

Pulse chase experiments followed by immunoprecipitation have determined some of the cleavage sites specific for each mammalian convertase, PC1/3 and PC2. The first cleavage to occur in proglucagon, regardless of enzyme, is at position Arg70 Arg71, between IP-1 and GLP-1, resulting in glicentin and the major proglucagon fragment (MPFG). This position seems to be the most available cut site, as almost any PC can successfully cleave in this position (Dhanvantari, S. et al. 1996). Following this initial cleavage, PC1/3 in intestinal L-cells, releases GRPP and oxyntomodulin from glicentin and biologically active GLP-1 and GLP-2 from the MPGF. This study settled a debate by

demonstrating that PC1/3 can cut at monobasic cut sites without assistance of another convertase or cofactor (Rouille, Y. et al. 1997b). PC2, in pancreatic α-cells, follows the initial cleavage with immediate processing of glicentin to glucagon through a short lived oxyntomodulin intermediate and MPGF is not further processed in the pancreas (Rouille, Y. et al. 1997a).

Besides the basic amino acids, there are no amino acids surrounding the dibasic sites that act as a consensus sequence for recognition of cleavage. Important factors in substrate recognition have been investigated for PC1/3 and appear to primarily involve the secondary structure of the peptide substrate. Opposing direction of the side chains of the two basic amino acids in the recognition sequence, as well as the presence of a β-turn in the region of cleavage and peptide flexibility were found necessary for optimum cleavage (Brakch, N. et al. 2000). Research with PC1/3 notwithstanding, understanding of convertase specific recognition of different dibasic cut sites leaves something to be desired.

PC2-type enzyme cDNAs have been sequenced in non-mammalian vertebrates and invertebrates such as frogs, anglerfish, chicken, the California sea hare, the green bottlefly and *C. elegans*. For some of these species, PC1 and furin-type enzyme sequences have also been described in nucleotide databases (Genbank). But characterization of these enzymes has only begun in a few species, including a frog *Rana ridibunda* and a teleost *Lophius americanus* (Gangnon, F. et al. 1999; Vieau, D. et al. 1998; Mackin, R.B. et al. 1991). In anglerfish, PC activity with similarity to PC2 has been described in the endocrine pancreas. Considering fish secrete GLP-1 as well as glucagon from the endocrine pancreas, it would be unlikely to find PC activity with the same site recognition characteristics as in mammalian PC2, which does not release processed GLP-1. Since teleosts release different patterns of peptides in both the endocrine pancreas and the intestine compared to mammals, it is expected the fish may have one or more prohormone convertases that are unique in cut site recognition, but still similar in catalytic function to the mammalian convertases. Alternatively, fish may merely possess additional convertases in addition to mammalian-like convertases.

Prohormone convertases play an important role in regulation of proglucagon derived peptides as they control maturation of the peptides. In some cases this means processing that leads to the absence of a peptide, as with GLP-1 (7-37) in mammalian pancreas, in others it is the release of one form of a peptide instead of another, as with oxyntomodulin, not glucagon, in intestine.

In addition to PCs, carboxypeptidase E (CPE also known as carboxypeptidase H) is involved in the maturation of proglucagon-derived peptides and many other peptide hormones and neurotransmitters. CPE removes the remaining two dibasic amino acids from peptides liberated by the prohormone convertases, which cleave prohormone precursors after the dibasic cut sites (Fricker, L.D. 1988). In CPE-deficient mice, the formation of mature amidated GLP-1 is significantly reduced, and maturation of proglucagon-derive peptides is inhibited in both intestine and pancreas (Friis-Hansen, L. et al. 2001). This interruption of CPE activity could result in incomplete peptide processing and disease.

Peptide Degradation

Another way to regulate this system is to remove one or more of the three peptides produced from the tissue of synthesis by degradation. Dipeptyl peptidase (DPP-IV) is an enzyme that cleaves GLP-1 and GLP-2 at the alanine in position 2, resulting in a biologically inactive peptide fragment. In rats, in particular, a high rate of degradation is observed in the plasma for GLP-1 and GLP-2. The half-life of GLP-1 in mammalian plasma is about 1 minute, while the half-life of GLP-2 is longer, at 7 minutes (Burrin, D.G. et al. 2001). In humans the rate of degradation in plasma after subcutaneous injection of GLP-2 is decreased with about 69 % remaining after one hour (Hartmann, B. et al. 2000). As both GLP-1 and GLP-2 have been considered hopeful treatments for type 2 diabetes mellitus and intestinal insufficiency (Drucker, D.J. 2001), respectively, effort has been put into the development of peptide analogs that are resistant to DPP-IV degradation. As a result, substitution of the alanine in position 2 of GLP-1 to a glycine prevents DDP-IV degradation thus increasing half-life potency of the peptides (Xiao, Q.

et al. 2001). Study of GLP-2 degradation products in human plasma has identified GLP-2 (3-33) as the only circulating form, suggesting DPP-IV is the primary degradation enzyme (Brubaker, P.L. et al. 1997). Yet, the kidney and liver are major factors involved in clearing peptides from circulation including GLP-1 and GLP-2, adding to DPP-IV action in the short half lives of these peptides (Plisetskaya, E.M. et al. 1996).

In some cases, the degradation of each peptide in a specific target organ is more important in regulation of proglucagon-derived peptides. Differential metabolism of GLP-1 has been demonstrated in pig tissues, suggesting that different tissues have different abilities to clear GLP-1, including DPP-IV concentrations (Deacon, C.F. et al. 1996).

Whether research on peptide degradation in mammals is applicable to non-mammalian vertebrates is unknown. Considering that position 2 is not conserved in these peptides within non-mammalian vertebrates, specificity of the DPP-IV equivalent enzymes must be altered to accomodate the different residues. As clearance of peptide hormones is an essential part of regulating their effect, DPP-IV counterpart enzymes most likely exist in non-mammalian vertebrates.

At the transcript level

In some non-mammalian vertebrates, regulation of peptide production is facilitated through two different forms of mRNA alternative splicing. In teleosts, chicken and a reptile, alternative splicing involves the retention of the intron following GLP-1 (Figure 5)(Irwin, D.M. et al. 1995; Chen, Y.E. et al. 1997). This intron contains a polyA signal thus retention of it leads to truncation of the mRNA, and prevents the expression and subsequent translation of GLP-2 in the pancreas. In two amphibian species, tiger frog (*Rana tigrina rigulosa*) and *X. laevis* (Yeung, C.M. et al. 2001; Irwin, D.M. et al. 1997) alternative splicing involves splicing out of an exon, containing either GLP-1 or GLP-2, preventing expression of that particular peptide. In this case, alternative splicing does not lead to a truncated mRNA, thus the 3' UTR of the alternatively spliced and non-alternatively spliced mRNAs are identical in sequence. When an intron is retained, this is

not the case as the intronic sequence then becomes the 3'UTR of the alternatively spliced mRNA. No evidence of alternative splicing in mammals has been reported.

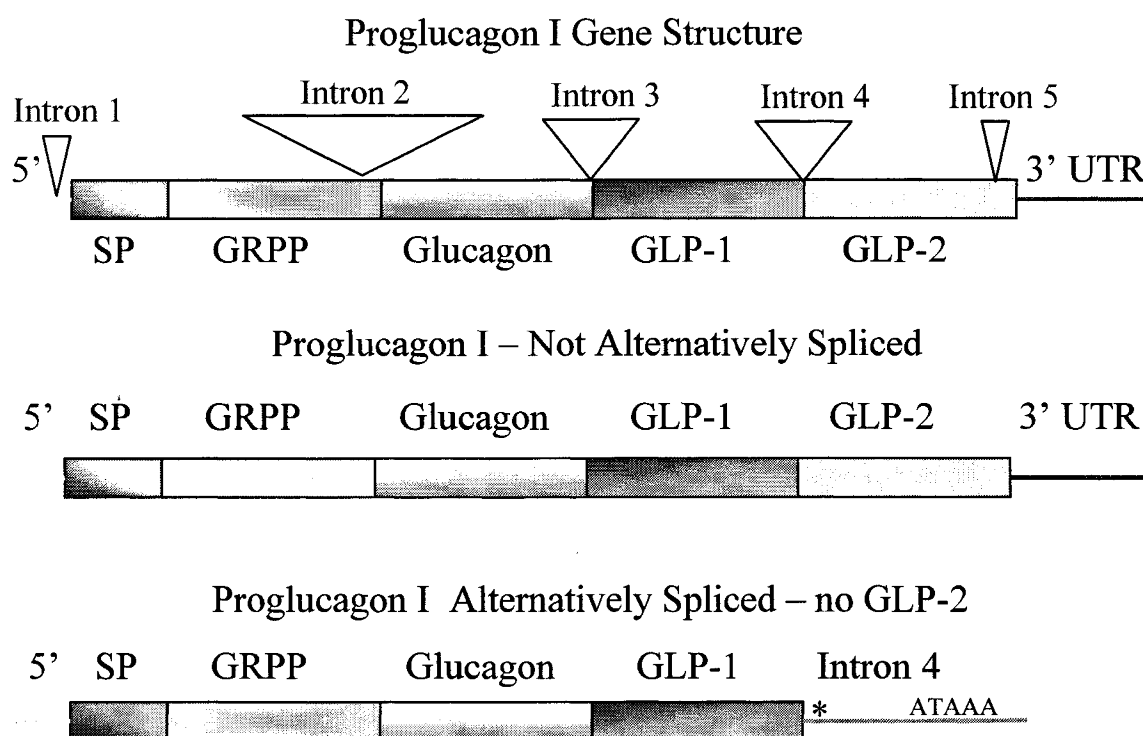

Figure 5. Schematic representation of alternative splicing in teleosts.

At the gene level

Transcription of the proglucagon gene is obviously necessary for the production of the peptides. Although regulation of expression cannot differentially affect the individual proglucagon-derived peptides, there are factors that affect transcription in a tissue specific manner. In the mammalian proglucagon gene, two of these factors are cAMP and insulin. Proglucagon gene expression is stimulated by activators of the protein kinase-A pathway (Drucker, D.J. et al. 1994), which is facilitated through a cAMP response element identified in the promoter region of the proglucagon gene (Knepel, W. et al. 1990). Considering that transcription, translation and secretion of proglucagon-derived peptides activate G_{α}-protein coupled receptors that lead to the production of cAMP, increased levels of cAMP in will lead to induction of the proglucagon gene in either an endocrine or paracrine fashion. Thus after the initial signal for proglucagon-

derived peptide production, peptide action will upregulate its own gene within that tissue. The production of GLP-1 in intestine leads to an increased release of insulin in the pancreas. Insulin triggers down-regulation of the proglucagon gene in pancreas (Philippe, J. 1991), leading to the inhibition of glucagon release. Nutrient-regulated gene expression has also been identified for the proglucagon gene in intestine (Hoyt, E.C. et al. 1996), and although it has been determined that nutrients don't directly stimulate expression, the pathway of activation remains uncertain (Nian, M. et al. 2002).

As mammalian proglucagon is encoded by a single gene, regulation is restricted to induction or inhibition of transcription of that gene. In contrast, in animals encoding two independent genes, such as the anglerfish or polyploidy species, expression of one gene without the other in different tissues provides an additional form of regulation.

Receptors

In certain ways, the presence of a peptide-specific receptor is the ultimate form of regulation, as it is a necessary part of peptide hormone signalling. In the proglucagon family, three distinct peptide hormones exist that show relatively little cross-reactivity to the other receptors (Moens, K. et al. 1996; Munroe, D.G. et al. 1999). Receptors for glucagon, GLP-1, and GLP-2 have been cloned, sequenced, and characterized in mammalian systems (Jelinek, L.J. et al. 1993; Thorens, B. 1992; Munroe, D.G. et al. 1999). They are all G-protein coupled glycosylated receptors characterized by seven transmembrane domains. Glucagon, GLP-1 and GLP-2 receptors are each linked to the G_s type α subunit, which stimulates activation of adenylyl cyclase and subsequently increases cAMP production (Jelinek, L.J. et al. 1993; Graziano, M.P. et al. 1993; Yusta, B. et al. 1999). Yet, some functions of GLP-1, such as the release of insulin, pancreatic differentiation and possibly a role for GLP-1 in pituitary, appear to be mediated via a PI3-kinase-dependent pathway (Buteau, J. et al. 1999; Satoh, F. et al. 2000; Wheeler, M.B. et al. 1993; Flamez, D. et al. 1999), correlated with an increase in intracellular calcium. Evidence suggesting a similar second pathway exists for the glucagon receptor, where interactions of low concentrations of glucagon with the receptor increase the concentrations of inositol triphosphate, and stimulate glycogenolysis and

gluconeogenesis in a cAMP-independent pathway (Wakelam, M.J. et al. 1986). Therefore, both the glucagon and GLP-1 receptor have the ability to couple to more than one G-protein.

Although increases in cAMP are usually associated with subsequent activation of protein kinase A, as with glucagon signalling, stimulation of insulin synthesis by GLP-1 and intestinotrophic effects of GLP-2 are mediated by distinct cAMP driven, protein kinase A-independent pathways (Skoglund, G. et al. 2000; Yusta, B. et al. 1999). It has become clear that this group of related receptors mediates their signals through a plethora of mechanisms, utilizing different G-proteins, leading to production of several intracellular messengers that can stimulate multiple downstream pathways, not all of which are identified.

Considering the complexity of these receptors, it is realistic to suggest that changes seen in function of the proglucagon-derived peptides among the vertebrates might be facilitated by changes in the receptors as well as changes in the peptide. In fact, the peptides are so highly conserved throughout vertebrates, it is likely that most of the functional diversity, especially in GLP-1 where a complete switch of function occurs between mammals and fish, is dependent on the receptors.

The only non-mammalian vertebrate glucagon receptor cDNA sequences identified to date are all from amphibians, *Xenopus laevis*, the leopard frog, (*Rana pipiens*) and the tiger frog (*Rana tigrina rigulosa*) (Sivarajah, P. et al. 2001; Ngan, E.S. et al. 1999). No GLP-1 or GLP-2 receptor sequences have yet been identified. Nevertheless, binding studies, target enzyme assays and detection of intracellular messages, such as cAMP and IP_3, have provided some information about glucagon, GLP-1, and GLP-2 receptor activation in target tissues of non-mammalian vertebrates.

In teleost hepatocytes, increases in cAMP are only observed at supraphysiological concentrations of glucagon, due to a lower sensitivity of fish liver to glucagon, yet evidence of activation of glycogenolysis by low concentrations of glucagon did not increase cAMP levels (Plisetskaya, E.M. et al. 1996), and does increase IP_3

concentrations (Moon, T.W. et al. 1997). This implies an IP_3/intracellular calcium message transduction pathway, may function at lower glucagon concentrations, as suggested for mammals (Navarro, I. et al. 1999). Interestingly, in two teleost species, the American eel (*Anguilla rostrata*) and the black bullhead catfish (*Ictalurus melas*), some receptors with low glucagon binding affinity and others with high binding affinity were identified. Possibly these distinct glucagon receptor groups may represent receptors linked to different signalling pathways. Additionally, glucagon binding to chicken hepatocytes occurs at lower affinity than in mammals, yet chicken heptaocytes demonstrate increased sensitivity to glucagon compensating for the low binding affinity (Navarro, I. et al. 1999).

A similar signalling scenario exists for the GLP-1 receptor in fish. Supraphysiological concentrations of GLP-1 are necessary to stimulate cAMP production, yet in this case, a cAMP-dependent protein kinase A antagonist reduced the glycogenolytic effect of GLP-1, suggesting that cAMP may play an indirect role in signalling at physiological GLP-1 concentrations (Plisetskaya, E.M. et al. 1996). Although IP_3 concentrations are not increased in teleost GLP-1 signalling, PLC activity is, suggesting this pathway may also be involved in signalling (Moon, T.W. et al. 1997). As most of the functions for GLP-1 in teleosts stimulate adenylyl cyclase, GLP-1 most likely acts primarily through this pathway, with the secondary pathway involving increases in intracellular calcium (Mommsen, T.P. et al. 2001).

To date, no evidence of GLP-2 receptors or activation of GLP-2 receptors has been presented for non-mammalian vertebrates.

Molecular Evolution

A phylogeny is the pattern of historical relationships between existing species or taxa, while a phylogenetic tree is an attempt to estimate the phylogeny using mathematical models. Reconstructing the phylogenetic relationships of a group of taxa is one of the main themes of molecular evolution, as molecular sequences, DNA and protein, represent the information used to estimate these evolutionary relationships. Many mathematical

methods exist for estimations of relationships, including two discrete methods, parsimony and maximum likelihood.

Discrete methods consider each nucleotide site (or a function of each site), and can interpret the nucleotide state of extinct ancestral nodes between extant taxa, unlike distance methods, which calculate pairwise distances in order to determine the minimum distance between sequences, but lose specific nucleotide site information. Thus, discrete methods are more powerful in predicting evolutionary trees. For this analysis, I have applied the maximum likelihood method, as it has advantages over the parsimony method. The maximum likelihood method applies an explicit model of evolution, chosen by the user, to compute likelihoods of substitutions occurring at a site to predict the most likely phylogeny (Page, R.D.M. et al. 1998). Parsimony uses a model with equal probability of changes at sites and draws trees with the fewest possible evolutionary changes between nodes. Thus, parsimony methods are simple and computationally easy, but may not be as accurate when rates of substitution are high or unevenly distributed over branches, which is more likely to occur when analyzing sequences spanning large evolutionary times and diverse groups of taxa, as is done here. Also, maximum likelihood can account for the possibility of multiple substitutions at one site between sequences, which is also likely to occur when analyzing taxa spanning large evolutionary times. The largest drawback to the maximum likelihood method is the almost prohibitive computational time needed to analyze large data sets. For this reason, I used a combination of methods. Maximum likelihood was used to calculated relationships and generate a distance matrix for each analysis, then a distance method, the neighbor-joining method was used to draw the trees.

In visual presentations of phylogenetic analyses, rooted or unrooted trees can be used. A rooted tree has a node identified as the root or ancestor, which applies a direction of time to the tree, allowing definition of ancestor-descendant relationships. In order to root a tree, alternate information about the evolutionary history is necessary or sequence from an outgroup taxon. In an unrooted tree, the branches indicate evolutionary relationships,

but do not define an evolutionary path or make assumptions about common ancestors. (Graur, D. et al. 2000).

Included in phylogenetic analyses is the use of an indicator of sampling error such as bootstrapping. This is a technique of pseudoreplication where the sites of a data set are randomly resampled and subjected to the phylogenetic analysis. As selection of each site in the random sampling is done on the complete number of sites, some sites might be resampled more than once, possibly changing the frequency of each site, and altering incidences of sampling error that may be inherent in the sequence. As the pseudoreplication in bootstrapping is repeated many times (between 100 and 1000), the consensus bootstrapping tree gives the investigator an indication of precision of the phylogenetic tree. If a certain branch exists in 999/1000 bootstrap trees, the confidence of precision of that particular node is strong. Some say only values above 95% are significant for strong confidence (Felsenstein, J. 1993) , but normally, values above 70% are accepted as instilling good confidence into the reader. Values below 50% are normally not included and the position of the branch should be interpreted very cautiously.

Applying all of the methods and techniques of phylogenetic analysis in molecular evolution, including proglucagon, a researcher can most likely accurately estimate the phylogeny of a group of taxa, yet a phylogenetic tree is merely a hypothesis of evolutionary relationships and will be truly accurate only when all evolutionary information for a group of species is identified, which is not likely to occur.

Research Objectives

As multiple proglucagon genes had been suggested for teleostean anglerfish (*Lophius americanus*), one of the objectives of this research was to determine if other teleostean species, specifically the copper rockfish (*Sebastes caurinus*) and channel catfish (*Ictalurus punctatus*) also encod and express multiple proglucagon genes. Once the presence of two transcripts in these fish was determined, I wanted to characterize tissue distribution of this expression and determine if the two genes were differentially regulated. My hypothesis was that expression of the two proglucagon geneswould be differentially regulated and that tissue distribution may vary under different physiological conditions. Although I wasn't able to show differential expression in starved versus fed catfish, differential tissue distribution of the two proglucagon genes was demonstrated in rockfish, suggesting the possibility of differential functions of the peptides derived from the two genes. Analysis of peptide production was performed on endocrine pancreas and demonstrated translation of both proglucagon genes in both rockfish and catfish. Unfortunately, partial peptide isolation and analysis from other proglucagon expressing tissues, brain and intestine, proved unattainable with the resources available.

Another objective of this research was to clarify the evolutionary changes in proglucagon structure with respect to function and phylogenetic relationships throughout vertebrates. Between mammals and teleost fish, the function of GLP-1 changes from insulinotropic to glycogenolytic and gluconeogenic, a complete reversal in function. With such a drastic change in function, evolutionary changes in the proglucagon sequence and structure were expected. By identifying proglucagon sequences from vertebrate groups not previously studied and performing phylogenetic analysis on the sequences, I hoped to reveal evolutionary mechanisms and relationships that would shed light on the evolutionary history of proglucagon.

Although not all of my research objectives have been fully completed, most have been successfully fulfilled and in the process have exposed unexpected aspects of the sequence, expression, regulation and evolution of the proglucagon genes.

Finally, any research in the field of proglucagon regulation is beneficial as the peptide hormones encoded in the gene are involved in critical aspects of healthy physiology. Glucagon is one of the main players in glucose metabolism, yet compared to insulin has been overlooked for many years. In recent years the full import of glucagon and the glucagon-like peptides has been touched on. As GLP-1, in mammals, stimulates the release of insulin, recent research has shown it may prove to be an excellent treatment for diabetes, a disease that affects over 2 million Canadians. Additionally, GLP-2 functions as an intestinal growth factor, and also may prove to be useful for treatment of patients recovering from bowel resections or fighting with inflammatory bowel disease. Currently, both of these peptides are involved in drug trials, with GLP-1 at level-2 testing for treatment of type 2 diabetes. Obviously, any additional information concerning all aspects of these three peptides, especially regulation, will contribute to our overall understanding of proglucagon and its role in metabolism.

Methodological Approach

A large portion of this research involved identification of proglucagon sequences from species not studied before. The primary mechanism of obtaining sequence applied in this work was the polymerase chain reaction (PCR), especially using mRNA as the template, known as reverse transcription (RT)-PCR, followed by DNA cloning and sequencing. As a first approach to a new sequence, degenerate primers, designed from the most closely related proglucagon sequences available, were utilized in RT-PCR reactions to amplify fragments of the proglucagon message. Although RT-PCR using degenerate primers can be tricky and usually involves several attempts at designing primers, the teleost sequences proved to be the least difficult, most likely as there were teleost proglucagon cDNA sequences available. The uniqueness of sequence, lack of closely related reported sequences, and low availability of animals and tissues made identification of proglucagon in the lungfish, elasmobranchs and even cane toad much more of a challenge. For these sequences many sets of degenerate primers were designed, combined with different types of PCR programs, such as touchdown PCR. Many of these attempts are not reported in this document as they were not successful. Some of these sequences were elusive for more than a year.

Once partial sequence was obtained, gene specific primers were designed and used in 3' and 5' Rapid Amplification of cDNA Ends (RACE)-PCR reactions, to provide complete coding sequences. As all the PCR products were cloned before sequencing, high quality sequences were provided with very few *taq* polymerase errors. Nevertheless, to ensure the accuracy of the sequences, clones from multiple independent PCR reactions were sequenced and compared. For the teleosts, at least three and often many more, and for all other sequences, at least two independent sequences were used to compile the reported consensus sequences. These sequences were then used to design gene specific primers for tissue screening, used to generate a DNA probe for the Southern blot, or used in the phylogenetic analysis.

Chapter 1 – Regulation of expression and translation of proglucagon genes in Copper Rockfish: variations on a theme.

Abstract

Proglucagon expression has been extensively studied in mammals, and proglucagon-derived peptide isolation has been performed, primarily on endocrine pancreas, in a vast number of non-mammalian vertebrates, including several different groups of fish. Yet, proglucagon expression research in teleosts has been limited, with only two species, anglerfish (*Lophius americanus*) and rainbow trout (*Oncorhynchus mykiss*) in which any mRNA work has been done. Thus, this chapter investigating proglucagon mRNA sequence, expression, and regulation, as well as evidence for peptide production in endocrine pancreas in the teleost, copper rockfish (*Sebastes caurinus*) is a welcome addition to the field. In copper rockfish, at least two independent genes encoding distinct proglucagon sequences were identified. These genes are differentially transcribed in endocrine pancreas, brain, and the gastrointestinal tract, and evidence of alternative splicing in one transcript increases the regulatory complexity of proglucagon transcription. Peptide isolation and mass spectrometric analysis of proglucagon-derived peptides in endocrine pancreas demonstrates both transcripts are translated and differential regulation of the peptides produced occurs. The complex differential regulation of the two proglucagon genes in copper rockfish suggests that individual roles may exist for the peptide products encoded in each gene.

Introduction

The vertebrate proglucagon gene encodes three main functional peptide hormones in a single transcript, namely glucagon, glucagon-like peptide 1 (GLP-1), and glucagon-like peptide 2 (GLP-2). This transcript is translated and proteolytically processed to release a combination of these functional peptides, depending on processing, regulation and tissue distribution of expression. In mammals, significant amounts of proglucagon transcripts are found in three tissues, the intestine, the endocrine pancreas and the brain. As is true for other peptides, transcription does not necessarily coincide with translation and in the

case of proglucagon, selective processing may lead to a plethora of combinations of peptides and fragments.

The main role of glucagon, in fish and mammals, is liberation of glucose from the liver, either from endogenous glycogen or through increased gluconeogenesis, with the goal to increase availability of glucose in the circulation. Other roles are not as well defined in function or locale, but include control of insulin production and release, decrease of lipolysis, and increase of amino acid metabolism (Plisetskaya, E.M. et al. 1996).

The function of GLP-1 varies among vertebrate groups. In contrast to mammals, where GLP-1 opposes glucagon by acting as an incretin hormone, fish GLP-1 complements glucagon by increasing blood glucose levels, as well as enhancing intestinal glucose transport. In all vertebrates analyzed to date, and this includes fish, the hormone may also play a role in food intake inhibition in the brain (Turton, M.D. et al. 1996; Furuse, M. et al. 1997; Silverstein, J.T. et al. 2001). In mammals GLP-2 has been identified as an important intestinal growth factor (Drucker, D.J. et al. 1996), and *in vitro* activation of brain membrane adenylyl cyclase has been described (Hoosein, N.M. et al. 1984a), while its role in fish has yet to be defined.

It has been shown for several vertebrate species, including teleosts, chicken, and a reptile, the gila monster, that proglucagon expressed in the endocrine pancreas does not include the sequence for GLP-2 (Lund, P.K. et al. 1983; Hasegawa, S. et al. 1990; Chen, Y.E. et al. 1997). Using rainbow trout as a model, Irwin (1995) indicated that the proglucagon transcript undergoes alternative splicing of the intron following GLP-1 (cf. Figure 5), which results in the retention of the intron and truncation of the prohormone after GLP-1. Considering this intron begins with a stop condon and contains a poly A signal, a mature mRNA still results. The subsequent transcript and consequent proprotein product, in Brockmann body, includes glucagon, GLP-1, but not GLP-2.

In birds, reptiles, and mammals, proglucagon is encoded by a single gene. So far, only one amphibian, *Xenopus laevis*, a tetraploid species, has been shown to possess two genes

for proglucagon. The situation in 'fish' may differ. For instance, there is mRNA sequence evidence suggesting the presence of two proglucagon genes in diploid anglerfish (Lund, P.K. et al. 1983) and lamprey (Irwin, D.M. et al. 1999).
One of the objectives of this study was to determine if duplicate proglucagon genes may be a common trait of all fish, or more specifically, all teleosts. Peptide isolation has provided some support for duplicate genes as more than one glucagon peptide has been isolated from several teleostean species.

In this chapter, I present mRNA and genomic DNA evidence for the existence of at least two proglucagon genes in teleosts, using the diploid marine copper rockfish (*Sebastes caurinus*) as the preferred model. As shown previously, copper rockfish liver cells respond sensitively to hormones (Busby, E.R. et al. 2002), including glucagon (Danulat, E. et al. 1990) and GLP-1 (Mommsen, T.P. et al. 1998). In the following, I describe several unique characteristics indicated in the genomic sequence of one of the rockfish proglucagon genes, and implication of multiple glucagons for some teleostean species. Further analysis of mRNA tissue distribution for the rockfish proglucagon genes reveals that these transcripts undergo alternative splicing, as well, other forms of transcript regulation. Finally, peptide extraction and mass spectroscopy analysis confirm production of peptides predicted by the mRNA analysis and demonstrate the surprising combinations of functional peptides produced in the endocrine pancreas of the rockfish.

The chapter on the rockfish scenario is followed by a brief chapter presenting mRNA evidence for two independent proglucagon sequences in another diploid teleost, the channel catfish, *Ictalurus punctatus* (Teleostei: Ictaluridae).

Methods

Fish

Copper rockfish (*Sebastes caurinus*) were caught by hook and line in Georgia Straight, BC, and maintained at the University of Victoria aquatics facility at 12°C, 2.9% salinity, under natural light conditions. Fish were killed by cervical dislocation and tissues were immediately sampled and frozen in liquid nitrogen. Gastrointestinal tissues were rinsed with saline prior to freezing.

Total RNA isolation

Tissues were flash frozen in liquid nitrogen or on dry ice and stored at -80°C. Alternatively, thin slices of freshly sampled tissue were stored in >5x volume of RNALater (Ambion) at -20°C. Between 10 and 100 mg of frozen tissues were ground under liquid nitrogen and the frozen powder was transferred to 1.0 ml of Trizol (Sigma) in a 1.5 ml microcentrifuge tube (MCT). Alternatively, between 10 and 100 mg of tissues stored in RNALater were homogenized in 1 ml of Trizol using a Dounce homogenizer. Samples were inverted and vortexed to fully resuspend frozen powder then left standing at room temperature for ten to fifteen minutes. 200 μl of chloroform were added and samples were mixed by inversion until the phases were uniform, which usually took 1 to 2 minutes. After 3 to 5 minutes at room temperature, samples were centrifuged at 12,000 x *g* for 10 to 15 minutes at 4°C. The aqueous phase was then transferred to a new 1.5 ml MCT and either 1 ml or 500 μl of isopropanol, for frozen or RNALater tissue, respectively, was added. Samples were inverted to mix and incubated at room temperature for 10 minutes or overnight at -80°C. Precipitated RNA was sedimented at 12,000 x *g* for 15 minutes at 4°C. The supernantant was decanted and the pellet was washed with 1 ml of 70% ethanol, followed by a 5 min centrifugation at 7500 x *g* at 4°C. The supernantant was decanted, tubes were tapped to remove excess ethanol and pellets were left to air dry at room temperature for 10 to 15 minutes. Pellets were not allowed to come to complete dryness. Pellets were dissolved in 5-50 μl of diethyl pyrocarbonate (DEPC)-treated sterile distilled H_2O (sdH_2O). Total RNA preparations were stored at -80°C.

cDNA synthesis

The cDNA synthesis reactions contained: up to 5 ug total RNA, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 0.8 mM dNTPs, 1.4 μM anchored polyT primer (AnchorT), 8 mM DTT, 5U of Ribonuclease inhibitor (GibcoBRL/Invitrogen), and 100U of SuperScript II RNase H^- Reverse Transcriptase (GibcoBRL/Invitrogen).
Total RNA was incubated with dNTPs and anchored polyT primer for 10 minutes at 65°C to reduce secondary structure, then immediately put on ice. The buffer, DTT, and RNase inhibitor were added and the samples were incubated at 42°C for 2 minutes. SuperScript II was then added and samples were incubated for 60 to 90 minutes at 42°C. After incubation, samples were heat denatured at 75°C for 15 minutes.

PCR

Primarily, the PCR cycles used in this work varied only in annealing temperatures and cycle times. The PCR reaction ingredients and volume (25 μl) remained constant as: 20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 2 mM dNTPs, 1 μM of each degenerate primer or 0.4 μM each non-degenerate primer, 0.025 U/μl (0.625 U/25μl) *taq* DNA polymerase, and 0.5 μl of template. PCR reactions were performed on either a DNA Engine™ thermal cycler (MJ Research Inc.), a PE9600 thermal cycler, a PE2400 thermal cycler (Perkin Elmer), or a Flexigene thermocycler (Techne). PCR products were visualized on 1% agarose gels stained with ethidium bromide. A 100 bp ladder (NEB) was used as a DNA marker.

PCR programs

Proglucagon teleost degenerate primer PCR program:

This program involves two rounds of amplification. The first round is performed with primers ZGluU1 (sense) and ZGluR4 (antisense). The second round is the same program as the first round, but uses 0.5 μl of the first round reaction product as template with primers ZGluU3 (sense) and ZGluR4 (antisense).

Initial denaturation: 94°C for 5 min
5 x denaturation: 94°C for 1 min
annealing: 50°C for 1 min
elongation: 72°C for 1:10 min
31 x denaturation: 94°C for 1 min
annealing: 55°C for 1 min
elongation: 72°C for 1:10 min
Final elongation: 72°C for 15 min

Standard PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 55°C for 30 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Rockfish Screening PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 58°C for 30 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Some 3'end sequences were obtained with a modified 3' RACE PCR technique in which we used an 'anchored' polyT primer for cDNA synthesis, then performed PCR reactions, using the Standard PCR program, a gene specific sense primer and the anchor (antisense) primer. Alternatively, standard 3' RACE PCR was performed with the 5' RACE PCR reactions using the kit protocol and usually a modified version of the kit PCR program.

Primers

Table 1. Names and sequences of primers used in the identification of copper rockfish proglucagon sequences.

Primer Name	Orientation	5'→ 3' Primer Sequence	Purpose
ZGlucU1	Sense	AARAGICAYTCIGARGGIACITT	Initial PCR
ZGlucU3	Sense	GGIACITTYTCIAAYGAYTAYAGYAARTA	Initial PCR
ZGlucR4	Antisense	TCRCTIGTRAARCTICCRTCIRCRTG	Initial PCR
progluc 32	Antisense	CCTGCTGAATGCCTCACCTCGCCA	PG I 5' RACE
progluc 33	Antisense	GGTCTTCAGCCAGTACACAAACTCCTTGGC	PG II 5' RACE
GluII51	Sense	CCTGGAGGAAAGGAAGGCGCAG	PG I 3' RACE
GluI51	Sense	TATCTGGAGACGAGAAGACGACAA	PG II 3' RACE PCR Walking Probe generation
Glu52	Sense	CTGAAACGGAATGGCGAGGTGAG	PG I 3' RACE
progluc51	Sense	GACGACCGACGTGAGCTCCTACCT	PCR Walking
progluc31	Antisense	GGCTCAAGTCTGGACAAGTCAGAAGAG	Probe generation
5UTRII2	Sense	GGAGAGAAAGAAGACAAGGTTAT	PG I Screen
3UTRII2	Antisense	TATATGGCTCTGTGTGCGGC	PG I Screen
3UTRII	Antisense	CTGGAATCAGACACAGGCCGATA	PG tI Screen
5UTRIII	Sense	CGGGCACCAGTCACCTACTT	PG II Screen
3UTRIII2	Antisense	TCACACAGAGCCAGTATGAATTG	PG II Screen
5' Actin	Sense	GTCGTCGACAACGGCTCCGGCATGTG	Actin Screen
3' Actin	Antisense	CATTGTAGAAGGTGTGGTGCCAGAT	Actin Screen

Table 2. Non-specific primers used in 5' and 3' RACE and PCR walking reactions.

Primer	Sequence 5'→ 3'
AnchorT	$GGCTCGAGCCCGGGAATTCCGT_{15}$
Anchor	GGCTCGAGCCCGGGAATTCCGT
CLONTECH Marathon cDNA Adaptor	CCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT H_2N-CCCGTCCA-PO_4
CLONTECH Adaptor Primer 1 (AP1)	CCATCCTAATACGACTCACTATAGGGC
CLONTECH Nested Adaptor Primer 2 (NAP2)	ACTCACTATAGGGCTCGAGCGGC
PPR1	PO_4-CTAGGGCCACCACG-NH_2
Pad1	GTAATACGACTCACTATAGGGCACGCGTGGTGGCC
PP1	GTAATACGACTCACTATAGGGC
PP2	ACTATAGGGCACGCGTGGT

mRNA Isolation

mRNA was isolated using Ambion's Micropoly(A)Pure™ mRNA Isolation Kit (Ambion Inc., Austin, TX), following the manufacturer's protocol. Briefly, fresh tissues were ground in lysis solution, containing guanidine thiocyanate, using a Dounce homogenizer, with 10 strokes of the loose pestle and 10 strokes of the tight pestle or until well homogenized. Frozen tissues were ground under liquid nitrogen in a ceramic mortar and pestle then the frozen powder was transferred to a Dounce homogenizer containing the lysis solution. Samples were homogenized by 10 strokes with each pestle. The homogenate was diluted with the dilution buffer and incubated with oligo dT cellulose resin shaking for 60 minutes at room temperature. The oligo dT cellulose, with now bound mRNA, was pelleted by centrifugation, washed with high salt buffers and collected in a column. The mRNA was eluted from the oligo dT cellulose resin with warm elution buffer.

5' and 3' RACE PCR

RACE PCR was performed using CLONTECH's Marathon cDNA Amplification Kit (CLONTECH Laboratories, Inc., Palo Alto, CA). The manufacturer's protocols were followed, with modifications made to the first strand cDNA synthesis and the PCR programs used. In the first strand cDNA synthesis, SuperScript II Reverse Transcriptase (GibcoBRL/Invitrogen) was substituted for the kit's AMV Reverse Transcriptase and the addition of DTT and RNase inhibitor. Modifications to the PCR program were made to suit the melting temperatures of the primers used for each transcript desired.

Briefly, mRNA, isolated as stated above, was transcribed into double stranded cDNA, using reverse transcriptase and T4 DNA Polymerase. Adaptors were then ligated to each end of the double stranded cDNA. Double stranded adaptors were supplied by the manufacturer (cf. Table 2 for adaptor primer sequences). This was used as the template in PCR reactions in which gene specific primers (GSP) were used with the adaptor primers to amplify the complete 5' or 3' end of the transcripts.

Cloning

Cloning of PCR products was done using Invitrogen's TOPO TA Cloning Kit, vector pCR® II-TOPO®. After PCR products were separated out on high purity agarose gel (GibcoBRL/Invitrogen), bands to be cloned were excised and DNA was extracted from gel plugs by repeated freeze-thaw cycles followed by centrifugation at >10,000 x *g* for 5 to 10 minutes at room temperature. The supernatant was removed and used directly in the cloning kit. If *taq* DNA polymerase was not the primary enzyme used in generating the PCR product, the supernatant of the freeze-thaw gel extraction step was incubated with PCR buffers, $MgCl_2$, dATP, and *taq* DNA polymerase at 72°C for 15 minutes, then used in the cloning kit. The manufacturer's protocol was followed for the cloning procedure, except 0.5 µl of TOPO® vector, instead of 1 µl, and half of the volume of competent cells was used for each reaction. Positive clones, identified by the use of X-gal in standard blue white screening, were screened using a PCR based system with *E.coli* broths as the template and M13F and M13R sequencing primers as the amplification primers. Positive clones were identified by presence and correct size of PCR product bands when run out on a 1% agarose gel.

Plasmid Preparations

Plasmid preparations of cloned PCR products were performed using either the Small-Scale Preparations of Plasmid DNA (Sambrook, J. et al. 1989) or Quiagen's QIAprep Spin Miniprep Kit (Qiagen, Mississauga, ON). Both protocols were followed as described. Plasmid concentrations were quantified spectrophotometrically at 260 nm and 280 nm.

Sequencing

Cloned products were sequenced by the Center for Environmental Health, University of Victoria, and processed on ABI DNA Sequencers, model 377 (Applied Biosystems, Foster City, CA) with fluorescent dye primers in a dideoxy method.

Genomic DNA Isolation

Genomic DNA was isolated from washed red blood cells (nucleated in fish). Blood was collected from a live anesthetized (clove oil) copper rockfish, treated with heparin and centrifuged (1000 x *g*) to separate red blood cells from plasma. Red blood cells were then washed with modified Hanks' solution (added 5 mM bicarbonate) pH 7.3 and centrifuged at 2500 x *g* for 7 minutes at room temperature. 100 µl of these washed red blood cells was incubated in digestion buffer (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA pH 8.0, 0.5 % SDS, 0.1 mg/ml Proteinase K) for 16 to 18 hours with shaking at 50°C. To accomplish complete digestion, the red blood cells were fully resuspended in the digestion buffer minus the Proteinase K and SDS, before these final components were added. After digestion, one equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the cells lysate. The mixture was inverted until the phases were mixed, then centrifuged at 2600 x *g* for four minutes at room temperature. The aqueous phase was removed using a wide bore pipette tip and this extraction was repeated two times. One extraction was done with chloroform:isoamyl alcohol (24:1). One half volume of 7.5 M ammonium acetate was added to the final aqueous phase. The solution was inverted to mix and two and a half volumes of 95% ethanol were added to precipitate the DNA. Long strands of DNA were scooped out using a bent Pasteur pipette. This DNA was washed in 1 ml of 70% ethanol by inverting and centrifuging at 1700 x *g* for two minutes at room temperature. This pellet was air dried for 10 to 15 minutes, but not taken to complete dryness. The pellet was resuspended in 100 µl of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). The concentration of DNA was 1.9 µg/µL with a 260nm/280nm ratio of 1.867.

Southern Blotting

Samples: 20 µg of copper rockfish (*Sebastes caurinus*) genomic DNA was digested with 20 Units of EcoRI or 22 Units of BglII (Amersham) restriction enzymes with manufacturer's 2X One Phor All Buffer + (20 mM Tris-HCl pH 7.5, 20 mM MgAcetate, 100 mM K acetate) and TE buffer for 18 hours at 37°C. After incubation, enzymes were heat denatured for 20 minutes at 80°C. To insure proper digestion, DNA and buffers

were mixed and allowed to equilibrate for 40 minutes at room temperature before enzyme was added and incubation began. After heat denaturation, 10 µg of the digested DNA samples were precipitated with one third volume of 10M ammonium acetate and 2 volumes of 95% ethanol, exposed to -80°C for 30 minutes, then centrifuged for 15 minutes at 12,000 x *g*, 4°C. The resulting pellets were washed with 1 ml of 70 % ethanol and centrifuged at 9,500 x *g*, 4°C for 5 minutes. Pellets were air dried for 10 minutes and then each resuspended in 10 µL of TE buffer pH 7.6. Gel: This was loaded and run out on a 0.8% agarose gel with 0.1 µg/µL ethidium bromide for visualization at 50 V for 2 hours. Transfer: The gel was transferred to positively charged nylon membrane (Roche) using a LKB Biotechnology VacuGene Blotting Unit vacuum apparatus (Phamacia). Nylon membrane and 3M filter paper, wet in distilled water, were placed under mask of vacuum apparatus, then gel was placed over mask with half of a centimeter overlap to create proper seal. The vacuum pressure was between 50 and 55 mbar. Depurination: ~50 mL of 0.2 N HCl on gel for 15 minutes (bromophenol blue dye goes yellow). Denaturation: ~50 mL of 0.5M NaOH, 0.5 M NaCl on gel for 15 minutes (bromophenol blue goes blue again). Neutralization: ~ 50 mL 1 M Tris pH 7.5, 1.5 M NaCl for 15 minutes. Transfer: ~600 mL 20X SSC for 45 minutes. After transfer, the wells, dye fronts and date were marked on the membrane. The DNA on the membrane was UV crosslinked using an automatic setting (1200 to 0) on a Stratalinker. Prehybridization and hybridization were done in a hybridization oven. Prehybriziation: 20 mL of standard hybridization buffer (5X SSC, 0.1% N-lauroylsarcosine, 0.02% SDS, 1% Blocking Reagent (Roche)) 65°C for 2 hours. Hybridization: Probe was heat denatured for 10 minutes at 95°C before use in hybridization. 10 mL standard hybridization buffer with the DIG-labelled probe (see probe generation below), 65°C 16 hours. Two five minute washes with 2X SSC, 0.1% SDS at room temperature were followed by two fifteen minute washes with prewarmed 0.5X SSC, 0.1% SDS at 65°C. Detection: All of the incubation steps of detection were done on a Belly Dancer shaker at medium rotation speed. One two minute wash in washing buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 0.3% Tween 20) was followed by 45 minutes of membrane blocking (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 1% Blocking Reagent (Roche)). After blocking, the membrane is incubated with 1/10,000 dilution of alkaline phosphatase labeled anti-DIG

antibody (Roche) in the blocking buffer for 30 minutes. Antibody incubation was followed with two fifteen minutes washes in washing buffer and one two minute wash in detection buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl). The membrane was then laid on a layer of Saran Wrap and covered (~15 drops) with chemoluminescent substrate, Ready-to-use CSPD (Roche) and covered with a second layer of Saran Wrap. The membrane was sealed and incubated at 37°C for fifteen minutes to develop the chemoluminescent reaction. Blot exposures were taken with a ChemiImager 4000 digital imaging and analysis system with shutter open for 20 minutes, using gel documentation software version 3.2 (Alpha Innotech Corporation, San Leandro, CA). A digital picture of the membrane was also taken at this time for measurement reference.

Probe generation: The probe was generated using PCR, sequence specific primers, DIG-labelled dUTP, and a copper rockfish proglucagon containing plasmid as template. The PCR reaction, with a final volume of 25 µl, included: 20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.4 µM of each primer, approximately 0.05 ng of plasmid template, 0.2 mM dG/C/ATP, 0.13 mM dTTP and 0.07 mM DIG-labelled dUTP, and 1.25 Units of *taq* polymerase (GibcoBRL/Invitrogen). The primers used were GlucI51 and progluc31 (cf. Table 1 for primer sequences) and the template was a clone containing glucagon (exon 3), intron 3, and GLP1 (exon 4) sequence (White, J.W. et al. 1986). A sample of the PCR product was run out on a 1% agarose gel, and showed one clean 450 bp band, therefore, there was no need for further clean up of the PCR product. 15 µl of this PCR product was put into 10 mL of standard hybridization buffer to make the hybridization probe. This probe was stored at -20°C between uses.

PCR Walking

This protocol is based on paper by Ziguo Zhang and Sarah Jane Gurr (Zhang, Z. et al. 2000). Three aliquots of 1 µg of CRF genomic DNA were digested with 10 units of one of three restriction enzymes, SpeI, XbaI, and NheI (NEB). Each reaction also included 1X NEB Buffer 2 and 100 µg/µl BSA in a final volume of 50 µl. The samples were incubated at 37°C for 12 hours. Each sample was then extracted twice with an equal volume of phenol:chloroform (1:1). The phases were separated by centrifugation at 2900 x *g* for five minutes at room temperature. The DNA was then precipitated with 1/5

volume of 10 M ammonium acetate and 2 volumes of 99% ethanol. The samples were stored at -20°C overnight and the precipitate was pelleted by centrifugation at 12,000 x *g* for 10 minutes at 4°C. The pellets were washed with 150 μl of 70 % ethanol by inversion and centrifugation at 12,000 x *g* for 5 minutes at 4°C. Pellets were allowed to air dry until most traces of ethanol were gone and resuspended in 10 μl of sdH_2O. Each aliquot of digested genomic DNA was then used in an adaptor ligation reaction that included 150 pmoles of adaptor 1 (cf. Table 2 for sequences, adaptors were made of two complementary primers, PPR1 and Pad1, by incubation of 10 nmoles of each primer with annealing buffer for 10 minutes at 65°C and allowed cool to room temperature), 1 X NEB buffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9), 1 mM ATP, 1 μg of digested genomic DNA, 6 units of restriction enzyme AvrII (New England Biolabs Inc., Beverly, MA), and 67 CEUnits (1 Weiss Unit) of T4 DNA Ligase (GibcoBRL/Invitrogen). The ligation was incubated in a DNA Engine thermal cycler (MJ Research Inc.) with the following programme: 6 cycles of 12°C 60 minutes, 25°C 20 minutes, followed by 120 minutes at 8°C and 6 minutes at 70°C. Once the incubations were complete, 180 μl of TE (pH 7.6) was added to each tube and these PCR walking templates were stored at – 20°C.

The PCR reaction: 25 μl volume, 1X *Pfx* Amplification buffer, 1 mM $MgSO_4$, 0.3 mM dNTPs each, 0.4 μM each primer, and 0.05 U/μl (1.25 U/25μ) Platinum® *Pfx* DNA polymerase. The primers used were GluI51 (Table 1) and PP1 (Table 2). The PCR program was: 1 X 94°C 3 minutes initial denaturation

3 X 94°C 2 seconds, 72°C 3 minutes
3 X 94°C 2 seconds, 70°C 3 minutes
3 X 94°C 2 seconds, 68°C 3 minutes
26 X 94°C 2 seconds, 66°C 20 seconds, 68°C 3 minutes
1 X 68°C 8 minutes final elongation

A second round of amplification was done by repeating the same PCR reaction and program, but the nested sense primer, progluc51 (Table 1) and the nested antisense primer, PP2 (Table 2) were used with the PCR product of the first PCR reaction as the template. Products from both the first round (~3000 bp) and second round (~2500 bp) of

amplification were cloned and sequenced as above and called AF42 and AF31, respectively.

Mass Spectrometry

Peptide Extraction

Brockmann body tissue (30 mg) from an individual copper rockfish (*S. caurinus*) was homogenized in 1.25 ml of acid ethanol (ethanol:0.7 M HCl 3:1, ice cold) with a 7 ml Dounce homogenizer until tissue was well homogenized. Homogenate was centrifuged at 2000 x *g* in a Hermle centrifuge at 1°C for 15 minutes. The supernatant was removed and the remaining pellet was resuspended in 0.75 ml of acid ethanol and ground in a Dounce homogenizer. The centrifugation step was repeated and supernatant was removed. Pooled supernatants were taken to dryness under reduced pressure and stored at 4°C.

C-18 SepPak

Dry pellets were resuspended in 1 ml 0.1 M HCl, with pipetting up and down, vortexing, and 7 sec sonication to resuspend. Insoluble matter was pelleted by centrifugation for 10 minutes at 12,000 x *g*. SepPak columns were conditioned with 10 bed volumes (8.5 ml) of acetonitrile followed by 10 bed volumes of dH_2O (8.5 ml) at a flow rate of 2 ml/min. 1.5 ml of resuspended sample was applied to a pre-conditioned SepPak column at a rate of 1 ml/min (gravity fed). SepPak was washed with 2 ml of 0.1% TFA and peptides were eluted with a three step elution profile: 2 ml of ACN:dH_2O:TFA (25:74.9:0.1); 2 ml of ACN:dH_2O:TFA (45:54.9:0.1); and 2 ml of ACN:dH_2O:TFA (70:29.9:0.1). Proglucagon derived peptides were expected to elute in the 45:54.9:0.1 fraction. Samples were stored at 4°C until the next purification step.

Sephadex G-50

Sephadex beads were hydrated and equilibrated (30-40 ml) in a 26 cm x 2 cm column at 1 ml/min with 0.1 M Acetic acid, which is also the running buffer. Approximately 2 ml of SepPak effluent was loaded and run at a flow rate of 1ml/min (gravity fed). Three ml

samples were collected and effluent absorbances were measured at 280 nm. Peaks with high absorbance were pooled and then taken to near dryness under reduced pressure in order to concentrate for loading on RP-HPLC.

C-18 Reverse Phase High Performance Liquid Chromatography

Buffer A: 0.1% TFA in dH_2O Buffer B: 0.06% TFA in acetonitrile. C-18 column was equilibrated with 80% Buffer B then 12.5 % Buffer B. The pooled Sephadex fractions were injected, the void volume was discarded and the program was started. Program: 1.5 minutes of 12.5% ACN (Buffer B), 50 minutes linear gradient from 12.5% ACN (Buffer B) to 50% ACN (Buffer B), 4 minutes at 50:50 0.1% TFA (Buffer A):ACN (Buffer B), then return to 12.5% ACN at a flow rate of 0.2 ml/min. Effluent absorbance at 230 nm was measured and peaks were collected. Depending on the volume of the peaks, samples were concentrated by centrifugation under reduced pressure for mass spectrometric analysis.

Mass Spectrometry Analysis

Samples were mixed 1:1 with matrix and spotted on sample grid block for crystallization, and the sample block was inserted into a Voyager-DE MALDI-TOF STR mass spectrometer (PerSeptive Biosystems). The samples were calibrated against CalMix 2 (PerSeptive Biosystems) which included angiotensin (1297.5 amu), ACTH (1-17) (2094.5 amu), ACTH (18-39) (2466.7 amu), ACTH (7-38) (3660.2 amu), and bovine insulin (5734.6 amu), with particular attention paid to the final three calibrant peptides. The matrix used was either 2,5-dihyroxybenzoic acid or α-cyano-4-hydroxy-cinnamic acid. The MALDI-TOF mass spectrometer operating conditions were: reflection mode of operation, delayed extraction mode, positive polarity, 20,000V accelerating voltage, 93 % grid voltage, 390 nanosecond extraction delay time, and 20 Hz laser rep rate.

Results

Proglucagon mRNA Sequences

Working with intestinal tissue, I used degenerate primers corresponding to the beginning of glucagon and the beginning of GLP-1/2 in RT-PCR to look for partial proglucagon sequences. Bands containing presumed proglucagon fragments were amplified, cloned and sequenced. Sequencing results indicated two distinct mRNA sequences for proglucagon in copper rockfish. As intestine, Brockmann body, and brain are the alleged primary tissues of proglucagon expression, they were initially screened using similar internal primers. All three tissues presented the same two distinct proglucagon fragment sequences. The sequence of one partial transcript (arbitrarily called transcript I) included most of glucagon, all of GLP-1 and the beginning of GLP-2, while transcript II was shorter and extended from the first third of glucagon to the beginning of GLP-1 only. Even with such short spans of partial sequence, the nucleotide sequences were different enough, 28 bp substitutions out of a total of 75 bp, to indicate that these transcripts were unlikely to be encoded by the same gene.

When I performed 3' RACE on Brockmann body and intestine samples, some unexpected sequences were revealed. Confirming results for rainbow trout (*O. mykiss*) (Irwin, D.M. et al. 1995) and anglerfish (*L. americanus*) (Lund, P.K. et al. 1983), in Brockmann body, I found that proglucagon is alternatively spliced to eliminate the GLP-2 sequence from the mRNA by retaining the intron following GLP-1. This intron includes a polyA signal and leads to truncation of the mRNA after GLP-1 (cf. Figure 5). As alternative splicing has not previously been shown in any other tissues, full length sequences including GLP-2 were expected in intestine. For proglucagon I, indeed, a full length sequence including GLP-2 was found in the intestine of rockfish. However, the alternatively spliced version (not including GLP-2) was also found for both proglucagon I and II, but a full length version of proglucagon II (including GLP-2) was not amplified. This implies that alternative splicing occurs in other tissues besides Brockmann body.

In order to clarify in which tissues alternative splicing occurs in rockfish, RT-PCR screening was done. These data are presented in the Tissue Distribution section of this chapter.

3'RACE of proglucagon II in Brockmann body identified three different lengths of 3' UTR. Besides the length of the 3' UTR, the sequences are identical. This suggests that during the creation of the proglucagon II message, different polyA signals are observed to generate different length UTRs, and three putative polyA signals exist within the longest UTR sequence (Figure 7). As the complete role of the 3'UTR is unclear, the role of different lengths is also unclear, yet it has been suggested that characteristics of the 3'UTR might affect the half-life of the message, suggesting a possible role for expressing proglucagon sequences with three different 3'UTR lengths.

5'RACE was then performed using sequence specific primers designed from the 3'ends of both sequences to ascertain the complete mRNA sequences shown in Figure 6 and Figure 7. As seen in these figures, both proglucagon sequences are similar to a typical proglucagon in structure. They both encode a signal peptide (assumed to target the endoplasmic reticulum), followed by glicentin-related polypeptide (GRPP), then glucagon, and GLP-1, separated by a short intervening peptide. In proglucagon I, GLP-2 is separated from GLP-1 by a second short intervening peptide. The total length of proglucagon I is 922 nucleotides, with an open reading frame encoding 176 amino acid residues (Figure 6). Proglucagon II is some what shorter with 805 nucleotides and an open reading frame of 123 amino acid residues (Figure 7). Most of the peptides included in this prohormone are surrounded by the dibasic prohormone convertase cleavage signal typical of prohormones.

```
                 5' UTR                                      Signal
                                                             M  M
1      gcagaggaactaacagcaccgttctttggagagaaagaagacaaggttataaccatgatg  60
        Peptide
       S  I  Q  S  L  A  G  I  L  L  V  L  G  L  V  Q  S  S  W  Q
61     agcatccaatccctggctggtatccttctggtccttggcttagtccagagcagctggcag  120
       GRPP
       V  P  L  Q  E  A  A  D  S  S  S  F  D  A  D  D  T  S  V  D
121    gttcctctgcaggaggctgctgacagctcaagcttcgatgcagacgacacatcagtggac  180
                                 Glucagon
       E  L  S  N  M  K  R  H  S  E  G  T  F  S  N  D  Y  S  R  Y
181    gagctgtcaaacatgaagagacactcggagggaactttctccaacgactacagcagatac  240

       L  E  E  R  K  A  Q  D  F  V  R  W  L  M  N  N  K  R  S  G
241    ctggaggaaaggaaggcgcaggacttcgttcggtggctgatgaacaacaagaggagcggt  300
                   GLP-1
       A  D  E  K  R  H  A  D  G  T  F  T  S  D  V  S  S  Y  L  K
301    gctgacgaaaagcgccacgccgacgggaccttcaccagcgacgtcagctcctacctcaag  360

       D  Q  A  I  K  D  F  V  N  R  L  K  S  G  Q  V  R  R  E  S
361    gaccaggcaatcaaagactttgtcaacaggctcaagtctggacaagtcagaagagaatct  420
                                                GLP-2
       E  T  E  W  R  G  E  A  F  S  R  R  H  V  D  G  S  F  T  S
421    gaaacggaatggcgaggtgaggcattcagcaggaggcatgtagatgggagcttcaccagc  480

       D  V  N  K  V  L  D  S  M  A  A  K  E  Y  L  L  W  V  M  T
481    gatgtgaacaaggtgctggactccatggctgccaaggaatatttactctgggtcatgacc  540
                                                      3'UTR
       S  K  P  S  G  E  S  K  K  R  Q  E  D  Q  *
541    tccaagccttcaggggagagcaagaaaagacaagaagaccaatgatactctcctcaagga  600
601    ctccctctgtgtgccgcacacagagccatatacaggaggaatagaaagccatactgttgt  660
661    aatgacaacactatcttgtttacttgctgtatgtgatgcaaagatacaaaaactatgact  720
721    cttccgtatggtgcaccagagagaaggaatgaaatcgagggcgtatttwttctacaaatw  780
781    taaattattttctatatttttaaaaacaagaacacaaaaagtgaaatctcaggtaggctt  840
841    atatatagaaaacacacgtagtgtagttggtattcattacatgttattaatccgctctgc  900
901    aaagataaagtttttcatcgtgaaaaaaaaaaaaaaaaaaaaaaaaaaaaa  952

Alternatively spliced 3'UTR
    GLP-1
       D  Q  A  I  K  D  F  V  N  R  L  K  S  G  Q  V  R  R  E  *
361    gaccaggcaatcaaagactttgtcaacaggctcaagtctggacaagtcagaagagagtag  420
421    gctgcctcttcctccttctcctcctcctcctcttcctcctcctcctacagtaccaaaacc  480
481    cctatcggcctgtgtctgattccagcgtgaccagagagacagagagaaagagctaatcaa  540
541    ctgcttcttgtgccaccttttttcctttctttgctttgtttcttgatgttccttgtgg  600
601    ttgtgcaaaaaataaatcacctgaacttttaaaaaaaaaaaaa  644
```

Figure 6. Complete cDNA sequence and putative translated amino acids for copper rockfish proglucagon I including the alternatively spliced 3'UTR. Underlined sequences correspond to peptides, glucagon, GLP-1, and GLP-2. Bold italicized nucleotides indicate polyA signals and dibasic cutsites are bold.

```
1     gagaagacgggcaccagtcaacctactttacaatccttctctcagacgctcatcagaaa  60
        5' UTR          Signal Peptide
                               M  K  S  A  H  S  L  A  G  V  L  L  L  I
61    ggagcacagaagtcaaacgatgaagagcgctcactctttggctggagtcctgctcctcat  120
                                     GRPP
       M  I  Q  S  S  W  Q  M  P  D  Q  D  T  D  R  N  P  M  L  L
121   catgatccaaagcagctggcagatgcctgaccaggacacagaccgaaaccccatgctatt  180
                                                            Glucagon
       A  P  N  S  M  L  A  E  P  I  E  L  P  N  M  K  R  H  S  E
181   ggctccaaactccatgttggccgaacccatcgagctcccaaacatgaagagacattcgga  240

       G  T  F  S  N  D  Y  S  K  Y  L  E  T  R  R  A  Q  D  F  V
241   ggggacgttttccaacgactacagtaaatatctggagacgagaagagcacaagactttgt  300
                                                      GLP-1
       Q  W  L  K  N  S  K  R  N  G  S  L  F  R  R  H  A  D  G  T
301   ccagtggctaaagaactcaaaaaggaacgggagcctatttagacgccatgcagacggcac  360

       Y  T  S  D  V  S  S  Y  L  Q  D  Q  A  A  K  E  F  V  Y  W
361   ctacaccagcgacgtgagctcctacctgcaggaccaggcagccaaggagtttgtgtactg  420
                                                 3'UTR
       L  K  T  G  R  G  R  R  E  *
421   gctgaagaccggccgaggcagaagagagtaaactcagccaactgctccataaataatgca  480
481   acaattcatactggctctgtgtgattatatgtttcacaaactgtactagactttttttct  540
541   tggtcatgctattaaattcccctcattgaaaagctctgtcggccttcagtggtggaagga  600
601   aacgttgtcaaaatgccaaaagttacagtatgtttgaagtttgtctcaatcataaatgct  660
661   gtatgacttttttttttaatcaaacattcacatttttcacattttggcacttttcccatc  720
721   ttccttctcatagaatttgtaagaatatgttacagtcacgttattgatctgaactgcaac  780
781   ttcgataattacgtctgtagatacctttttcttattattcagataaaagtcatatttagtt  840
841   ttcaataaagaaatgtgtacaatctaaaaaaaaaaaaaaa  880
```

Figure 7. Complete cDNA sequence and putative translated amino acids for copper rockfish proglucagon II. Underlined sequences correspond to peptides, glucagon and GLP-1. Bold italicized nucleotides indicate polyA signals and dibasic cutsites are bold.

Although the corresponding glucagon and GLP-1 amino acid sequences appear quite similar between the two sequences, the complete cDNA sequences for proglucagon I and II are appreciably different. Just comparing the nucleic sequence for glucagon through GLP-1, the most conserved region, the identity between the two sequences is 73.2%. When the entire comparable open reading frame is considered, this value decreases to 69.4% including insertions and deletions (indels) created by the best alignment. As this value is considerably lower than the generally accepted allelic variation, I propose that these sequences are encoded on independent genes in rockfish.

Genomic DNA Sequences and Southern Blot for Proglucagon Genes

The next step was to confirm the hypothesis on the existence of two proglucagon genes by nucleotide sequence analysis of rockfish genomic DNA and Southern blot analysis.

Both proglucagon gene sequences have the same exon/intron structure, similar to that found in other proglucagon genomic sequences, such as human and rat (White, J.W. et al. 1986). In the 5'end of the rockfish genes, the first intron is found just before the open reading frame begins, and the second intron is located in the middle of GRPP. Introns 3 and 4 directly follow glucagon position 29 and GLP-1 position 34, respectively. In proglucagon I only, intron 5 follows position 35 of GLP-2. Figure 8 is a schematic representation of the rockfish proglucagon genes and the sequences of these genes are shown in Figure 9 and Figure 10, with noteworthy features marked.

Like other prohormones in the glucagon superfamily, such as secretin, VIP and PACAP, proglucagons contain functional units that are contained in single exons (Sherwood, N.M. et al. 2000). This is seen in rockfish with exons 3, 4, and 5 that completely contain glucagon, GLP-1 and GLP-2, respectively (Figure 8). Considering proglucagons display this type of gene structure and sequence similarity between glucagon, GLP-1 and GLP-2 is high, the latter peptides most likely originated through exon duplication of glucagon.

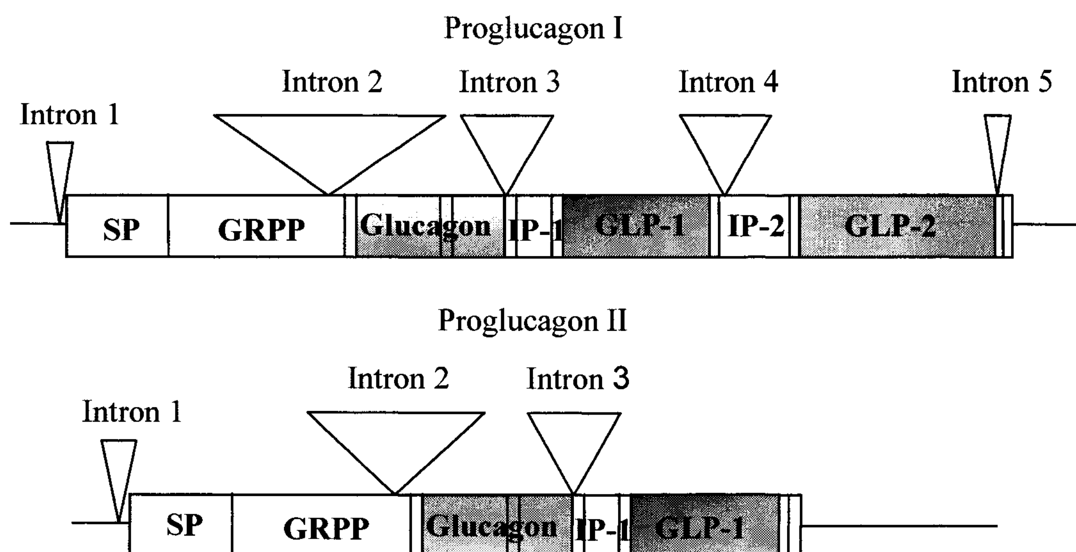

Figure 8. A schematic representation of the rockfish proglucagon genes.

Figure 9 also shows an alignment of the proglucagon I genomic intron 4 sequence and the 3' UTR of the alternatively spliced proglucagon I mRNA found in Brockmann body. As the sequences are identical, this figure demonstrates how alternative splicing leads to the retention of this intron, which, in essence, becomes the 3'UTR of the mRNA sequence. As seen in this intron, and noted on the figures, an interesting characteristic of these genomic sequences is the existence of a stop codon at the beginning of several introns directly following glucagon, GLP1, or GLP2. Therefore, if the intron is not spliced out of the sequence, translation of a truncated, but still functional, precursor occurs. Another characteristic that only shows up in intron 4 is the presence of a polyA signal, underlined in the Figure 9. This polyA signal is necessary for transforming this intron into a 3'UTR in the case of alternative splicing.

```
            5'UTR                Intron 1
1       ggagagaaagaagacaaggttattttgttgcaatggtgaaacactttagctttctgatc  60
                                                          Signal
                                                          M  M  S
61      cgatatgcaagctctttctgtgtgtgtgatgttctgtggtaggttgtaaccatgatga  120
          Peptide
          I  Q  S  L  A  G  I  L  L  V  L  G  I  V  Q  G  S  W  Q  V
121     gcatccaatccctggctggtatccttctggtccttggcatagtgcaaggcagctggcagg  180
          GRPP I
          P  L  Q  E  A  A  D  S  S  S      Intron 2
181     ttcctctgcaggaggctgctgacagctcaaggtacagaaatcacatctaaacggatgagg  240
241     ttatgtctttgtctaaggaccgcaggttcactacttgtttactgtgtctaaatcggacta  300
301     aatcacaagaaaaattgtgggtaattttgggttacaatccaaatggcaacattcataaga  360
361     ttagaagttactcaaactgaaattttgttactacaatgaccccaaaatagtatgtcgaaa  420
421     aaagtaataaaacattcatagtatattatgtcaaaaaagtaataaaaagtcacagtatat  480
481     tatgtcgtaaaagtaaaaaaacaaaaaattcatggtatagtatgttgaaaaatgtgataa  540
541     aaaattcatagtatagtgtgttaaaaaaagtcataaaaagtcacagtatagtatgttgta  600
601     aaagtattaaaaaaaaggcatagtatagtacgttgaaaacagtgatacaaaaaaagtcat  660
661     agtgtagtatgttaaaaaaaatttggaagtctgttgttgagaaaagaacaaattgtagca  720
721     aagagtaaaagatataattaaaatgtgatgttgcttaacacctgtggtgatgtttggaag  780
781     atagagatgctgcattttattaccccaaattaaaatactacatgtgattcatctgcatca  840
841     tctatcattccgtatattcatattggattggtgtgaactcaaagtgattgtgttctccct  900
```

```
                          GRPP I
      Intron 2      F  D  A  D  D  T  S  V  D  E  L  S  N  M  K  R
901   cctgcgctgcagcttcgatgcagacgacacatcagtggacgagctgtcaaacatgaagag  960
      Glucagon I
       H  S  E  G  T  F  S  N  D  Y  S  R  Y  L  E  E  R  K  A  Q
961   acactcggagggaactttctccaacgactacagcagatacctggaggaaaggaaggcgca  1020

       D  F  V  R  W  L  M  N  N  K  R  S  G     Intron 3
1021  ggacttcgttcggtggctgatgaacaacaagaggagcgggtcagtgtgggatcaatttgg  1080
1081  aaaagtgatcatacgaatctttattgtaaaaagctttgacttgtttaaacttaattaatg  1140
1141  aaacactggttatattcaagctcaggatgttatgatgtggtggagtttttacagcaatgt  1200
1201  ataagtaaaagtcctgcattcaatatgttaaacaagtacagaagtgttatcagctaaata  1260
1261  tactcaaccataggggtctcaacctaggggtcgggcccctctaaagggtcacaagataaa  1320
1321  tctgagaggtcatgagatgattaataagagaggaaagaagaaagttctgctacacaaata  1380
1381  tatattaattaatctcgttaactttgacagacgttattttcacattttcgcgttggctac  1440
1441  tgtcattctcctacacgtttggaacatgggagtaaggtggttgcaatctgcaaccgcacc  1500
1501  actaaatgccgccaaatcctacacagcgcacctttaatgtaaaatcttgatctgaaaagt  1560
1561  aatcattcagataaatgttaaaatgtaaatgtaaaattttaaattataccaaatatctc  1620
                                            GLP-1 I
          Intron 3      A  D  E  K  R  H  A  D  G  T  F  T  S
1621  cctcactgtgtgtttttcagtgctgacgaaaagcgccacgccgacgggaccttcaccagc  1680

      D  V  S  S  Y  L  K  D  Q  A  I  K  D  F  V  N  R  L  K  S
1681  gacgtcagctcctacctcaaggaccaggcaatcaaagactttgtcaacaggctcaagtct  1740

      G  Q  V  R  R  E  *   Intron 4 aligned with alt spliced cDNA
1741  ggacaagtcagaagagagtaggctgcctcttcctccttctcctcctcctcctcctcctcc  1800
417    Alt cDNA 3'UTR  gtaggctgcctcttcctccttctcctcctcctcctcttcctcc  cDNA

1801  tcctcctacagtaccaaaacccctataggcctgtgtctgattccagcgtgaccagagaga  1860
477   tcctcctacagtaccaaaacccctatcggcctgtgtctgattccagcgtgaccagagaga  cDNA

1861  cagagagaaagagctaatcaactgcttcttgtgccaccttttttcctttctttgctttg  1920
537   cagagagaaagagctaatcaactgcttcttgtgccaccttttttcctttctttgctttg  cDNA

1921  ttttcttgatgttccttgtggttgtgcaaaaaataaatcacctgaacttttggcactgc  1980
597   ttttcttgatgttccttgtggttgtgcaaaaaataaatcacctgaacttttaa-poly A  cDNA

1981  gttgattgatgattgtgtaaaatgattatttgtgatacaaaagcggacaatgaagaaatc  2040
2041  aagggatgatcagagcatcacataatcagaaataacaacaataaaatgatgattttctga  2100
2101  tgaataatggccaaaaattggagctccaatggtaggtttccacatcgacccttaatgggc  2160

                     Intron 4             S  E  T  E  W  R  G  E
2161  agcgtttcacattgtgcgcctgtgtacatactcagatctgaaacggaatggcgaggtgag  2220
                     GLP-2
      A  F  S  R  R  H  V  D  G  S  F  T  S  D  V  N  K  V  L  D
2221  gcattcagcaggaggcatgtagatgggagcttcaccagcgatgtgaacaaggtgctggac  2280
      S  M  A  A  K  E  Y  L  L  W  V  M  T  S  K  P  S  G  E  S
2281  tccatggctgccaaggaatatttactctgggtcatgacctccaagccttcaggggagagg  2340
        Intron 5
2341  taaaaaaagtcagtcacatgatttgataattgccagaaattctaggctcggcgtgattgt  2400
                               K  E  R  Q  E  D  Q  *   3'UTR
2401  gatttctctcttttgcagcaagaaaagacaagaagaccaatgatactctcctcaaggac  2460
2461  tccctctgtgtgccgcacacagagccatata  2491
```

Figure 9. Genomic DNA sequence for proglucagon I, including an alignment with the 3'UTR of the alternatively spliced proglucagon I cDNA. Dibasic sites are bold, putative peptides are underlined, putative splice junction sequences are italicized bold.

```
                                       5'UTR      Intron 1
1       ttacaatccttctctcagacgctcatcggaaaggagcacaggtgagaacttttgtacctt  60
61      tcctccttttcaaatccagactgtctatactatgaaattgtattttattatgctgtatt  120
                                              Signal Peptide
                  Intron 1            5'UTR   M   K   S   A   H   S   L
121     gaaatatcttttctctgctcatcttttagaagtcaaacgatgaagagcgctcactcttt  180
                                                   GRPP II
         A   G   V   L   L   L   I   M   I   Q   S   S   W   Q   M   P   D   Q   D   T
181     ggctggagtcctgctcctcatcatgatccaaagcagctggcagatgcctgaccaggacac  240

         D   R   N   P   M   *    Intron 2
241     agaccgaaaccccatgtaagagcaactctcatttaaaaatctcccctatgatcctttatg  300
301     agaggattaggaaggtctcaccatgagagtatttggtgacaccatcccgagctgcaaaaa  360
361     aatttagcaaatcttttgagtgttgcagcagaatagcactttttaagcttaaatctttt  420
421     tttctttttgttaaattcatctttattttgtgtttgtttttggttgagagggctccag  480
481     ttggccagacatttctaaaggaagacagacaaagaatacttgtgttggcagaccgatttg  540
541     ccattgacagctccatgtgttgcattcaaagtaccagctgctgacattagaggaaaagga  600
601     tctccagtgcaggggttgttatgtgagatgatgtttttatgtttcatgtattttaatgcc  660
661     ttcctagtattatttgcatgtattgccattttatacttatatgttttagttataacgtgg  720
721     tgtctcttgtagtgcttgctgatgcaaaatgtgtggtgtaattgaatctctggaacattc  780
781     aatttcatacgcaagaatcacagcatcagcgagatccaaatttacaaatgtatgagcagt  840
841     gatcgatggatgagaatggtgaagcaggataaacagttgaggtgtggtgaaggaacacct  900
901     ctaggggaccaagagagagaaaggacagaaataaaaaggtcagataggcaaggacatgaa  960
961     aaacggaaggcatttacagatcggaaggggagtgtttgatatgagactgtgtaagctttc  1020

         Intron 2     L   L   A   P   N   S   M   L   A   E   P   I   E   L   P   N
1021    ttatcatttctaggctattggctccaaactccatgttggccgaacccatcgagctcccaa  1080
                      Glucagon II
          M   K   R   H   S   E   G   T   F   S   N   D   Y   S   K   Y   L   E   T   R
1081    acatgaagagacattcggaggggacgttttccaacgactacagtaaatatctggagacga  1140

          R   A   Q   D   F   V   Q   W   L   K   N   S   K   R   N   G   *   Intron 3
1141    gaagagcacaagactttgtccagtggctaaagaactcaaaaaggaacgggtgagtgtgcc  1200
1201    ctttcttacttgatcatttcaacttggtctcatagtctgtaatatattgagaaaaccaca  1260
1261    agtctggttttgattgcattaatactgtatatagctgtgtgcctattggatcagacaaca  1320

                        Intron 3                  S   L   F   R   R   H   A
1321    cgattacatatacagttattgtgactgtgtctctacaggagcctatttagacgccatgca  1380

         D   G   T   Y   T   S   D   V   S   S   Y   L   Q   D   Q   A   A   K   E   F
1381    gacggcacctacaccagcgacgtgagctcctacctgcaggaccaggcagccaaggagttt  1440

         V   Y   W   L   K   T   G   R   G   R   R   E   *     3'UTR
1441    gtgtactggctgaagaccggccgaggcagaagagagtaaactcagccaactgctccataa  1500
1501    ataatgcaacaattcatactggctctgtgtgattatatgtttcacaaactgtactagact  1560
1561    tttttcttggtcatgctattaaattcccctcattgaaaagctctgtcggccttcagtgg  1620
1621    tggaaggaaacgttgtcaaaatgccaaaagttacagtatgtttgaagtttgtctcaatca  1680
1681    taaatgctgtatgacttttttttaatcaaacattcacatttttttcacattttggcactt  1740
1741    ttcccatcttccttctcatagaatttgtaagaatatgttacagtcacgttattgatctga  1800
1801    actgcaacttcgataattacgtctgtagatacctttttcttattattcagataaaagtcat  1860
1861    atttagttttcaataaagaaatgtgtacaatctatgagttccctctcgaagaaaattaat  1920
1921    ccattttgaaatattggtttgacaggatcactcttggcaaggatctacatctcagtgtgc  1980
1981    acatggatatc  1991
```

Figure 10. Genomic DNA sequence for proglucagon II. Dibasic sites are bold, putative peptides are underlined, putative splice junction sequences are italicized bold.

As indicated from the different intron sizes and drastically different intronic sequence between the two proglucagons, it is highly unlikely that they are allelic. To obtain further confirmation, a Southern blot was performed. Rockfish genomic DNA, digested by two infrequent cutting restriction enzymes, EcoR I and Bgl II, was probed with a proglucagon I and II hybrid probe using a low stringency wash. This was to ensure that both genes would be visualized on this blot. Also, positive and negative controls of plasmid DNA containing the probe, proglucagon I, II and unrelated rockfish cDNA were included to ensure accurate interpretation of the blot (Figure 11). Unfortunately, because the concentration of the plasmids used for controls was greater than optimal, some non-specific binding to plasmid DNA by the probe occurred and the bands of the positive controls are vastly brighter than the bands seen in the genomic DNA lanes (Figure 11). However, this does not detract from the conclusion that the probe binds selectively to proglucagon sequences on the genomic DNA. EcoR I and Bgl II were chosen for genomic DNA digestion because they do not cut within the rockfish proglucagon gene sequences. This ensures that each band seen in the blot represents a different gene. Therefore, as indicated in the EcoR I digest lane, rockfish have at least two distinct proglucagon genes. There is one broad band visible in the Bgl II digest lane. This doesn't not counter the results seen with EcoR I as the location of the Bgl II cut sites in the genomic DNA are not known. One band may result as there may be no Bgl II cut sites between the two proglucagon genes on the chromosome leading to one piece of DNA forming a large band. Alternatively, two bands may exist, but may be very similar in size, thus overlapping each other and appearing as one band. This may explain why the Bgl II band appears broader than the EcoR I bands.

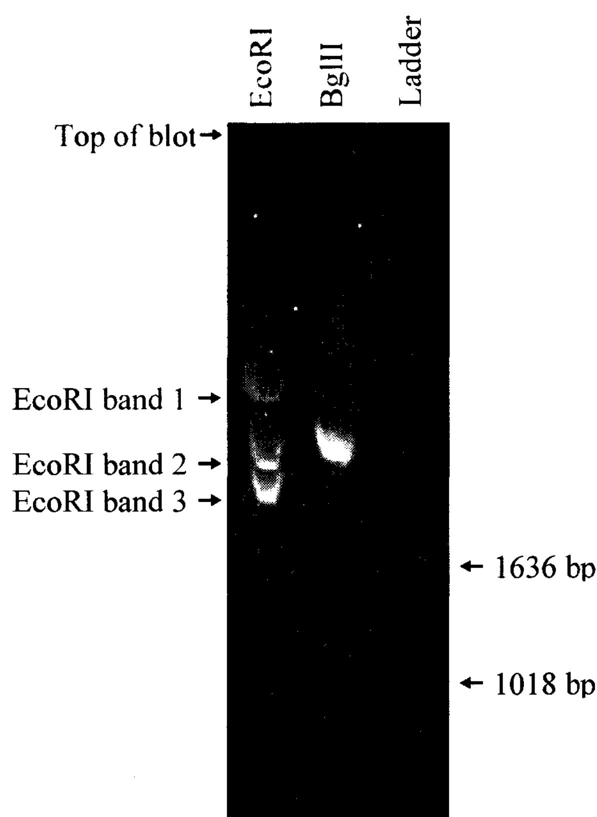

Figure 11. Southern Blot for the Proglucagon gene in Copper Rockfish DNA digested by EcoRI and BglII.

Since I found no evidence of a proglucagon II mRNA encoding a GLP-2 sequence, PCR based genome walking was used to determine the 3' flanking region of proglucagon II, to identify if there was a GLP-2 encoded in this gene. Approximately 3000 base pairs of genomic sequence were acquired on the 3' end of the reported proglucagon II sequence, and no GLP-2 sequence was identified (Figure 10). This indicates that this proglucagon gene indeed lacks a GLP-2 sequence.

Tissue Distribution of Proglucagon Messenger RNAs

Proglucagon, processed into its various active components, has been shown to be expressed in or can be extracted from three major tissues in fish, namely Brockmann body, intestine, and brain. These tissues from copper rockfish, as well as some related tissues, were screened for expression of proglucagon I, the alternatively spliced version

of proglucagon I (also referred to as the truncated form of proglucagon I), proglucagon II, and β-actin. This was done by RT-PCR with primers designed from the 5' and 3' UTRs, which ensures complete specificity for each proglucagon amplification. β-Actin served as a positive control for the RT-PCR and an indicator of the quality of the mRNA. As shown in the following figures, the expected sizes and lane markers for each PCR reaction are:

β-actin [A]	280 bp
proglucagon I (including GLP-2) [I]	605 bp
truncated proglucagon I (no GLP-2) [tI]	478 bp
proglucagon II [II]	497 bp

For each lane a single band was expected.

The β-actin RT-PCR products also indicate the presence of genomic DNA in the mRNA preparation as the primers were designed to span an intron. If genomic DNA is present in the template sample, a 320 bp band containing 40 bp of intronic sequence will occur together with the 280 bp band that corresponds to mRNA. Although the β-actin primers were originally designed from zebrafish (*Danio rerio*) sequence, they can be used to amplify β-actin in many different teleost species with great efficiency, including rockfish.

In Figure 12 the proglucagon transcripts amplified by RT-PCR are depicted with the primer positions indicated.

Full length Proglucagon I

605 bp

5UTRII2

3UTRII2

SP GRPP Glucagon I IP-1 GLP-1 I IP-2 GLP-2

Alternatively spliced Proglncagon I

478 bp

5UTRII2

3UTRII

*

SP GRPP Glucagon I IP-1 GLP-1 I

3' UTR from Intron 4

Proglucagon I including intron

1042 bp

5UTRII2

3UTRII2

GLP-2

*

SP GRPP Glucagon I IP-1 GLP-1 I

Intron 4

IP-2

Proglucagon II

497 bp

5UTRIII

3UTRIII2

SP GRPP Glucagon II IP-1 GLP-1 II

Figure 12. Schematic representation of the proglucagon transcripts amplified by RT-PCR.

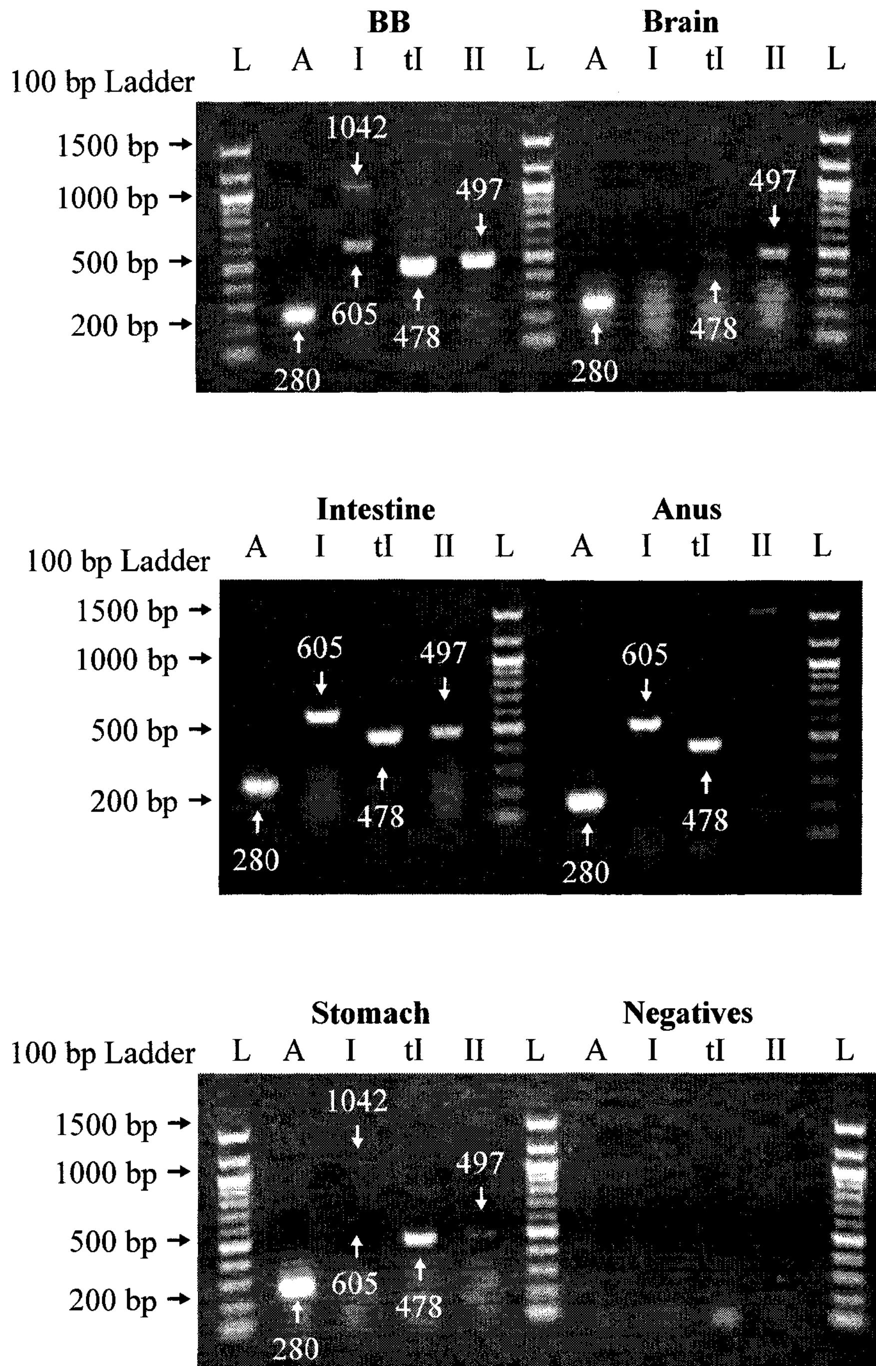

Figure 13. Tissue Distribution of Proglucagon transcripts in copper rockfish tissues, screened by RT-PCR of mRNA. Amplified products include β-actin [A], 280 bp, proglucagon I [I], 605 bp, truncated proglucagon I [tI], 478 bp, and proglucagon II [II], 497 bp and were visualized with ethidium bromide on a 1% agarose gel.

Brockmann body (endocrine pancreas)

From published rainbow trout data (Irwin, D.M. et al. 1995) and our initial 3'RACE work on rockfish, only the truncated form of proglucagon I and proglucagon II transcripts were expected in the PCR from endocrine pancreas. Surprisingly, the proglucagon I lane, I, in Figure 13 clearly shows a 605 bp band, indicating that a transcript including GLP-2 is transcribed in Brockmann body. Also in this lane, a larger band of 1042 bp, (Figure 13), was found in every Brockmann body mRNA tested. After cloning and sequencing, it was determined to be proglucagon I with retention of intron 4, GLP-2 and the full length 3'UTR. Thus alternative splicing occurred, but the polyA signal in the intron was not recognized and the rest of the proglucagon message was transcribed. Although this PCR was not designed to provide quantitative results, the intensity of these proglucagon I transcripts is markedly lower than that for truncated proglucagon I and proglucagon II, suggesting that the transcripts not encoding GLP-2 may be the primary proglucagon transcripts in Brockmann body.

Brain

While the role of proglucagon derived peptides in fish brain is not yet clear, some recent evidence points toward a function for GLP-1 in food intake inhibition (Silverstein, J.T. et al. 2001). In addition, a role for food intake inhibition for both GLP-1 and GLP-2 in mammals has been proposed (Turton, M.D. et al. 1996; Tang-Christensen, M. et al. 2000). Because of this, full length proglucagon I message, including GLP-2, was expected in rockfish brain. This was not the case. Although I screened cDNAs prepared from brains of nine individual copper rockfish, no full length proglucagon I transcript was amplified. In fact, the only proglucagon I transcript amplified was the truncated form (Figure 13, lane tI). This was unexpected because, to date, alternative splicing had only been reported for the Brockmann body.

Also, in the nine individual rockfish screened, the truncated proglucagon I message was not always amplified. In contrast, proglucagon II transcript is consistently found in rockfish brain, (Figure 13, lane II), indicating that proglucagon II transcript is likely the predominant proglucagon transcribed in rockfish brain.

Gastrointestinal tract

Proglucagon derived peptides are known to play an instrumental role in the gastrointestinal tract of mammals by affecting nutrient uptake in intestine, gastric emptying, and intestinal growth (Drucker, D.J. 2002). Some of these functions may also occur in fish, especially considering proglucagon transcripts have been shown in intestine of several fish (Irwin, D.M. et al. 1995; Youson, J.H. et al. 2001). Since some of the proposed functions of proglucagon derived peptides may act on other GI organs, such as the stomach or anus, these tissues were also screened for proglucagon transcript production. As seen in Figure 13, all three proglucagon transcripts were amplified from rockfish stomach cDNA with proglucagon II showing a significantly brighter band than either form of proglucagon I. cDNA prepared from a randomly positioned piece of rockfish intestine also provided template for amplification of all forms of proglucagon. In contrast, proglucagon II was not amplified from rockfish anus. Of all the rockfish tissues scanned, the anal region of the gastrointestinal tract is the only tissue devoid of transcript for proglucagon II.

Unlike the mammalian digestive system, the fish GI-tract is not segmented into small intestine and colon. Although it is anatomically one organ, it may not necessarily function as one. Support for this has been indicated by zonation of teleostean intestinal enzyme activities from the distal to proximal ends of the intestine (Mommsen, T.P. et al. 2003b). Zonation indicates physiological differences in regions of the intestine and probably different functions in these regions. Suspecting zonation could occur for proglucagon expression, we sampled intestinal segments from the distal to proximal ends of the intestine to identify differences in proglucagon expression. Throughout the intestinal segments, proglucagon I, in both full and truncated forms, is consistently expressed. In contrast to proglucagon I, the proglucagon II message shows a trend of decreased expression towards the anal portion of the intestine (Figure 14). As is obvious from the figure, this trend does not occur in the β-actin lanes. Also, this trend is consistent with my previous observation, that proglucagon II is expressed in stomach (Figure 13, lane II), but not in anus (Figure 13, lane II).

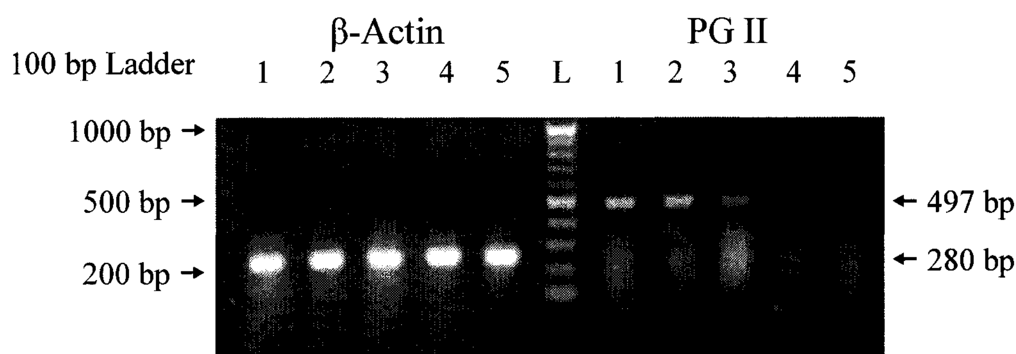

Figure 14. Differential expression of proglucagon II along the intestine of copper rockfish. Amplified products include β-actin [A], 280 bp, and proglucagon II [II], 497 bp and were visualized with ethidium bromide on a 1% agarose gel.

From the evidence above, I conclude that rockfish possess two proglucagon genes and that both are expressed as mRNA concurrently in various rockfish tissues. Further, I demonstrated alternative splicing and differential tissue expression of proglucagon mRNA. The next logical step was to perform peptide isolation and mass spectrometry analysis to establish or refute the presence of peptides postulated from the molecular data.

Mass Spectrometry Data

After partial purification, including subsequent separation on a Sephadex G-50 column and a C-18 reverse phase HPLC (Figure 16), peptides from the copper rockfish Brockmann body were analyzed by MALDI-TOF mass spectrometry. Theoretical and experimental masses for the rockfish proglucagon derived peptides are reported in Table 3 and the mass spectrograms follow, Figure 17 - Figure 21.

All of the peptides identified by mass spectrometry had experimental masses within 0.5 atomic mass units of the calculated masses determined from the putative amino acid sequences. No post-translational modifications were expected for these peptides, except possible amidation of peptides with C-terminal glycine residues. C-terminal amidation affects the mass by -0.98 amu under acidic conditions. The only rockfish peptide potentially undergoing amidation is GLP-1 II (31 aa), which is identified with low intensity in HPLC peak 11 (Figure 16 and Figure 20).

Considering the major proglucagon transcripts in the Brockmann body are the alternatively spliced version of proglucagon I and proglucagon II, which is truncated after GLP-1, two possible lengths of GLP-1 peptide could be expected. If the precursor is processed at all the double cut sites, a 31 amino acid GLP-1 peptide is released. This is the most commonly reported form of GLP-1. A 34 amino acid GLP-1 peptide can also be released if the double cut site following GLP-1 is not cleaved as in both sequences a stop codon follows amino acid 34 (Figure 6 and Figure 7). If the full-length proglucagon I transcript is translated it would produce only the 31 amino acid GLP-1.

Two forms of glucagon, glucagon I and glucagon II were identified in Brockmann body by mass spectrometry. In addition, four forms of GLP-1 were identified; two 31 amino acid versions, one from each transcript and two 34 amino acid peptides, again one from each transcript (Table 3). Thus the four GLP-1s identified are: GLP-1 I (31aa), GLP-1 I (34 aa), GLP-1 II (31 aa), and GLP-1 II (34 aa). No GLP-2 was detected in the rockfish Brockmann body by mass spectrometric analysis.

Two fractions were detected and collected from the Sephadex G-50 column. Each fraction was separated on the C-18 RP-HPLC column. A comparison of Figure 15 and Figure 16 indicates that the separation profiles of the two Sephadex fractions were very similar, with only one major peak showing in Sephadex fraction 1 that is not in Sephadex fraction 2. This peak was analyzed by mass spectrometry but did not contain any proglucagon derived peptides. Due to the similarity in profiles and therefore probably similarity in composition, data from Sephadex fraction 2 were used for all further analysis.

A sample of each Sephadex fraction was also analyzed by mass spectrometry before separation by HPLC. These data are designated as "whole fraction". Every significant peak from the RP-HPLC (Figure 16) was analyzed by mass spectrometry. Peaks containing proglucagon derived peptides are numerically labelled in Figure 16. Most of the proglucagon derived peptides were identified in or around the major peptide peak of

the HPLC separation (peak10), which eluted around 42% acetonitrile. This is within the expected range for proglucagon derived peptides of 35-45% (Plisetskaya, E.M. et al. 1986; Cutfield, S.M. et al. 1993; Conlon, J.M. et al. 1998).

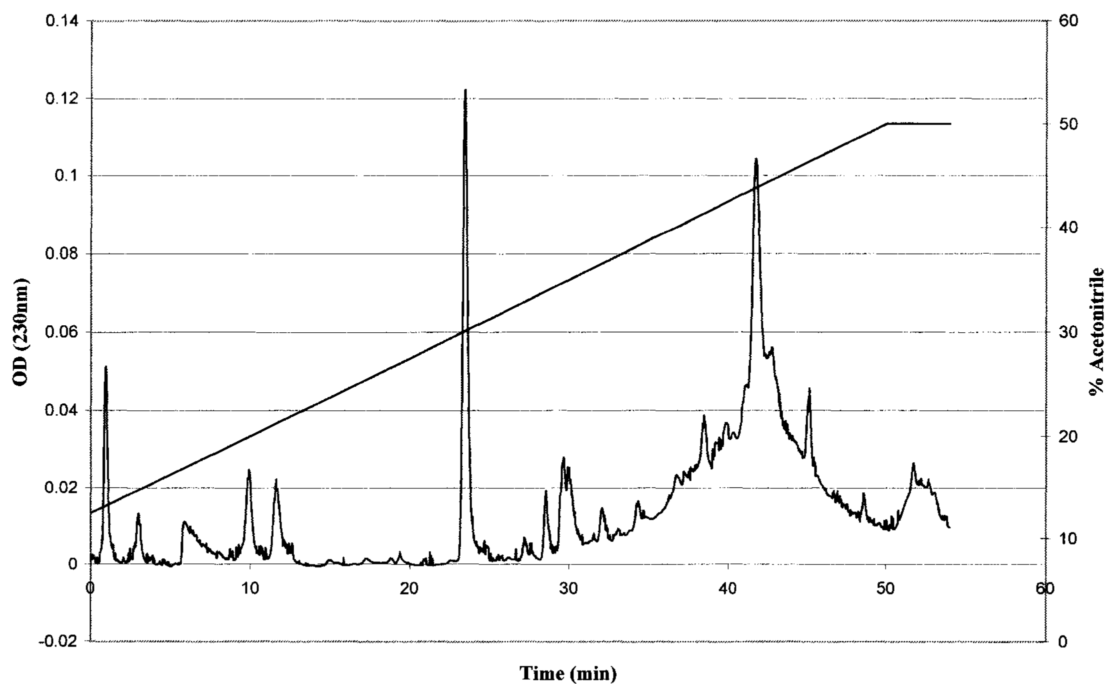

Figure 15. Partial purification of pancreatic peptides. Absorbance profile at 230 nm of Sephadex G-50 BB fraction 1 sample separated by C-18 Reverse Phase HPLC.

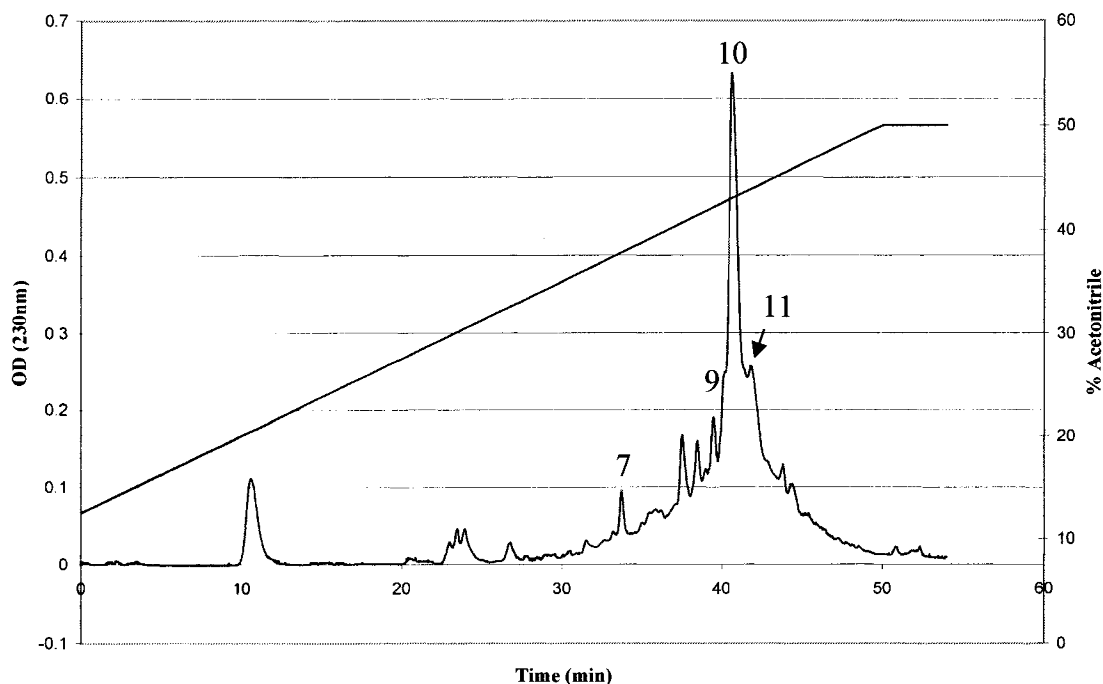

Figure 16. Partial purification of pancreatic peptides. Absorbance profile at 230 nm of Sephadex G-50 BB fraction 2 sample separated by C-18 Reverse Phase HPLC.

Table 3. Theoretical and experimental masses for copper rockfish proglucagon-derived peptides. Numbers in brackets indicate peptide length, NF indicates peptide not found.

Peptide	**Calculated Average Mass**	**Experimental Average Mass**	**HPLC peak**
Glucagon I (29)	3593.89	3594.57	whole
Glucagon II (29)	3507.78	3507.83	7
GLP-1 I (31)	3427.78	3428.14	10
GLP-1 II (31)	3494.78	3494.64	11
GLP-1 I (34)	3869.26	3869.16	7
GLP-1 II (34)	3936.27	3936.00	9
GLP-2 I	3830.30	NF	

GLP1 I (31aa) and GLP1 II (34aa) were clearly identified in HPLC peaks 10 and 9, respectively (Figure 17 and Figure 18). Glucagon II and GLP-1 I (34aa) were determined to co-elute in HPLC peak 7 (Figure 19). GLP1 II (31aa) and Glucagon I were more

difficult to identify as they generated low intensity mass spectrometry peaks in HPLC peak 11 and the Sephadex whole fraction 2, respectively (Figure 20 and Figure 21). The low intensity of the mass spectrometry analysis may indicate low amounts of these particular peptides in rockfish Brockmann body.

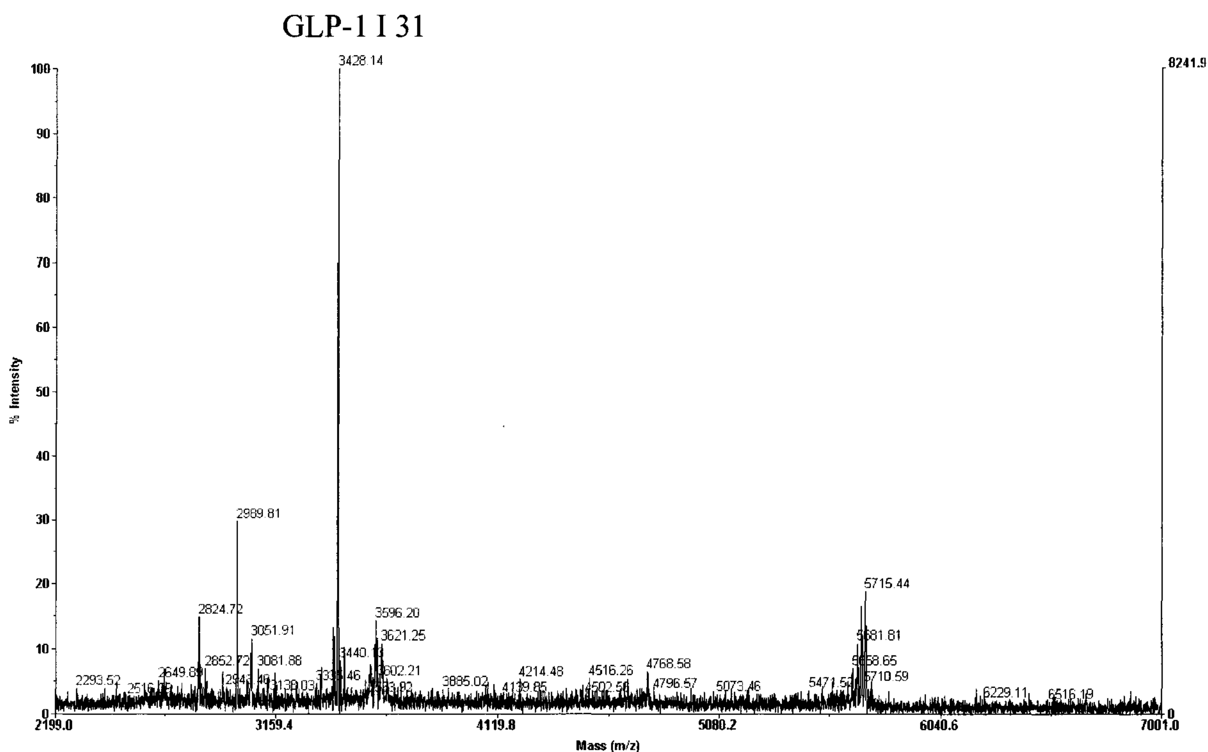

Figure 17. Mass spectrometry data for HPLC peak 10 identifying GLP1 I 31aa.

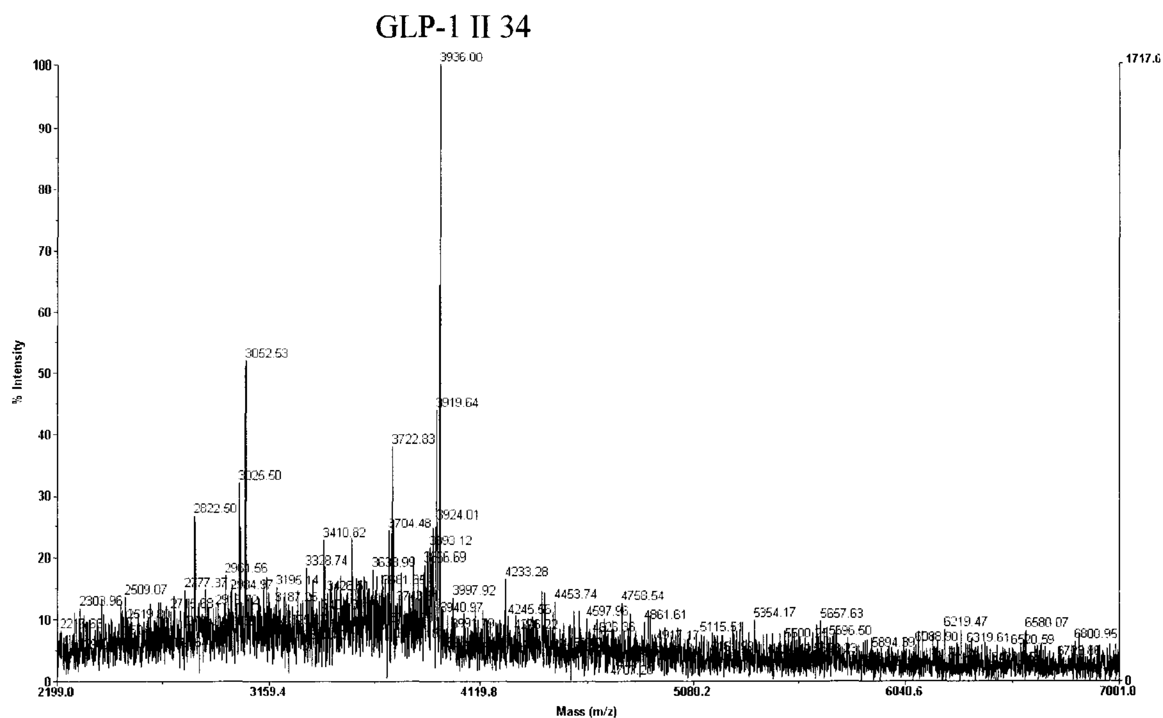

Figure 18. Mass spectrometry data for HPLC peak 9 identifying GLP1 II 34aa.

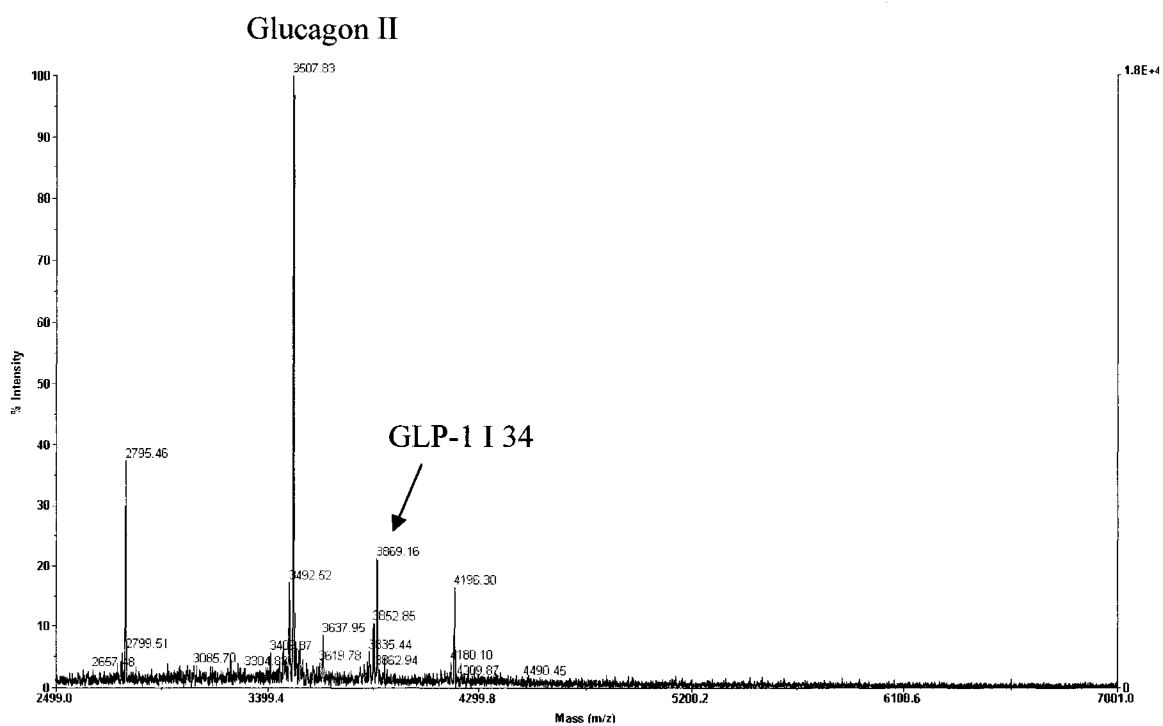

Figure 19. Mass spectrometry data for HPLC peak 7 identifying glucagon II and GLP1 I 34 aa.

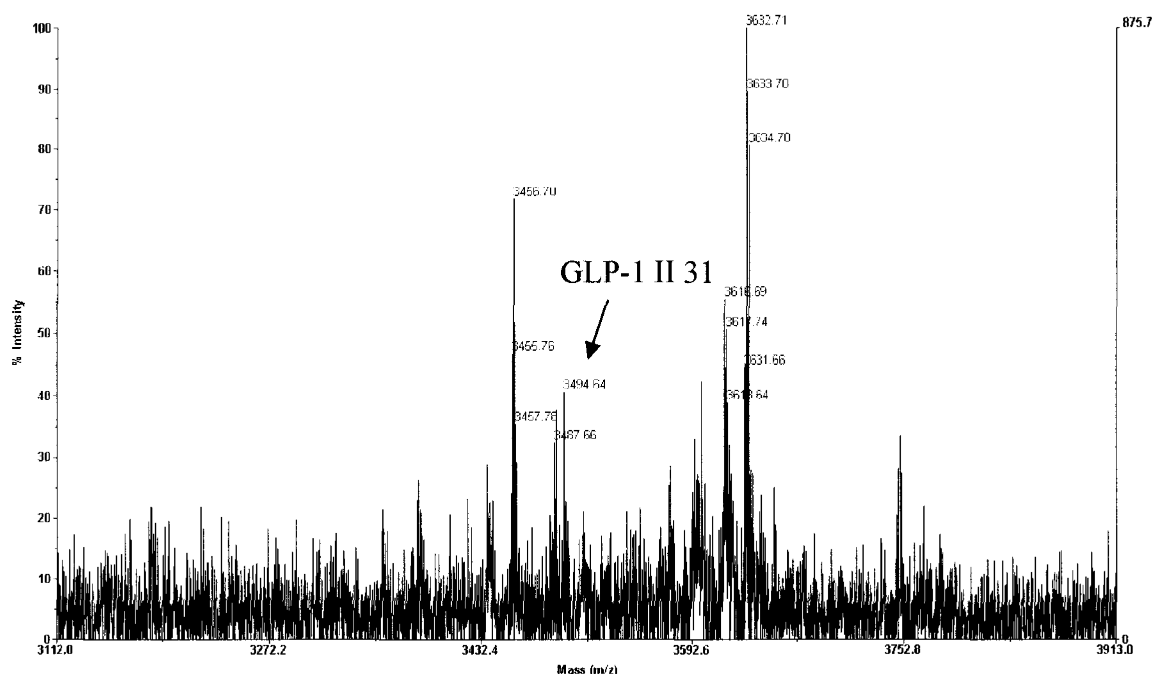

Figure 20. Mass spectrometry data for HPLC peak 11 identifying GLP1 II 31aa.

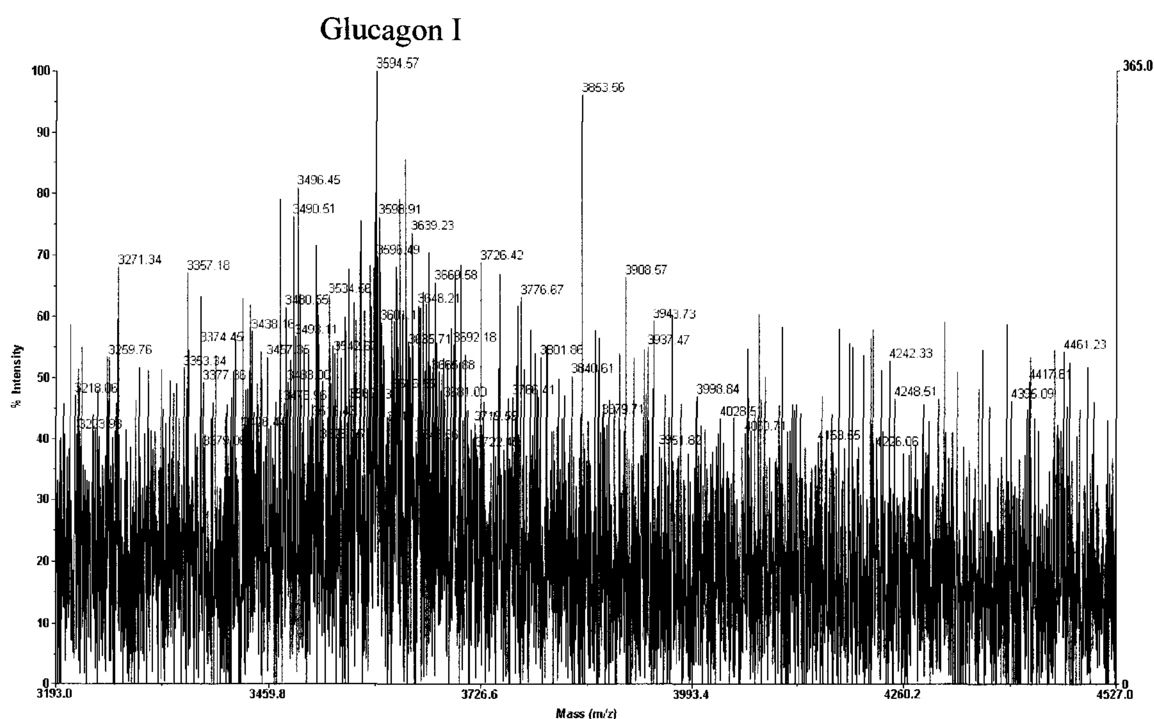

Figure 21. Mass spectrometry data for Sephadex peak 2 whole identifying glucagon I.

Considering differential ionization occurs with MALDI-TOF mass spectrometry, it is difficult to accurately quantify peptides. However, since the rockfish glucagons and

GLP-1s show relative similarity in size and charge, I would expect similar ionization of these peptides. Thus, the drastically different peak sizes seen in the mass spectrometric analysis would indicate a difference in peptide concentration. The major peptides present in rockfish Brockmann body are GLP-1 I (31 aa), glucagon II, GLP-1 I (34 aa), and GLP1 II (34 aa). The largest peak of the HPLC column (peak 10) contains GLP-1 I (31 aa) supporting the idea that it is one of the prominent peptides produced in the rockfish endocrine pancreas.

Discussion

In this work, I have identified at least two proglucagon genes in copper rockfish (*S. caurinus*), one of them encoding a truncated proglucagon sequence missing GLP-2. Both genes are differentially expressed throughout the three major proglucagon-expressing tissues, endocrine pancreas, gastrointestinal tract, and brain, giving some of these tissues unique proglucagon expression patterns. Besides differential expression, rockfish proglucagon I sequence can be alternatively spliced to express only glucagon and GLP-1, without GLP-2. Unexpectedly, this was found to occur in all three tissue types. Finally, partial peptide purification and mass spectrometric analysis of copper rockfish endocrine pancreas demonstrates the production of some, but not all, of the possible proglucagon-derived peptides expressed in this organ.

Comparison of the two rockfish proglucagon sequences to anglerfish

Two proglucagon sequences have previously been reported for the diploid anglerfish (Lund, P.K. et al. 1983). The authors suggested, from the differences noted between the two sequences, that they are derived from different genes. Comparison of the rockfish proglucagon sequences with the anglerfish indicates an interesting relationship between the four sequences. Comparing proglucagon open reading frame nucleic acid sequences, less similarity is seen within the rockfish than when comparing rockfish to the anglerfish. There is 86.3% identity in proglucagon I sequence between rockfish and anglerfish. And 87.5% identity is seen in proglucagon II for the same comparison. However, only 69.4% identity is noted between rockfish proglucagons I and II. The corresponding comparison

between anglerfish proglucagon I and II gives a nucleotide identity of 68.0%. Higher similarity is seen between proglucagon I of the two species than between two proglucagons of one species suggesting the gene duplication that created the two proglucagon genes occurred before the speciation of rockfish and anglerfish, and possibly before the plethora of speciation in the teleost group.

Different putative peptide sequences

The occurrence of two proglucagon genes allows the possible translation of two of each glucagon and GLP-1, with differing amino acid sequence. Thus the two glucagon or GLP-1 sequences may have different biological activities due to the amino acid differences between the duplicate peptides. Analysis of the substitutions between each rockfish peptide pair, taken with structure-function experimental data previously reported, albeit all for mammalian peptides, will give some insight as to possible differences in functionality between the rockfish peptides.

Within the two rockfish glucagon amino acid sequences, six substitutions exist, in positions 12, 16, 18, 24, 27, and 29 (Figure 22). Some of these substitutions are of similar biochemical nature, such as positions 12 and 18, which are both arginine vs. lysine, and position 24, which is arginine vs. glutamine. Yet other substitutions might have significant effect on the peptide overall binding affinity or activation of its receptor.

Receptor binding and activation assays with alanine substituted mammalian glucagon peptides have identified serine in position 16 as the only one of these positions required for potent receptor activation (Unson, C.G. et al. 1994). Considering this position is not highly conserved outside of mammals, and it is, in fact, one of the most variable positions throughout vertebrates, (Irwin, D.M. 2001) the difference seen in position 16 between the two rockfish glucagons, between a glutamate and a threonine, is probably not significant in creating differential efficacies of these two peptides. Actually, by analyzing the peptide sequence alone, with the information that is known about glucagon structure function relationships, there are no drastic differences between the two rockfish peptides to suggest they would have differing binding capacities or activation abilities. Of course,

as the glucagon receptor and specific positions of the peptide involved in binding has not been extensively studied in most fish, never mind rockfish specifically, it is difficult to make any statements about differences in alleged binding for these glucagon peptides.

```
Rockfish Glucagon I       HSEGTFSNDY SRYLEERKAQ DFVRWLMNN
Rockfish Glucagon II      HSEGTFSNDY SKYLETRRAQ DFVQWLKNS

Anglerfish Glucagon I     HSEGTFSNDY SKYLEDRKAQ EFVRWLMNN
Anglerfish Glucagon II    HSEGTFSNDY SKYLETRRAQ DFVQWLKNS

Leopard frog Glucagon     HSQGTFTSDY SKYLDSRRAQ DFVQWLMNS
Human Glucagon            HSQGTFTSDY SKYLDSRRAQ DFVQWLMNT
```

Figure 22. Amino acid alignment of some glucagon sequences, including the two putative rockfish sequences.

In the rockfish GLP-1 sequences nine substitutions exist, in positions 6, 15, 19, 21, 24, 25, 28, 30 and 31 (Figure 23), some of which are in positions previously described in mammalian GLP-1 as involved in receptor binding and activation. Of the positions that vary between the rockfish sequences, positions 6 and 25 have been identified, in studies done with mammalian GLP-1, as residues that interact with the receptor and are necessary for proper binding (Adelhorst, K. et al. 1994; Gallwitz, B. et al. 1994; Parker, J.C. et al. 1998; DaCambra, M.P. et al. 2000). Position 6 in the rockfish sequences varies between phenylalanine and tyrosine, two homologous amino acids, suggesting that this substitution is not likely to greatly affect the activity of the GLP-1 peptide. Position 25, on the other hand, is very highly conserved among vertebrates and is almost always a tryptophan residue (Irwin, D.M. 2001). In rockfish the GLP-1 I sequence, this position is a biochemically distinct residue, arginine (Figure 23). Although, this substitution is also seen in one of the anglerfish sequences, the effect of this substitution on functionality of the peptide is difficult to predict.

Positions 19 and 21 are in a region of the mammalian GLP-1 that has been identified as not important for receptor binding (Montrose-Rafizadeh, C. et al. 1997a), and considering this region of GLP-1 is relatively similar between the mammal and one rockfish sequence, this may hold true for the rockfish sequences. Alternatively, this region may

be more important for binding to the rockfish GLP-1 receptor than seen in mammals, hence the existence of two substitutions in this region may provide differing binding capacity for each rockfish peptide. Position 28 and 29 have been described as necessary for proper conformation for recognition by the mammalian receptor (Adelhorst, K. et al. 1994) and work with GLP-1/glucagon chimeras indicate that the C-terminal end of GLP-1 is involved in specific receptor recognition (Hjorth, S.A. et al. 1994). Thus variation of this region between the rockfish species may also provide varying levels of specificity or affinity for the rockfish GLP-1 receptor. Obviously, as GLP-1 peptides from different vertebrate species can activate membrane adenylate cyclase of different vertebrate species (Plisetskaya, E.M. et al. 1996; Conlon, J.M. et al. 1993), the changes that are seen across vertebrates must not completely prohibit receptor binding, yet I would propose that the difference in peptide sequence will accent or attenuate binding and activation ability of each peptide for its specific receptor. This is shown for the three xenopus GLP-1 peptides which have many substitutions throughout the sequence, and demonstrate different abilities to bind and activate human and fish receptors (Irwin, D.M. et al. 1997; Mommsen, T.P. et al. 2001). Thus, there is a possibility that the peptide residue variation between the rockfish GLP-1 peptides may be responsible for different binding efficiencies and thus potencies, especially considering the positions of some of the substitutions. The production of two biologically distinct GLP-1 or glucagon peptides increases the regulatory ability of this system.

```
Rockfish GLP1 I       HADGTFTSDV SSYLKDQAIK DFVNRLKSGQ V
Rockfish GLP1 II      HADGTYTSDV SSYLQDQAAK EFVYWLKTGR G

Anglerfish GLP1 I     HADGTFTSDV SSYLKDQAIK DFVDRLKAGQ V
Anglerfish GLP1 II    HADGTYTSDV SSYLQDQAAK DFVSWLKAGR G

Leopard frog GLP1a    HAEGTYTNDV TQFLEEKAAK EFIDWLIKGK P
Leopard frog GLP1b    HADGTFTSDM SSYLEEKAAK EFVDWLIKGR Q
Human GLP1            HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR G
```

Figure 23. Amino acid alignment of some GLP-1 sequences, including the two putative rockfish sequences.

At least two proglucagon genes identified in rockfish

Genomic DNA sequences for two proglucagon genes and multiple bands on the Southern blot demonstrate the presence of at least two distinct proglucagon genes. As mentioned before, multiple proglucagon genes have been reported for another diploid teleost species, anglerfish (Lund, P.K. et al. 1983) and multiple glucagon peptides have been isolated from several other diploid fish (Irwin, D.M. 2001). This taken with the newly reported rockfish proglucagon genes and additional reports of multiple proglucagon sequences from two other diploid fish, channel catfish and fugu (cf. Chapters 2 and 3), it is quite likely that at least two proglucagon genes is the expected state of teleosts. Two proglucagon genes have also been reported for rainbow trout (Irwin, D.M. et al. 1995), but considering this species is tetraploid and the two sequences show a high degree of sequence similarity, these genes are most likely the product of the quite recent genome duplication during the tetraploid event (Allendorf, F.W. et al. 1984).
To date, there have been no reports of more than two proglucagon genes in a diploid species, yet the Southern blot for copper rockfish indicated the possibility of a third proglucagon gene, partial gene, or pseudogene. Currently, without further work, it is difficult to determine if there is a third rockfish proglucagon gene, but a blast search of the *Takifugu rubripes* genomic DNA database did identify two proglucagon genes as well as an additional glucagon-like genomic sequence (cf. Chapter 3). It is conceivable to suggest that a partial gene like this may also exist in rockfish and be responsible for the third band in the Southern blot.

Truncated proglucagon II gene lacking GLP-2

When proglucagon cDNA sequence was first identified in fish, the transcript from anglerfish endocrine pancreas did not contain a GLP-2 sequence (Lund, P.K. et al. 1982). Identification of the proglucagon sequence from a teleostean intestine indicated teleost the proglucagon gene did contain GLP-2 sequence, but it was not expressed in endocrine pancreas due to alternative splicing of the intron preceding GLP-2 (Irwin, D.M. et al. 1995). Now I have identified that one of two rockfish proglucagon genes does not

encode GLP-2. This may be the case in all teleosts, suggesting that alternative splicing in the endocrine pancreas only occurs in one of the two proglucagon genes.

Alternative splicing in proglucagon I

Alternative splicing, in the case of glucagon, provides an increased level of regulation for this complex sequence. Considering that three distinct biologically functional peptides are co-encoded in the same precursor, it is difficult to express or produce one peptide without the others. One mechanism to assist in differential regulation of production of these peptides is alternative splicing, and as these peptides have different functions in different tissues, it may be desirable under particular physiological conditions to have differential production of these peptides.

Previously, alternative splicing, leading to the lack of GLP-2 encoded in the transcript, had only been reported in endocrine pancreas of teleosts (Irwin, D.M. et al. 1995). Here we have described that this regulatory mechanism also occurs in segments of the gastrointestinal tract and the brain of copper rockfish. Also, rockfish demonstrate how this mechanism may be important for regulation, as in the brain, only the alternatively spliced version of proglucagon I is expressed as well as the permanently truncated proglucagon II. Thus in the physiologically normal copper rockfish brain, no transcript for GLP-2 is expressed and glucagon and GLP-1 can be translated without GLP-2.

Alternative splicing is not new to the proglucagon gene, as typical alternative splicing, where a whole exon is spliced out, is reported in some, but not all, amphibian species. Alternative splicing removes GLP-1a from the transcript of tiger frog (*Rana tigrina rugulosa*) brain, colon, and pancreas (Yeung, C.M. et al. 2001), as well as splicing out GLP-2 in *Xenopus laevis* pancreas and stomach (Irwin, D.M. et al. 1997), yet no alternative splicing, in any position is reported for leopard frog (*Rana pipiens*) (Irwin, D.M. et al. 2000). This shows that alternative splicing, in whatever form, is most likely an important mechanism of transcript regulation in proglucagon, as it is with many other peptide prohormones, for example GnRH (Von Schalburg, K.R. et al. 1999).

Considering alternative splicing could play an integral role in expression of one proglucagon-derived peptide without another and the position of classical alternative splicing varies within the amphibians, there is merit in analyzing the teleostean proglucagon structure for possible other sites of alternative splicing. Copper rockfish proglucagon introns separate biologically functional units and some introns start immediately with a stop codon, as is seen in intron 4, the intron retained in rockfish proglucagon I alternative splicing. Although the intron directly following glucagon does not contain a polyA signal as does the intron following GLP-1 in proglucagon I, this intron starts with a stop codon, indicating that if retained, the stop codon would directly follow position 29 of glucagon, thus translation would produce a biologically intact peptide. If alternative splicing occurred in this location of either gene, production of glucagon without GLP-1 would be possible. At this point, no clear evidence of this in rockfish exists, but this form of regulation may be necessary only under certain physiological conditions.

Differential tissue expression of the two proglucagon genes

Between alternative splicing of proglucagon I and the presence or absence of each gene in each tissue, differential tissue expression was identified for copper rockfish proglucagon. In some tissues a complete absence of amplified transcript was identified indicating a lack of expression of that transcript. In other tissues, amplified products from each transcript were visible, yet drastic differences in band intensities indicate possible differences in relative transcript abundance. In either case, intricate regulation of the two proglucagon transcripts occurs to provide differential expression, which would most likely translate to different peptide profiles for each tissue depending on the functional needs of that tissue.

Brockmann Body - Proglucagon I transcript encoding GLP-2 identified

Two unexpected amplified products were identified from the rockfish Brockmann body, a non-alternatively spliced proglucagon I encoding GLP-2 and a proglucagon I including both intron 4 and GLP-2 sequence (the 1042 bp band). Both of these transcripts were

amplified with gene specific primers designed from the UTR sequences, but were not amplified in the 3'RACE PCR, which employs only one gene specific primer. This discrepancy is probably because the 3'RACE primers are not as specific and therefore not as sensitive. Also, as these bands were quite faint, the corresponding transcripts are likely to be of low abundance in the total RNA sample, and would not be amplified unless the PCR reaction has been optimized, which is rarely done for 3'RACE. If these transcripts are rare in the total RNA, the question is what is their purpose? Including cDNA data alone, it is impossible to determine if the non-alternatively spliced version of proglucagon I transcript, including GLP-2, is translated into peptide in the Brockmann body. The peptide mass spectrometric data did not confirm the translation of GLP-2, yet it is difficult to absolutely say that it is not translated. It is possible that a small portion of proglucagon I message is not alternatively spliced, yet also not translated. An explanation for the 1042 bp band including intron 4 and GLP-2 may be that during transcription of the alternatively spliced version, in a small percentage of the messages, the polyA signal in the middle of intron 4 is missed and the second polyA signal in the 3' UTR following GLP-2 is recognized. In sequences that include multiple polyA signals, which is seen with the rockfish proglucagon II sequence, the first one is not always observed by the enzyme (cf. Figure 7). Alternatively, this large band may represent the small percentage of messages that are being processed and the intron detected is the last one to be spliced out, revealing the full length proglucagon I. If this is the case, it indicates that intron 4 is the final intron to be spliced from the proglucagon RNA precursor, as the only intermediate message amplified contained intron 4. The fact that this potential intermediate mRNA form was the only intermediate detected may also suggest that perhaps pre-mRNA is kept at this stage to allow quick production of mature mRNA. The fact that the 1042 bp band only appears in tissues where the truncated form of proglucagon I is significantly brighter than the full length version, Brockmann body and stomach, supports the first hypothesis, in which the first polyA signal was not recognized, as the relative amount of this product would increase if the abundancy of the alternatively spliced proglucagon I was also high.

Brain - Absence of messenger RNA for GLP-2

mRNA encoding proglucagon has been detected in specific cell bodies of the rat brain stem, the nucleus of the solitary tract (NTS), suggesting *de novo* synthesis of proglucagon-derived peptides only occur in this part of the rat brain (Larsen, P.J. et al. 1997a; Han, V.K. et al. 1986). Yet, immunoreactive evidence in neurons for glucagon (Conlon, J.M. et al. 1979; Lui, E.Y. et al. 1990), oxyntomodulin, glicentin (Tager, H. et al. 1980), GLP-1 (Larsen, P.J. et al. 1997a), and GLP-2 (Tang-Christensen, M. et al. 2000) has been shown throughout mammalian brain, with highest concentrations in the hypothalamus. Central administration of GLP-1 (Turton, M.D. et al. 1996), GLP-2 (Tang-Christensen, M. et al. 2000) and oxyntomodulin (Dakin, C.L. et al. 2001) in rat brain leads to inhibition of food intake, suggesting these peptides may play a role in appetite regulation and satiation. This is supported by the fact that the region of the brain responsible for these functions is the periventricular nucleus of the hypothalamus where the highest concentrations of proglucagon-derived peptides were found.

Considering two forms of GLP-1 are expressed in rockfish brain, GLP-1 immunoreactivity has been detected in several parts of a catfish (*Clarias batrachus*) brain, including the hypothalamus (Sarkar, S. et al. 2001) and Silverstein *et al* (2001) have shown a distinct inhibitory effect of centrally administered GLP-1 on channel catfish (*Ictalurus punctatus*) food intake, it can clearly be suggested that GLP-1 may have a function in teleost brain. These points suggest that the role of GLP-1 is most likely similar to mammals and acts as an inhibitor of food intake due to signals of either satiety, the presence of nutrient in the gut, or interoceptive stress (McMahon, L.R. et al. 1998; Havel, P.J. 2001; van Dijk, G. et al. 1999).

The question is whether GLP-2, as has been suggested for mammals, also serves a function in the teleostean brain. In rockfish brain, there is no mRNA message for GLP-2, as is found in mammals, making the production of this peptide in the brain impossible. Thus, either GLP-1 fills the regulation of food intake role for teleosts, peripheral sources of GLP-2 reach the brain and assist GLP-1 in food intake regulation, or the function of GLP-2 is not maintained from mammals to fish and GLP-2 has no role in teleost brain.

At this point it is difficult to determine which scenario is the case, yet it is generally considered that fish brain is less complex than mammals, indicating that they may have a more simple food intake regulation system that relies more directly on gastrointestinal feedback and does not include GLP-2 production in the brain. Two peripheral sources of GLP-2 may complement the lack of GLP-2 production in rockfish brain, circulating and intestinal GLP-2. Two regions of the mammalian brain, the area postrema and the subfornical organ are outside the blood brain barrier and also have been shown to specifically bind GLP-1 (Goke, R. et al. 1995; Orskov, C. et al. 1996). Also, GLP-2 receptor message has been identified generally in the rodent brain stem (Yusta, B. et al. 2000), suggesting that some areas outside of the blood brain barrier may be able to respond to peripheral GLP-2. Also, GLP-1 has been demonstrated to be able to cross the blood brain barrier by passive diffusion (Kastin, A.J. et al. 2002) and possibly GLP-2 may as well. Thus circulating GLP-2 could reach the brain, by binding to receptors outside the blood brain barrier, or possibly crossing similar to GLP-1. Another mechanism GLP-2 may use to signal the brain could be through the vagus nerve that connects the intestine to the brain stem and is necessary for GLP-1 stimulation of CNS involved in GLP-1-induced inhibition of gastric mobility (Imeryuz, N. et al. 1997) or a signal of interoceptive stress (van Dijk, G. et al. 1999). Considering GLP-2 also stimulates the gut in a fashion that involves the CNS, signalling most likely also occurs through enteric neurons and the vagus nerve (Bjerknes, M. et al. 2001). Between the vagus nerve, circulating GLP-2, and perhaps a more simple form of food intake regulation, GLP-2 production in the brain may not be necessary for functions that involved the central nervous system in teleosts.

Gastrointestinal tract - Zonation of proglucagon II expression in intestine

The rockfish gastrointestinal tract shows differential expression of the proglucagon II in a zonational manner from the proximal to the distal, with the stomach expressing proglucagon II and the anus not expressing proglucagon II. Zonational expression suggests that rockfish intestine, although not obviously separated into distinct morphological segments has different functional regions, as has been indicated by expression and activity of some gastrointestinal enzymes (Mommsen, T.P. et al. 2003b;

Mommsen, T.P. et al. 2003a). Differential expression of proglucagon also suggests different functions, or at least different potencies, for the proglucagon-derived peptides from each gene throughout the intestine of copper rockfish. Specifically, there must be intestinal requirements for the proglucagon II products, oxyntomodulin or GLP-1, but not GLP-2 as it is not encoded in proglucagon II, in the proximal, but not the distal, end of the digestive tract. Although in mammals, the proglucagon-derived peptide most important for the gastrointestinal tract is GLP-2, which is involved in intestine epithelial mucosal growth (Drucker, D.J. et al. 1996), in teleosts the roles for oxyntomodulin or GLP-1 may be more significant, especially in the distal end of the intestine.

This does not preclude a function for GLP-2 in teleostean intestine, as proglucagon I, in both full length form, including GLP-2 and the alternatively spliced form are expressed in all segments of rockfish intestine and anus, suggesting that GLP-2 may also act as an intestinal growth factor in non-mammalian vertebrates.

Expression of two forms of the same gene in the same tissue

An interesting characteristic of intestinal and anal samples is the presence of both truncated and full-length versions of proglucagon I transcripts in the same tissue. If the full-length version is expressed, no different peptides would be produced by the concurrent expression of the truncated version. Therefore, a direct purpose of expressing both is not clear. As the peptide products are the same, it is impossible to identify if both transcripts are translated. Yet if translation of both leads to production of more peptide, this may be an alternate method of producing more oxyntomodulin and GLP-1 than GLP-2. Again, this may suggest a more significant role than expected for oxyntomodulin or GLP-1, which in mammals are involved in inhibition of gastric acid secretion and gastric motility. It is possible that oxyntomodulin or GLP-1 may be more involved in regulation of nutrient uptake and glucose homeostasis, especially as intestinal GLP-1 regulates the β-cell of the endocrine pancreas (Mojsov, S. 2000).

Another possibility is that both transcripts are not translated at the same time, but the mRNA of each transcript is available for a switch between translation of one with GLP-2

and one without GLP-2. In this case, induction of translation of the GLP-2 encoding message would most likely be triggered by a physiological event. And the constitutive expression and translation of oxyntomodulin and GLP-1, indicates an intestinal maintenance role for these proglucagon-derived peptides. In this case, mass spectrometric analysis of intestinal peptides should confirm the presence of GLP-2.

The presence of mRNA transcripts from full-length and alternatively spliced transcripts in the same tissue is not unique to rockfish intestine. This has also been described for proglucagon in gila monster (*Heloderma suspectum*) pancreas (Chen, Y.E. et al. 1997), *Xenopus laevis* pancreas (Irwin, D.M. et al. 1997), and tiger frog (*Rana tigrina rugulosa*) brain, colon, and pancreas (Yeung, C.M. et al. 2001). Because this phenomenon is widespread throughout non-mammalian vertebrates, there is most likely a physiological function for it, as opposed to a technicality of the identification method.

Identification of proglucagon derived peptides in pancreas

Peptides from both proglucagon transcripts were identified. This provides evidence that both transcripts were translated into the prohormone precursor and processed to release mature peptides.

Even though differential ionization can occur with MALDI-TOF mass spectrometry, all of the proglucagon derived peptides have similar charge, mass, and hydrophobicity, thus I would expect them to ionize in a similar fashion. Therefore, the great differences in intensity of ionization observed are most likely due to differences in concentration of the peptides in the samples. The physico-chemical similarity of the peptides is also supported by the fact that all the proglucagon-derived peptides identified eluted from the reverse phase HPLC column closely together.

The major peptides identified were GLP-1 I 31aa, GLP-1 II 34aa, glucagon II, and to a lesser extent GLP-1 I 34 aa. Very low levels of glucagon I and GLP-1 II 31 aa were also identified.

Thus, the fact that not all of the proglucagon derived peptides were found in equal amounts suggests regulation at the peptide level is also occurring in the complex regulation of proglucagon. This is most likely due to differential degradation of the peptides. Dipeptylpeptidase-IV (DPP-IV) is a hydrolytic enzyme that has been shown to inactivate mammalian glucagon, GLP-1, and GLP-2 by removing the first two amino acids of the peptide. A differential rate of inactivation has been shown for glucagon versus GLP-1, primarily due to the substitution of a serine residue for an alanine residue in position 2 of these peptides, respectively. Considering position two does not vary with the two rockfish glucagon or GLP-1 sequences, a teleostian version of DPP-IV could not be responsible for the differential degradation. I would suggest that other hydrolytic enzymes, as of yet unidentified, are responsible for the differential degradation of these very similar peptides.

Another possible explanation for the presence of only one of two possible peptides produced from one transcript is differential secretion. It is possible that some of the peptides are immediately released from the Brockmann body, therefore they have already been removed before peptide isolation begins.

Summary – It all comes back to regulation

In rockfish, and teleosts in general, the presence of two proglucagon genes that encode peptides with different amino acid sequences increases the complexity and versatility of possible proglucagon regulation. Differential tissue distribution of the two gene transcripts allows the production of one gene product without the other between tissues where the functions of each peptide may vary. Differential expression may be an indication that the gene products of each proglucagon gene differ enough to have individual potencies or roles in their function of each tissue. Another aspect of peptide functionality depends on the presence and efficient binding of the peptide to its receptor, as well as the message transduction system linked to each receptor.

The fact that the proglucagon II gene doesn't encode GLP-2 and the role of alternative splicing in expressing proglucagon I transcript without GLP-2 allow the transcription and

possible translation of some proglucagon-derived peptides without others. This adds another level of regulation to the proglucagon system, especially as alternative splicing occurs in many rockfish tissues.

Analysis of peptide production indicated unequal levels of peptides derived from the same prohormone precursor. This suggests regulation also occurs at the peptide level in either the form of differential peptide degradation or secretion. As only one tissue was analyzed for peptide production, we cannot determine if this is a tissue specific phenomenon, but expect that the degree of peptide regulation would vary from tissue to tissue, as other mechanisms of regulation, such as differential expression have.

Thus, we have demonstrated four levels of proglucagon regulation 1) the presence of two genes that encode different peptide products which may have different functional efficiencies 2) differential distribution of these genes allows production of one precursor without the other 3) alternative splicing allows the transcription of glucagon and GLP-1 without GLP-2 and 4) differential peptide concentrations from peptides derived in the same precursor. Other aspects of proglucagon regulation have not been investigated here, such as the presence and regulation of the specific peptide receptors and differential binding affinities for the peptide receptors in the target tissue. All of these mechanisms of regulation taken together provide a complex and intriguing system of regulation for a family of peptides involved in several aspects of metabolism.

Chapter 2 – Proglucagon Expression and Distribution in Channel Catfish and Identification of a Glucagon-like Receptor in Teleosts

Abstract

Two distinct proglucagon sequences were identified in channel catfish, *Ictalurus punctatus*, that are most likely encoded on independent genes. Including catfish, four teleost species have been reported encoding two independent proglucagon sequences, suggesting probably all teleosts have duplicate proglucagon genes. RT-PCR was employed to screen three tissues, brain, Brockmann body, and gastrointestinal tract, for the expression and possible differential tissue distribution of the two catfish proglucagon sequences. Both transcripts were expressed in endocrine pancreas, all parts of the brain, and all parts of the gastrointestinal tract, thus no evidence for differential tissue distribution was observed. Yet, some indication of differential message concentrations in various tissues was noted for each transcript. Also, peptide isolation and analysis by mass spectrometry indicates differential translation as peptides from only one transcript were identified in the endocrine pancreas. Therefore, this system demonstrates several modes of regulation for the proglucagon genes. In order to further elucidate regulatory aspects of proglucagon, I attempted to identify and characterize the message for putative GLP-1 receptors from several teleosts. Partial glucagon-like receptor sequences were identified in copper rockfish (*Sebastes caurinus*) and rainbow trout (*Oncorhynchus mykiss*). Further analysis of channel catfish proglucagon sequences and all teleost receptor sequences will answer some of the mysteries of teleost glucagon and GLP-1 regulation.

Introduction

Channel catfish (*Ictalurus punctatus*) is an important economical species, forming the basis of a large aquaculture industry in the United States, especially in warmwater aquaculture. Because of its commercial importance, its wide availability and other reasons, channel catfish has become a well studied model of fish metabolism and

genetics. Of course, as understanding food and nutrient intake and how it translates to fish growth is central to maximizing aquaculture yields, knowledge of hormones involved in these processes is essential. This naturally leads to the study of proglucagon derived peptides as they have been shown to play a role in regulation of glucose metabolism and uptake, as well as regulation of food intake in channel catfish (Silverstein, J.T. et al. 2001). Also, a role for glucagon-derived peptides in nutrient uptake and intestinal growth in mammals has been postulated, if not yet entirely supported experimentally (Lund, P.K. et al. 1993).

Teleostean proglucagon coencodes glucagon, GLP-1, and GLP-2, that are released from the precursor upon proteolytic processing, most likely by prohormone convertases. Currently, in teleosts, evidence has been presented for proglucagon gene transcription in three tissues, namely the brain, the endocrine pancreas, and the gastrointestinal tract (cf. Chapter 1)(Irwin, D.M. et al. 1995; Lund, P.K. et al. 1983). Additional evidence suggests proglucagon translation in the brain of another catfish species, *Clarias batrachus*, and endocrine pancreas of channel catfish (Sarkar, S. et al. 2001; Andrews, P.C. et al. 1985). In fish, glucagon and GLP-1 are directly involved in regulation of plasma glucose and glucose turnover by increasing the release of glucose from the liver through either gluconeogenesis or glycogenolysis. To date, the role for GLP-2 in fish has yet to be determined, but it functions as an intestinal growth factor in mammals (Drucker, D.J. et al. 1996).

Thus far, multiple proglucagon genes have been described for three teleostean fish, rockfish (*Sebastes caurinus*), rainbow trout (*Oncorhynchus mykiss*), and anglerfish (*Lophius americanus*) (cf. Chapter 1) (Irwin, D.M. et al. 1995; Lund, P.K. et al. 1983). For a number of additional species, two glucagon peptides have been isolated, also suggesting the presence of multiple proglucagon genes (Plisetskaya, E.M. et al. 1996; Irwin, D.M. 2001). As described below, channel catfish do not deviate from this picture, displaying two distinct proglucagon sequences with enough dissimilarity to suggest that channel catfish indeed also encode proglucagon on two independent genes.

Alternative splicing, differential gene expression, and differential peptide degradation are processes involved in regulating tissue distribution of the proglucagon derived peptides in other teleosts, namely rockfish (*S. caurinus*) and rainbow trout (*O. mykiss*) (cf. Chapter 1) (Irwin, D.M. et al. 1995). As I will describe here, some, but not all, of these processes, are involved in regulation of these peptide hormones in channel catfish.

In addition to the hormones themselves, regulation by these peptides can be controlled at the receptor level. The specific receptors for hormones of the glucagon-family are glycosylated seven transmembrane proteins associated with G-protein coupled transduction pathways. To date, non-mammalian glucagon receptors have been cloned and sequenced only from three amphibians, *X. laevis*, the leopard frog (*Rana pipiens*), and the tiger frog (*Rana tigrina rugulosa*) (Ngan, E.S. et al. 1999; Sivarajah, P. et al. 2001), while no non-mammalian vertebrate GLP-1 or GLP-2 receptor sequences have been reported. Some information about receptor activity has been determined from binding studies and measurement of physiological effects caused by treatment of target tissues with these peptide hormones in teleostean systems. In American eel (*Anguilla rostrata*) and black bullhead catfish (*Ictalurus melas*) hepatocytes, glucagon, at physiological concentrations, mediates glucose release primarily through activation of phospholipase C (PLC) and subsequent increases in IP_3 and intracellular calcium (Moon, T.W. et al. 1997; Plisetskaya, E.M. et al. 1996). Yet the glucagon receptor, when responding to higher concentrations of glucagon, responds with increases in cAMP (Plisetskaya, E.M. et al. 1996). GLP-1 stimulates the hepatic release of glucose and increases intestinal glucose uptake primarily through activation of adenylyl cyclase and production of cAMP, with a secondary involvement of the PLC/IP_3/intracellular calcium pathway (Moon, T.W. et al. 1997; Mommsen, T.P. 2000). Thus, teleost glucagon and GLP-1 receptors are most likely able to associate with more than one G-protein alpha subunit, G_s or G_q. Alternatively, as suggested by some mammalian GLP-1 receptor data, two molecular forms of each receptor could exist, each associated with a different signal transduction pathway (Montrose-Rafizadeh, C. et al. 1997b; Yang, H. et al. 1998).

Regardless, identifying cDNA and putative protein sequence for glucagon and GLP-1 receptors in a teleost is the first step in characterizing this receptor system. In this chapter, I will present partial cDNA sequences for a glucagon-like receptor from two teleosts, copper rockfish (*S. caurinus*) and rainbow trout (*O. mykiss*), beginning the process of molecular characterization of this receptor.

Methods

Fish

Channel catfish (*Ictalurus punctatus*) were maintained at the University of Victoria aquatics facility at 18°C under natural light conditions. Fish were killed by an overdose of anesthetic and tissues were immediately sampled and frozen in liquid nitrogen or stored in RNALater. Gastrointestinal tissues were rinsed with saline prior to freezing. Dissected catfish brains, including starved and refed brain tissues, were kindly sent frozen by Dr. J.T. Silverstein (USDA/ARS Catfish Genetics Research Unit, Stoneville, Mississippi, USA). Copper rockfish (*Sebastes caurinus*) and rainbow trout (*Oncorhynchus mykiss*) tissues were dissected from fish maintained at the University of Victoria aquatics facility, immediately frozen in liquid nitrogen, and stored at -80°C.

Total RNA isolation

Tissues were flash frozen in liquid nitrogen or on dry ice and stored at -80°C. Alternatively, thin slices of freshly sampled tissue were stored in >5x volume of RNALater (Ambion) at -20°C. Between 10 and 100 mg of frozen tissues were ground under liquid nitrogen and the frozen powder was transferred to 1.0 ml of Trizol (Sigma) in a 1.5 ml microcentrifuge tube (MCT). Alternatively, between 10 and 100 mg of tissues stored in RNALater were homogenized in 1 ml of Trizol using a Dounce homogenizer. Samples were inverted and vortexed to fully resuspend frozen powder then left standing at room temperature for ten to fifteen minutes. 200 µl of chloroform were added and samples were mixed by inversion until the phases were uniform, which usually took 1 to 2 minutes. After 3 to 5 minutes at room temperature, samples were centrifuged at 12,000 x *g* for 10 to 15 minutes at 4°C. The aqueous phase was then transferred to a

new 1.5 ml MCT and either 1 ml or 500 μl of isopropanol, for frozen or RNALater tissue, respectively, was added. Samples were inverted to mix and incubated at room temperature for 10 minutes or overnight at -80°C. Precipitated RNA was sedimented at 12,000 x *g* for 15 minutes at 4°C. The supernantant was decanted and the pellet was washed with 1 ml of 70% ethanol, followed by a 5 min centrifugation at 7500 x *g* at 4°C. The supernantant was decanted, tubes were tapped to remove excess ethanol and pellets were left to air dry at room temperature for 10 to 15 minutes. Pellets were not allowed to come to complete dryness. Pellets were dissolved in 5-50 μl of diethyl pyrocarbonate (DEPC)-treated sterile distilled H_2O (sdH_2O). Total RNA preparations were stored at -80°C.

cDNA synthesis

The cDNA synthesis reactions contained: up to 5 ug total RNA, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 0.8 mM dNTPs, 1.4 μM anchored polyT primer (AnchorT), 8 mM DTT, 5U of Ribonuclease inhibitor (GibcoBRL/Invitrogen), and 100U of SuperScript II RNase H^- Reverse Transcriptase (GibcoBRL/Invitrogen).

Total RNA was incubated with dNTPs and anchored polyT primer for 10 minutes at 65°C to reduce secondary structure, then immediately put on ice. The buffer, DTT, and RNase inhibitor were added and the samples were incubated at 42°C for 2 minutes. SuperScript II was then added and samples were incubated for 60 to 90 minutes at 42°C. After incubation, samples were heat denatured at 75°C for 15 minutes.

PCR

Primarily, the PCR cycles used in this work are based on the Standard PCR program and were varied only in annealing temperatures and cycle times. The PCR reaction ingredients and volume (25 μl) remained constant as: 20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 2 mM dNTPs, 1 μM of each degenerate primer or 0.4 μM each non-degenerate primer, 0.025 U/μl (0.625 U/25μl) *taq* DNA polymerase, and 0.5 μl of template. PCR reactions were performed on either a DNA Engine™ thermal cycler (MJ

Research Inc.), a PE9600 thermal cycler, a PE2400 thermal cycler (Perkin Elmer), or a Flexigene thermocycler (Techne). PCR products were visualized on 1% agarose gels stained with ethidium bromide. A 100 bp ladder (NEB) was used as a DNA marker.

PCR programs

Standard PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 55°C for 30 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Catfish Tissue Screening PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 61°C for 30 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Primers

Table 4. Names and sequences of primers used in identification of channel catfish proglucagon sequences.

Primer Name	Orientation	5'→ 3' sequence	Purpose
BBGlu51	Sense	CGGGCTCAGGACTTCGT	Initial PCR
BBGlu31	Antisense	TCGTCTCAACAGCATCTGGTTT	Initial PCR
CC13'UTR31	Antisense	GGAACGGGAAGCACTCTGTCT	PG I 5'RACE
CC3GLP131	Antisense	TCTCTTCTCCCTCTCCCCATT	PG II 5' RACE
ccGRPP51	Sense	CGAGTCTCCTGGACGAGAGGA	PG I 3' RACE PG I Screen
cc35UTR51	Sense	AGCCTGATAAGACGCCAGCAT	PG II 3' RACE PG II Screen
ccGLP231	Antisense	CCAAGCTGTCCAGCACCTTGTT	PG I Screen
cc33'UTR31	Antisense	AGGCAATGACACCACAGTTCC	PG II Screen
ccGluII5'out	Antisense	GTTCATGAGCCACTGAACGAAG	PG aI Screen
5' Actin	Sense	GTCGTCGACAACGGCTCCGGCATGTG	Actin Screen
3' Actin	Antisense	CATTGTAGAAGGTGTGGTGCCAGAT	Actin Screen
Anchor	Antisense	GGCTCGAGCCCGGGAATTCCG	3' RACE
5' RACE inner	Sense	GCTGATGGCGATGAATGAACACTG	Ambion 5' RACE
5' RACE outer	Sense	CGCGGATCCGGAACACTGCGTTTGC TGGCTTTGATG	Ambion 5' RACE

Table 5. Names and sequences of primers used in identification of glucagon-like receptor sequences.

Primer Name	Orientation	5'→ 3' sequence	Purpose
GLPR5282	Sense	CACTGCATGAGGAACARCATCCACATG	Initial PCR
GLPR3642	Antisense	CTGTACTGCATCATCACCAT	Initial PCR

mRNA Isolation

mRNA was isolated using Ambion's Micropoly(A)Pure™ mRNA Isolation Kit (Ambion Inc., Austin, TX), following the manufacturer's protocol. Briefly, fresh tissues were ground in lysis solution, containing guanidine thiocyanate, using a Dounce homogenizer, with 10 strokes of the loose pestle and 10 strokes of the tight pestle or until well homogenized. Frozen tissues were ground under liquid nitrogen in a ceramic mortar and pestle then the frozen powder was transferred to a Dounce homogenizer containing the lysis solution. Samples were homogenized by 10 strokes with each pestle. The homogenate was diluted with the dilution buffer and incubated with oligo dT cellulose resin shaking for 60 minutes at room temperature. The oligo dT cellulose, with now bound mRNA, was pelleted by centrifugation, washed with high salt buffers and collected in a column. The mRNA was eluted from the oligo dT cellulose resin with warm elution buffer.

5' and 3' RACE PCR

RACE PCR was performed using Ambion's First Choice™ RLM-RACE Kit (Ambion Inc., Austin, TX). The manufacturer's protocols were followed, with modifications made to the first strand cDNA synthesis and the PCR programs used. In the first strand cDNA synthesis, SuperScript II Reverse Transcriptase (GibcoBRL/Invitrogen) was substituted for the kit's AMV Reverse Transcriptase and the addition of DTT and RNase inhibitor. Modifications to the PCR program were made to suit the melting temperatures of the primers used for each species in which RACE was applied. For the catfish proglucagon 5' RACE PCR reactions, the standard PCR program was run with annealing temperatures from 57-60 °C.

Briefly , mRNA was treated with calf intestinal phosphatase (CIP) to inactivate any sequences that have been degraded or are not full length, then tobacco acid pyrophosphatase (TAP) was used to remove the cap of full length sequences. Adaptors were ligated to decapped transcripts, which were then made into cDNA with reverse transcriptase. Double stranded adaptors were supplied by the manufacturer (cf. Table 4 for adaptor primer sequences). This kit can also be used to perform 3' RACE as the primer used to synthesize cDNA contains a 3' RACE Adapter sequence.

Cloning

Cloning of PCR products was done using Invitrogen's TOPO TA Cloning Kit, vector pCR® II-TOPO®. After PCR products were separated out on high purity agarose gel (GibcoBRL/Invitrogen), bands to be cloned were excised and DNA was extracted from gel plugs by repeated freeze-thaw cycles followed by centrifugation at >10,000 x *g* for 5 to 10 minutes at room temperature. The supernatant was removed and used directly in the cloning kit. If *taq* DNA polymerase was not the primary enzyme used in generating the PCR product, the supernatant of the freeze-thaw gel extraction step was incubated with PCR buffers, $MgCl_2$, dATP, and *taq* DNA polymerase at 72°C for 15 minutes, then used in the cloning kit. The manufacturer's protocol was followed for the cloning procedure, except 0.5 µl of TOPO® vector, instead of 1 µl, and half of the volume of competent cells was used for each reaction. Positive clones, identified by the use of X-gal in standard blue white screening, were screened using a PCR based system with *E.coli* broths as the template and M13F and M13R sequencing primers as the amplification primers. Positive clones were identified by presence and correct size of PCR product bands when run out on a 1% agarose gel.

Plasmid Preparations

Plasmid preparations of cloned PCR products were performed using either the Small-Scale Preparations of Plasmid DNA (Sambrook, J. et al. 1989) or Quiagen's QIAprep Spin Miniprep Kit (Qiagen, Mississauga, ON). Both protocols were followed as

described. Plasmid concentrations were quantified spectrophotometrically at 260 nm and 280 nm.

Sequencing

Cloned products were sequenced by the Center for Environmental Health, University of Victoria, and processed on ABI DNA Sequencers, model 377 (Applied Biosystems, Foster City, CA) with fluorescent dye primers in a dideoxy method.

Peptide Extraction

Brockmann body tissue (96 mg) from an individual channel catfish (*I. punctatus*) was homogenized in 1.0 ml of acid ethanol (ethanol:0.7 M HCl 3:1, ice cold) with a 7 ml Dounce homogenizer until tissue was well homogenized. Homogenate was centrifuged at 2000 x *g* in a Hermle centrifuge at 1°C for 15 minutes. The supernatant was removed and the remaining pellet was resuspended in 0.75 ml of acid ethanol and ground in a Dounce homogenizer. The centrifugation step was repeated and supernatant was removed. Pooled supernatants were taken to dryness under reduced pressure and stored at 4°C.

C-18 SepPak

Dry pellets were resuspended in 1 ml 0.1 M HCl, with pipetting up and down, vortexing, and 7 sec sonication to resuspend. Insoluble matter was pelleted by centrifugation for 10 minutes at 12,000 x *g*. SepPak columns were conditioned with 10 bed volumes (8.5 ml) of acetonitrile followed by 10 bed volumes of dH_2O (8.5 ml) at a flow rate of 2 ml/min. 1.5 ml of resuspended sample was applied to pre-conditioned SepPak columns at a rate of 1 ml/min (gravity fed). SepPaks were washed with 2 ml of 0.1% TFA and peptides were eluted with a three step elution profile: 2 ml of ACN:dH_2O:TFA (25:74.9:0.1); 2 ml of ACN:dH_2O:TFA (45:54.9:0.1); and 2 ml of ACN:dH_2O:TFA (70:29.9:0.1). Proglucagon derived peptides were expected to elute in the 45:54.9:0.1 fraction. Samples were stored at 4°C until the next purification step.

Sephadex G-50

Sephadex beads were hydrated and equilibrated (30-40 ml) in a 26 cm x 2 cm column at 1ml/min with 0.1 M acetic acid, which is also the running buffer. Column size: gravity flow at 1ml/min. Approximately 2 ml of SepPak effluent was loaded and run at a flow rate of 1ml/min (gravity fed). 3 ml samples were collected and effluent absorbances were measured at 280 nm. Peaks with high absorbance were pooled and then taken to near dryness under reduced pressure in order to concentrate for loading on RP-HPLC.

C-18 Reverse Phase High Performance Liquid Chromatography

Buffer A: 0.1% TFA in dH_2O Buffer B: 0.06% TFA in acetonitrile. C-18 column was equilibrated with 80% Buffer B then 12.5 % Buffer B. The pooled Sephadex fractions were injected, the void volume was discarded and the program was started. Program: 1.5 minutes of 12.5% ACN (Buffer B), 50 minutes linear gradient from 12.5% ACN (Buffer B) to 50% ACN (Buffer B), 4 minutes at 50:50 0.1% TFA (Buffer A):ACN (Buffer B), then return to 12.5% ACN at a flow rate of 0.2 ml/min. Effluent absorbance at 230 nm was measured and peaks were collected. Depending on the volume of the peaks, samples were concentrated by centrifugation under reduced pressure for mass spectrometric analysis.

Mass Spectrometry Analysis

Samples were mixed 1:1 with matrix and spotted on sample grid block for crystallization and sample block was inserted into MALDI-TOF MS. The matrix used was either 2,5-dihyroxybenzoic acid or α-cyano-4-hydroxy-cinnamic acid. Mass spectrometric analysis was performed at two locations, the University of Victoria Proteomic facility, and with Dr. Michael Conlon at the Regulatory Peptide Center – Creighton University Medical School, Omaha.

Results

Sequence Data

Using specific primers (Table 4) and RT-PCR on mRNA isolated from channel catfish intestine and Brockmann body, two partial proglucagon sequences were identified. The two proglucagon sequences, called proglucagon I and proglucagon II based on similarity to the rockfish proglucagon I and proglucagon II sequences, were sufficiently distinct in sequence, with merely 61.6% nucleotide identity in the comparable coding regions, to suggest they originated from separate genes.

To determine the full length sequence of each transcript, gene specific primers, and adaptor primers were used in 5' and 3' RACE-PCR reactions. Most of the RACE-PCR reactions were performed on intestinal mRNA, but a few reactions on Brockmann body mRNA were also performed to obtain the 3' UTR of the alternatively spliced sequences. Results confirmed that the 3' UTR sequence in the full length proglucagon I sequence differed from the 3' UTR of the alternatively spliced proglucagon I sequence.

The complete sequence for channel catfish proglucagon I comprises 1028 nucleotides with an open reading frame containing 173 amino acids (Figure 24). The somewhat shorter alternatively spliced version of proglucagon I contains a sequence of 779 base pairs and an open reading frame of 122 amino acids (Figure 24). Proglucagon II was always identified without the GLP-2 sequence and also lacks the intervening sequence that is normally located between the C-terminal of GLP-1 and the N-terminal histidine of GLP-2. As a result, the cDNA sequence is 652 nucleotides long, including an open reading frame of 120 amino acids (Figure 25).

```
        5’UTR
1       aaagtcagaacagaaagcaagaagggaggccacggagcttacgcacacacacacacacac  60
                                                   Signal Peptide
                                                    M  K  G  T  H  F
61      acacacacaagcacacagactcgtctacgcaccggtgctcaagatgaaaggcacacactt  120
                                                         GRPP-I
          S  A  G  L  V  L  L  L  V  L  V  Q  S  S  L  E  L  P  L  Q
121     cagcgctggtctcgtcctgctcctcgtcctcgtccagagcagcctggagctcccgctaca  180

          D  D  T  T  S  S  E  T  D  A  S  L  L  D  E  R  K  D  T  M
181     agacgacacgaccagctcagagacagacgcgagtctcctggacgagaggaaagacacgat  240
                     Glucagon-I
          R  A  R  R  H  S  E  G  T  F  S  N  D  Y  S  K  Y  L  E  T
241     gcgggcaaggagacattcagaggggactttctctaacgactacagtaaatacctggagac  300
                                                IP1
          R  R  A  Q  D  F  V  Q  W  L  M  N  S  K  R  S  D  S  P  A
301     acggcgggcgcaggacttcgttcagtggctcatgaactccaagaggagcgacagcccggc  360
              GLP1-I
          R  R  H  A  D  G  T  Y  T  S  D  V  S  S  Y  L  Q  D  Q  A
361     ccggcgccacgctgacggcacctacaccagcgacgtcagctcgtacctgcaggaccaggc  420
                                                IP2
          A  K  D  F  I  T  W  L  K  S  G  Q  P  K  P  D  A  V  E  T
421     ggctaaagacttcatcacctggctgaaaagcggccagcccaaaccagatgccgttgagac  480
                                    GLP2
          R  S  V  D  T  L  R  R  R  H  V  D  G  S  F  T  S  D  V  N
481     gaggagcgtggacacgttacgcaggagacacgtggacggcagcttcaccagcgacgtcaa  540

          K  V  L  D  S  L  A  A  K  E  Y  L  Q  W  V  M  N  S  P  S
541     caaggtgctggacagcttggccgccaaggagtacctccagtgggtcatgaactctccgtc  600

          S  A  S  G  K  R  H  *  3’UTR
601     gtcggctagcggcaaacgacactgaagtcccggagaaagacagagtgcttcccgttcccg  660
661     gagtgaacttttaggatgtcgttcttcaccctgagaatcttcaaatcatggaaagccata  720
721     gtccttagccagagctgatacaacacaggaaatccaaagccatatatatatatatatata  780
781     tatataatgttcaatagtttgtactgcgagtatttatttgccagagatttgtgctacgcc  840
841     tgaaaagtcaaggatatgaaaaagaaatatttttctaatagatttactctatttttaatg  900
901     cttaatatttataaaatggttatgaaactcaggtaaaattttcaaaactctcaggttta  960
961     gtgagtgtcaaagaaaatgtagtatcaaaaatggtgccctgctctgcaaaataaaacttc 1020
1021    atgttcaataaaaaaaaaaaaaaa  1044

alternate splicing 3’ UTR starting at the last 13 aa of GLP-1

          A  K  D  F  I  T  W  L  K  S  G  Q  P  K  P  E  *
421     ggctaaagacttcatcacctggctgaaaagcggccagcccaaaccagagtaggcatcacc  480
481     accaccaccaccaccatcatcatcatcatcgccacggcaacctcacctactgtgcct  540
541     tatggagaaagaaacagataaagagatttaatagatagaaataacgagagatatttataa  600
601     caaaataaacccttatgcttctttttttatttcattttattttgactttcccttttcttt  660
661     tcttttttttttcctcttctttttttatccttctttggccttgtatctttttttccctttttg  720
721     attaccctgtcctaacccccttcttgctaagaagtggccacgctaataaatcgctgaatta  780
781     aaaaaaaaaaaaaaaaaaaa  800
```

Figure 24. Complete cDNA sequence and putative amino acid sequence for channel catfish (*Ictalurus punctatus*) proglucagon I. The sequence for the alternatively spliced 3’UTR follows the full length sequences beginning with the final 13 amino acids of GLP-1. Underlined sequences correspond to peptides, glucagon, GLP-1 and GLP-2. Bold italicized nucleotides indicate polyA signals and dibasic cutsites are bold.

```
        5′ UTR
1       aaaaaggaaacgctcctgggagcctgataagacgccagcatacaccagcaccctttctct  60
                                                         Signal Peptide
                                                         M  K  S  I
61      atcatccttctgtttcttcctttcggcagtcgatctacagggagcagaatgaagagcatc  120
                                                                GRPP-II
        Q  Y  L  A  G  L  L  L  L  I  I  V  Q  S  S  W  Q  I  P  I
121     cagtatttggctggacttttactcctcatcattgtacaaagcagttggcaaattcccatc  180

        D  D  T  E  E  N  S  S  V  M  L  E  D  T  L  L  N  N  P  T
181     gatgacacagaggaaaattccagtgtaatgttagaagatacgctactcaataatccaacc  240
                    Glucagon-II
        Y  M  K  R  H  S  E  G  T  F  S  N  D  Y  S  K  Y  L  E  T
241     tacatgaagaggcactcagaaggaacattctcaaatgattacagcaaatatctggagacc  300

        R  R  A  Q  D  F  I  Q  W  L  M  N  S  K  R  S  G  S  H  T
301     agaagagctcaggatttcatacagtggttgatgaactcaaagagaagtggttcccacaca  360
              GLP1-II
        R  R  H  A  D  G  T  Y  T  S  D  V  S  S  Y  M  Q  D  Q  A
361     agacggcatgcagacggcacatacaccagcgatgtgagctcatacatgcaggaccaagca  420
                                                                3′ UTR
        A  K  E  F  V  S  W  L  K  M  G  R  G  R  R  E  *
421     gccaaggagtttgtgtcctggctgaaaatggggagagggagaagagagtagagaagttcc  480
481     cttcagcaaaagagaagttctttatacctttttacacaggaaagagacattctgcaacact  540
541     cttttggaccatcctattccatcacaccatattttataaggaactgtggtgtcattgcct  600
601     aaaacaaaatgtgtttgaaataaatctctgctgaatctaagcactgtgatataaaaaaa  660
661     aaaaaaa   667
```

Figure 25. Complete cDNA sequence and putative amino acid sequence for channel catfish (*Ictalurus punctatus*) proglucagon II. Underlined sequences correspond to peptides, glucagon and GLP-1. Bold italicized nucleotides indicate polyA signals and dibasic cutsites are bold.

The two full-length channel catfish sequences differ significantly in nucleic sequences, with identity in the comparable open reading frames of 61.6%. Although the nucleotide identity is higher at 71.3%, when only the region from the N-terminal histidine of glucagon to the C-terminal of GLP-1 is taken into consideration, there is enough difference to suggest they are derived from two independent proglucagon genes.

Since genomic DNA sequence from two other teleosts, rockfish and fugu (*Takifugu rubripes*), had indicated the proglucagon II gene does not encode GLP-2, our expectation was that channel catfish would follow the same pattern and lack GLP-2 on gene II. In absence of studies involving genomic DNA for catfish, we approached this issue by utilizing multiple 3' RACE-PCR reactions with different gene specific primers. Using our data on rockfish as a guide, we focused on brain and intestine mRNA – the two tissue

most likely to express full length versions of proglucagon II. Confirming our idea, indeed, proglucagon II encoding GLP-2 was not identified in any tissue.

Tissue Distribution

During the process of sequence identification, full length proglucagon I sequence was found in intestine, and truncated versions of both proglucagons were found in Brockmann body. To determine more broadly the tissue distribution of both proglucagon transcripts, RT-PCR with gene specific primers was performed on separated brain parts, Brockmann body, stomach, intestine segments, and anus. Primers were designed either in regions of dissimilarity between the two sequences or in the UTR region of the sequences to ensure specificity. The primers were tested for gene specificity by PCR on diluted plasmid that included proglucagon sequence. β-Actin was used as a positive control for the RT-PCR reactions and also an indicator of the quality of the mRNA. The expected sizes and lane markers for the gel pictures of the RT-PCR reactions are:

β-Actin [A] 280 bp
Proglucagon I [I] 351 bp
Proglucagon II [II] 580 bp

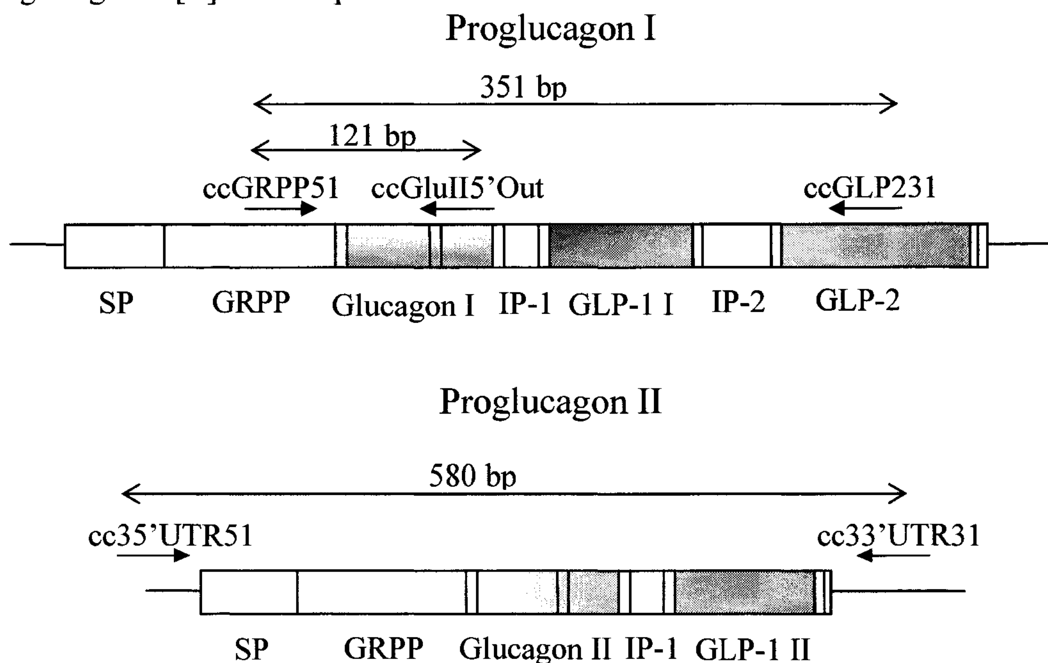

Figure 26. A schematic diagram of the catfish proglucagon transcripts, indicating the location of the gene specific primers used in RT-PCR reactions.

The antisense primer used in the RT-PCR reactions for proglucagon I is designed towards the center portion of the GLP-2 sequence (Figure 26), thus this primer will only amplify the full length version of proglucagon I that includes GLP-1 and will not amplify the alternatively spliced version of proglucagon I. When initial screening was performed, all tissues expressed the full-length form of proglucagon I. As a result, we did not anticipate any need to look for the truncated version, and no primers specific for the alternatively spliced 3' UTR were designed.

All tissues tested showed amplification of both proglucagon transcripts, but differences may exist in mRNA quantities of each transcript in the different tissues. Even though these RT-PCR reactions were not designed to be quantitative, drastic difference in relative band brightness are suggestive of differences in message composition, especially as the reactions are performed on the same cDNA, with the same PCR reagents under identical reaction conditions. For example, in catfish anus, the band corresponding to proglucagon I is much brighter than the band for proglucagon II, suggesting proglucagon I may be the predominant proglucagon message in this tissue. In contrast, whole brain RT-PCR indicates a proglucagon II band that is much brighter then the proglucagon I band, suggesting the opposite.

In RT-PCR reactions where little proglucagon II transcript was present to act as template, some assumed non-specific binding occurred as excess primers bind to unrelated template message. This resulted in a few larger bands of approximately 1000 bp appearing on the gels. In samples where the amplified band for proglucagon II is bright these non-specific bands do not appear, as ideal PCR reaction template to primer ratios occur.

Brockmann Body

A 351 bp band, corresponding to the full-length proglucagon I, and a 580 bp band, corresponding to proglucagon II, were amplified from the Brockmann body mRNA sample (Figure 27). This clearly demonstrates that the full-length version of proglucagon I, including GLP-2, is transcribed in the channel catfish Brockmann body. The 280 bp

band corresponding to β-actin confirms the quality of the cDNA and can be used as a coarse measure of relative abundance of messages. Also, unlike rockfish where the message for full-length proglucagon I seemed to be rather rare, the intensity of the band suggests that proglucagon I may be a relatively abundant species in the channel catfish Brockmann body. However, this postulate needs to be supported with additional data.

Considering the alternatively spliced version of proglucagon I was expected as one of the major transcripts in Brockmann body, RT-PCR reactions were also performed with primers that would amplify both versions of proglucagon I. This primer combination amplified a band of 121 bp, which is also seen in the [aI] lane of the Brockmann body gel. Unfortunately, because both of the proglucagon I bands are very bright, a significant difference in intensity cannot be determined, thus it is difficult to identify the presence of the truncated proglucagon I. I am confident the truncated version of proglucagon I is present in Brockmann body as it was amplified in the RACE PCR, but this PCR reaction was most likely saturated and producing as much product as possible. Thus, if there was a difference, the conditions of the reaction would limit its detection.

In several PCR reactions, a larger band, around 600 bp, was also amplified with the 351 bp band of the full length proglucagon I. Considering the primers in this PCR reaction show no evidence of non-specific binding under many conditions, this band could be the equivalent to what was seen in rockfish, i.e. a full length proglucagon I including the intron between GLP-1 and GLP-2. As no information about the gene is available or on the length of this intron, we cannot confidently state that this band is the full length version of proglucagon I including an intron.

Brain

RT-PCR on mRNA isolated from whole ground channel catfish brains amplified both proglucagon transcripts. In preparations from eight individual fish, the band corresponding to proglucagon II was notably brighter than the proglucagon I band, suggesting it may be the predominant proglucagon message in brain.

Channel catfish brains were dissected into telencephalon, hind brain, olfactory tract, pituitary, and hypothalamus. Proglucagon I and II mRNA messages were amplified from all of these sections. In some brain parts the bands corresponding to the proglucagon transcripts are faint, but this is also true of the β-actin bands of the same samples (Figure 27). β-actin may not be the best reference transcript, as it may not be evenly expressed throughout all parts of the brain.

Figure 27. Tissue distribution of proglucgon transcripts in channel catfish (*Ictalurus punctatus*) Brockmann body and brain as identified by RT-PCR. Whole and brain parts were analyzed. The brain parts are abbreviated as hind brain, Hind; olfactory tract, OT; telencephalon, Tel; pituitary, Pit, and hypothalamus, Hyp. Amplified products include β-actin [A], 280 bp, proglucagon I [I], 351bp, all forms of proglucagon I [aI], 121 bp, and proglucagon II [II], 580 bp and were visualized with ethidium bromide on a 1% agarose gel.

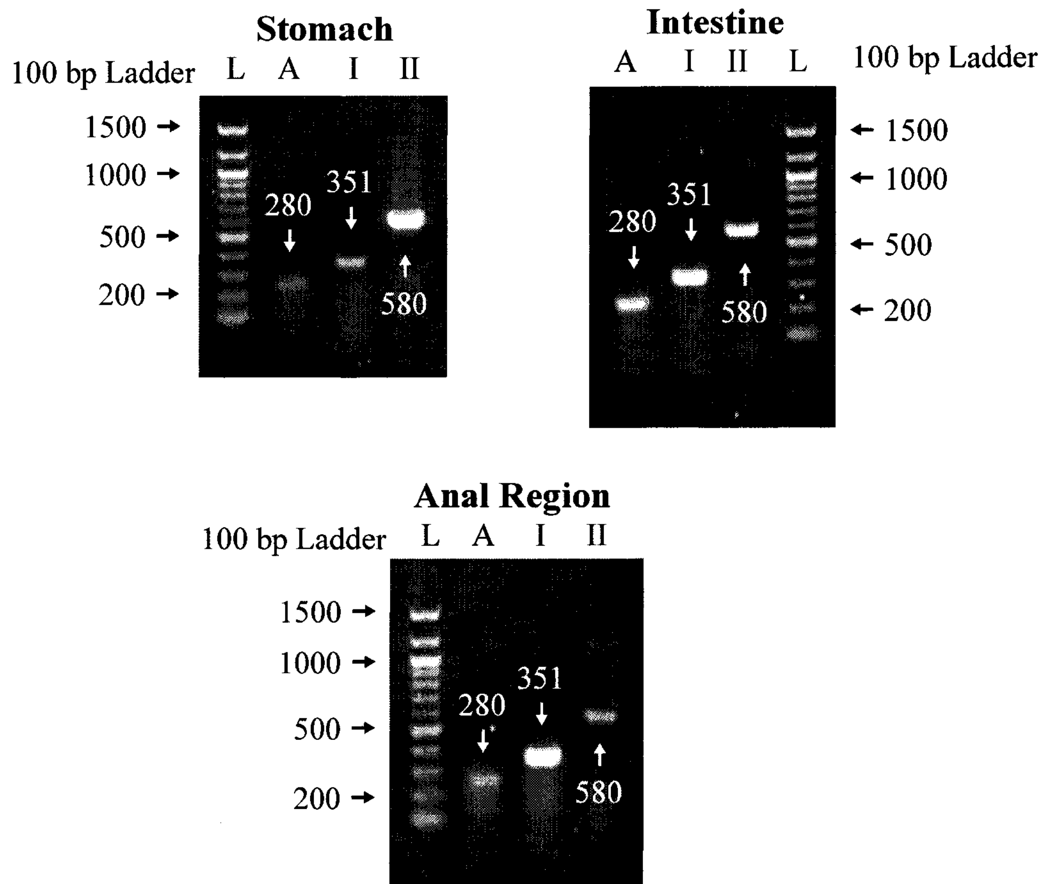

Figure 28. Tissue distribution of proglucgon transcripts in channel catfish (*Ictalurus punctatus*) gastrointestinal tract as identified by RT-PCR. Amplified products include β-actin [A], 280 bp, proglucagon I [I], 351bp, and proglucagon II [II], 580 bp and were visualized with ethidium bromide on a 1% agarose gel.

Gastrointestinal tract

Stomach

Segments of the stomach (proximal, central, and distal regions) were screened for proglucagon transcripts. No quantitative or significant quantitative differences were noted between the regions of the stomach. The bands produced as a result of RT-PCR for the posterior stomach are shown in Figure 28, with a bright band corresponding to proglucagon II and a notably lighter band for proglucagon I.

Intestine

Similarly, distal to proximal segments of intestine were screened for proglucagon transcripts, and as in stomach, both transcripts were detected in all segments. However, in contrast to stomach where proglucagon II predominated, similar intensity was seen for the bands corresponding to proglucagon I and proglucagon II in intestine (Figure 28). No differential expression of either gene was detected through the length of channel catfish intestine.

Anus

Both proglucagon transcripts were detected in the anal region of the gastrointestinal tract. The proglucagon I band was always appreciably brighter than the band corresponding to the proglucagon II (Figure 28), a situation opposite of what was found in the stomach.

Starved and Refed Experiments

Since glucagon-related peptides are intricately linked to the feeding status of animals, I also analyzed potential changes in expression of proglucagons in starved and refed catfish. This experiment was made possible through the availability of farmed catfish. From these animals, mRNA was isolated from brain and Brockmann body and screened for the proglucagon transcripts with RT-PCR. No difference in expression pattern or indication of different levels of proglucagon transcripts was detected between starved and refed animals for either tissue. The patterns seen did not differ from those presented in Figure 27.

Peptide Identification by Mass Spectrometry

Partial peptide isolation was performed on channel catfish Brockmann body tissue, to reveal the production of the following proglucagon-derived peptides from each transcript. Glucagon I, and GLP-1 I 34 aa were the main products isolated from channel catfish endocrine pancreas. Lower levels of a formylated GLP-1 I 34 and an unknown glucagon peptide were also identified. Formylation of GLP-1 is a byproduct of the isolation procedure when formic acid is used in the extraction solution (Andrews, P.C. et al. 1985).

The unknown glucagon peptide has an atomic mass different than predicted for either glucagon sequence, yet partial N-terminal sequencing of the isolated peptide identified it as a glucagon sequence. Thus, this peptide may be a modified version of either glucagon I or glucagon II, and unfortunately, the N-terminal sequencing didn't provide enough sequence to identify it as glucagon I or II, as the first unique residue is position 23.

As seen in rockfish and as expected by the lack of reports in other teleost species, no GLP-2 peptide was identified in the endocrine pancreas. As an alternatively spliced version of proglucagon I is most likely expressed in the Brockmann body and there is no evidence that the full length version, including GLP-2, is translated in Brockmann body, it is possible that this peptide is not produced in this tissue.

Not including the unknown glucagon peptide, as its source is unknown, only peptides from proglucagon I were identified, thus it is possible that proglucagon II in channel catfish is not translated. From the results seen for copper rockfish (cf. Chapter 1), this would be an unexpected result. Also, considering the channel catfish mass spectrometry analysis was done in a different institution, with a less sensitive mass spectrometry machine, it is possible some of the channel catfish proglucagon II derived peptides were not detected.

Table 6. Calculated and Experimental Masses for channel catfish (*Ictalurus punctatus*) proglucagon-derived peptides isolated from the Brockmann body and analyzed by MALDI-TOF mass spectrometry.

Peptide	Calculated Average Mass (a.m.u.)	Experimental Average Mass (a.m.u.)
Glucagon I	3510.80	3510.19
Glucagon II	3524.83	
Unknown glucagon		3541
GLP-1 I 31	3430.68	
GLP-1 I 34	3785.10	3784.43
Formylated GLP-1 I 34	3813.12	3813.47
GLP-1 II 31	3466.80	
GLP-1 II 34	3908.29	
GLP-2 35	3741.10	

Receptor sequences

Using RT-PCR and degenerate primers, we were able to amplify, clone, and sequence partial sequences for a putative GLP-1 receptor in both copper rockfish (*Sebastes caurinus*) and rainbow trout (*Oncorhynchus mykiss*). As initial assessment of these sequences showed similarity to teleost GLP-1 receptor sequence (Dr. S. Mojsov, unpublished data), we identified them as GLP-1 receptor sequences. However, further analysis and an additional amino acid sequence for a goldfish (*Carassius auratus*) glucagon receptor (Dr. S. Mojsov, unpublished data) has introduced some uncertainty into this identification. An alignment of these sequences with previously reported mammalian GLP-1 receptors, mammalian glucagon receptors and amphibian glucagon receptors (Figure 29) demonstrates regions of high similarity between glucagon, GLP-1 and our teleost sequences. Also, nucleotide percent identities between our teleost sequences and some other teleost glucagon and GLP-1 receptor nucleotide sequences (S. Mojsov, unpublished data) (Table 7) identifies the rockfish and trout sequences as putative glucagon or GLP-1 receptors. Additionally, transmembrane domain analysis of the partial rockfish and trout sequences, performed by TMAP software (Persson, B. et al. 1994), identifies the first two putative transmembrane domains, characteristic of this family of receptors (Figure 30), in the same positions as seen in the amphibian glucagon receptor (Sivarajah, P. et al. 2001).

However, analysis of informative amino acid sites, defined as an amino acid site where the state is conserved for the teleost (goldfish, zebrafish, and channel catfish) GLP-1 receptor sequences yet differs from the goldfish glucagon receptor sequence, and a peptide phylogenetic tree, indicate equal similarity between the new teleost partial sequences and the teleost glucagon or GLP-1 receptors. Analysis of informative sites identified 6 sites in the copper rockfish sequence that are identical to the glucagon receptor sequence and 6 sites identical to the GLP-1 receptor sequences, while 7 of each type of site were identified in the rainbow trout receptor sequence, indicating no overall higher identity for either glucagon or GLP-1 receptor sequence. In the phylogenetic tree (Figure 31), drawn from a Clustal W alignment of amino acid sequences at Biology Workbench website (http://biowb.sdsc.edu) (Thompson, J.D. et al. 1994), the rockfish

and rainbow trout sequences show similar distances to both teleost glucagon and GLP-1 receptors. Although the sequences appear more similar to the goldfish (*C. auratus*) glucagon sequence, without proper phylogenetic analysis and bootstrapping values, it is impossible to confirm or refute this identification.

Therefore, at this time we can merely identify the new rockfish and rainbow trout sequences as belonging to the proglucagon-derived peptide family of receptors.

```
               109
Onmy R         PQKNTSECDAND---PAQPHYGQILKKFRTMYTVGYSLSLGALVLALGILITFRK
Seca R         NTKNTSECDPND---RSLQYYGHI----RIMYTVGYSLSLVALVLALGILISFRK
HosaGlucR      PWRDASQCQMDGEEIEVQKEVAKMYSSFQVMYTVGYSLSLGALLLALAILGGLSK
XelaGlucR      SNRDARECELNKEDAEVEQNFFKTYERFKIVYTVGYSVSLGTLILAFAILVGFSK
RapiGlucR      RQNETLECQRDLEDLKEEEIFFKTYENFKVMYTVGYSSSVAALILALAILVGFSK
HosaGLP-1R     PWRDLSECEESKR--GERSSPEEQLLFLYIIYTVGYALSFSALVIASAILLGFRK
                 .:  :*: .           .       :*****: *. :*::* .**  : :

               164
OnmyR          LHCMRNNIHMNLFASFILRALSILIKDAMLEA--PSVNIGQDQE-----VTDYEV
SecaR          LHCMRNNIHMNLFASFILRALSILIKDALLEANIPSQDLSSDQEQGFPQASMPPV
HosaGlucR      LHCTRNAIHANLFASFVLKASSVLVIDGLLRTR-YSQKIGDD----------LSV
XelaGlucR      LHCMRNYIHMNLFASFILKSLSVLVQDILLKTR-YTKKIED------------EV
RapiGlucR      LHCMRNYIHINLFISFILRGVGILVSDTLLKTS-YS-KID------------FEH
HosaGLP-1R     LHCTRNYIHLNLFASFILRALSVFIKDAALKWM-YSTAAQQH-----------QW
               *** ** ** *** **:*:. .::: *  *.    :

               205
Onmy R         EWLVNNETAVGCRIA
Seca R         ELLVNNETAVGCRIA
HosaGlucR      STWLSDGAVAGCRVA
XelaGlucR      HSRLSSEALAGCRAA
RapiGlucR      HVEISNET--GCRVA
HosaGLP-1R     DGLLSYQDSLSCRLV
                  :.     .** .
```

Figure 29. Amino acid alignment of mammalian and amphibian glucagon and mammalian GLP-1 receptors with putative rockfish and rainbow trout GLP-1 receptor sequences. The alignment was drawn with Clustal W and position numbers are numbers corresponding to the *Rana pipiens* sequence (Sivarajah, P. et al. 2001).

Table 7. Percent similarities of partial regions of receptor cDNA sequences between several teleost glucagon and GLP-1 receptor sequences. Gaps created by maximum alignment of sequence were not included as matches or mismatches in analysis. See page 134 for species common names.

	Caau GLP-1R	Dare GLP-1R	Icpu GLP-1R	Xela GLP-1R	Rapi GLP-1R	Onmy R
Seca R	67.8 %	67.4 %	67.6 %	56.6 %	59.1 %	77 %
Onmy R	72.4 %	70.7 %	69.1 %	61 %	55.8 %	N/A

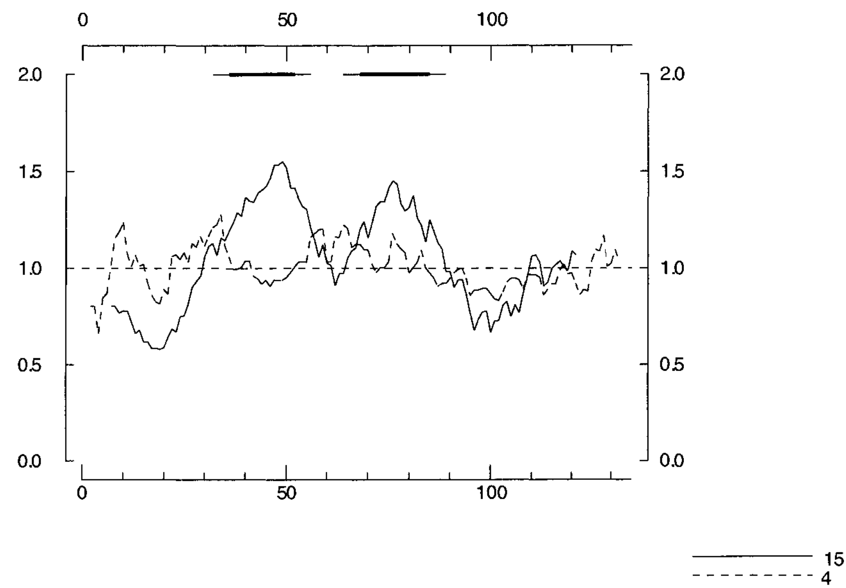

Figure 30. Transmembrane domain prediction for the partial glucagon-like receptor amino acid sequence from copper rockfish (*S. caurinus*). Transmembrane domain predictions performed using TMAP algorithm, available at Biology Workbench website.

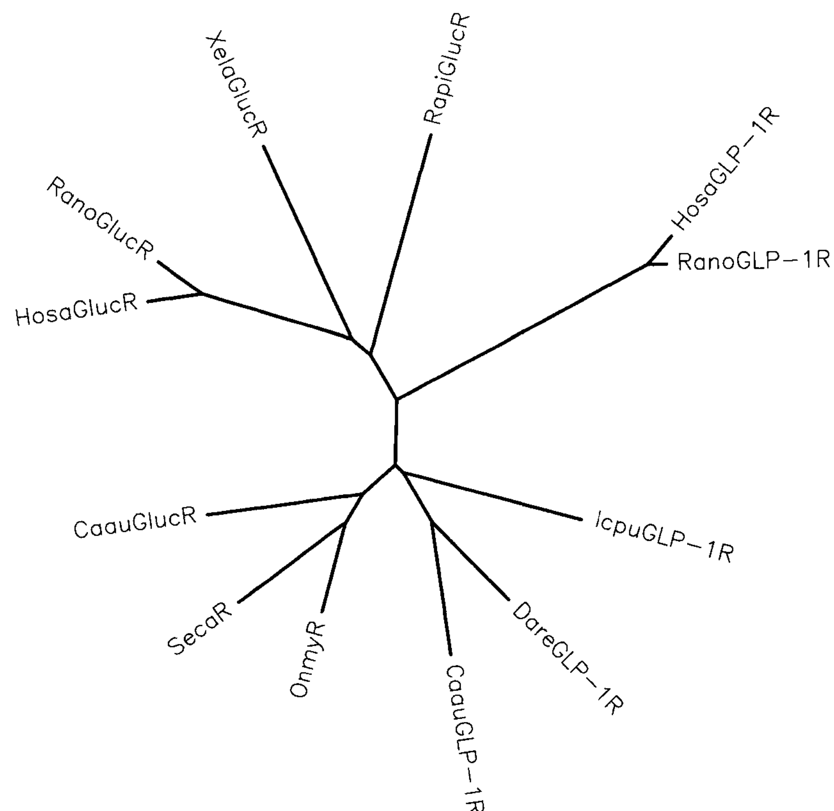

Figure 31. ClustalW phylogenetic tree for teleost, mammalian, and amphibian Glucagon and GLP-1 Receptors.

Discussion

Two catfish proglucagon sequences encode distinct peptides

The existence of two distinct sequences is becoming the norm for teleosts. I therefore venture to postulate that all teleosts have two proglucagon genes, but for some species the second sequence has yet to be identified. Sequence evidence clearly shows that two proglucagon genes also exist for channel catfish, as the nucleotides sequences significantly differ between the two channel catfish sequences identified.

Due to the conserved nature of glucagon, very little variation in the peptide sequence occurs between the two catfish transcripts. Glucagon I and glucagon II only differ by one amino acid, in position 23. The substitution seen in this position maintains the biochemical nature of the peptide, as it is from one aliphatic residue (valine) to another (isoleucine). Besides, position 23 of glucagon has not been identified by mammalian or evolutionary studies as a position that is expected to affect the binding to its receptor or biological action of glucagon (Unson, C.G. et al. 1998; Irwin, D.M. 2001).

In contrast, GLP-1 shows more variation in peptide sequence with seven amino acid differences between channel catfish GLP-1 I and GLP-1 II. As seen in Figure 32, all seven differences occur in the C-terminal half of the peptide with many of them being between biochemically similar amino acids, such as leucine to methionine (position 14), aspartate to glutamate (position 21), isoleucine to valine (position 23) and threonine to serine (position 24). More drastic amino acid changes occur in the final few amino acids of GLP-1, namely serine to methionine (position 28), glutamine to arginine (position 29), and proline to glycine (position 31). Studies done with mammalian GLP-1 indicate position 23 as the only of these positions that are critical for mammalian GLP-1 receptor binding and activation (Adelhorst, K. et al. 1994; Gallwitz, B. et al. 1994). Yet, as seen by evolutionary comparison many vertebrates possess valine instead of the mammalian isoleucine in position 23 and the similarity of the biochemical nature of these two amino acids suggests this substitution is unlikely to deleteriously affect receptor binding or activation. As the N-terminal end of GLP-1 is directly involved in receptor activation

and signal transduction, it has been suggested that the C-terminal portion of GLP-1 is more involved in forming the conformational structure involved in specific receptor binding (Adelhorst, K. et al. 1994). Thus, the C-terminal amino acid variation seen between the two catfish peptides may create different affinities to the GLP-1 receptor or may, in fact, suggest the existence of more than one GLP-1 specific receptor.

```
                           10         20         30
Icpu GLP-1 I    HADGTYTSDV SSYLQDQAAK DFITWLKSGQ P
Icpu GLP-1 II   HADGTYTSDV SSYMQDQAAK EFVSWLKMGR G
                ********** ***:****** :*::*** *:
```

Figure 32. Alignment of the amino acid sequences for the channel catfish GLP-1s.

Tissue Distribution

Brockmann Body

Considering the alternatively spliced version of proglucagon I and proglucagon II were the only expected transcripts of the Brockmann body, it was surprising to find a nicely amplified full length version of proglucagon I. For this reason, the need for primers that would specifically amplify the alternatively spliced version of proglucagon I was not anticipated. Regardless, it is clear that channel catfish transcribe a GLP-2 sequence in the Brockmann body. However, at this point, evidence indicating this transcript is actually translated into the GLP-2 peptide is still absent.

Brain

The fact that proglucagon transcripts are distributed throughout the catfish brain is not surprising. Although work on murine brain has identified proglucagon message only in the hypothalamus and brain stem (Drucker, D.J. et al. 1988), GLP-1 and GLP-2 immunoreactivity has been demonstrated in neural fibers leading from the brain stem to the thalamus, hypothalamus, cortex and pituitary. As well, GLP-1 and GLP-2 receptors were identified throughout the rodent brain (Lovshin, J. et al. 2001). In another species of catfish, *Clarias batrachus*, glucagon-like immunoreactivity has been identified throughout the olfactory portion of the brain, including the medial olfactory tract that lead to the telencephalon and preoptic area, parts of the hypothalamus, and the pituitary gland

(Sarkar, S. et al. 2001). Expression and immunoreactivity of proglucagon derived products and their receptors throughout the brain suggests a role for these peptides in the brain. This is supported by the work done with injections of GLP-1 into the periventricular nucleus of the hypothalamus or third ventricle leading to inhibition of food intake in rodents and channel catfish (Turton, M.D. et al. 1996; McMahon, L.R. et al. 1998; Silverstein, J.T. et al. 2001). Together, this evidence suggests that proglucagon-derived peptides may function throughout the catfish brain possibly regulating food intake in some capacity.

Surprisingly, proglucagon expression in the brain differs between rockfish and channel catfish. Rockfish do not express the full length version of proglucagon I and thus do not have expression of GLP-2 sequence (cf. Chapter 1). In channel catfish, the full length version of proglucagon I is expressed, albeit possibly not as the primary transcript. Nevertheless, the ability to translate GLP-2 in catfish brain clearly exists. This indicates a possible difference in the GLP-2 role or regulation between rockfish and catfish, as the rockfish must either rely solely on GLP-1 or peripheral sources of GLP-2 for feeding regulation in the brain (Kastin, A.J. et al. 2002), whereas channel catfish have the ability to produce GLP-2 in the brain and use it in a paracrine or autocrine fashion.

Gastrointestinal tract

The expression of proglucagon transcripts throughout the channel catfish gastrointestinal tract is very consistent. Both transcripts are expressed from the anterior end of the stomach through to the anus. Unlike rockfish, no differential tissue distribution of proglucagon II is seen in channel catfish gastrointestinal-tract, yet if quantitative mRNA analysis techniques were used, a similar trend might be detected. In channel catfish stomach, proglucagon II appears to be the dominant transcript and in anus it is proglucagon I, indicating that as hypothesized in rockfish, the roles for the two proglucagon sequences may vary throughout the length of the digestive system. It is not surprising that the degree of this trend is different between these two teleosts, as the diet and growth pattern, and therefore metabolic needs differ between the marine rockfish and freshwater channel catfish.

Starved and Refed Catfish

The lack of difference in expression patterns of the two proglucagon transcripts was unexpected for animals that had been starved compared to animals that were starved and then refed before tissue sampling, especially in brain where proglucagon derived peptides have been proposed to play a role in food intake inhibition. I anticipated observing a change in the expression patterns of the two proglucagon sequences under starving conditions compared to either untreated fish or starved and refed fish, as the two sequences encode peptides with varying amino acid compositions that may produce peptides of different efficiencies or possibly physiological roles.

Although unexpected, the lack of differential tissue distribution between the two experimental groups does not imply that proglucagon expression is not affected by starvation or that the proglucagons do not play differing roles in channel catfish tissues. Several factors may interfere with observation of a role for proglucagon in starvation. Firstly, the tissues were sampled four hours after refeeding in order to allow enough time for gene transcription to be turned on after starvation was ended. Although four hours is usually enough time to initiate gene transcription, this process for the proglucagon genes has never been studied and may diverge from the norm. Secondly, as both proglucagon transcripts are expressed in Brockmann body and brain of untreated fish, it may be necessary to utilize more sensitive techniques, such as quantitative real time PCR or RNase protection assays, to identify differences between the two proglucagon sequences. Third, considering the alternatively spliced version of proglucagon I was detected in these rockfish tissues, it is possible that the alternatively spliced version of proglucagon I is differentially expressed in these tissues of starved channel catfish. Finally, as mRNA message is only part of the story, different levels of the proglucagon-derived peptides may result from differential peptide processing or degradation in the brain, Brockmann body, or intestine of channel catfish that have been starved compared to refed animals. Ideally, peptides should be isolated from starved and fed tissues, especially brain and intestine, and analyzed by mass spectrometry to identify the protein profile of each tissue under different physiological conditions. Unfortunately, that was beyond the scope of this project at this time.

Proglucagon-derived peptide receptors

Analysis of the receptor sequences determined from copper rockfish (*S. caurinus*) and rainbow trout (*O. mykiss*) demonstrates higher similarity to other teleost glucagon and GLP-1 receptors than to either amphibian glucagon or mammalian glucagon or GLP-1 receptor sequences, as indicated by higher percent identities observed between the teleost sequences (Table 7, Figure 29 and Figure 31). From these data, some features of our teleost sequences are revealed: 1) the newly identified partial receptor sequences from rainbow trout and copper rockfish belong to the glucagon receptor family, 2) at this point, we can not identify whether the new sequences are glucagon receptor sequences or GLP-1 receptor sequences 3) higher similarity (77%) is seen between the rockfish and trout sequences than with the other three teleost GLP-1 sequences (67.6 - 70.7%), suggesting that rockfish and trout are the same type of receptor and 4) the teleost glucagon receptor sequence is more similar to the teleost GLP-1 receptor sequences than to other vertebrate glucagon receptor sequences. Yet, considering the drastic change in function of GLP-1 seen between mammals and teleosts and the similarity in function of teleost glucagon and GLP-1, this is not surprising. In fact, this supports the hypothesis that the receptor sequences as well as location of expression is responsible for changes in functionality of the proglucagon-derived peptides seen across vertebrates (Sivarajah, P. et al. 2001).

As this project continues and full length sequences as well as possibly other glucagon family receptor sequences are obtained, complete phylogenetic analysis will provide the identity of these teleost receptors. Also, analysis of receptor tissue distribution as well as stable expression for binding studies and analysis of the message transduction system will assist in characterization of these receptors.

Further characterization of this receptor family is desperately needed as currently, successful binding studies on *in situ* teleost GLP-1 receptors have been elusive (Plisetskaya, E.M. et al. 1996). Also, glucagon receptor populations with different binding affinities have been described (Navarro, I. et al. 1999), and evidence suggests glucagon and GLP-1 receptors can associate with different G-proteins to activate different signal transduction pathways (Moon, T.W. et al. 1997). Taken together, this

evidence suggests the possibility of multiple glucagon or GLP-1 receptor sequences in teleosts. This is reasonable as other peptides, such as NPY, in which multiple forms exist, activate multiple receptors (Gehlert, D.R. 1998). Also, considering the proglucagon sequences indicate an ancestral genome duplication in teleosts, more than one set of receptor genes are likely to exist, increasing the possibility of multiple glucagon-family specific receptors.

Summary

Two distinct proglucagon sequences, most likely encoded by separate genes, have been identified in channel catfish, supporting the hypothesis of all teleosts carrying two copies of the proglucagon gene. RT-PCR analysis of the tissue distribution of these transcripts identified both sequences in all proglucagon expressing tissues, namely Brockmann body, brain and intestine, thus unlike copper rockfish, no differential expression may occur in channel catfish. However, there is some indication that the level of each message expressed in the various tissues may not be equal between the two genes. Also, it is possible that differential translation may occur in channel catfish as peptide isolation and mass spectrometry in Brockmann body identified peptides derived from proglucagon I only. Additional peptide analysis would contribute greatly to the understanding of the regulation of proglucagon-derived peptides in other tissues, including catfish brain and intestine.

As channel catfish have been shown to reduce food intake in response to third ventricle injections with GLP-1 (Silverstein, J.T. et al. 2001) and GLP-1 immunoreactivity has been identified in another teleost, *Clarias batrachus* (Sarkar, S. et al. 2001), we attempted to obtain putative GLP-1 receptor sequence from several teleost species. We obtained partial sequences from two teleosts, copper rockfish and rainbow trout, with sufficient similarity to receptors from other teleost, amphibians, and mammals to identify them as proglucagon-derived peptide receptor sequences. However, with only partial sequences we cannot identify the new teleost sequences as glucagon receptor or GLP-1 receptor sequences. Continued work on this project will unravel this mystery and hopefully contribute to characterization of teleost glucagon and GLP-1 receptors.

Chapter 3 – Molecular Evolution of Proglucagon in Non-Mammalian Vertebrates

Abstract

The 29 amino acid peptide hormone glucagon is one of the primary hormones involved in the regulation of vertebrate glucose metabolism. It is essential in the precise regulation of circulating glucose in mammals and the availability of glucose as an oxidative substrate in non-mammalian vertebrates. Glucagon is encoded and translated as a precursor prohormone, proglucagon, that includes two additional biologically active hormones, namely glucagon-like peptide-1 (GLP-1) and glucagon-like peptide-2 (GLP-2). GLP-1 and GLP-2 are similar to glucagon in structure, but not in function. Proglucagon is multifaceted as expression, translation, and regulation of the three main peptides vary among tissues and species. Also, a change in the function of GLP-1 and possibly GLP-2 is seen between mammals and non-mammalian vertebrates. Here, we describe an evolutionary analysis of the proglucagon gene, using new data sets from cartilaginous fishes, several teleosts, two species of lungfish, an amphibian and a reptile, and compare them with published data on other vertebrate taxa, including birds and mammals. Using the neighbor-joining distance method for phylogenetic analysis of nucleic acid sequences, we identify the evolutionary stages of the proglucagon gene. My analysis reveals numerous fascinating alterations in structure and sequence indicating an active evolutionary history for proglucagons. The prohormone constitutes an intriguing model of a plethora of evolutionary processes that result in gene modification. Gene duplication, exon duplication, whole and partial exon loss, and increased evolutionary diversification facilitated by relaxation of functional constraints, all emerge as part and parcel in the evolution of the proglucagon gene in vertebrates.

Introduction

Proglucagon is a precursor that generally co-encodes the peptides hormones glucagon, glucagon-like peptide (GLP)-1, and GLP-2. In all vertebrates studied this far glucagon is involved in the regulation of glucose concentration and metabolism by stimulating

glucose release from the liver through gluconeogenesis and glycogenolysis. It is primarily produced in the α-cells of the endocrine pancreas.

Two other biologically active peptides released after proteolytic processing of proglucagon, GLP-1 and GLP-2, are not as well studied as glucagon. In mammals, GLP-1 is an incretin hormone that stimulates the production and release of insulin from pancreatic β-cells, thus opposing the function of glucagon. A similar insulinotropic action for GLP-1 is also noticed in some non-mammalian vertebrates, such as amphibians (Irwin, D.M. et al. 1997; Mommsen, T.P. 2000). Yet, in other non-mammalian vertebrates, such as teleostean fish, GLP-1 acts in a similar fashion to glucagon by stimulating the hepatic release of glucose either through gluconeogenesis or glycogenolysis. In these fish, GLP-1 is very potent at stimulating this effect and has been referred to as 'a better glucagon' (Plisetskaya, E.M. et al. 1996). No clear cut metabolic actions, other than via insulin, have been noted for GLP-1 in mammals, amphibians or reptiles (Mommsen, T.P. 2000).

Considering the opposing roles of glucagon and GLP-1 in mammals and amphibians, regulatory mechanisms exist to prevent the peptides from functionally canceling each other. One of these mechanisms is the presence of a six amino acid extension on the N-terminal of GLP-1 that masks the biological activity of the peptide. Only once this N-terminal extension is proteolytically cleaved, the peptide is fully biologically active. In teleostean fish, where the function of GLP-1 is similar to glucagon, the N-terminal extension is not found in the proglucagon structure.

To date, a role for GLP-2 has been identified in mammals only. It has been described as an intestinal growth factor as it stimulates epithelial proliferation in the intestine (Drucker, D.J. et al. 1996), and may play a role in regulating food intake (Tang-Christensen, M. et al. 2000), as GLP-2 has been shown to activate adenylyl cyclase in isolated rat brain membranes (Hoosein, N.M. et al. 1984a).

In most vertebrates, proglucagon is encoded on a single gene; this holds true for all mammals, chicken, gila monster (a reptile), and diploid amphibians (MacNeil, D.J. et al. 1994; Hasegawa, S. et al. 1990; Irwin, D.M. et al. 1995; Chen, Y.E. et al. 1997; Irwin, D.M. et al. 2000). In contrast, several teleosts as well as lamprey have been shown to encode proglucagon on two distinct genes (cf. Chapters 1 and 2) (Lund, P.K. et al. 1983; Irwin, D.M. et al. 1999). Multiple genes would be expected to increase the diversity of the sequences as duplication provides potential freedom from functional constraints. Such a theory could explain some of sequence and functional variation observed in proglucagon-derived peptides, especially considering both exon and gene duplication occurs in proglucagon throughout the vertebrates. This is demonstrated by the evolutionary analysis presented in this chapter.

Previous molecular evolutionary analysis of proglucagon derived peptides has shown high conservation in glucagon with less found in GLP-1 and GLP-2, respectively (Irwin, D.M. 2001). The above mentioned evolutionary analysis was performed on peptides sequences as currently more peptide sequences are available, compared to cDNA sequences. The work presented in this chapter reveals identification of additional cDNA sequences, some from vertebrate classes not previously represented, and an evolutionary analysis on nucleotide data including sequences from many vertebrate groups. One of the advantages of analyzing cDNA sequences is the ability to study structural aspects of proglucagon that are not indicated by peptide sequences alone. Also, as these peptides are highly conserved in function and length, cDNA sequence provides increased resolution in sequence variation. For instance, this consideration is very important with glucagon, where the peptide sequence is invariant across the mammals and among some teleosts.

Methods

Fish

African lungfish, *Protopterus dolloi*, were obtained live and maintained at the University of Victoria aquatics facility. Animals were killed by an overdose of anesthetic. Tissues

were dissected and immediately frozen in liquid nitrogen. Frozen tissues of Australian lungfish, *Neoceratodus forsteri*, were kindly supplied by Dr. Jean Joss (Macquarie University). Goby, *Gillichthys mirabilis*, tissues were sent frozen by Dr. Kevin Kelley from University of California at Long Beach. Spiny dogfish, *Squalus acanthias*, and ratfish, *Hydrolagus colliei*, tissues were collected at the Bamfield Marine Station. Tissues were frozen on dry ice immediately. The cane toads, *Bufo marinus*, were purchased live from W. Lemberger Co., Oshkosh, WI, USA and tissues were dissected from pithed animals and frozen in liquid nitrogen. Cottonmouth snake, *Agkistrodon piscivorus*, tissues were kindly supplied in RNALater by Stephen Secor (University of Alabama).

Total RNA isolation

Tissues were flash frozen in liquid nitrogen or on dry ice and stored at -80°C. Alternatively, thin slices of freshly sampled tissue were stored in >5x volume of RNALater (Ambion) at -20°C. Between 10 and 100 mg of frozen tissues were ground under liquid nitrogen and the frozen powder was transferred to 1.0 ml of Trizol (Sigma) in a 1.5 ml microcentrifuge tube (MCT). Alternatively, between 10 and 100 mg of tissues stored in RNALater were homogenized in 1 ml of Trizol using a Dounce homogenizer. Samples were inverted and vortexed to fully resuspend frozen powder then left standing at room temperature for ten to fifteen minutes. 200 µl of chloroform were added and samples were mixed by inversion until the phases were uniform, which usually took 1 to 2 minutes. After 3 to 5 minutes at room temperature, samples were centrifuged at 12,000 x *g* for 10 to 15 minutes at 4°C. The aqueous phase was then transferred to a new 1.5 ml MCT and either 1 ml or 500 µl of isopropanol, for frozen or RNALater tissue, respectively, was added. Samples were inverted to mix and incubated at room temperature for 10 minutes or overnight at -80°C. Precipitated RNA was sedimented at 12,000 x *g* for 15 minutes at 4°C. The supernantant was decanted and the pellet was washed with 1 ml of 70% ethanol, followed by a 5 min centrifugation at 7500 x *g* at 4°C. The supernantant was decanted, tubes were tapped to remove excess ethanol and pellets were left to air dry at room temperature for 10 to 15 minutes. Pellets were not allowed to come to complete dryness. Pellets were dissolved in 5-50 µl of diethyl pyrocarbonate

(DEPC)-treated sterile distilled H_2O (sdH_2O). Total RNA preparations were stored at -80°C.

cDNA synthesis

The cDNA synthesis reactions contained: up to 5 ug total RNA, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 0.8 mM dNTPs, 1.4 μM anchored polyT primer (AnchorT), 8 mM DTT, 5U of Ribonuclease inhibitor (GibcoBRL/Invitrogen), and 100U of SuperScript II RNase H^- Reverse Transcriptase (GibcoBRL/Invitrogen).
Total RNA was incubated with dNTPs and anchored polyT primer for 10 minutes at 65°C to reduce secondary structure, then immediately put on ice. The buffer, DTT, and RNase inhibitor were added and the samples were incubated at 42°C for 2 minutes. SuperScript II was then added and samples were incubated for 60 to 90 minutes at 42°C. After incubation, samples were heat denatured at 75°C for 15 minutes.

PCR

Primarily, the PCR cycles used in this work varied only in annealing temperatures and cycle times. The PCR reaction ingredients and volume (25 μl) remained constant as: 20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 2 mM dNTPs, 1 μM of each degenerate primer or 0.4 μM each non-degenerate primer, 0.025 U/μl (0.625 U/25μl) *taq* DNA polymerase, and 0.5 μl of template. PCR reactions were performed on either a DNA Engine™ thermal cycler (MJ Research Inc.), a PE9600 thermal cycler, a PE2400 thermal cycler (Perkin Elmer), or a Flexigene thermocycler (Techne). PCR products were visualized on 1% agarose gels stained with ethidium bromide. A 100 bp ladder (NEB) was used as a DNA marker.

PCR programs

Standard PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 57°C for 30 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Modified Standard PCR program:

Initial denaturation: 94°C for 3.5 min
35 x denaturation: 94°C for 30 sec
annealing: 55°C for 45 sec
elongation: 72°C for 1 min
Final elongation: 72°C for 15 min

Primers

For most of the successful PCR reactions used in this project, annealing temperatures between 55 °C and 58 °C in both the Standard and Modified Standard PCR programs were used with combinations of the following primers. Although tissue screening was not done for these species, β-actin was still used as a measure of mRNA quality, thus the primer sequences are found in Table 8.

Table 8. Names and sequences of primers used in the identification of various non-mammalian vertebrate proglucagon sequences. For brevity, abbreviations were used in the table: LF for both lungfish species, RF for ratfish, DF for dogfish, BM for *Bufo marinus*, the cane toad, and CM for the cottonmouth snake.

Primer Name	Orientation	5'→ 3' Primer Sequence	Purpose
XFGlu51d	Sense	GACTTYGTYCAGTGGYTAATGAAC	LF 3' RACE CM 3'RACE
GRPP52	Sense	WTMGTICARRGCAGCTKGCA	LF glucagon PCR RF glucagon PCR
LFGlu31	Antisense	CGTCCCTTCAKTAACCATCTG	LF glucagon PCR
DFGlu51	Sense	ACRAGCGACTACAGTAAGTATATG	DF & RF Initial PCR
DFGlu31	Antisense	GGTGTAGGTNCCRTCRGCRTG	DF & RF Initial PCR
DFGlu32	Antisense	GYCATTTTGACGCTTGTTGTAG	RF glucagon PCR
DFGlu52	Sense	CAGAAGGGAGCTACCGTGACA	DF 3' RACE
XFGlu51	Sense	ACRAGCGACTACAGKAARTAYATG	BM Initial PCR
XRGlu31	Antisense	GWGCACGCMWTWATATGKAGGRC	BM Initial PCR
BMGlu51	Sense	GACTTTGCCCAGTGGTTAATGA	BM 3'RACE
Anchor	Antisense	GGCTCGAGCCCGGGAATTCCG	All 3' RACE
5' Actin	Sense	GTCGTCGACAACGGCTCCGGCATGTG	Actin Test
3' Actin	Antisense	CATTGTAGAAGGTGTGGTGCCAGAT	Actin Test

3'RACE

Some 3'end sequences were obtained with a modified 3' RACE PCR technique in which we used an 'anchored' polyT primer for cDNA synthesis, then performed PCR reactions, using the Standard or Modified Standard PCR program, a gene specific sense primer and the anchor (antisense) primer.

Cloning

Cloning of PCR products was done using Invitrogen's TOPO TA Cloning Kit, vector pCR® II-TOPO®. After PCR products were separated out on high purity agarose gel (GibcoBRL/Invitrogen), bands to be cloned were excised and DNA was extracted from gel plugs by repeated freeze-thaw cycles followed by centrifugation at >10,000 x *g* for 5 to 10 minutes at room temperature. The supernatant was removed and used directly in the cloning kit. If *taq* DNA polymerase was not the primary enzyme used in generating the PCR product, the supernatant of the freeze-thaw gel extraction step was incubated with PCR buffers, $MgCl_2$, dATP, and *taq* DNA polymerase at 72°C for 15 minutes, then used in the cloning kit. The manufacturer's protocol was followed for the cloning procedure, except 0.5 µl of TOPO® vector, instead of 1 µl, and half of the volume of competent cells was used for each reaction. Positive clones, identified by the use of X-gal in standard blue white screening, were screened using a PCR based system with *E.coli* broths as the template and M13F and M13R sequencing primers as the amplification primers. Positive clones were identified by presence and correct size of PCR product bands when run out on a 1% agarose gel.

Plasmid Preparations

Plasmid preparations of cloned PCR products were performed using either the Small-Scale Preparations of Plasmid DNA (Sambrook, J. et al. 1989) or Quiagen's QIAprep Spin Miniprep Kit (Qiagen, Mississauga, ON). Both protocols were followed as described. Plasmid concentrations were quantified spectrophotometrically at 260 nm and 280 nm.

Sequencing

Cloned products were sequenced by the Center for Environmental Health, University of Victoria, and processed on ABI DNA Sequencers, model 377 (Applied Biosystems, Foster City, CA) with fluorescent dye primers in a dideoxy method.

Databases

Proglucagon sequence was found in the pufferfish, *Takifugu rubripes*, genomic DNA database, http://www.fugu-sg.org/tools/description.html, by searching with the BlastN and tBlastX algorithms. Copper rockfish, *Sebastes caurinus*, glucagon sequence was used as a query sequence.

With permission of Dr. Ben Koop at the University of Victoria, the local Atlantic salmon, *Salmo salar*, EST database was searched with BlastN and tBlastx algorithms (Altschul, S.F. et al. 1997) for proglucagon, also using copper rockfish glucagon as the query sequence.

Using keyword searches, the GenBank database was searched for previously published proglucagon sequences. Partial or complete proglucagon cDNA sequences were found and used in phylogenetic analysis, for the following species: leopard frog, *Rana pipiens*, complete cds; claw-toed frog, *Xenopus laevis*, two sequences complete cds; gila monster, *Heloderma suspectum*, complete cds; rainbow trout, *Oncorhynchus mykiss*, two full sequences, complete cds; anglerfish, *Lophius americanus*, two truncated sequences, complete cds; zebrafish, *Danio rerio*, truncated sequence, complete cds; goldfish, *Carassius auratus*, truncated sequence, complete cds; rock bass *Ambloplites rupestris*, partial, GLP-1 only; chinook salmon, *Oncorhynchus tshawytscha*, partial, GLP-1 only; sea lamprey, *Petromyzon marinus*, two sequences complete cds.

Proglucagon sequences for copper rockfish, *Sebastes caurinus*, and channel catfish, *Icatalurus punctatus*, previously reported by us, were also included in the analysis.

Phylogenetic Analysis

In nucleic sequence alignments, spaces were inserted either by eye based on the amino acid alignments or with Clustal W (Thompson, J.D. et al. 1994). Distance matrices were calculated with maximum likelihood method under typical parameters using the PHYLIP software package (Felsenstein, J. 1993). Phylogenetic trees were drawn from the distance matrices using the neighbor joining method with randomized input orders. Bootstrap values for branches were determined using the same combination of methods with 1000 repetitions and computing the best tree as the consensus tree. Bootstrap values over 50% are reported. The bootstrap consensus tree agrees with the presented tree on all branches with values above 50%, but some of the other branches have minor, insignificant variations in topology.

Phylogenetic trees were also drawn using the maximum likelihood method, but this method was too CPU intensive for determining bootstrap analysis in cases with many sequences, e.g. GLP-1. The neighbor joining method resulted in trees that were very similar to trees drawn with the maximum likelihood method, yet bootstrap analysis was computationally possible. Therefore the neighbor joining method was used for phylogenetic tree analysis.

All of the phylogenetic trees presented are unrooted trees. In order to generate a rooted phylogenetic tree, an outgroup is necessary for defining the root of the tree. As this analysis encompasses all vertebrate groups, a proglucagon sequence from a suitable outgroup does not exist. Therefore, to avoid making possibly incorrect assumptions about the root of this diverse group of vertebrates, all phylogenic analysis are presented as unrooted trees.

Results

The results of this chapter involve the description of cDNA sequences identified for proglucagon from several species representing different vertebrate groups, followed by presentation of the phylogenetic analysis of these and other previously reported proglucagon sequences in the form of five phylogenetic trees. In the discussion, new information pertaining to proglucagon evolution in three of the vertebrate groups, lungfish, elasmobranch, and teleost, will be discussed, followed by insights provided by analysis of the proglucagon sequences on topics such as exon duplication, increased diversity, and finally gene duplication.

cDNA sequences

The newly reported proglucagon sequences are shown with the putative amino acid sequences in Figure 33 to Figure 38.

The two lungfish sequences show similar proglucagon structures, both encoding a signal peptide, GRPP, glucagon, GLP-1 and GLP-2. Unlike teleosts and more like reptiles, birds, and mammals, the lungfish proglucagon contains a single GLP-1 with the 6 amino acid N-terminal extension. Although some variation occurs in the sequence of the N-terminal extension, especially in the amphibians, the lungfish N-terminal extension sequences bears significant similarity to the reptile and bird sequences. Throughout this analysis there have been no indications of more than one proglucagon gene in either lungfish species.

```
     Part of GRPP
     P   V   Q   D   A   E   N   E   S   R   L   L   N   T   S   E   M   E   A   V
1    cctgttcaagatgcagaaaatgaaagcaggttactcaatacttcagagatggaagcagta  60
                                             Glucagon
     D   E   S   R   E   H   T   E   A   K   R   H   S   Q   G   T   F   M   S   D
61   gatgaatctagagaacacacagaagctaagcgtcattcacaaggcacattcatgagtgat  120

     Y   A   K   L   L   E   A   K   H   A   L   D   L   V   Q   R   L   M   S   T
121  tatgctaaattactagaagccaaacatgcgctagaccttgtccagcggctaatgagcaca  180
                                          N-term extension  GLP-1
     K   R   T   G   A   V   S   K   R   H   N   E   I   E   R   H   A   E   G   T
181  aagcgaactggagcagtatcaaaacgtcacaatgaaattgaaaggcatgccgagggaaca  240

     Y   T   S   D   I   S   S   Y   L   E   G   Q   A   V   N   E   F   I   R   W
241  tataccagtgatatcagctcttacctggaaggacaggctgtaaatgaatttatcagatgg  300

     L   M   K   G   R   G   R   R   D   F   S   E   T   D   I   E   E   M   D   R
301  ttaatgaagggacgagggcgaagagattttcagaaactgatattgaagaaatggacaga  360
       GLP-2
     R   H   A   D   G   T   V   T   S   D   I   N   S   V   L   E   S   I   A   T
361  agacatgctgatggaacagtgactagcgatatcaacagtgtgttagaaagcattgcaacc  420
                                                        3'UTR
     R   E   F   L   N   W   L   I   N   A   K   S   S   *
421  agagaattcttaaactggcttatcaatgccaagagttcttaaacaagggatttcatcagt  480
481  gtgtgcaccaaagaaggaatgcattgttacttaagatatgtcatctcccaacatccactt  540
541  caaaaaatcttgaaaaacgtgttttgaatttataacctgtgccatatacatcaggctatc  600
601  atgtatgtcatagaattactttagtatattgctgccttgccaaagaaattatacaagga  660
661  ataaaatatggatgaatattaaggttttattttgaaatatgaagtacaacctcgtcaaat  720
721  ttacatatgcataagccaaaaaaaaagactactcacacggtatcacattttctgaaacta  780
781  tttttctactcaactgatttgtaaattttatatacctgcataaatgggtcagttgatgtc  840
841  aacgttactagtaacaagtgggtggtaaaccaatcagaaacacactgttattaaaatatg  900
901  aaaagcttttcagatgttacatgatagacaaaataaatatttttcatgttaaaaaaaaa  960
961  aaaaaaaaaaaaa   974
```

Figure 33. Partial proglucagon cDNA sequence from African lungfish (*Protopterus dolloi*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italacized and bold.

```
        Part of GRPP
         Y  R  L  F  N  A  S  E  M  E  P  V  D  A  S  R  E  H  T  E
1       aatacaggttattcaatgcttcagagatggaaccagtagatgcatctagagaacacacag  60
                    Glucagon
         A  K  R  H  S  Q  G  T  F  M  S  D  Y  A  K  L  L  E  A  R
61      aagctaaacgccattcacaaggcacatttatgagtgattatgctaaattactagaggcca  120

         H  A  L  D  F  V  Q  R  L  M  N  T  K  R  N  G  G  V  S  K
121     gacatgctctagactttgtacagcggctaatgaacacaaagagaaatggaggagtttcaa  180
             N-term extension   GLP-1
         R  H  S  E  I  E  R  H  A  E  G  T  Y  T  S  D  I  S  S  Y
181     agcgtcactctgaaattgaaaggcatgctgaagggacctacactagtgacatcagctctt  240

         L  E  G  Q  A  A  N  E  F  V  R  W  L  L  K  G  R  G  R  R
241     acttggaaggacaggctgcaaacgaatttgtcagatggttactgaagggacgaggtagaa  300
                                                       GLP-2
         D  F  S  D  T  D  A  E  E  M  G  R  R  H  A  D  G  T  I  T
301     gagatttttcagacactgatgctgaagaaatgggaagaagacatgcagatgggactatca  360

         N  D  M  N  N  V  L  E  S  I  A  T  R  E  F  L  N  W  L  I
361     ctaatgatatgaacaatgttttagaaagcattgctacaagagaattcctaaactggctaa  420
                          3'UTR
         N  S  K  G  S  *
421     tcaactccaaaggttcttaaacaagtgatttcttcagtgaattccagaaaaggaatgcat  480
481     tgttactcaagatttgtcgtctcccaacatttgccagccacttgaaaaatcttgaaatat  540
541     gtgttttgtaaatgataacaggtgtcacataaagccacgctgctttgcatgtcacagaat  600
601     gactcatattatattgctctgttgctaaagggattatataaggaataaaatattgttaaa  660
661     ggatagttaaggttttaccttgaaatacaaagtgccacctaatcaaatttcaatatattc  720
721     aagccaaaaaaaagaactacctgtctgtttacacattttatgaaactatttttctataca  780
781     attgatttgtatatgtatatatgaataattaagtgcaatcaattactgcttacgttatag  840
841     caaaaagggcgtggtaaaaagaagtcttctgaaaatcaaaattcactgttaaaacattgg  900
901     ttgtcaaacaacaacagtgaaaagcttttaggatattaaatgttagcttaaataaacttt  960
961     tttcacgttaaaaaaaaaaaaaaa  984
```

Figure 34. Partial proglucagon cDNA sequence from Australian lungfish (*Neoceratodus forsteri*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italacized and bold.

An unexpected result was the proglucagon structure found in the elasmobranch species, spiny dogfish (*Squalus acanthias*) and spotted ratfish (*Hydrolagus colliei*). In these species, one proglucagon sequence was identified that encoded two consecutive GLP-1 sequences, similar to what has been previously reported in most amphibians, specifically the *Rana* species (Irwin, D.M. et al. 2000; Yeung, C.M. et al. 2001). Unlike the frog sequences, where only the first GLP-1 is N-terminally extended, the elasmobranchs have the N-terminal extension of GLP-1 on only the second GLP-1. In the elasmobranchs, like the lungfish, only a single proglucagon gene was identified.

```
        C-terminal part of Glucagon
        D  N  R  R  A  K  D  F  V  Q  W  L  M  S  T  K  R  N  G  D
1       gataatagacgtgcaaaagactttgtacaatggctgatgagcacaaaacgaaatggtgac   60
             GLP-1a
        K  T  K  R  H  A  E  G  T  Y  T  S  D  V  D  S  L  S  D  Y
61      aaaaccaagaggcatgctgaaggaacttacacaagtgatgtagattctctttctgactac   120

        F  K  A  K  R  F  V  D  S  L  T  S  Y  N  K  R  Q  N  G  R
121     ttcaaagcaaaacgctttgtagattcgcttacaagctacaacaagcgtcaaaatggcagg   180
                                N-term extension  GLP-1b
        S  I  S  K  R  H  V  E  S  T  R  H  T  E  G  S  Y  R  D  I
181     agcatttctaagagacacgttgaatctaccagacacacagaagggagctaccgtgacata   240

        S  S  Y  L  E  A  K  A  A  K  D  F  I  N  W  L  I  K  G  R
241     agttcttacctagaggcgaaggcagcgaaggactttattaactggctaataaagggacgt   300

        G  R  R  E  F  P  E  E  S  K  E  I  V  N  E  V  I  P  E  E
301     ggcagaagagagtttccagaagaaagtaaggaaattgtgaatgaagtgatccctgaggaa   360

        M  D  R  R  H  A  D  G  T  F  T  S  E  I  N  I  V  L  D  T
361     atggacagaagacatgctgatggcaccttcaccagtgaaattaatattgttttggacacc   420

        I  A  A  K  E  F  L  N  W  I  L  N  S  K  T  I  Q  S  R  D
421     attgctgccaaggagtttttgaactggatcttaaactccaaaaccatccagtcaagggac   480
                                              3'UTR
        S  E  I  E  F  F  N  E  Y  N  *
481     tctgaaatagaattcttcaacgagtacaactaaaagaagagactgatttcaatttagcat   540
541     tttgtgttgcaacattggctacgtcttgattaaatatttacagctttttgttccgcaagt   600
601     gacacagaatgttatatgaagccataatgctttgcatgcaaaagaattgactcaaaatgt   660
661     aatgtgctgttttgaagtaggatttgaagtaaataaatgttatcaaaggaaaaaaaaaa   720
721     aaaaaaaaaaaa
```

Figure 35. Partial proglucagon cDNA sequence from spiny dogfish (*Squalus acanthias*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italacized and bold.

```
          Part of GRPP
      T  P  V  Q  D  S  D  S  S  R  A  L  E  M  A  E  K  E  S  V
1     gactccagtgcaggactctgattcctctagggcattggagatggcagagaaagaatctgt  60
                                            Glucagon
      S  N  P  S  T  V  T  D  V  K  R  H  T  D  G  I  F  S  S  D
61    gtctaaccccagcacagtgacagatgtcaaacgccacaccgacgggatcttcagcagcga  120

      Y  S  K  Y  L  D  N  F  R  T  K  D  F  V  Q  W  L  L  S  T
121   ctacagtaaatacttggacaaccggcgcacgaaagactttgtgcaatggctgttgagcac  180
                                                GLP-1a
      K  R  N  G  A  N  T  D  K  T  K  R  H  A  D  G  I  Y  T  S
181   caaacgtaatggtgccaacactgacaaaaccaagagacacgccgacggtatttacactag  240

      D  V  A  S  L  T  D  Y  L  K  S  K  R  F  V  E  S  L  S  N
241   tgatgtggcatctctcaccgactacctcaaatcaaagcgatttgtggaatcactcagcaa  300
                                                N-terminal extension
      Y  N  K  R  Q  N  D  R  R  M  S  K  R  H  V  E  G  A  R  R
301   ctacaacaagcgtcaaaatgacaggaggatgtcaaagagacacgtagagggagccaggcg  360
      GLP-1b
      T  E  E  D  Y  P  N  D  F  S  S  Y  L  E  A  K  A  A  K  D
361   cacagaggaagattatcccaatgatttcagctcctacctggaggccaaagccgccaaaga  420

      F  I  D  W  L  I  K  G  Q  G  R  R  D  F  A  E  E  S  R  E
421   cttcatcgactggttgataaaggacaaggaaggagagactttgccgaggagagcaggga  480

      I  E  N  E  V  I  A  E  E  L  G  R  R
481   gatcgaaaatgaagtgatcgctgaggagctgggcagaagg   520
```

Figure 36. Partial proglucagon cDNA sequence from spotted ratfish (*Hydrolagus colliei*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italacized and bold.

The proglucagon sequences determined for the three teleosts confirmed the presence (cf. Chapters 1 and 2) of two distinct genes with one gene containing the signal peptide, GRPP, glucagon, a single GLP-1 lacking an N-terminal extension, and GLP-2. The second gene contains all the same portions of the precursor, with unique sequences, except it lacks sequence for GLP-2.

The cane toad and the cottonmouth snake showed similar structures to previously reported amphibians and reptiles, respectively.

```
     C-terminal end of Glucagon                          N-terminal
     D  F  A  Q  W  L  M  N  S  K  R  S  G  G  M  S  R  R  N  V
1    agactttgcccagtggttaatgaactcaaaaagaagcggaggaatgtcaagacgtaatgt  60
     extension    GLP-1a
     Q  F  E  R  H  A  E  G  T  Y  T  N  D  V  T  Q  F  L  E  E
61   tcagtttgaaagacatgctgaaggaacatataccaatgatgtaacccaattttggaaga  120

     K  A  A  K  E  F  I  D  W  L  L  K  G  I  P  K  K  Q  R  L
121  aaaggcagccaaagaatttattgattggttgttaaaaggaataccaaagaaacagagact  180
           GLP-1b
     S  R  H  A  E  G  T  F  T  S  D  M  T  S  F  L  E  E  K  A
181  ctcacggcacgcagagggaactttcactagtgatatgaccagcttttggaagaaaaagc  240

     A  K  E  F  V  D  W  L  I  K  G  R  P  K  R  N  F  S  D  V
241  agccaaagaatttgttgattggttaattaaaggaagaccaaagagaaatttctcagatgt  300
                                    GLP-2
     S  A  A  D  D  M  D  R  R  H  A  D  G  S  F  T  S  D  F  N
301  ttccgctgcagatgacatggacagaagacacgctgatggcagttttaccagtgacttcaa  360

     K  A  L  D  I  K  A  A  Q  E  F  L  D  W  I  I  N  T  P  V
361  caaagccctcgatatcaaggctgcccaagaattttggactggattatcaacaccccagt  420
                                    3'UTR
     K  E  R  D  L  L  E  E  Q  *
421  taaggaaagggatttgctggaagaacaataagaaaactgaagatctctaccaaggaacca  480
481  cccggctgtcacttgaaaatgatgactttgtgttgtgtaaattcaagcaggtgtatcata  540
541  aagccatacagtatagctttgcatgcaaagtaattgaaattcacgatagtgttttcttgt  600
601  cggggaatgacacttagagcaaattctactaacatacacatcaaccataaaaggtttcat  660
661  gctttcggtaatggctggactagtattgttcttagcacatatgctttggctgtccttcca  720
721  gaactacgagggctttaaatccagaagtgccattttctcacccaattatttagagaaaaa  780
781  aaaaaaaaaa  790
```

Figure 37. Partial proglucagon cDNA sequence from cane toad (*Bufo marinus*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italacized and bold.

```
                  IP-1
       T  K  R  S  G  Q  Q  G  L  E  D  R  E  K  E  N  F  L  D  Q
1      actaaaagaagcggacaacaaggacttgaagacagagagaaagaaaatttcctggaccag  60
                                       N-term extension  GLP-1
       L  A  S  N  G  L  A  R  P  H  A  E  Y  E  R  H  A  D  G  T
61     cttgcaagtaatggacttgcccggcctcatgccgaatacgaaagacatgctgacggcacc  120

       Y  T  S  D  I  S  S  Y  L  E  G  Q  A  A  K  E  F  I  A  W
121    tataccagtgacatcagctcttacttggaaggtcaagctgctaaggagttcattgcttgg  180

       L  V  N  G  R  G  R  R  D  F  S  E  G  A  H  T  G  E  D  L
181    ttagtgaatgggcgaggaagaagagatttctcggaaggagcccatacaggtgaagatctt  240
                  GLP-2
       G  R  R  H  A  D  G  T  F  T  S  D  Y  N  K  L  L  D  D  I
241    ggtcgaagacatgcagatggcactttcaccagtgattacaacaaactcctggatgacata  300

       A  T  Q  E  F  L  K  W  L  I  N  Q  K  V  T  Q  R  D  L  Q
301    gctacccaggaattcttgaagtggttgattaaccaaaaagttacccagagggaccttcag  360
                      3'UTR
       G  E  Y  Q  *
361    ggagagtaccaataagatgcagtaaaaacatggcatcaaggtcttctctatatccactct  420
421    cagccactggaatttattttgaattttcatatttttttctgtattctaaaaccacaga  480
481    tcttgacatgcaatgaaatgaatttcactatttgatcctactatggagggattttaaaat  540
541    gagtgtctccttgtcagctggaaataactaaaatcacaatattaccagtcaaaaggttca  600
601    gttttcagatctttctgaaagatctttaaggaagttctcaaaatgggacccaagcaaatg  660
661    tattgacaagaccattgctttattttattttctaaaaagataatatgtatatagaaatga  720
721    tgcacccaaccaatatgaaccaatatcaacagcaagactggcagtttaaattgcggggaa  780
781    atgtccttcttgaaacctttgttataaaaaggctccattctccagaccatcagtgaatag  840
841    cctttaataaatgaatgaataaataaaacaaattatactgctaaaaaaaaaa  892
```

Figure 38. Partial proglucagon cDNA sequence from cottonmouth snake (*Agkistrodon piscivorus*). Putative peptides are underlined, dibasic residues are bold, and putative polyA signals are italicized and bold.

Fugu database search.

Three proglucagon like sequences were identified after searching the pufferfish (*T. rubripes*) genomic DNA database with the BlastN algorithm. Scaffolds 504, 38, and 6491 contained proglucagon sequences I, II, and III respectively. Only scaffold 504 was annotated as glucagon.

Scaffold 504 contains 1.35 kb of proglucagon which includes the signal peptide, GRPP, Glucagon, GLP-1 and GLP-2 with 4 introns interspersed in the same pattern as seen in my copper rockfish proglucagon I.

Pufferfish proglucagon II was found on scaffold 38, containing signal peptide, GRPP, glucagon and GLP-1, but not GLP-2, with similar gene layout to copper rockfish proglucagon II. Using copper rockfish GLP-2 as the query in a BlastN and tBlastX search, Scaffold 38 was found to be devoid of GLP-2 sequence. This confirms my findings in copper rockfish proglucagon II and suggests that all teleosts may have one proglucagon gene that does not encode GLP-2.

Analysis of scaffold 6491 identified a glucagon sequence that was not followed by either GLP-1 or GLP-2. BlastN and tBlastX searches with GLP-1 and GLP-2 as the query failed to identify either peptide in the genomic DNA sequence of scaffold 6491. Due to a stretch of poor sequence in this region of the genome, the lack of conservation seen in GRPP, and the fact that GRPP is interrupted by an intron, it was not possible to identify GRPP in scaffold 6491. It is possible that this glucagon sequence in the pufferfish is the result of a single exon duplication and relocation, leading to a portion of the genome containing only glucagon and none of the framework necessary for proper translation. Also, the sequence of this glucagon is quite varied compared to other teleost glucagons, to a point where it may be a psuedogene. The only evidence suggesting that this glucagon may actually be functional is that the putative peptide sequence maintains all of the necessary residues for a functional glucagon and some invariant residues, such as a His^{1}, Gly^{4}, Phe^{6}, Asp^{9}, Typ^{25} and Leu^{26}. There is currently no evidence to suggest that this sequence is transcribed or translated, but for completeness and comparison it has been included in the phylogenetic analysis of glucagon (cf. Figure 42).

Signal Peptide
M R S S L A G I L L L L G F I Q N S W Q
1 atgagaagctccctggctggtatcctcctgcttctcggctttattcagaacagctggcag 60
GRPP
V P L L E T G D S S S L A A D D T L Q D
61 gttcccctgctcgagaccggcgacagctcgagtctggctgctgatgacacattacaggac 120
Glucagon I
D S R E L S N M K R H S E G T F S N D Y
121 gattccagggagctgtcgaacatgaagaggcactcggaggggactttctccaacgactac 180

S K Y L E D R K A Q D F V R W L M N N K
181 agcaaatatctggaggacaggaaggcgcaggactttgtccgatggctgatgaacaacaag 240
GLP-1 I
R S G T A E K R H A D G T F T S D V S S
241 aggagcggcacagcagaaaagcgtcacgctgacgggaccttcaccagtgacgtgagctcc 300

Y L K D Q A I K D F V A R L K A G Q V R
301 tacctcaaggaccaggcaatcaaagactttgtggccaggctcaaagctggacaagttcga 360
GLP-2
R E S E T N R R V E A F N R R H V D G T
361 agagaatctgagacaaacaggagagtggaagcattcaacaggaggcatgtagatggaacc 420

F T S D V N K V L D S M A A K E Y L L W
421 ttcaccagcgacgtgaacaaggtgctggactccatggcggccaaagaatatttactgtgg 480

V M A S K P S G E R *
481 gtgatggcctccaaaccttcgggggagaggtaa 513

Figure 39. Proglucagon I cDNA sequence for pufferfish, *Takifugu rubripes*, determined from the genomic DNA sequence database.

Signal Peptide
M N S A Y S L A G L L L F I T I Q S S W
1 atgaatagtgcttactctttggctggactcctgctcttcatcaccatccaaagcagctgg 60
GRPP
Q F P E Q D R N S I R L L N E N S V L T
61 cagtttcccgaacaggacagaaactccatcaggctactgaatgaaaactcggtgctgact 120
Glucagon II
E P S E L S N M **K** **R** H S E G T F S N D Y
121 gaacccagtgagctctcaaacatgaagagacattcagaggggaccttttccaacgactac 180

S K Y L E T R R A Q D F V Q W L K N S **K**
181 agtaaatacctcgagacgaggagggcgcaggactttgtccagtggctgaagaactcaaag 240
GLP-1 II
R N G S L F **R** **R** H A D G T Y T S D V S T
241 agaaatgggagcctgtttagacggcacgctgacggcacctacaccagcgacgtcagcacc 300

Y L Q D Q A A K E F V S W L K T G P G **R**
301 tacctgcaggaccaggcagccaaggagtttgtgtcctggctgaagaccggcccaggcagg 360

R E *
361 agagagtga 369

Figure 40. Proglucagon II cDNA sequence for pufferfish, *Takifugu rubripes*, determined from genomic DNA sequence.

```
                                          Glucagon III
     V  G  K  G  G  N  T  F  H  V  L  K  R  H  S  E  G  T  F  S
1    gtgggaaaaggaggcaataccttccacgtcctcaaaagacattctgaggggacctttagc  60

     H  D  F  S  R  Y  L  D  K  I  K  T  K  A  F  V  E  W  L  A
61   catgattttagccgctatctggataaaatcaagactaaagcttttgtggagtggctggcc  120

     S  T  K  E  R  *
121  agcactaaggagaggtga  138
```

Figure 41. Glucagon III cDNA sequence for pufferfish, *Takifugu rubripes*, determined from genomic DNA sequence.

Phylogenetic Analysis

Nucleic acid sequence was used for construction of the phylogenetic trees. Because the proglucagon derived peptides are relatively short and quite highly constrained by function, nucleotide, as opposed to amino acid sequence, provides more information of evolutionary relationships.

Because exon duplication of GLP-1, leading to multiple consecutive GLP-1 sequences, has occurred in some vertebrates included in this analysis, comparison of the complete coding region becomes difficult due to drastic differences in sequence length. Also, as the biological functionality of the precursor is restricted to distinct portions of the precursor, there are advantages to analyzing proglucagon in segments. For these two reasons sequences from each peptide are analyzed individually.

Glucagon phylogenetic tree

Phylogenetic analysis of glucagon was performed on the 87 base pairs corresponding to the 29 amino acid peptide. All the sequences analyzed were of this length to generate the phylogenetic tree for glucagon found in Figure 42. As is seen by the low bootstrap numbers this peptide is highly conserved, indicating that the sequences are so similar that the branching pattern of the phylogenetic tree is not always the same. Nevertheless, branches leading to the teleosts and amphibian sequences are clearly separate in this tree.

Somewhat unexpected, the two lungfish sequences branch off a node located between the amphibians and mammals, far from the teleost sequences. The lengths of the branches in these trees are indicative of the genetic distance between the compared sequences. Thus the long branch of the extant lungfish species signifies many nucleotide changes compared to glucagon sequences from other species. In contrast, the branch lengths observed for *R. pipiens* and *X. laevis* are shorter than that of the lungfish, indicating higher sequence similarity with species on neighboring branches.

Other sequences that show long branch lengths are one of two lamprey (*P. marinus*) glucagon sequences and ratfish (*H. colliei*) glucagon, both of which are quite ancient fish. The final sequence demonstrating immense distance with long branch lengths is the third pufferfish (*T. rubripes*) glucagon sequence, which as mentioned before, could possibly be a pseudogene. Yet, even if it is a functional glucagon sequence, the presence of two other presumably functional pufferfish glucagons has decreased the functional constraint for the third sequence and allowed the nucleotide sequence to undergo numerous nucleotide changes.

With respect to the teleosts in the glucagon tree, some of the species, such as anglerfish, rockfish, and pufferfish, fall into two clear groups containing the glucagon sequences from each proglucagon gene, yet other teleosts don't fall into these same groups. The picture is obscured by the fact that two genes have not yet been identified for each teleost species, making it difficult to identify which gene is present, especially by analyzing glucagon sequences alone. For this reason, further analysis of the two teleost genes is reserved for the tree generated from longer pieces of proglucagon sequence.

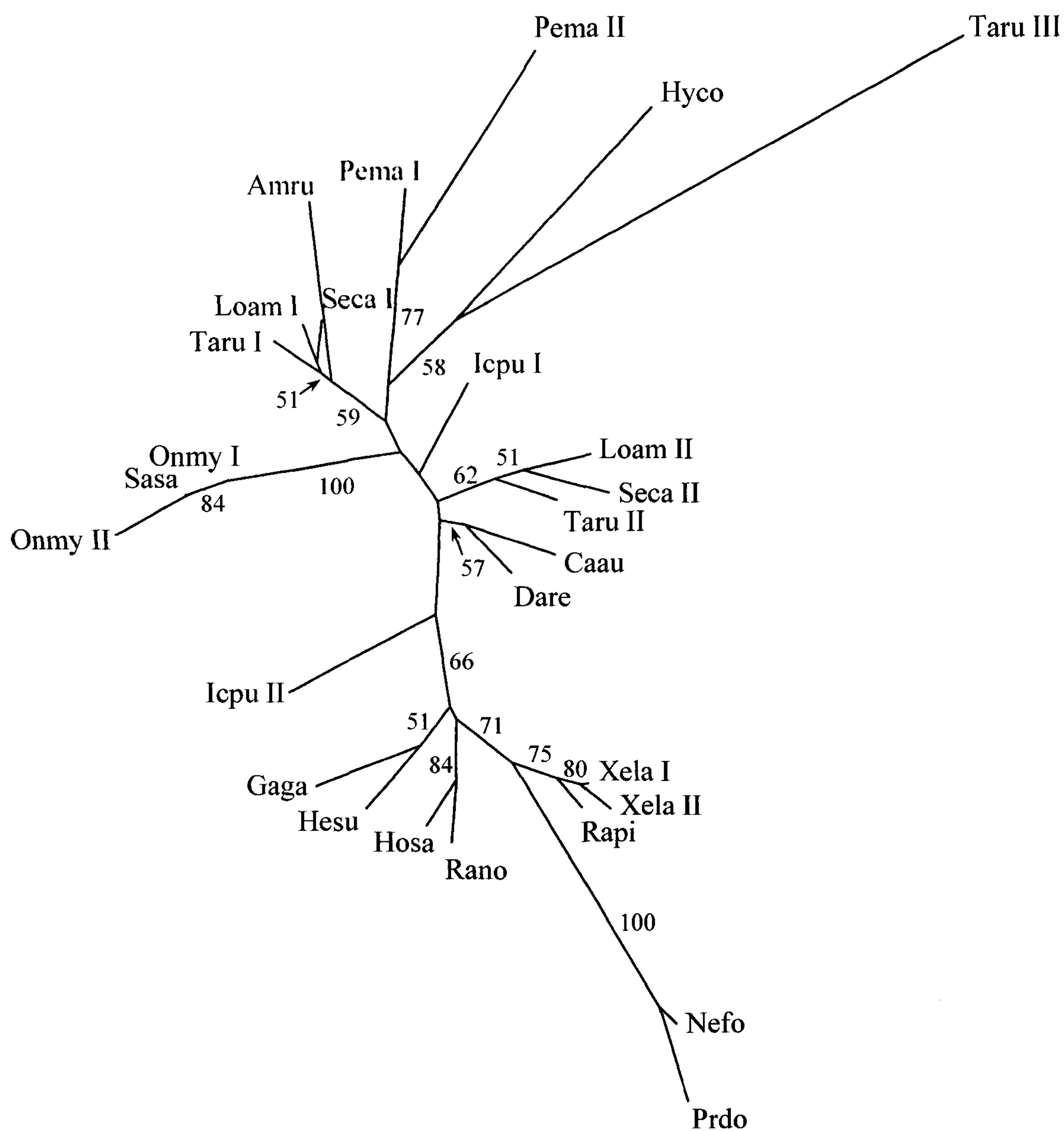

Figure 42. Phylogenetic tree for glucagon. Drawn using the neighbor joining method with branch lengths representing genetic distance. Branches with bootstrap values exceeding 50% are indicated. Sequences are identified by the first two letters of the genus and the first two letters of the species.

Abbreviations used to indicate the species for each sequence in the phylogenetic trees are comprised of the first two letters from genus and the first two letters from the species:

Agpi cottonmouth snake, *Agkistrodon piscivorus*
Amru rock bass *Ambloplites rupestris*
Buma cane toad, *Bufo marinus*
Caau goldfish, *Carassius auratus*
Dare zebrafish, *Danio rerio*
Gaga chicken, *Gallus gallus*
Gimi California goby, *Gillichthys mirabilis*
Hesu gila monster, *Heloderma suspectum*
Hosa human, *Homo sapiens*
Hyco spotted ratfish, *Hydrolagus colliei*
Icpu channel catfish, *Ictalurus punctatus*
Loam anglerfish, *Lophius americanus*
Nefo Australian lungfish, *Neoceraodus forsteri*
Onmy rainbow trout, *Oncorhynchus mykiss*
Onts chinook salmon, *Oncorhynchus tshawytscha*
Pema sea lamprey, *Petromyzon marinus*
Prdo African lungfish, *Protopterus dolloi*
Rano rat, *Rattus norvegicus*
Rapi leopard frog, *Rana pipiens*
Sasa Atlantic salmon, *Salmo salar*
Seca copper rockfish, *Sebastes caurinus*
Sqac spiny dogfish, *Squalus acanthias*
Taru pufferfish, *Takifugu rubripes*
Xela claw-toed frog, *Xenopus laevis*

GLP-1 phylogenetic tree

The N-terminal extension was not used in the analysis of the GLP-1 sequences. Primarily, the sequence length of 93 nucleotides, corresponding to the 31 amino acids of the biologically functional peptide, was used to generate the phylogenetic tree. Three exceptions exist. The dogfish GLP-1b sequence has a three base deletion in the corresponding amino acid position seven, resulting in a 30 residue GLP-1b peptide. Also, both ratfish and dogfish GLP-1a sequences are only 90 base pairs long, as determined by a dibasic cut site immediately following the corresponding amino acid 30. Thus a three base pair gap was added to the end of these sequences for the alignment. When gaps exist in the alignment, this merely means that those positions do not contribute to determining the genetic distance of that sequence compared to others. Therefore, gaps are not detrimental to the analysis, but do allow for a better alignment of sequence and thus a more accurate distance matrix is calculated.

The phylogenetic tree for GLP-1 displays three distinct clades, one including the teleost sequences, one including the lungfish, reptile, bird and mammal sequences, and a final one including the amphibian GLP-1 sequences (Figure 43). The elasmobranch branches root at nodes between these clades, but a great distance from each other, with the GLP-1b sequences showing high similarity to the amphibian GLP-1s. The elasmobranch GLP-1a and lamprey sequences possess long branch lengths indicating large genetic distance between sequences.

Analysis of the amphibian sequences in the tree can be used to pinpoint the order of duplication events. Higher similarity is seen between *R. pipiens* and *B. marinus* GLP-1b and *X. laevis* GLP-1c and similarly between *R. pipiens* and *B. marinus* GLP-1a and *X. laevis* GLP-1b. This suggests that the exon duplication resulting in *X. laevis* GLP-1a must have been the most recent duplication. The glycogenolytic potency of *X. laevis* GLP-1s on teleost hepatocytes indicates a similar trend. *X. laevis* GLP-1c is consistently more potent at stimulating the release of glucose compared to *X. laevis* GLP-1b, while *X. laevis* GLP-1a shows the smallest activation. Thus the oldest GLP-1 sequence, GLP-1c, maintains the highest functionality of GLP-1, glycogenolysis in teleost liver (Mommsen, T.P. et al. 2001).

Two proglucagon genes exist for *X. laevis*, which is not surprising as this animal is generally considered to be tetraploid. The GLP-1s from each gene are more similar to each other than to either the *R. pipiens* or *B. marinus* sequences. I conclude this duplication must be a fairly recent event and most likely due to the tetraploidization event, thus it is not a proglucagon specific gene duplication.

Interestingly, the same trends are identified by phylogenetic analysis and from analysis of proglucagon structure. For example, the possession of an N-terminal extension on the lungfish GLP-1, suggested similarity to the mammalian, reptile and bird sequences, and the phylogenetic trees identified little genetic distance among these groups. Also, the existence of multiple GLP-1 sequences in the elasmobranch proglucagon sequences

suggested similarity to the amphibian structures, while the GLP-1 phylogenetic tree grouped the elasmobranch GLP1b sequences close to the amphibian GLP-1 clade. This relationship is not seen in either phylogenetic tree for glucagon or GLP-2, which may indicate a similarity in function of GLP-1 that is not seen with glucagon and GLP-2.

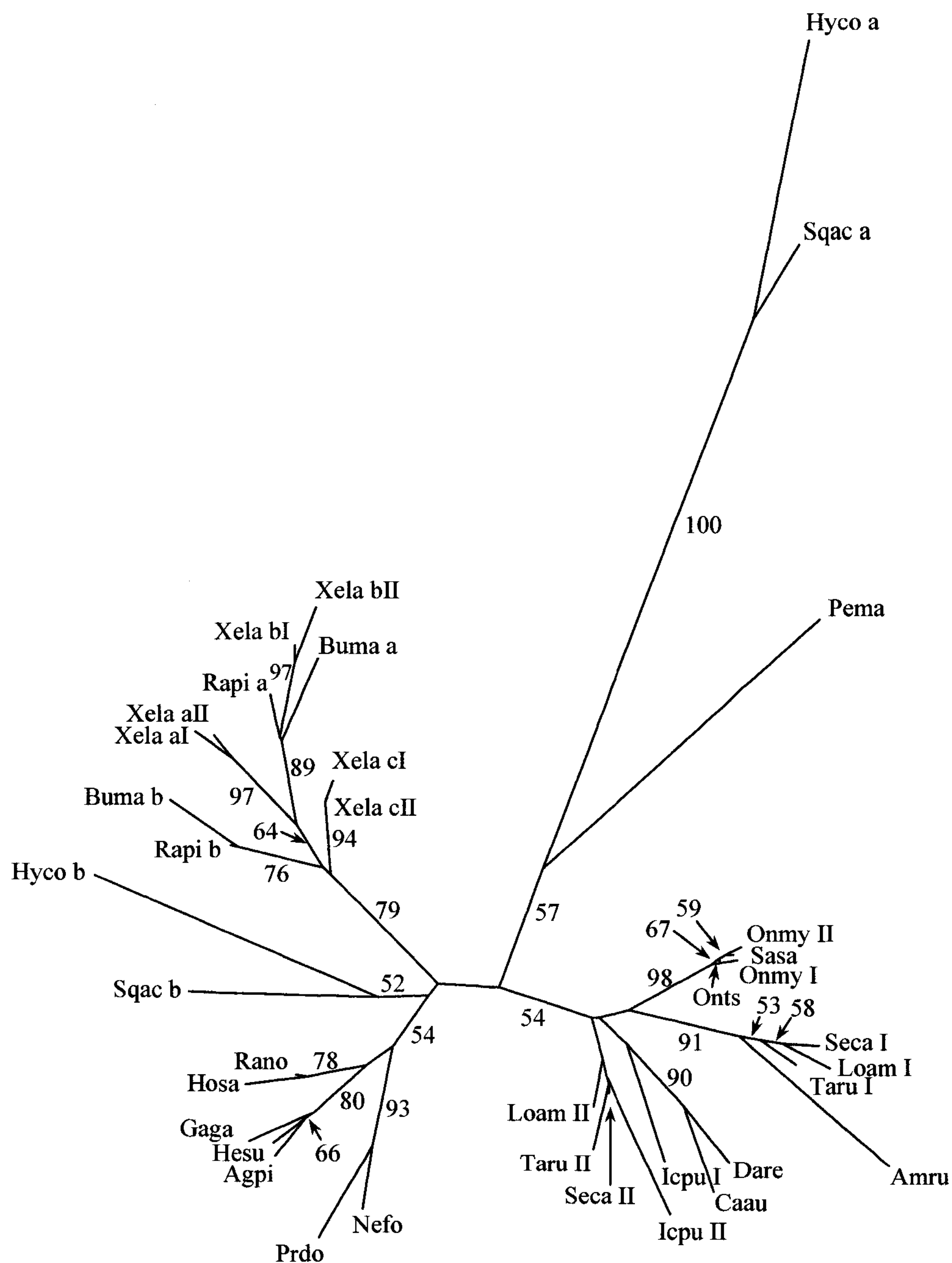

Figure 43. Phylogenetic tree for GLP-1. drawn using the neighbor joining method. Branches with bootstrap values exceeding 50% are indicated. Sequences are identified by the first two letters of the genus and the first two letters of the species.

GLP-2 phylogenetic tree

The length of GLP-2, as defined by the dibasic cut site or stop codon following it in the precursor sequence, is variable among the vertebrates. Mammals, reptiles, and chicken have 99 base pairs, or 33 amino acids, while lungfish and *X. laevis* have 96 base pairs. Some teleosts, pufferfish (*T. rubripes*) and rainbow trout (*O. mykiss*), are 96 base pairs while GLP-2 in rockfish and catfish possess 105 base pairs. One lamprey GLP-2 sequence has a 15 base pair deletion in the position of amino acids 4-8 most likely creating a non-functional peptide (Irwin, D.M. et al. 1999), thus it has not been included in this analysis. The other lamprey GLP-2 sequence is only 32 amino acids, 96 base pairs, with a deletion in position 27, a non-essential position. For this analysis, the two longer teleosts were trimmed to 99 base pairs, and end gaps were added to the five sequences that were only 96 base pairs, thus the standard size of GLP-2 for this analysis was 99 base pairs.

GLP-2 is not as highly conserved as either glucagon or GLP-1, as indicated by the high bootstrap values in the phylogenetic tree (Figure 44). The increased variation provides a more distinct and thus consistent branching pattern.

Still, similar trends are discernible in the GLP-2 tree as in the other trees. Again, the sequences from the two lungfish species are found within the same clade as the mammals and sharing an ancestral node. The teleost sequences, along with the lamprey GLP-1 form a clade distinct from the other vertebrate sequences. The position of the lamprey sequence within the clade is on the branch farthest from the tetrapod clade, similar to that seen in the glucagon phylogenetic tree. Unfortunately, as fewer teleost GLP-2 sequences have been identified, partially because it is only encoded on one of the proglucagon genes and is not expressed in all tissues due to alternative splicing (cf. Chapter 1) (Irwin, D.M. et al. 1995), a thorough analysis of teleostean sequences is difficult.

An interesting feature of this tree is the exceptional length of the branch leading to the mammals, with an assigned bootstrap value of 100%, indicating very strong confidence in the branch. The length of this branch indicates a larger genetic distance between

mammals and other vertebrates, compared to that seen in the glucagon and GLP-1 trees, suggesting more change in the mammalian GLP-2 sequence relative to the other taxa. This idea has been presented before, with the suggestion of an increased rate of evolution in the GLP-2 sequences between the ancestral and extant mammals (Irwin, D.M. 2001). It is interesting to see that a similar result is found using nucleotides sequences and a different method of phylogenetic analysis (Dr. Irwin used parsimony to generate phylogenetic trees).

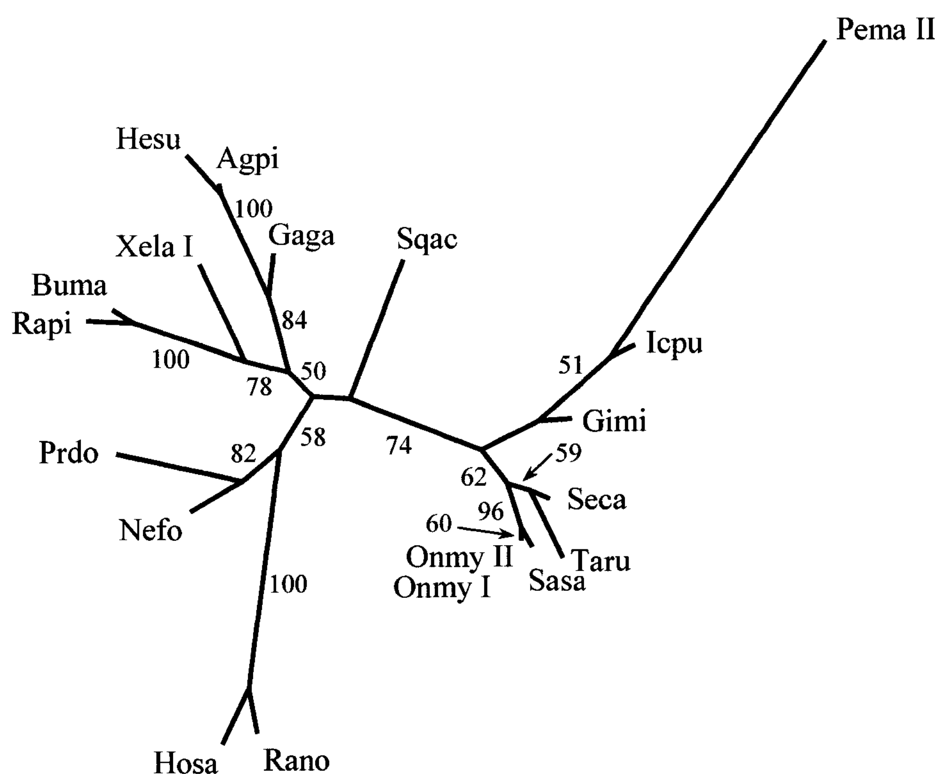

Figure 44. Phylogenetic tree for GLP-2. Drawn using the neighbor joining method. Branches with bootstrap values exceeding 50% are indicated. Sequences are identified by the first two letters of the genus and the first two letters of the species.

Teleost phylogenetic tree

Considering the number of sequences and diversity seen within the teleost group, a more in depth analysis is necessary for these species. I did this by using a larger portion of the precursor sequence shared by all the species, i.e. the nucleotide sequence from the beginning of glucagon to the end of GLP-1. The phylogenetic analysis combines the trends seen individually in glucagon and GLP-1, increasing the resolution of the trees.

For five species included in the analysis, two proglucagon gene sequences exist, while for four other species there is only one, and at first glance, no phylogenetic trend appears between the two sets of sequences. However, the proglucagon II sequences from four species, *S. caurinus*, *L. americanus*, *T. rubripes*, and *I. punctatus* form a clade, while the proglucagon I sequence from almost the same group of species, *S. caurinus*, *L. americanus*, *T. rubripes*, and *A. rupestris* form a second clade. The complement of these two clades suggest the proglucagon I sequences display similarity to each other and likewise for the proglucagon II sequences. Perhaps the clades represent each duplicate proglucagon gene seen in teleosts. From the tree alone, it is not clear to which group the remaining sequences belong, or why they do not fall within the clades.

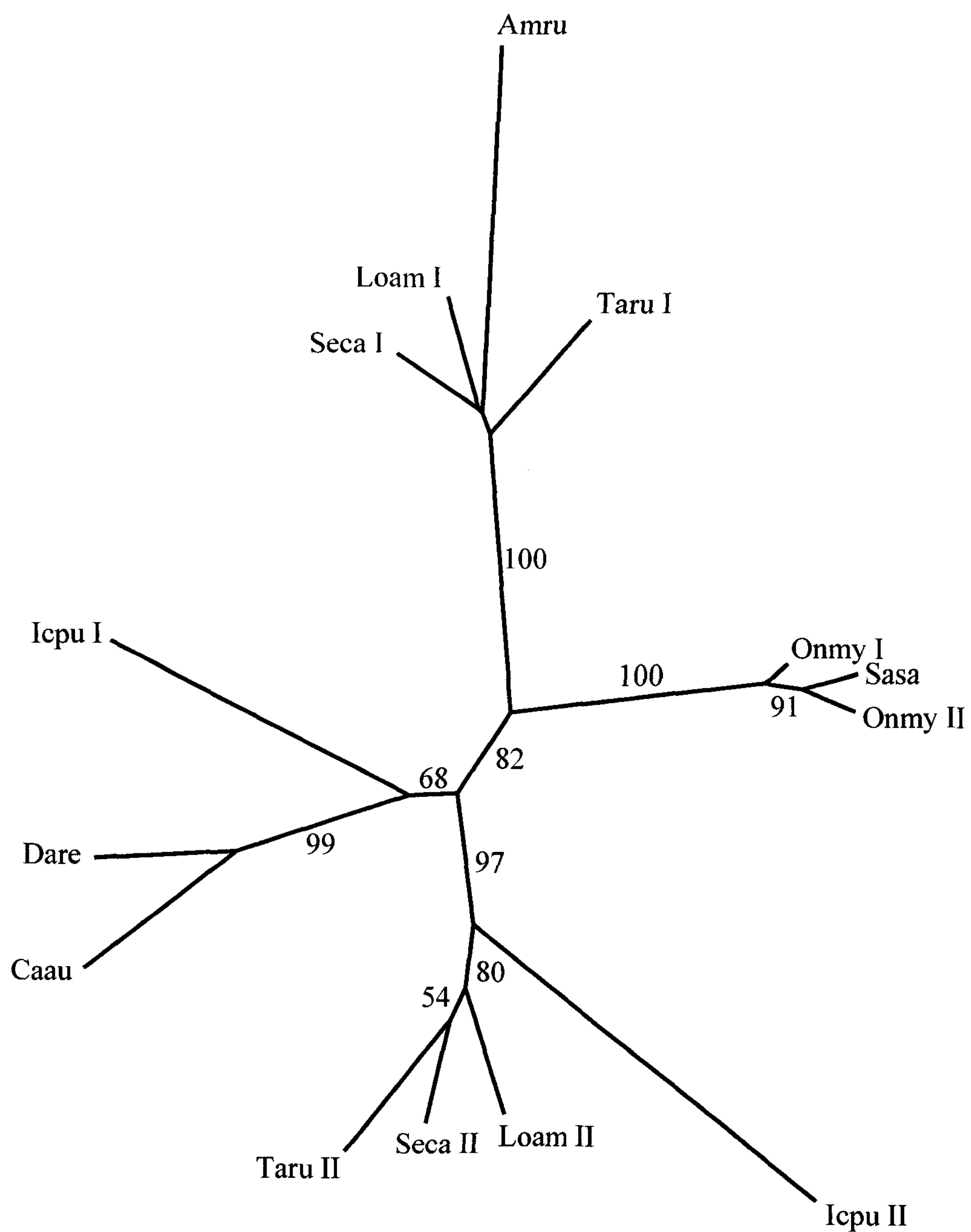

Figure 45. A phylogenetic tree for teleosts from glucagon through to the end of GLP-1. Drawn using the neighbor joining method. Branches with bootstrap values higher thaqn 50% are indicated. Sequences are identified by the first two letters of the genus and the first two letters of the species.

A Phylogenetic tree including all three peptides

Each phylogenetic tree presented was generated independently; therefore the scale of each tree is different. To give an idea of how the genetic distances compare from peptide to peptide, a single tree was generated that includes most of the nucleotide sequences for all three peptide trees. For simplicity of the alignment, only the first 87 base pairs of each sequence were used in the analysis.

Obviously, most of the sequences from each peptide fall into three distinct clades (Figure 46). Also, the branch lengths and thus the genetic distances for GLP-2 are larger than those of glucagon or GLP-1, again indicating the smaller degree of conservation seen in this peptide. Four sequences seem to fall close to the borders of each peptide group. Three of these sequences, the *T. rubripes* glucagon III and the two elasmobrach GLP-1a sequences, come from animals that have had peptide duplication, allowing increased freedom of mutation. The final sequence, the lamprey *P. marinus* GLP-1, comes from a fish whose ancestors had some 400 million years to evolve from an ancestral GLP-1.

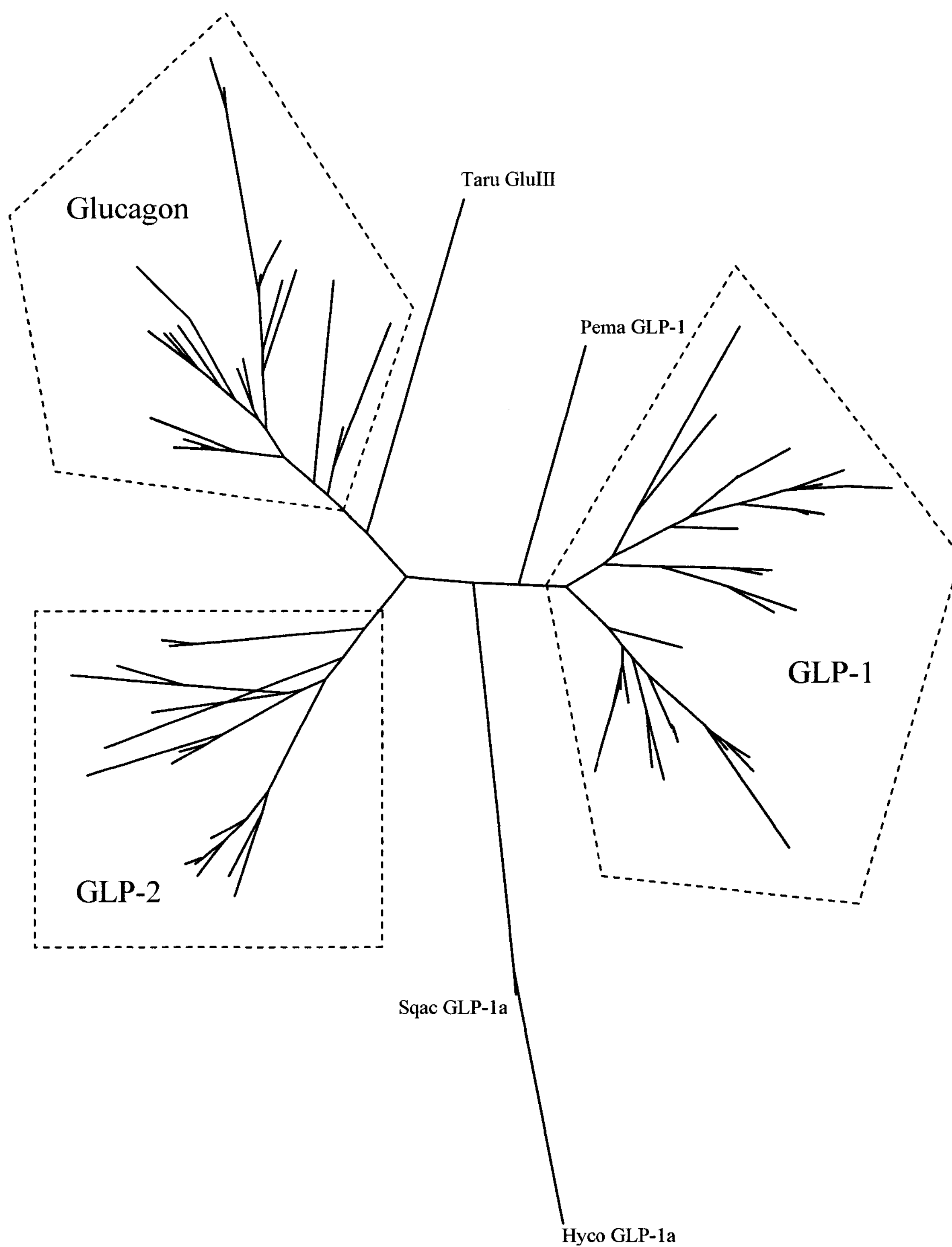

Figure 46. A phylogenetic tree for glucagon, GLP-1 and GLP-2. Drawn using the neighbor joining method.

Discussion

Lungfish

Lungfish are considered the most highly evolved fish that supposedly bridge the gap between fish and tetrapods (Zardoya, R. et al. 1998). Thus, it was expected the lungfish would show some similarity to the tetrapods, including amphibians, and some similarity to the teleosts. Surprisingly, the lungfish display more similarity to the reptilian and avian tetrapods than to either the amphibians or the teleosts. This is seen in all three peptides with lungfish branching off an ancestral line that leads to reptiles, birds and mammals, and clearly not on an ancestral line leading to amphibians. Observing the proglucagon structure alone, a similar conclusion is determined as the lungfish are most similar to reptiles, birds, and mammals, by containing only one GLP-1 that is N-terminally extended and encoded on a single proglucagon gene.

Also, it has been previously suggested that African lungfish are more derived than Australian lungfish (Randal, D.J. et al. 1981). Analysis of proglucagon from these two species indicates that more nucleotide substitutions have occurred in African lungfish, as the branches in the trees for all three peptides are longer for African lungfish indicating larger distances. However, this does not indicate that this species or the genus *Protopterus* is more similar to tetrapods, more likely that it has undergone more change since the tetrapods and lungfish last shared an ancestor.

Elasmobranchs

In phylogenetic analysis, the elasmobranch peptide relationships to other vertebrates are not consistent. In the glucagon tree, the elasmobranch sequence shows similarity with the lamprey sequence, which roots in amongst the teleost sequences. In the GLP-1 tree, strong similarity with the amphibians is seen for elasmobranch GLP-1b. In GLP-2, the elasmobranch sequence falls between the tetrapod/lungfish clade and the teleost/lamprey clade. However, in all three cases the bootstrap values for the branches surrounding the elasmobranch branch are low, indicating that the topology of these surrounding branches is not strongly supported by the data. In phylogenetic analysis, if a sequence possesses

some but not significant levels of homology with the other aligned sequences, it becomes more difficult to make solid statements about evolutionary relationships. In the case of the elasmobranchs, a relatively ancient fish, the genetic distance between it and other vertebrate sequences is not as low as seen with lamprey (*P. marinus*), thus it is not always positioned as an outside taxon in the tree, as is lamprey. Nor does it consistently cluster with any one vertebrate group because it is old enough to have incurred significant genetic change. Also, considering we are studying an extant species which has gone through extensive genetic variation since it last shared a common ancestor with other vertebrates in this analysis, many nucleotide substitutions at a single site may have occurred, which is difficult to predict with current algorithms. This will of course lead to more variable results in the phylogenetic analysis and low bootstrap values.

Teleosts

The teleostean species show a great deal of variation in phenotype and genotype, as these species have undergone considerable diversification since the adaptive radial speciation, 300 to 450 mya (Taylor, J.S. et al. 2001). Thus teleosts, and fish in general, are ideal for studying evolutionary changes in genes.

It is obvious that proglucagon in many teleosts has undergone gene duplication. In fact, for all of the teleost species in which a proglucagon-derived peptide has been identified, either cDNA or peptide, half of those species demonstrate evidence for two genes. Since those species with evidence for two genes are dispersed throughout all teleosts, it could be suggested that all teleosts have undergone proglucagon gene duplication and the second sequence has merely not yet been identified in several species. The question is whether this duplication is specific to the glucagon gene or applies to the whole genome. It has been proposed that an early ancestor of Actinopterygii, before the teleost speciation, underwent a whole genome duplication (Aparicio, S. 2000; Van de, P.Y. et al. 2002), which would explain the observed proglucagon gene duplication. If this is the case, we can predict the teleost proglucagon genes should separate into two groups: the genes within one group should show more similarity across species than with the other

proglucagon gene from the same species, indicating that the gene duplication event occurred before the speciation event.

This is clearly seen in the proglucagon II clade with the sequences from *S. caurinus*, *L. americanus*, *T. rubripes*, and *I. punctatus* (Figure 45). However, it is not as clear with the proglucagon I clade for one reason, *I. punctatus* does not fall within this clade, otherwise consisting of *S. caurinus*, *L. americanus*, *T. rubripes*, and *A. rupestris*. This suggests that perhaps all of the sequences outside the proglucagon II clade are the proglucagon I sequences for each species. This hypothesis is confirmed by analysis of the peptide sequences for glucagon and GLP-1 from these species as well as species for which only peptides sequences have been obtained. Classification of the proglucagon sequences according to the amino acids they possess in positions 16, 18, 24 and 29 of glucagon and positions 6, 15, 19, or 25 identify the peptides as either a proglucagon I or a proglucagon II sequence. Sequences with D or E, K, R, N in glucagon and F, K, I, R or K in GLP-1 identifies a proglucagon I sequence, while sequences with T, R, Q, S in glucagon and Y, Q, A, W in GLP-1 identifies a sequence as proglucagon II. As seen in Figure 47, the ambiguous sequences in the tree fall into the category of proglucagon II sequences. Interestingly, peptide sequences from three Actinopterygii species, North American paddlefish (*Polydon spathula*), bowfin (*Amia calvia*), and alligator gar (*Lepisosteus spatula*), which diverged from the common ancestor just before teleosts, have residues in the above positions that classify them as proglucagon II sequences. Considering only one proglucagon sequence has been reported for these species thus far, I would suggest that proglucagon II is the more ancient form of the teleost sequences.

```
        Glucagon
                          1618     24   29
LoamI   HSEGTFSNDY SKYLEDRKAQ EFVRWLMNN
TaruI   HSEGTFSNDY SKYLEDRKAQ DFVRWLMNN
SecaI   HSEGTFSNDY SRYLEERKAQ DFVRWLMNN
AmRu    HSQGTFTNDY TNYLEDRQAQ DFIRWLMNN
IcpuII  HSEGTFSNDY SKYLETRRAQ DFIQWLMNS
LoamII  HSEGTFSNDY SKYLETRRAQ DFVQWLKNS
TaruII  HSEGTFSNDY SKYLETRRAQ DFVQWLKNS
SecaII  HSEGTFSNDY SKYLETRRAQ DFVQWLKNS
Dare    HSEGTFSNDY SKYLETRRAQ DFVQWLMNA
Caau    HSEGTFSNDY SKYLETRRAQ DFVEWLMNS
IcpuI   HSEGTFSNDY SKYLETRRAQ DFVQWLMNS
OnmyI   HSEGTFSNDY SKYQEERMAQ DFVQWLMNS
OnmyII  QSEGTFSNYY SKYQEERMAR DFLHWLMNS
Sasa    HSEGTFSNHY SKYQEERMAR DFVQWLMNT

        GLP-1
             6          15  19       25
LoamI   HADGTFTSDV SSYLKDQAIK DFVDRLKAGQ V
TaruI   HADGTFTSDV SSYLKDQAIK DFVARLKAGQ V
SecaI   HADGTFTSDV SSYLKDQAIK DFVNRLKSGQ V
AmRu    HADGTFTDDA SSDFYDQAIK DFVAKLKSGQ D
IcpuII  HADGTYTSDV SSYMQDQAAK EFVSWLKMGR G
LoamII  HADGTYTSDV SSYLQDQAAK DFVSWLKAGR G
TaruII  HADGTYTSDV STYLQDQAAK EFVSWLKTGP G
SecaII  HADGTYTSDV SSYLQDQAAK EFVYWLKTGR G
Dare    HAEGTYTSDV SSYLQDQAAQ RFVARLKSGQ P
Caau    HAEGTYTSDI SSFLRDQAAQ NFVAWLKSGQ P
IcpuI   HADGTYTSDV SSYLQDQAAK DFITWLKSGQ P
OnmyI   HADGTYTSDV STYLQDQAAK DFVSWLKSGR A
OnmyII  HADGTYTSDV STYLQDQAAK DFVSWLKSGP A
Sasa    HADGTYTSDV STYLQDQAAK DFVSWLKSGP A
```

Figure 47. Amino acid alignment of teleost glucagon and GLP-1 sequences.

A difference in branch length is seen between the proglucagon II sequences that form a clade, and the other putative proglucagon II sequences, for many of which only one sequence is present. A possible explanation for this is the order in which speciation occurred among teleosts. *D. rerio*, *C. auratus*, *I. punctatus*, *O. mykiss*, and *S. salar* all belong to teleost orders that are considered to have undergone speciation longer ago than *S. caurinus*, *L. americanus*, and *T. rubripes* (Nelson, J.S. 1994), thus these sequences will

have larger distances due to longer evolutionary time. Ideally, the best way to identify which proglucagon group the unidentified sequences belong to is to obtain sequence for the second proglucagon gene for each teleost. Until this is done, a phylogenetic analysis of teleost proglucagons will be incomplete.

As usual the rainbow trout falls into its own category. Originally, two proglucagon sequences had been identified in rainbow trout and had been interpreted as two separate genes (Irwin, D.M. et al. 1995). My analysis clearly shows that the two trout proglucagon genes reveal much more similarity than seen in duplicate proglucagon genes in other species. Considering *O. mykiss* is a salmonid fish species, and an ancestor of salmonid fish are believed to have undergone a tetraploidization event between 25-100 mya in which the whole genome is duplicated, four proglucagon sequences are expected for *O. mykiss*. Also, as the tetraploidization event occurred more recently than the speciation of teleost, duplicates that arise from the tetraploidization would display more similarity than seen in duplicates of other teleosts. Thus, I would postulate that the other two genes for proglucagon have yet to be identified in *O. mykiss*.

Additionally, an interesting anomaly appears in the phylogenetic comparison of another salmonid fish, Atlantic salmon (genus *Salmo*) with the rainbow trout (genus *Oncorhynchus*). For glucagon and GLP-1, the Atlantic salmon (*S. salar*) falls within the pair of rainbow trout (*O.mykiss*) sequences, but not for the GLP-2 tree. If all salmonids have undergone tetraploidization, one would expect the result seen in the glucagon and GLP-1 trees, where the *S. salar* sequence is more similar to one of the *O. mykiss* sequences (proglucagon II) as the speciation event would occur after the duplication event. The GLP-2 tree implies that this is not the case. Luckily, inspection of the GLP-2 sequences provides insight and possible explanations. The two *O. mykiss* sequences differ by only one base pair within GLP-2, while the *S. salar* sequence compared to *O. mykiss* proglucagon II differ by 3 base pairs, which is more typical of the rest of the sequence. This could be the result of convergent evolution, although there is no other evidence for convergent evolution in proglucagon. Or this could be the result of partial

gene conversion within the *O.mykiss*, creating a region of unexpected high sequence similar between the two *O.mykiss* sequences and not the *S. salar* sequence.

Exon duplication

The two consecutive GLP-1s found in elasmobranchs and amphibians suggest relatively recent occurrence of exon duplication. Considering the generally accepted evolutionary scheme of non-mammalian vertebrates (Nelson, J.S. 1994), it is most likely that the exon duplication responsible for duplicate GLP-1s in amphibians and elasmobranchs occurred independently in each lineage.

The presence or absence of the N-terminal extension gives some indication of the mechanism of the exon duplication. In elasmobranchs the GLP-1 containing the N-terminal extension is second in position, whereas in *Rana* and *Bufo* species, the N-terminally extended GLP-1 is first. Considering that this process most likely occurs by unequal crossing over of chromosomes, it is clear that the position of the newly duplicated GLP-1 is random. Also, as the N-terminal extension is not present in every copy of the exon, it is possible that it may be lost during exon duplication. We know it is not always lost upon exon duplication as *X. laevis* proglucagon, which includes three consecutive GLP-1 sequences, sheds more information on the situation. As previously mentioned, comparison of *R. pipiens* and *B. marinus* with *X. laevis* GLP-1s (Figure 43) indicates that *X. laevis* GLP-1a is the result of the most recent exon duplication to occur in *X. laevis*. Yet, the first two *X. laevis* GLP-1s, GLP-1a and GLP-1b, retain the N-terminal extension during exon duplication, suggesting that during the first exon duplication, which most likely occurred in all amphibians, the N-terminal extension is lost, but in the second exon duplication, limited to *Xenopus*, the N-terminal extension was retained. This indicates that the consecutive GLP-1 exon duplications both occurred from the same original exon, otherwise retention of the N-terminal extension in the final duplication of *X. laevis* would have been impossible.

This same mechanism may have occurred in the exon duplication that created GLP-1 and GLP-2 in the ancestral proglucagon sequence. Support for this is seen in Figure 46,

where less overall distance is seen between GLP-2 and glucagon than between GLP-2 and GLP-1, indicating that GLP-2 is most likely a duplication of glucagon and not a duplication of GLP-1.

Duplication leads to diversification of function

GLP-1 is an ideally suited peptide for this type of analysis as some species possess multiple GLP-1 sequences due to gene duplication or exon duplication. Indeed, GLP-1 is truly an excellent model for demonstrating how diversification can occur as a result of duplication. A unique scenario exists in species with multiple GLP-1 sequences, as demonstrated by the amphibian and elasmobrach GLP-1 sequences. When multiple exons exist, the functional constraint is reduced, as at least one peptide is present to fill the existing biological role of the peptide. Therefore, if additional sequences, often expressed concurrently, are subject to evolutionary change, metabolic alterations are not observed, allowing mutations to be fixed. Considering elasmobranchs are ancient fish, this effect is compounded by the long period of evolutionary time GLP-1a has had to diversify. This becomes evident in the phylogenetic analysis including glucagon and GLP-2, as the elasmobranch GLP-1a sequences do not fall within one of the three peptide clades (Figure 46), suggesting sufficient nucleotide mutation has occurred to make it difficult to truly classify them as GLP-1 sequences. It is possible this is how the original duplication of glucagon led to the evolution of GLP-1 and GLP-2, peptides with distinct physiological functions. It is also possible that the elasmobranch GLP-1a sequence no longer functions as a GLP-1 hormone and could have or will develop a new function.

Throughout the vertebrates, the function of glucagon is constant, while the function of GLP-1 reveals a drastic difference between tetrapods and teleosts. In Figure 43, it is obvious that the GLP-1 sequences for teleosts have more similarity with each other than with the tetrapod GLP-1 sequences. Yet these sequences are still recognizable as GLP-1 sequences and do not contain the level of change expected with a full role reversal. This is primarily because the change in function of GLP-1 is associated with the localization of the GLP-1 receptor and the co-evolution of peptides and receptors (Sivarajah, P. et al. 2001). As mentioned earlier, lungfish and elasmobranch GLP-1s show more similarity to

the tetrapods than to the teleosts. Also, these "fish" contain the N-terminal extended version of GLP-1, a feature associated with the insulinotropic role of GLP-1 in mammals. The question arises whether GLP-1 in elasmobranchs and lungfish function like a mammalian GLP-1 as an incretin hormone, or like a teleost GLP-1, as a "better glucagon". If, as the sequences indicate, they function more similar to the tetrapods, this would suggest that teleost glucose metabolism regulation and the role of GLP-1 in glucose metabolism is unique within teleostean fish. At the very least, this subject deserves more investigation.

To date, the role for GLP-2 is only known in mammals, but there is molecular evidence to suggest that the function for this peptide may not have been static in vertebrate evolution (Irwin, D.M. 2001). Unlike glucagon or GLP-1, a surprisingly large genetic distance is seen when comparing mammalian GLP-2 with non-mammalian vertebrates, such as reptiles and birds. This may suggest that the role of GLP-2 has changed in this period of intense genetic change during the evolution of mammals. Thus, it is possible that GLP-2 does not function as an intestinal growth factor in non-mammalian vertebrates, yet the GLP-2 sequence is still conserved in non-mammalian vertebrates implying a biological function is associated with it in these animals. Therefore, it is possible a change in function is observed in GLP-2 as well as GLP-1 between mammals and non-mammalian vertebrates. Of course this cannot be confirmed until a function for GLP-2 is determined in non-mammalian vertebrates.

Gene duplication

In most cases, proglucagon gene duplication has been part of a whole genome duplication of the animal. This is responsible for duplicate proglucagon genes in all teleosts, plus in *O. mykiss* and *X. laevis* for which tetraploidization events have occurred. Two proglucagon genes exist in *P. marinus*, where whole genome duplication has not yet been suggested, thus it is possible independent proglucagon duplication may have occurred. Although, a haploid chromosome count for *P. marinus* is 84, which may indicate full or multiple genome duplications, especially considering the haploid chromosome count for *Sebastes* species is 24, for *Ictalurus* species is 29, for elasmobranch species is between 29

and 31, for *O. mykiss*, for which a relatively recent tetraploidization event is generally accepted, is between 54-60, and for the many species of polyploid sturgeon (genus *Acipenser*), varies between 50 and 130 (www.fishbase.org) (Thorgaard, G.H. 1983). Therefore to date all incidences of proglucagon gene duplication have most likely occurred with a whole genome duplication event.

Whether the presence of two proglucagon genes is a result of whole genome duplication or independent proglucagon duplication, some fluidity of the gene exists. The GLP-2 exon has been lost in one of the two proglucagon genes in at least two teleost species (cf. Chapter 1), and the GLP-1 exon has been lost in lamprey (*P. marinus*), (Irwin, D.M. et al. 1999). Yet, many teleosts have not completely lost the duplicate genes. Possibly these animals have maintained both copies of proglucagon for a regulatory reason. Analysis of sequences from *T. rubripes* may provide some insights as to whether this is true.

The extensive genetic distance that *T. rubripes* glucagon III displays compared to any other glucagon suggests that it is truly superfluous in function as a regulator of glucose metabolism. Comparing this sequence to the two other *T. rubripes* glucagon sequences demonstrates the two other sequences are quite conserved; suggesting *T. rubripes* and most likely other teleosts have developed individual roles for each gene product, and perhaps even acquired differentially regulated roles for each gene product. This is supported by the frequent fixation and conservation of the duplicate sequences after a unique role for each gene is developed, such as the differential expression of proglucagon seen in the teleost, *S. caurinus* (cf. Chapter 1) (Aparicio, S. 2000).

Summary

A schematic cladogram for the evolution of proglucagon genes in vertebrates is in Figure 48, which summarizes in what lines various modifications of the gene occur and how this has appeared in the extant species studied in this thesis. The presence of the N-terminal extension in various different species as well as GLP-1 exon duplication has been instrumental in determining the evolutionary history of proglucagon.

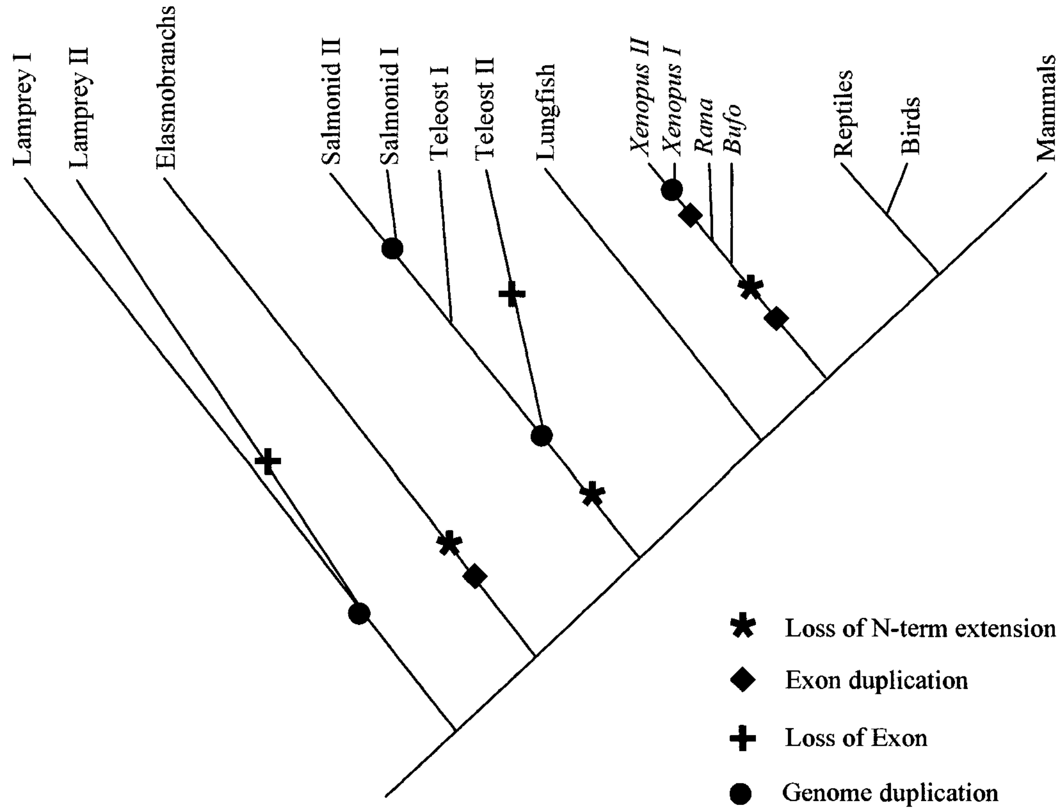

Figure 48. A schematic cladogram for the evolution of proglucagon genes in vertebrates.

Considering that exon duplication, partial or full exon loss, gene duplication, and increased sequence diversification due to duplication are all demonstrated in the proglucagon gene, it could be considered to be quite a genetically fluid gene and at the very least is an excellent model for studying evolutionary processes.

Conclusions and Future Work

The research objectives of this project have included investigation of the expression, regulation and evolution of the proglucagon gene in non-mammalian vertebrates. When I started this research project, four years ago, I had a few hypotheses as to what I would find once investigation began. Once evidence for two rockfish proglucagon sequences developed, and it became clear that each transcript encoded sequence for physiochemically distinct peptides, I expected to find differential tissue expression of these transcripts in both teleosts studied, copper rockfish and channel catfish. Considering several distinct functions have been demonstrated for proglucagon-derived peptides in mammals, which require several levels of regulation as all three peptides are encoded in the same precursor, it is expected that differential tissue distribution of the transcripts might be a necessary aspect of this regulation in teleosts. What surprised me about the tissue distribution expression patterns identified in copper rockfish was the complexity of levels of potential regulation observed at the mRNA level. Proglucagon genes undergo regulation of transcription through 1) differential tissue distribution of the two genes throughout the three primary proglucagon expressing tissues, endocrine pancreas, brain, and gastrointestinal tract 2) expression of truncated sequences in which not all three peptides are expressed 3) expression of two forms of the same sequence, possibly to provide the ability to increase peptide levels by concurrent expression of two forms of the same transcript, alternatively, a mechanism for rapid transition between translation of two forms of the same sequence and 4) in some tissues, a possible indication of different levels of transcript was observed. Also surprising was that not all of these aspects of transcript regulation, such as differential expression, were obvious in a different teleost, channel catfish. This suggests that the degree of proglucagon transcript regulation may vary between species, perhaps depending on the metabolic needs involving proglucagon-derived peptides of the species.

Regulation of proglucagon continues at the peptide level of teleosts, as indications of differential transcript translation or differential peptide degradation were demonstrated in endocrine pancreas of both teleosts studied. Identifying and measuring the levels of

peptides produced from the different proglucagon transcripts is essential to fully understanding regulation of peptide function in teleost. At this point, peptide analysis has only been performed on one of three interesting tissues of synthesis. Further analysis of peptide production in teleost brain and gastrointestinal tract, as well as plasma peptide levels involved in the endocrine characteristic of these peptide hormones, will provide insights into the regulation and functions of proglucagon-derived peptides in teleost.

Preliminary work on the teleostean receptors of proglucagon, that are also involved in regulation of peptide function, provided, along with indications from other physiological work, the suggestion that more than one receptor subtype for each peptide may exist in teleosts. Obviously, as this work is preliminary, further molecular and physiological characterization of these receptors will confirm or refute this hypothesis.

With respect to the evolutionary aspects of proglucagon, several evolutionary processes that contribute to the fluidity of the proglucagon gene were identified, including proglucagon gene duplication, as a part of whole genome duplication, exon duplication, whole or partial exon loss, and increased sequence and implied functional diversity as a result of relaxed functional constraint associated with duplication. Unexpectedly, several of these processes occur within the proglucagon gene multiple times in independent lines of evolution, such as the exon duplication leading to multiple GLP-1 sequences. Another unexpected result of this work was the significant sequence similarity of GLP-1 sequences between species that encode GLP-1 peptides with an N-terminal extension. As a result, a hypothesis was formulated indicating that in all species that possess an N-terminally extended GLP-1 sequence, GLP-1 functions as an insulinotropic peptide hormone. Function of the peptide as an hepatic metabolic hormone, hallmark of the non-N-terminally extended GLP-1 found in teleosts, may represent a unique situation, restricted to the teleosts and reflect an ancient loss of function and peptide processing. Future work measuring the physiological effect of GLP-1 on hepatocytes and β-cells in a species such as spiny dogfish or African lungfish will confirm or disprove this hypothesis.

Although not all of the original objectives of this project, such as peptide analysis from teleostean brain and intestine, were met, research on the successful objectives led to other intriguing and unexpected discoveries, such as the presence of GLP-2 transcript in teleost endocrine pancreas. The research also provided multifacetted insights into the amazing variability of a prohormone that produces allegedly 'simple' straight-chain peptide hormones.

Acknowledgements

This research was monetarily supported by an NSERC research grant to TPM and a PEO doctoral grant to ERB, scientifically supported by colleagues and collaborators, Dr. Svetlana Mojsov, Rockefeller U, NY; Dr. Jeff Silverstein, USDA, WV; Dr. Tom Moon, U of Ottawa; and Dr. Kris von Schalburg, UVic, technically and socially supported by many members of the Mommsen lab, especially GAC, GRS, NKH, DRB, JCE, HLB, and LJB, as well as members of the Department of Biochemistry and Microbiology, including my peers and supervisory committee. Special thanks go to Gord Brown for his phylogenetic expertise and CPU (mental and computer) time and Ian for everlasting support in all aspects of life. Last but not least, with utmost respect and friendship, I thank my supervisor, Dr. Thomas P. Mommsen, for all I have learned under his guidance.

References

Abraham,E.J., Leech,C.A., Lin,J.C., Zulewski,H., and Habener,J.F., 2002. Insulinotropic hormone glucagon-like peptide-1 differentiation of human pancreatic islet-derived progenitor cells into insulin-producing cells. Endocrinology 143, 3152-3161.

Adelhorst,K., Hedegaard,B.B., Knudsen,L.B., and Kirk,O., 1994. Structure-activity studies of glucagon-like peptide-1. J. Biol. Chem. 269, 6275-6278.

Al Mahrouki,A.A. and Youson,J.H., 1998. Immunohistochemical studies of the endocrine cells within the gastro-entero-pancreatic system of Osteoglossomorpha, an ancient teleostean group. Gen. Comp Endocrinol. 110, 125-139.

Allendorf,F.W. and Thorgaard,G.H., 1984. In: Turner,B.J. (ed.), Tetraploidy and the Evolution of Salmonid Fishes, Plenum Press, New York, pp. 1-46.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., and Lipman,D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Andrews,P.C. and Ronner,P., 1985. Isolation and structures of glucagon and glucagon-like peptide from catfish pancreas. J. Biol. Chem. 260, 3910-3914.

Aparicio,S., 2000. Vertebrate evolution: recent perspectives from fish. Trends Genet. 16, 54-56.

Bjerknes,M. and Cheng,H., 2001. Modulation of specific intestinal epithelial progenitors by enteric neurons. Proc. Natl. Acad. Sci. U. S. A 98, 12497-12502.

Brakch,N., Rholam,M., Simonetti,M., and Cohen,P., 2000. Favourable side-chain orientation of cleavage site dibasic residues of prohormone in proteolytic processing by prohormone convertase 1/3. Eur. J. Biochem. 267, 1626-1633.

Brubaker,P.L., Crivici,A., Izzo,A., Ehrlich,P., Tsai,C.H., and Drucker,D.J., 1997. Circulating and tissue forms of the intestinal growth factor, glucagon-like peptide-2. Endocrinology 138, 4837-4843.

Burrin,D.G., Petersen,Y., Stoll,B., and Sangild,P., 2001. Glucagon-like peptide 2: a nutrient-responsive gut growth factor. J. Nutr. 131, 709-712.

Busby,E.R., Cooper,G.A., and Mommsen,T.P., 2002. Novel role for prostaglandin E2 in fish hepatocytes: regulation of glucose metabolism. J. Endocrinol. 174, 137-146.

Buteau,J., Roduit,R., Susini,S., and Prentki,M., 1999. Glucagon-like peptide-1 promotes DNA synthesis, activates phosphatidylinositol 3-kinase and increases transcription factor pancreatic and duodenal homeobox gene 1 (PDX-1) DNA binding activity in beta (INS-1)-cells. Diabetologia 42, 856-864.

Cavanaugh,E.S., Nielsen,P.F., and Conlon,J.M., 1996. Isolation and structural characterization of proglucagon-derived peptides, pancreatic polypeptide, and somatostatin from the urodele Amphiuma tridactylum. Gen. Comp Endocrinol. 101, 12-20.

Chen,Y.E. and Drucker,D.J., 1997. Tissue-specific expression of unique mRNAs that encode proglucagon-derived peptides or exendin 4 in the lizard. J. Biol. Chem. 272, 4108-4115.

Collie,N.L., Walsh,J.H., Wong,H.C., Shively,J.E., Davis,M.T., Lee,T.D., and Reeve,J.R., Jr., 1994. Purification and sequence of rat oxyntomodulin. Proc. Natl. Acad. Sci. U. S. A 91, 9362-9366.

Conlon,J.M., Fan,H., and Fritzsch,B., 1998. Purification and structural characterization of insulin and glucagon from the bichir Polypterus senegalis (Actinopterygii: Polypteriformes). Gen. Comp Endocrinol. 109, 86-93.

Conlon,J.M., Samson,W.K., Dobbs,R.E., Orci,L., and Unger,R.H., 1979. Glucagon-like polypeptides in canine brain. Diabetes 28, 700-702.

Conlon,J.M., Youson,J.H., and Mommsen,T.P., 1993. Structure and biological activity of glucagon and glucagon-like peptide from a primitive bony fish, the bowfin (Amia calva). Biochem. J. 295 (Pt 3), 857-861.

Curtis,K.S., Sved,A.F., Verbalis,J.G., and Stricker,E.M., 1994. Lithium chloride-induced anorexia, but not conditioned taste aversions, in rats with area postrema lesions. Brain Res. 663, 30-37.

Cutfield,S.M. and Cutfield,J.F., 1993. A second glucagon in the pancreatic islets of the daddy sculpin Cottus scorpius. Gen. Comp Endocrinol. 91, 281-286.

DaCambra,M.P., Yusta,B., Sumner-Smith,M., Crivici,A., Drucker,D.J., and Brubaker,P.L., 2000. Structural determinants for activity of glucagon-like peptide-2. Biochemistry 39, 8888-8894.

Dakin,C.L., Gunn,I., Small,C.J., Edwards,C.M., Hay,D.L., Smith,D.M., Ghatei,M.A., and Bloom,S.R., 2001. Oxyntomodulin inhibits food intake in the rat. Endocrinology 142, 4244-4250.

Dalle,S., Fontes,G., Lajoix,A.D., LeBrigand,L., Gross,R., Ribes,G., Dufour,M., Barry,L., LeNguyen,D., and Bataille,D., 2002. Miniglucagon (glucagon 19-29): a novel regulator of the pancreatic islet physiology. Diabetes 51, 406-412.

Danulat,E. and Mommsen,T.P., 1990. Norepinephrine: a potent activator of glycogenolysis and gluconeogenesis in rockfish hepatocytes. Gen. Comp Endocrinol. 78, 12-22.

Deacon,C.F., Pridal,L., Klarskov,L., Olesen,M., and Holst,J.J., 1996. Glucagon-like peptide 1 undergoes differential tissue-specific metabolism in the anesthetized pig. Am. J. Physiol 271, E458-E464.

Dhanvantari,S., Seidah,N.G., and Brubaker,P.L., 1996. Role of prohormone convertases in the tissue-specific processing of proglucagon. Mol. Endocrinol. 10, 342-355.

Drucker,D.J., 2001. Minireview: the glucagon-like peptides. Endocrinology 142, 521-527.

Drucker,D.J., 2002. Gut adaptation and the glucagon-like peptides. Gut 50, 428-435.

Drucker,D.J. and Asa,S., 1988. Glucagon gene expression in vertebrate brain. J. Biol. Chem. 263, 13475-13478.

Drucker,D.J., Ehrlich,P., Asa,S.L., and Brubaker,P.L., 1996. Induction of intestinal epithelial proliferation by glucagon- like peptide 2. Proc. Natl. Acad. Sci. (USA) 93, 7911-7916.

Drucker,D.J., Jin,T., Asa,S.L., Young,T.A., and Brubaker,P.L., 1994. Activation of proglucagon gene transcription by protein kinase-A in a novel mouse enteroendocrine cell line. Mol. Endocrinol. 8, 1646-1655.

Dunphy,J.L., Taylor,R.G., and Fuller,P.J., 1998. Tissue distribution of rat glucagon receptor and GLP-1 receptor gene expression. Mol. Cell. Endocrinol. 141, 179-186.

Edwards,C.M., Todd,J.F., Mahmoudi,M., Wang,Z., Wang,R.M., Ghatei,M.A., and Bloom,S.R., 1999. Glucagon-like peptide 1 has a physiological role in the control of postprandial glucose in humans: studies with the antagonist exendin 9-39. Diabetes 48, 86-93.

Eissele,R., Goke,R., Willemer,S., Harthus,H.P., Vermeer,H., Arnold,R., and Goke,B., 1992. Glucagon-like peptide-1 cells in the gastrointestinal tract and pancreas of rat, pig and man. Eur. J. Clin. Invest 22, 283-291.

Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.5c. 1993. Distributed by the author.

Flamez,D., Gilon,P., Moens,K., Van Breusegem,A., Delmeire,D., Scrocchi,L.A., Henquin,J.C., Drucker,D.J., and Schuit,F., 1999. Altered cAMP and Ca2+ signaling in mouse pancreatic islets with glucagon-like peptide-1 receptor null phenotype. Diabetes 48, 1979-1986.

Flint,A., Raben,A., Astrup,A., and Holst,J.J., 1998. Glucagon-like peptide 1 promotes satiety and suppresses energy intake in humans. J. Clin. Invest 101, 515-520.

Flint,A., Raben,A., Ersboll,A.K., Holst,J.J., and Astrup,A., 2001. The effect of physiological levels of glucagon-like peptide-1 on appetite, gastric emptying, energy and substrate metabolism in obesity. Int. J. Obes. Relat Metab Disord. 25, 781-792.

Fricker,L.D., 1988. Carboxypeptidase E. Annu. Rev. Physiol 50, 309-321.

Friis-Hansen,L., Lacourse,K.A., Samuelson,L.C., and Holst,J.J., 2001. Attenuated processing of proglucagon and glucagon-like peptide-1 in carboxypeptidase E-deficient mice. J. Endocrinol. 169, 595-602.

Furuse,M., Matsumoto,M., Okumura,J., Sugahara,K., and Hasegawa,S., 1997. Intracerebroventricular injection of mammalian and chicken glucagon-like peptide-1 inhibits food intake of the neonatal chick. Brain Res. 755, 167-169.

Gallwitz,B., Witt,M., Paetzold,G., Morys-Wortmann,C., Zimmermann,B., Eckart,K., Folsch,U.R., and Schmidt,W.E., 1994. Structure/activity characterization of glucagon-like peptide-1. Eur. J. Biochem. 225, 1151-1156.

Gangnon,F., Danger,J.M., Jegou,S., Vieau,D., Seidah,N.G., and Vaudry,H., 1999. Molecular cloning, characterization of cDNA, and distribution of mRNA encoding the frog prohormone convertase PC1. J. Comp Neurol. 405, 160-172.

Geary,N., Le Sauter,J., and Noh,U., 1993. Glucagon acts in the liver to control spontaneous meal size in rats. Am. J. Physiol 264, R116-R122.

Gehlert,D.R., 1998. Multiple receptors for the pancreatic polypeptide (PP-fold) family: physiological implications. Proc. Soc. Exp. Biol. Med. 218, 7-22.

Ghatei,M.A., Goodlad,R.A., Taheri,S., Mandir,N., Brynes,A.E., Jordinson,M., and Bloom,S.R., 2001. Proglucagon-derived peptides in intestinal epithelial proliferation: glucagon-like peptide-2 is a major mediator of intestinal epithelial proliferation in rats. Dig. Dis. Sci. 46, 1255-1263.

Goke,R., Larsen,P.J., Mikkelsen,J.D., and Sheikh,S.P., 1995. Distribution of GLP-1 binding sites in the rat brain: evidence that exendin-4 is a ligand of brain GLP-1 binding sites. Eur. J. Neurosci. 7, 2294-2300.

Gomez-visus,I., Garcia-Hernandez,M.P., Lozano,M.T., and Agulleiro,B., 1998. Glucagon- and NPY-related peptide-immunoreactive cells in the gut of sea bass (Dicentrarchus labrax L.): a light and electron microscopic study. Gen. Comp Endocrinol. 112, 26-37.

Graur,D. and Li,W.H., 2000. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, Mass.

Graziano,M.P., Hey,P.J., Borkowski,D., Chicchi,G.G., and Strader,C.D., 1993. Cloning and functional expression of a human glucagon-like peptide-1 receptor. Biochem. Biophys. Res. Commun. 196, 141-146.

Groff,K.E. and Youson,J.H., 1997. An immunohistochemical study of the endocrine cells within the pancreas, intestine, and stomach of the gar (Lepisosteus osseus L.). Gen. Comp Endocrinol. 106, 1-16.

Gros,L., Thorens,B., Bataille,D., and Kervran,A., 1993. Glucagon-like peptide-1-(7-36) amide, oxyntomodulin, and glucagon interact with a common receptor in a somatostatin-secreting cell line. Endocrinology 133, 631-638.

Gutzwiller,J.P., Drewe,J., Goke,B., Schmidt,H., Rohrer,B., Lareida,J., and Beglinger,C., 1999a. Glucagon-like peptide-1 promotes satiety and reduces food intake in patients with diabetes mellitus type 2. Am. J. Physiol 276, R1541-R1544.

Gutzwiller,J.P., Goke,B., Drewe,J., Hildebrand,P., Ketterer,S., Handschin,D., Winterhalder,R., Conen,D., and Beglinger,C., 1999b. Glucagon-like peptide-1: a potent regulator of food intake in humans. Gut 44, 81-86.

Han,V.K., Hynes,M.A., Jin,C., Towle,A.C., Lauder,J.M., and Lund,P.K., 1986. Cellular localization of proglucagon/glucagon-like peptide I messenger RNAs in rat brain. J. Neurosci. Res. 16, 97-107.

Hartmann,B., Harr,M.B., Jeppesen,P.B., Wojdemann,M., Deacon,C.F., Mortensen,P.B., and Holst,J.J., 2000. In vivo and in vitro degradation of glucagon-like peptide-2 in humans. J. Clin. Endocrinol. Metab 85, 2884-2888.

Hasegawa,S., Terazono,K., Nata,K., Takada,T., Yamamoto,H., and Okamoto,H., 1990. Nucleotide sequence determination of chicken glucagon precursor cDNA. Chicken preproglucagon does not contain glucagon-like peptide II. FEBS Lett. 264, 117-120.

Havel,P.J., 2001. Peripheral signals conveying metabolic information to the brain: short-term and long-term regulation of food intake and energy homeostasis. Exp. Biol. Med. (Maywood.) 226, 963-977.

Hjorth,S.A., Adelhorst,K., Pedersen,B.B., Kirk,O., and Schwartz,T.W., 1994. Glucagon and glucagon-like peptide 1: selective receptor recognition via distinct peptide epitopes. J. Biol. Chem. 269, 30121-30124.

Holst,J.J., 1997. Enteroglucagon. Annu. Rev. Physiol 59, 257-271.

Holst,J.J., Bersani,M., Johnsen,A.H., Kofod,H., Hartmann,B., and Orskov,C., 1994. Proglucagon processing in porcine and human pancreas. J. Biol. Chem. 269, 18827-18833.

Holst,J.J., Orskov,C., Nielsen,O.V., and Schwartz,T.W., 1987. Truncated glucagon-like peptide I, an insulin-releasing hormone from the distal gut. FEBS Lett. 211, 169-174.

Hoosein,N.M. and Gurd,R.S., 1984a. Human glucagon-like peptides 1 and 2 activate rat brain adenylate cyclase. FEBS Lett. 178, 83-86.

Hoosein,N.M. and Gurd,R.S., 1984b. Identification of glucagon receptors in rat brain. Proc. Natl. Acad. Sci. U. S. A 81, 4368-4372.

Hoyt,E.C., Lund,P.K., Winesett,D.E., Fuller,C.R., Ghatei,M.A., Bloom,S.R., and Ulshen,M.H., 1996. Effects of fasting, refeeding, and intraluminal triglyceride on proglucagon expression in jejunum and ileum. Diabetes 45, 434-439.

Imeryuz,N., Yegen,B.C., Bozkurt,A., Coskun,T., Villanueva-Penacarrillo,M.L., and Ulusoy,N.B., 1997. Glucagon-like peptide-1 inhibits gastric emptying via vagal afferent-mediated central mechanisms. Am. J. Physiol 273, G920-G927.

Irwin,D.M., 2001. Molecular evolution of proglucagon. Regul. Pept. 98, 1-12.

Irwin,D.M., Huner,O., and Youson,J.H., 1999. Lamprey proglucagon and the origin of glucagon-like peptides. Mol. Biol. Evol. 16, 1548-1557.

Irwin,D.M., Satkunarajah,M., Wen,Y., Brubaker,P.L., Pederson,R.A., and Wheeler,M.B., 1997. The *Xenopus* proglucagon gene encodes novel GLP-1-like peptides with insulinotropic properties. Proc. Natl. Acad. Sci. (USA) 94, 7915-7920.

Irwin,D.M. and Sivarajah,P., 2000. Proglucagon cDNAs from the leopard frog, *Rana pipiens*, encode two GLP-1-like peptides. Mol. Cell. Endocrinol. 162, 17-24.

Irwin,D.M. and Wong,J., 1995. Trout and chicken proglucagon: alternative splicing generates mRNA transcripts encoding glucagon-like peptide 2. Mol. Endocrinol. 9, 267-277.

Jarrousse,C., Bataille,D., and Jeanrenaud,B., 1984. A pure enteroglucagon, oxyntomodulin (glucagon 37), stimulates insulin release in perfused rat pancreas. Endocrinology 115, 102-105.

Jelinek,L.J., Lok,S., Rosenberg,G.B., Smith,R.A., Grant,F.J., Biggs,S., Bensch,P.A., Kuijper,J.L., Sheppard,P.O., Sprecher,C.A., and ., 1993. Expression cloning and signaling properties of the rat glucagon receptor. Science 259, 1614-1616.

Jin,S.L., Han,V.K., Simmons,J.G., Towle,A.C., Lauder,J.M., and Lund,P.K., 1988. Distribution of glucagonlike peptide I (GLP-I), glucagon, and glicentin in the rat brain: an immunocytochemical study. J. Comp Neurol. 271, 519-532.

Kastin,A.J., Akerstrom,V., and Pan,W., 2002. Interactions of glucagon-like peptide-1 (GLP-1) with the blood-brain barrier. J. Mol. Neurosci. 18, 7-14.

Kervran,A., Blache,P., and Bataille,D., 1987. Distribution of oxyntomodulin and glucagon in the gastrointestinal tract and the plasma of the rat. Endocrinology 121, 704-713.

Kieffer,T.J., McIntosh,C.H., and Pederson,R.A., 1995. Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 in vitro and in vivo by dipeptidyl peptidase IV. Endocrinology 136, 3585-3596.

Knepel,W., Chafitz,J., and Habener,J.F., 1990. Transcriptional activation of the rat glucagon gene by the cyclic AMP-responsive element in pancreatic islet cells. Mol. Cell Biol. 10, 6799-6804.

Ku,S.K., Lee,H.S., Lee,J.H., and Park,K.D., 2001. An immunohistochemical study on the endocrine cells in the alimentary tract of the red-eared slider (Trachemys scripta elegans). Anat. Histol. Embryol. 30, 33-39.

Larsen,P.J., Tang-Christensen,M., Holst,J.J., and Orskov,C., 1997a. Distribution of glucagon-like peptide-1 and other preproglucagon-derived peptides in the rat hypothalamus and brainstem. Neuroscience 77, 257-270.

Larsen,P.J., Tang-Christensen,M., and Jessop,D.S., 1997b. Central administration of glucagon-like peptide-1 activates hypothalamic neuroendocrine neurons in the rat. Endocrinology 138, 4445-4455.

Layer,P., Holst,J.J., Grandt,D., and Goebell,H., 1995. Ileal release of glucagon-like peptide-1 (GLP-1). Association with inhibition of gastric acid secretion in humans. Dig. Dis. Sci. 40, 1074-1082.

Ling,Z., Wu,D., Zambre,Y., Flamez,D., Drucker,D.J., Pipeleers,D.G., and Schuit,F.C., 2001. Glucagon-like peptide 1 receptor signaling influences topography of islet cells in mice. Virchows Arch. 438, 382-387.

Lovshin,J., Estall,J., Yusta,B., Brown,T.J., and Drucker,D.J., 2001. Glucagon-like peptide (GLP)-2 action in the murine central nervous system is enhanced by elimination of GLP-1 receptor signaling. J. Biol. Chem. 276, 21489-21499.

Lui,E.Y., Asa,S.L., Drucker,D.J., Lee,Y.C., and Brubaker,P.L., 1990. Glucagon and related peptides in fetal rat hypothalamus in vivo and in vitro. Endocrinology 126, 110-117.

Lund,P.K., Goodman,R.H., Dee,P.C., and Habener,J.F., 1982. Pancreatic preproglucagon cDNA contains two glucagon-related coding sequences arranged in tandem. Proc. Natl. Acad. Sci. U. S. A 79, 345-349.

Lund,P.K., Goodman,R.H., Montminy,M.R., Dee,P.C., and Habener,J.F., 1983. Anglerfish islet pre-proglucagon II. Nucleotide and corresponding amino acid sequence of the cDNA. J. Biol. Chem. 258, 3280-3284.

Lund, P.K., Hoyt, E., Simmons, J.G., and Ulshen, M.H. Regulation of intestinal glucagon gene expression during adaptive growth of small intestine. Digestion 54, 371-373. 1993.
Ref Type: Abstract

Mackin,R.B., Noe,B.D., and Spiess,J., 1991. Identification of a somatostatin-14-generating propeptide converting enzyme as a member of the kex2/furin/PC family. Endocrinology 129, 2263-2265.

MacNeil,D.J., Occi,J.L., Hey,P.J., Strader,C.D., and Graziano,M.P., 1994. Cloning and expression of a human glucagon receptor. Biochem. Biophys. Res. Commun. 198, 328-334.

Mallat,A., Pavoine,C., Dufour,M., Lotersztajn,S., Bataille,D., and Pecker,F., 1987. A glucagon fragment is responsible for the inhibition of the liver Ca2+ pump by glucagon. Nature 325, 620-622.

McGregor,G.P., Goke,R., and Goke,B., 1998. Biological actions of glucagon-like peptide(GLP)-2 revealed--how pluripotential is the glucagon gene? Exp. Clin. Endocrinol. Diabetes 106, 25-28.

McMahon,L.R. and Wellman,P.J., 1998. PVN infusion of GLP-1-(7-36) amide suppresses feeding but does not induce aversion or alter locomotion in rats. Am. J. Physiol 274, R23-R29.

Mills,J.C. and Gordon,J.I., 2001. The intestinal stem cell niche: there grows the neighborhood. Proc. Natl. Acad. Sci. U. S. A 98, 12334-12336.

Moens,K., Heimberg,H., Flamez,D., Huypens,P., Quartier,E., Ling,Z., Pipeleers,D., Gremlich,S., Thorens,B., and Schuit,F., 1996. Expression and functional activity of glucagon, glucagon-like peptide I, and glucose-dependent insulinotropic peptide receptors in rat pancreatic islet cells. Diabetes 45, 257-261.

Mojsov,S., 2000. Glucagon-like peptide-1 (GLP-1) and the control of glucose metabolism in mammals and teleost fish. Amer. Zool. 28, 246-258.

Mommsen,T.P., 2000. Glucagon-like peptides in fishes: the liver and beyond. Amer. Zool. 40, 259-268.

Mommsen,T.P., Busby,E.R., von Schalburg,K.R., Evans,J.C., Osachoff,H.L., and Elliot,M.E., 2003a. Glutamine synthetase in tilapia gastrointestinal tract: zonation, cDNA and induction by cortisol. J. Comp. Physiol. (in press).

Mommsen,T.P., Conlon,J.M., and Irwin,D.M., 2001. Amphibian glucagon family peptides: potent metabolic regulators in fish hepatocytes. Regul. Pept. 99, 111-118.

Mommsen,T.P. and Mojsov,S., 1998. Glucagon-like peptide-1 activates the adenylyl cyclase system in rockfish enterocytes and brain membranes. Comp Biochem. Physiol B Biochem. Mol. Biol. 121, 49-56.

Mommsen,T.P., Moon,T.W., and Walsh,P.J., 1994. Hepatocytes: isolation, maintenance and utilization. In: Hochachka,P.W. and Mommsen,T.P. (eds.), Biochemistry and Molecular Biology of Fishes, Elsevier Biomedical, Amsterdam, New York, pp. 355-373.

Mommsen,T.P., Osachoff,H.L., and Elliot,M.E., 2003b. Metabolic zonation in gastrointestinal tract: effects of fasting and cortisol. J. Comp. Physiol. (in press).

Montrose-Rafizadeh,C., Yang,H., Rodgers,B.D., Beday,A., Pritchette,L.A., and Eng,J., 1997a. High potency antagonists of the pancreatic glucagon-like peptide-1 receptor. J. Biol. Chem. 272, 21201-21206.

Montrose-Rafizadeh,C., Yang,H., Wang,Y., Roth,J., Montrose,M.H., and Adams,L.G., 1997b. Novel signal transduction and peptide specificity of glucagon-like peptide receptor in 3T3-L1 adipocytes. J. Cell Physiol 172, 275-283.

Moon,T.W., Busby,E.R., Cooper,G.A., and Mommsen,T.P., 1999. Fish hepatocyte glycogen phosphorylase - a sensitive indicator of hormonal activation. Fish Physiol. Biochem. 21, 15-24.

Moon,T.W., Gambarotta,A., Capuzzo,A., and Fabbri,E., 1997. Glucagon and glucagon-like peptide signaling pathways in the liver of two fish species, the American eel and the black bullhead. J. Exp. Zool. 279, 62-70.

Munroe,D.G., Gupta,A.K., Kooshesh,F., Vyas,T.B., Rizkalla,G., Wang,H., Demchyshyn,L., Yang,Z.J., Kamboj,R.K., Chen,H., McCallum,K., Sumner-Smith,M., Drucker,D.J., and Crivici,A., 1999. Prototypic G protein-coupled receptor for the intestinotrophic factor glucagon-like peptide 2. Proc. Natl. Acad. Sci. U. S. A 96, 1569-1573.

Murayama,Y., Kawai,K., Suzuki,S., Ohashi,S., and Yamashita,K., 1990. Glucagon-like peptide-1(7-37) does not stimulate either hepatic glycogenolysis or ketogenesis. Endocrinol. Jpn. 37, 293-297.

Naslund,E., Barkeling,B., King,N., Gutniak,M., Blundell,J.E., Holst,J.J., Rossner,S., and Hellstrom,P.M., 1999. Energy intake and appetite are suppressed by glucagon-like peptide-1 (GLP-1) in obese men. Int. J. Obes. Relat Metab Disord. 23, 304-311.

Navarro,I., Leibush,B., Moon,T.W., Plisetskaya,E.M., Banos,N., Mendez,E., Planas,J.V., and Gutierrez,J., 1999. Insulin, insulin-like growth factor-I (IGF-I) and glucagon: the evolution of their receptors. Comp Biochem. Physiol B Biochem. Mol. Biol. 122, 137-153.

Nelson,J.S., 1994. Fishes of the World. J. Wiley, New York.

Ngan,E.S., Chow,L.S., Tse,D.L., Du,X., Wei,Y., Mojsov,S., and Chow,B.K., 1999. Functional studies of a glucagon receptor isolated from frog Rana tigrina rugulosa: implications on the molecular evolution of glucagon receptors in vertebrates. FEBS Lett. 457, 499-504.

Nguyen,T.M., Mommsen,T.P., Mims,S.M., and Conlon,J.M., 1994. Characterization of insulins and proglucagon-derived peptides from a phylogenetically ancient fish, the paddlefish (Polyodon spathula). Biochem. J. 300 (Pt 2), 339-345.

Nian,M., Gu,J., Irwin,D.M., and Drucker,D.J., 2002. Human glucagon gene promoter sequences regulating tissue-specific versus nutrient-regulated gene expression. Am. J. Physiol Regul. Integr. Comp Physiol 282, R173-R183.

Orskov,C., Holst,J.J., Knuhtsen,S., Baldissera,F.G., Poulsen,S.S., and Nielsen,O.V., 1986. Glucagon-like peptides GLP-1 and GLP-2, predicted products of the glucagon gene, are secreted separately from pig small intestine but not pancreas. Endocrinology 119, 1467-1475.

Orskov,C., Holst,J.J., Poulsen,S.S., and Kirkegaard,P., 1987. Pancreatic and intestinal processing of proglucagon in man. Diabetologia 30, 874-881.

Orskov,C., Poulsen,S.S., Moller,M., and Holst,J.J., 1996. Glucagon-like peptide I receptors in the subfornical organ and the area postrema are accessible to circulating glucagon-like peptide I. Diabetes 45, 832-835.

Orskov,C., Rabenhoj,L., Wettergren,A., Kofod,H., and Holst,J.J., 1994. Tissue and plasma concentrations of amidated and glycine-extended glucagon-like peptide I in humans. Diabetes 43, 535-539.

Page,R.D.M. and Holmes E.C., 1998. Molecular Evolution. A Phylogenetic Approach, Blackwell Science Ltd., Oxford, London, pp. 179-201.

Parker,D.B., Coe,I.R., Dixon,G.H., and Sherwood,N.M., 1993. Two salmon neuropeptides encoded by one brain cDNA are structurally related to members of the glucagon superfamily. Eur. J. Biochem. 215, 439-448.

Parker,J.C., Andrews,K.M., Rescek,D.M., Massefski,W., Jr., Andrews,G.C., Contillo,L.G., Stevenson,R.W., Singleton,D.H., and Suleske,R.T., 1998. Structure-function analysis of a series of glucagon-like peptide-1 analogs. J. Pept. Res. 52, 398-409.

Patzelt,C. and Schiltz,E., 1984. Conversion of proglucagon in pancreatic alpha cells: the major endproducts are glucagon and a single peptide, the major proglucagon fragment, that contains two glucagon-like sequences. Proc. Natl. Acad. Sci. U. S. A 81, 5007-5011.

Persson,B. and Argos,P., 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol. 237, 182-192.

Philippe,J., 1991. Insulin regulation of the glucagon gene is mediated by an insulin-responsive DNA element. Proc. Natl. Acad. Sci. U. S. A 88, 7224-7227.

Plisetskaya,E.M. and Mommsen,T.P., 1996. Glucagon and glucagon-like peptides in fishes. Int. Rev. Cytol. 168, 187-257.

Plisetskaya,E.M., Pollock,H.G., Rouse,J.B., Hamilton,J.W., Kimmel,J.R., and Gorbman,A., 1986. Isolation and structures of coho salmon (*Oncorhynchus kisutch*) glucagon and glucagon-like peptide. Regulat. Pept. 14, 57-67.

Randal,D.J., Burggren,W.W., Farrell,A.P., and Haswell,M.S., 1981. The Evolution of Air Breathing in Vertebrates. Cambridge University Press, New York.

Read,N., French,S., and Cunningham,K., 1994. The role of the gut in regulating food intake in man. Nutr. Rev. 52, 1-10.

Rocca,A.S. and Brubaker,P.L., 1999. Role of the vagus nerve in mediating proximal nutrient-induced glucagon-like peptide-1 secretion. Endocrinology 140, 1687-1694.

Rouille,Y., Bianchi,M., Irminger,J.C., and Halban,P.A., 1997a. Role of the prohormone convertase PC2 in the processing of proglucagon to glucagon. FEBS Lett. 413, 119-123.

Rouille,Y., Kantengwa,S., Irminger,J.C., and Halban,P.A., 1997b. Role of the prohormone convertase PC3 in the processing of proglucagon to glucagon-like peptide 1. J. Biol. Chem. 272, 32810-32816.

Sambrook,J., Fritsch,E.F., and Maniatis,T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Sarkar,S. and Subhedar,N., 2001. Glucagon-like immunoreactivity in the forebrain and pituitary of the teleost, Clarias batrachus (Linn.). Gen. Comp Endocrinol. 121, 23-31.

Satoh,F., Beak,S.A., Small,C.J., Falzon,M., Ghatei,M.A., Bloom,S.R., and Smith,D.M., 2000. Characterization of human and rat glucagon-like peptide-1 receptors in the neurointermediate lobe: lack of coupling to either stimulation or inhibition of adenylyl cyclase. Endocrinology 141, 1301-1309.

Scrocchi,L.A., Brown,T.J., MaClusky,N., Brubaker,P.L., Auerbach,A.B., Joyner,A.L., and Drucker,D.J., 1996. Glucose intolerance but normal satiety in mice with a null mutation in the glucagon-like peptide 1 receptor gene. Nat. Med. 2, 1254-1258.

Seino,S., Blackstone,C.D., Chan,S.J., Whittaker,J., Bell,G.I., and Steiner,D.F., 1988. Appalachian spring: variations on ancient gastro-entero-pancreatic themes in New World mammals. Horm. Metab Res. 20, 430-435.

Sherwood,N.M., Krueckl,S.L., and McRory,J.E., 2000. The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. Endocr. Rev. 21, 619-670.

Shimizu,I., Hirota,M., Ohboshi,C., and Shima,K., 1987. Identification and localization of glucagon-like peptide-1 and its receptor in rat brain. Endocrinology 121, 1076-1082.

Silverstein,J.T., Bondareva,V.M., Leonard,J.B., and Plisetskaya,E.M., 2001. Neuropeptide regulation of feeding in catfish, Ictalurus punctatus: a role for glucagon-like peptide-1 (GLP-1)? Comp Biochem. Physiol B Biochem. Mol. Biol. 129, 623-631.

Sivarajah,P., Wheeler,M.B., and Irwin,D.M., 2001. Evolution of receptors for proglucagon-derived peptides: isolation of frog glucagon receptors. Comp Biochem. Physiol B Biochem. Mol. Biol. 128, 517-527.

Skoglund,G., Hussain,M.A., and Holz,G.G., 2000. Glucagon-like peptide 1 stimulates insulin gene promoter activity by protein kinase A-independent activation of the rat insulin I gene cAMP response element. Diabetes 49, 1156-1164.

Soengas,J.L. and Moon,T.W., 1998. Transport and metabolism of glucose in isolated enterocytes of the black bullhead ictalurus melas: effects of diet and hormones. J. Exp. Biol. 201 (Pt 23), 3263-3273.

Tager,H., Hohenboken,M., Markese,J., and Dinerstein,R.J., 1980. Identification and localization of glucagon-related peptides in rat brain. Proc. Natl. Acad. Sci. U. S. A 77, 6229-6233.

Tager,H.S. and Markese,J., 1979. Intestinal and pancreatic glucagon-like peptides. Evidence for identity of higher molecular weight forms. J. Biol. Chem. 254, 2229-2233.

Tagliafierro,G., Carlini,M., Faraldi,G., Morescalchi,A.M., Putti,R., Della,R.A., Fasulo,S., and Mauceri,A., 1996. Immunocytochemical detection of islet hormones in the digestive system of Protopterus annectens. Gen. Comp Endocrinol. 102, 288-298.

Tang-Christensen,M., Larsen,P.J., Goke,R., Fink-Jensen,A., Jessop,D.S., Moller,M., and Sheikh,S.P., 1996. Central administration of GLP-1-(7-36) amide inhibits food and water intake in rats. Am. J. Physiol 271, R848-R856.

Tang-Christensen,M., Larsen,P.J., Thulesen,J., Romer,J., and Vrang,N., 2000. The proglucagon-derived peptide, glucagon-like peptide-2, is a neurotransmitter involved in the regulation of food intake. Nat. Med. 6, 802-807.

Tang-Christensen,M., Vrang,N., and Larsen,P.J., 2001. Glucagon-like peptide containing pathways in the regulation of feeding behaviour. Int. J. Obes. Relat Metab Disord. 25 Suppl 5, S42-S47.

Taylor,J.S., Van de,P.Y., Braasch,I., and Meyer,A., 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos. Trans. R. Soc. Lond B Biol. Sci. 356, 1661-1679.

Thiele,T.E., Seeley,R.J., D'Alessio,D., Eng,J., Bernstein,I.L., Woods,S.C., and van Dijk,G., 1998. Central infusion of glucagon-like peptide-1-(7-36) amide (GLP-1) receptor antagonist attenuates lithium chloride-induced c-Fos induction in rat brainstem. Brain Res. 801, 164-170.

Thiele,T.E., van Dijk,G., Campfield,L.A., Smith,F.J., Burn,P., Woods,S.C., Bernstein,I.L., and Seeley,R.J., 1997. Central infusion of GLP-1, but not leptin, produces conditioned taste aversions in rats. Am. J. Physiol 272, R726-R730.

Thompson,J.D., Higgins,D.G., and Gibson,T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Thorens,B., 1992. Expression cloning of the pancreatic beta cell receptor for the gluco-incretin hormone glucagon-like peptide 1. Proc. Natl. Acad. Sci. U. S. A 89, 8641-8645.

Thorgaard,G.H., 1983. Chromosomal differences among rainbow trout populations. Copeia 1983, 650-662.

Toft-Nielsen,M.B., Madsbad,S., and Holst,J.J., 1999. Continuous subcutaneous infusion of glucagon-like peptide 1 lowers plasma glucose and reduces appetite in type 2 diabetic patients. Diabetes Care 22, 1137-1143.

Tsai,C.H., Hill,M., and Drucker,D.J., 1997. Biological determinants of intestinotrophic properties of GLP-2 in vivo. Am. J. Physiol 272, G662-G668.

Turton,M.D., O'Shea,D., Gunn,I., Beak,S.A., Edwards,C.M., Meeran,K., Choi,S.J., Taylor,G.M., Heath,M.M., Lambert,P.D., Wilding,J.P., Smith,D.M., Ghatei,M.A., Herbert,J., and Bloom,S.R., 1996. A role for glucagon-like peptide-1 in the central regulation of feeding. Nature 379, 69-72.

Unger,R.H., 1976. The Banting Memorial Lecture 1975. Diabetes and the alpha cell. Diabetes 25, 136-151.

Unson,C.G. and Merrifield,R.B., 1994. Identification of an essential serine residue in glucagon: implication for an active site triad. Proc. Natl. Acad. Sci. U. S. A 91, 454-458.

Unson,C.G., Wu,C.R., Cheung,C.P., and Merrifield,R.B., 1998. Positively charged residues at positions 12, 17, and 18 of glucagon ensure maximum biological potency. J. Biol. Chem. 273, 10308-10312.

Uttenthal,L.O., Toledano,A., and Blazquez,E., 1992. Autoradiographic localization of receptors for glucagon-like peptide-1 (7-36) amide in rat brain. Neuropeptides 21, 143-146.

Van de,P.Y., Taylor,J.S., Joseph,J., and Meyer,A., 2002. Wanda: a database of duplicated fish genes. Nucleic Acids Res. 30, 109-112.

van Dijk,G. and Thiele,T.E., 1999. Glucagon-like peptide-1 (7-36) amide: a central regulator of satiety and interoceptive stress. Neuropeptides 33, 406-414.

Vieau,D., Gangnon,F., Jegou,S., Danger,J.M., and Vaudry,H., 1998. Characterization of the cDNA encoding the prohormone convertase PC2 and localization of the mRNA in the brain of the frog Rana ridibunda. Brain Res. Mol. Brain Res. 63, 1-13.

Von Schalburg,K.R. and Sherwood,N.M., 1999. Regulation and expression of gonadotropin-releasing hormone gene differs in brain and gonads in rainbow trout. Endocrinology 140, 3012-3024.

Wakelam,M.J., Murphy,G.J., Hruby,V.J., and Houslay,M.D., 1986. Activation of two signal-transduction systems in hepatocytes by glucagon. Nature 323, 68-71.

Wang,Z., Wang,R.M., Owji,A.A., Smith,D.M., Ghatei,M.A., and Bloom,S.R., 1995. Glucagon-like peptide-1 is a physiological incretin in rat. J. Clin. Invest 95, 417-421.

Wettergren,A., Wojdemann,M., Meisner,S., Stadil,F., and Holst,J.J., 1997. The inhibitory effect of glucagon-like peptide-1 (GLP-1) 7-36 amide on gastric acid secretion in humans depends on an intact vagal innervation. Gut 40, 597-601.

Wheeler,M.B., Lu,M., Dillon,J.S., Leng,X.H., Chen,C., and Boyd,A.E., III, 1993. Functional expression of the rat glucagon-like peptide-I receptor, evidence for coupling to both adenylyl cyclase and phospholipase-C. Endocrinology 133, 57-62.

White,J.W. and Saunders,G.F., 1986. Structure of the human glucagon gene. Nucleic Acids Res. 14, 4719-4730.

Xiao,Q., Giguere,J., Parisien,M., Jeng,W., St Pierre,S.A., Brubaker,P.L., and Wheeler,M.B., 2001. Biological activities of glucagon-like peptide-1 analogues in vitro and in vivo. Biochemistry 40, 2860-2869.

Yang,H., Egan,J.M., Wang,Y., Moyes,C.D., Roth,J., Montrose,M.H., and Montrose-Rafizadeh,C., 1998. GLP-1 action in L6 myotubes is via a receptor different from the pancreatic GLP-1 receptor. Am. J. Physiol 275, C675-C683.

Yeung,C.M. and Chow,B.K., 2001. Identification of a proglucagon cDNA from Rana tigrina rugulosa that encodes two GLP-1s and that is alternatively spliced in a tissue-specific manner. Gen. Comp Endocrinol. 124, 144-151.

Youson,J.H., Al Mahrouki,A.A., Naumovski,D., and Conlon,J.M., 2001. The endocrine cells in the gastroenteropancreatic system of the bowfin, Amia calva L.: an immunohistochemical, ultrastructural, and immunocytochemical analysis. J. Morphol. 250, 208-224.

Yusta,B., Huang,L., Munroe,D., Wolff,G., Fantaske,R., Sharma,S., Demchyshyn,L., Asa,S.L., and Drucker,D.J., 2000. Enteroendocrine localization of GLP-2 receptor expression in humans and rodents. Gastroenterology 119, 744-755.

Yusta,B., Somwar,R., Wang,F., Munroe,D., Grinstein,S., Klip,A., and Drucker,D.J., 1999. Identification of glucagon-like peptide-2 (GLP-2)-activated signaling pathways in baby hamster kidney fibroblasts expressing the rat GLP-2 receptor. J. Biol. Chem. 274, 30459-30467.

Zardoya,R., Cao,Y., Hasegawa,M., and Meyer,A., 1998. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. Mol. Biol. Evol. 15, 506-517.

Zhang,Z. and Gurr,S.J., 2000. Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. Gene 253, 145-150.