INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

STUDIES OF CHROMOGRANIN A

by

Yu Hou

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment

of the requirements for the degree of

Master of Science

Department of Medicine

Division of Experimental Medicine

McGill University

Montreal, Quebec, Canada

Abstract

Chromogranin A (CgA) is an acidic glycoprotein that is specifically expressed in neuroendocrine tissue. It is the best characterized member of the granin family, which includes proteins proposed to be involved in the regulation of peptide hormone and neurotransmitter secretion in the regulated secretory pathway by helping granule formation, targeting peptide hormones and neurotransmitters to the granules, affecting peptide hormone processing and controlling secretion through a feedback mechanism.

Although elaborate *in vitro* studies have been conducted on CgA, its precise neuroendocrine function and the complete elements that control its neuroendocrine specific expression remain unclear. This project applied transgenic technology to study CgA regulation and function *in vivo*. For the study of the regulation of CgA gene expression, four constructs, containing one of two promoter portions of the CgA gene (the 184bp proximal promoter portion and the 6kb 5'-flanking region) in two reporter systems (β-galactosidase and green fluorescent protein systems), were constructed. The feasibility of the strategy was confirmed by *in vitro* detection of the reporter activity that is higher in neuroendocrine than in nonendocrine cell lines. These constructs can be used for pronuclei microinjection to generate transgenic mice harboring the promoter-reporter cassettes to be used for promoter activity analysis. In addition, a new strategy of introducing the promoter-reporter cassettes into the Hprt locus was proposed and some preliminary work has been done for constructing this targeting vector. For the CgA gene function study, targeted disruption of the CgA gene in mice was conducted by transfection of the CgA targeting vector to embryonic stem (ES) cells and positive recombinant ES cell lines were used to generate CgA knockout mice. Several

heterozygous and homozygous CgA gene ablation mice were obtained. They are to be analyzed for further study.

In the meanwhile, the entire human and mouse CgA gene sequences including the flanking regions were analyzed and compared to gain insight into gene regulation and the evolution of interspersed repeat elements within these regions. In addition, an attempt was made to map the CgA gene in the zebrafish genome by the LN54 radiation hybrid cell panel.

Résumé

La chromogranine A (CgA) est une glucoprotéine acide exprimée de manière spécifique dans les tissus neuroendocriniens. Elle comporte les meilleures caractéristiques de la famille des granines et, selon toute vraisemblance, possèdent des protéines qui jouent un rôle dans la régulation de la sécrétion des hormones peptidiques et des neurotransmetteurs dans la voie de sécrétion régulée en contribuant à la formation des granules, en guidant les hormones peptidiques et les neurotransmetteurs vers les granules, en agissant sur le traitement des hormones peptidiques et en contrôlant la sécrétion par le biais d'un mécanisme de rétroaction.

Même si des études *in vitro* très élaborées ont été entreprises sur CgA, sa fonction neuroendocrine précise et les éléments qui contrôlent son expression neuroendocrine spécifique restent obscurs. Dans le cadre de ce projet, nous avons appliqué la technique transgénique à l'étude de la régulation et de la fonction de CgA *in vivo*. Pour l'étude de la régulation de CgA, quatre gènes hybrides, contenant une des deux régions promotrices du gène CgA (portion du promoteur proximal 184bp et région flanquante 6kb 5') dans deux systèmes rapporteurs (β-galactosidase et protéine fluorescente verte) ont été mis au point. La faisabilité de la stratégie a été confirmée par détection *in vitro* de l'activité des rapporteurs qui est plus haute dans les lignées cellulaires neuroendocrines que dans les lignées cellulaires non endocrines. Ces gènes hybrides peuvent être utilisés pour des micro-injections de pronucléus dans le but de générer des souris transgéniques porteuses de cassettes promoteur-rapporteur en vue de leur utilisation pour l'analyse de l'activité des promoteurs. Une nouvelle stratégie visant l'introduction de cassettes promoteur-rapporteur dans le locus Hprt a été proposée et des travaux préliminaires ont été effectués

pour construire ce vecteur de ciblage. Pour l'étude de la fonction du gène CgA, la perturbation ciblée du gène CgA chez la souris a été entreprise par transfection du vecteur de ciblage CgA dans les cellules souches embryonnaires et dans des lignées de cellules souches embryonnaires recombinantes positives qui ont été utilisées pour la production de souris privées de CgA (souris knockout). Plusieurs souris privées du gène CgA hétérozygote et homozygote ont été obtenues. Celles-ci feront l'objet d'autres études.

En attendant, l'intégralité des séquences du gène CgA chez l'homme et la souris y compris les régions flanquantes ont été analysées et comparées afin de mieux comprendre la régulation de ce gène et l'évolution d'éléments répétés intercalés dans ces régions. De plus, une tentative a été faite en vue de cartographier le gène CgA dans le génome du poisson zébré au moyen d'une méthode de cartographie par hybride d'irradiation LN54.

Acknowledgements

Many thanks to:

Dr.Geoffrey N. Hendy for his patience, understanding and supervision throughout my graduate studies.

Lucie Canaff and Bing Yang for their helpful suggestions throughout my training. Richard Feldstein for construction and purification of the targeting vector. Dr.Andrew C. Karaplis and members of his laboratory for their help with generation of ES cell clones. Dr.Tong Li for her help in doing ES cell clone confirmation by southern blot (Fig 4.7 at p84). Dr. Michel Tremblay and members of his laboratory for blastocyst injection and generation of the heterozygote [CgA (+/-)] mice.

Dr. Hugh P. J. Bennett and Ben Cadieux for their help with organizing the zebrafish CgA mapping.

The Royal Victoria Hospital Research Institute, The Department of Medicine of McGill University and The Medical Research Council (Operating Grant to Dr.Geoffrey Hendy) for providing funding during my training.

Dedication

I dedicate this thesis to my parents and husband for their support, love and encouragement throughout my studies

Table of Contents

Title page I

Abstract II

Resume IV

Acknowledgement VI

Dedication VII

Table of Contents VIII

Abbreviations XI

List of Figures XIV

List of Tables XVI

Part I: Introduction and Literature Review 1

Chapter 1: Chromogranin A 2

1.1 The granins 3

1.1.1 Introduction 3

1.1.2 Functions of the granins 6

1.2 Chromogranin A 6

1.2.1 Biochemistry and structural features 6

1.2.2 Post-translational modifications 8

- Phosphorylation
- Glycosylation
- Sulphation

1.2.3 Proteolytic processing 10

1.2.4 Putative function 12

1.2.4.1 Extracellular functions 13

- Regulation of the resident hormone secretion
- Modulation of cell adhesion
- Antibacterial and antifungal activities

1.2.4.2 Intracellular functions 16

- Packing of peptide hormones and neuropeptides to secretory granules
- Regulating the proteolytic processing of the resident prohormones
- Regulating intracellular calcium flow as an IP3-sensitive intracellular calcium-binding protein

1.2.5 CgA gene structure 19
1.2.6 Distribution and ontogeny 21
- Distribution
- Ontogeny

1.2.7 Regulation of biosynthesis and secretion 23
1.2.8 CgA as a neuroendocrine tumor marker 24

Chapter 2 Transgenic approaches for gene function and regulation study 27

2.1 Transgenic technology 28
2.1.1 Introduction 28
2.1.2 Development of transgenic technology 28
2.2 Promoter activity study *in vivo* 30
2.2.1 Reporter systems 30
- β- gal reporter system
- GFP reporter system

2.2.2 Conventional approach 32
2.2.3 Introducing a single copy of transgene cassette to a chosen site 32
2.3 Gene functional analysis *in vivo* 34
2.3.1 Gain-of-function analysis 34
2.3.2 Loss-of-function analysis 34
- Conventional approach
- Tissue-specific inducible gene knockout 36

2.3.3 Introducing mutations into the target gene 41

Part II Experimental section 45

Chapter 3 Construction of constructs to study neuroendocrine specific regulation of CgA gene *in vivo* 46

3.1 In vitro study of neuroendocrine specific expression of CgA gene 47
3.1.1 Introduction 47
3.1.2 DNA response elements in the CgA 5' flanking region 48
3.2 Strategies of studying neuroendocrine specific expression of CgA gene *in vivo*
3.2.1 Promoter portions studied 51
3.2.2 Reporter systems used 52
3.2.3 Introducing a single copy of the CgA promoter-reporter cassette to the mouse genome 52
3.3 Material and methods 53
3.3.1 β-gal reporter system construction 53
3.3.2 GFP reporter system construction 56
3.3.3 Testing the feasibility of the promoter reporter cassettes 58

3.3.4 Preparation to generate the targeting vector for introducing a single copy of the promoter-reporter cassette to the Hprt gene locus 58
3.4 Results 60
3.4.1 β-gal reporter constructs 60
3.4.2 GFP reporter constructs 63
3.4.3 β-gal assay of the β-gal construct transfected cells 64
3.4.4 Green fluorescent microscopy of the GFP construct transfected cells 65
3.4.5 Introducing a NotI site to the β-gal construct I 66
3.5 Discussion 67

Chapter 4. Functional analysis of CgA gene *in vivo* 72

4.1 Introduction and strategy 73
4.2 Material and methods 77
4.2.1 Generation of CgA gene ablation ES cell lines 77
4.2.1.1 ES cell culture 77
4.2.1.2 Transfection and colony selection 77
4.2.1.3 Screening of the colonies 78
- ES cell genomic DNA extraction
- DNA digestion with XhoI
- Southern blot analysis
- Probes generation and labeling

4.2.1.4 Culture of selected ES cell lines 81
4.2.2 CgA gene knockout mice generation 82
4.3 Results 82
4.3.1 Screening results 82
4.3.2 Recombination confirmation 85
4.3.3 Transgenic mice generation 87
4.4 Discussion 87
4.4.1 ES cell manipulation 87
4.4.2 Screening strategy 89
4.4.3 CgA gene knockout outcome expectation 90

Chapter 5. CgA gene study

Preface 95
5.1 Zebrafish CgA gene mapping 96
5.1.1 Introduction of Zebrafish genetic mapping
5.1.2 Strategy of zebrafish CgA gene mapping
- LN54 RH panel
- Primer design & PCR amplification

5.1.3 Results and discussion 100
5.2 Comparison of human and mouse CgA gene 104
5.2.1 Methods of comparing hCgA and mCgA gene 104
5.2.2 Discussion 108
5.2.2.1 Interspersed repetitive elements in the hCgA gene and mCgA gene

5.2.2.2 Transcription regulation elements in the conserved regions.....112

Part III References 114

Abbreviation

1,25 $(OH)_2D$	1,25-dihydroxy-vitamin D
bp	base pairs
Ca2+	calcium ion
cAMP	cyclic adenosine-3 ',5 '-monophosphate
cDNA	complementary DNA
CgA	Chromogranin A
CgB	Chromogranin B
CgC	Chromogranin C
cpm	counts per minute
CRE	cAMP response element
CREB	cAMP response element binding protein
DNA	deoxyribonucleic acid
ECL	enterochromaffin cell line
EN-1 & 2	Engrailed 1 and 2
ERK	extracellular regulated kinase
ERE	estrogen response element
ES cell	embryonic stem cell
EST	expressed sequence tag
FBS	fetal bovine serum
FIAU	fialuridine

FLAM	free left Alu monomer
FRAM	free right Alu monomer
FSH	follicle stimulating hormone
g	gram
GFP	green fluorescent protein
GRE	glucocorticoid response element
hCgA	human chromogranin A
IR	interspersed repeat (or insulin receptor in chapter 2)
JAK	Janus kinase
kb	kilobase pair
kDa	kilodalton
LIF	leukemia inhibitory factor
LH	luteinizing hormone
LINE	long interspersed nuclear element
LTR	long terminal repeat
MAPK	mitogen-activated protein kinase
M	molar
mCgA	mouse chromogranin A
MIR	mammalian interspersed repeat element
ml	millilitre
mM	millimole
mRNA	messenger RNA
neo^r	neomycin resistant gene

NGF	nerve growth factor
PBS	phosphate buffered saline
PC1	prohormone convertase 1
PC2	prohormone convertase 2
PCR	polymerase chain reaction
PKA	protein kinase A
PKC	protein kinase C
POMC	pro-opiomelanocortin
PTH	parathyroid hormone
PTHrP	parathyroid hormone-related protein
RAPD	random amplified polymorphic DNA
RH panel	radiation hybrid panel
RLEs	retrovirus-like elements
RSK	ribosomal protein S6 kinase
SgI	secretogranin I (CgB)
SgII	secretogranin II (CgC)
SgIII	secretogranin III (1B *1075)*
SgIV	secretogranin IV (HISL-19)
SgV	secretogranin V (protein 7B2)
SINE	short interspersed nuclear element
SRP	signal recognition particle
SSLP	simple sequence length polymorphism
STAT	signal transducer and activator of transcription

VDRE vitamin D response element

List of Figures

Fig1.1 Structure of the granin family members 5

Fig 1.2 CgA gene structure 20

Fig 2.1 Mechanism of introducing the transgene into the Hprt gene locus 33

Fig 2.2 Targeted homologous recombination 35

Fig 2.3 Mechanism of Cre/lox site-specific recombination 38

Fig 2.4 The mechanism of tetracycline-inducible binary system 40

Fig 2.5 The structure of the Cre-ER^T transgene 41

Fig 2.6 Introducing a point mutation into the α-MHC gene 42

Fig 2.7 Targeting strategy for substitution of En-2 for En-1 44

Fig 3.1 β-gal reporter system construction 55

Fig 3.2 GFP reporter system construction 57

Fig 3.3 Polylinker site of the vector pMP8SKB 59

Fig 3.4 β-gal construct I and GFP construct I 61

Fig 3.5 β-gal construct II and GFP construct II 62

Fig 3.6 β-gal assay result of the β-gal constructs 64

Fig 3.7 Transient transfection result of the GFP constructs 65

Fig. 3.8 Introducing a Not I site to the β-gal construct I 66

Fig. 4.1 Targeting vector and targeting event for CgA gene 74

Fig. 4.2 Generation of CgA gene knockout mice 76

Fig 4.3 Screening strategy of CgA gene knockout in ES cells 83

Fig 4.4 Southern blot screening of the survived ES cell colonies 84

Fig 4.5 Neomycin gene reprobe of the screening blot 84

Fig 4.6 Strategy of confirming the ES cell lines containing targeted allele 86

Fig 4.7 Southern blot confirmation of the ES cell lines containing targeted allele 86

Fig 5.1 Sequence and electrophoresis of Zebrafish CgA gene intron1 101

Fig 5.2 PCR of Z-CgA-In1 in LN 54 RH panel 102

Fig 5.3 IRs in the hCgA & mCgA gene coding region 106

Fig 5.4 Highly conserved regions between hCgA & mCgA 107

Fig 5.5 The evolution of Alu family members 111

List of Tables

Table 1.1 General properties of the granin family members . 4

Table 1.2 CgA derived peptides and their biological effects 10

Table 1.3 The putative functions of chromogranin A 12

Table 3.1 Putative cis-acting response elements in the 5' flanking region of the human chromogranin A gene 50

Table 5.1 Mapping result of zebrafish CgA gene 102

Part I

Introduction and Literature Review

Chapter 1

Chromogranin A

1.1 The Granins

1.1.1 Introduction

The granins are a family of acidic secretory proteins that are co-stored with resident peptide hormones, neurotransmitters, or amines in secretory granules of a wide variety of endocrine cells and neurons[67]. Presently, seven members of the granin family have been described: chromogranin A[158] , chromogranin B[158], secretogranin II (SgII)[157], 1B1075 gene product (SgIII)[111], HISL-19 antigen (SgIV)[79], 7B2(SgV)[127] and NESP55 (SgVI?)[71]. A comparison of the primary structure reveals that chromogranin A and chromogranin B are the most closely related members of the granin family[10].

Table1.1 shows several properties of the granins. The high percentage of acidic amino acids (glutamic acid, Glu and aspartic acid, Asp) contributes to their low isoelectric point (PI). The heat stability is due to their high hydrophilic nature. At least three of the granins (CgA, CgB, SgII) have been shown to have the capacity to bind calcium. CgA, CgB and 7B2 possess a consensus disulphide loop structure. The most prominent feature of the granins is that they all contain multiple sites with two or more adjacent basic residues that are potential sites for proteolytic processing. (Fig1.1)

Granin[2]	Mass (kDa)	mRNA (kB)	Chromosome Assignment	pI	% Glu/Asp	Dibasic Sites	Heat Stability	Calcium Binding	Disulphide Loop
CgA	49	2.1	14q32.2-32.3	4.9	25	10	Yes	Yes	Yes
CgB	76	2.5	20pter-p12	5.1	24	15	Yes	Yes	Yes
SgII	68	2.5	ND[3]	5.0	20	9	Yes	Yes	No
SgIII (1B1075)	57	2.2	9*	5.1	19	10	ND	ND	No
SgIV (HISL-19)	ND	ND	ND	5.6	ND	ND	ND	ND	ND
SgV (7B2)	21	1.4	15q11-q15	5.2	16	3	Yes	ND	Yes
SgVI? (NESP55)	55	0.7	ND	4.8	21	5	Yes	ND	ND

Table 1.1 General properties of the Granin Family members

1. Adapted from Huttner et al., (1991), Fisher-Colbrie et al., (1997) and Feldstein (1998)
2. Data shown is from human granins, except for SgIII (from rat and mouse*) and SgVI (from bovine)
3. ND means "not determined"

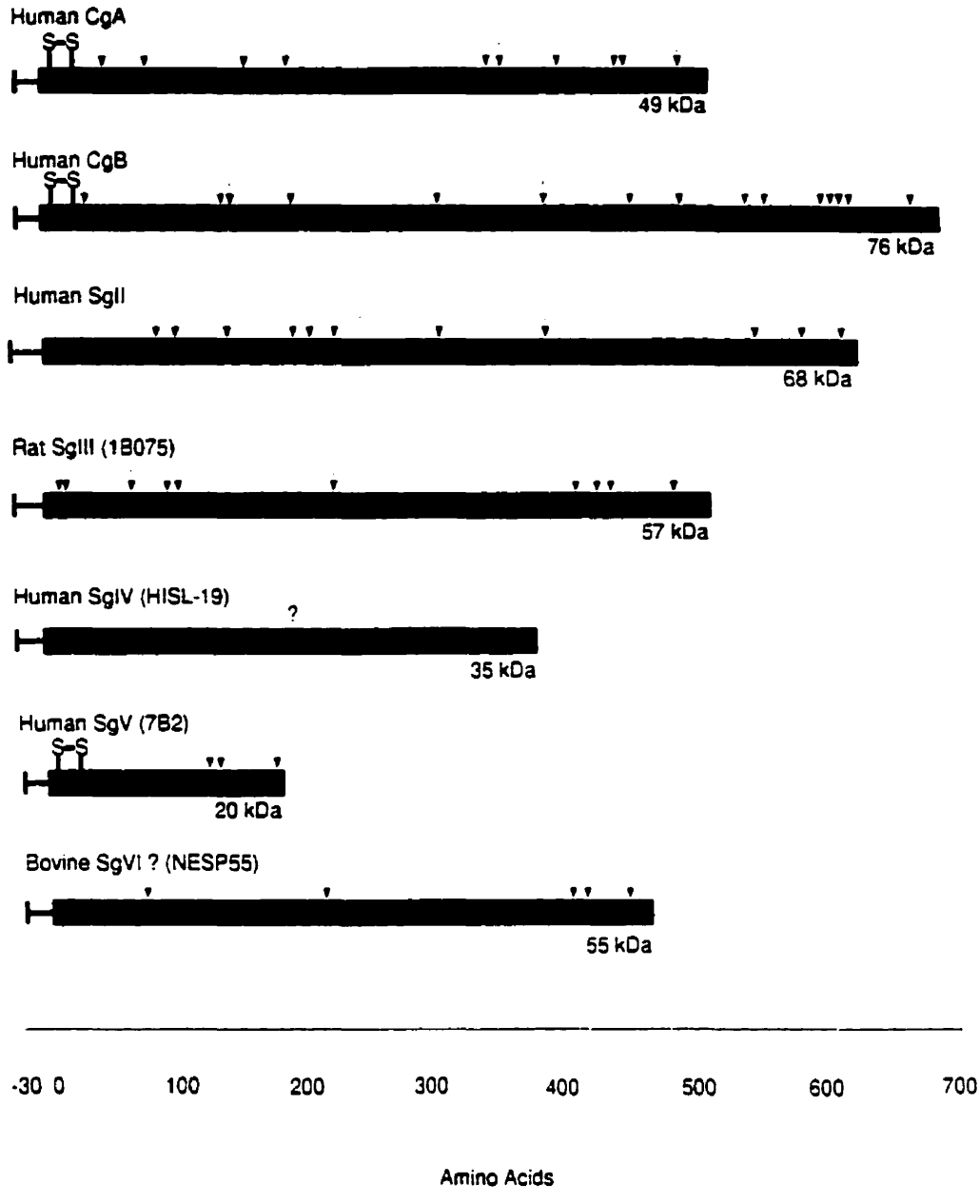

Fig1.1 Structure of the granin family members. (Taken from Feldstein, 1998) The human granins are shown, except for SgIII (from rat) and SgVI (from bovine). The arrows represent pairs of basic amino acid residues for proteolytic processing. In CgA, CgB and 7B2, there is a disulphide loop present at the amino-terminal region.

1.1.2 Functions of the granins

The physiological function of the granins is still an open question. However, granins have been proposed to play multiple roles in the secretory process, and these functions can be divided into two categories: extracellular and intracellular. Extracellularly, they serve as precursors to biologically active peptides and regulate the secretion of resident hormones or neuropeptides. Intracellularly, they play a role as helper proteins to sort hormones or neuropeptides into regulated secretory granules; and as a prohormone, they are substrates of the prohormone convertases themselves and function as regulators of the resident hormone processing in a competitive manner. Furthermore, they may also function as intracellular calcium binding proteins and play some role in calcium related actions.

1.2 Chromogranin A

Chromogranin A is the major member of the granin family. It is the first member of this family to be identified. The name derives from its original discovery in the catecholamine-containing chromaffin granules of the adrenal medulla. The research work of this thesis is focused on chromogranin A. Therefore, the following sections limit the discussion to the structure, function and regulation of CgA.

1.2.1 Biochemistry and structural features

The calculated molecular weight of CgA is approximately 49 kDa. However, the apparent Mr on SDS-polyacrylamide gels is 75-80 kDa. The migration aberrance is due to the unusually high content of glutamic and aspartic acid residues (about 25%) and of

proline (10%), which contribute to its acidic pI of 4.5-5 and its extreme hydrophilicity. Moreover, post-translational modifications may also contribute to the aberrant migration.

The primary product of CgA mRNA translation contains an 18 amino acid signal sequence. The mature CgA protein varies in length from 430 amino acids in the pig to 445 in the mouse. The amino acid sequence of CgA contains 8-10 pairs (depending upon the species) of basic amino acid residues, which may serve as targets for endoproteolytic attack by convertases. Several biologically active peptides are putatively derived from the CgA molecule: β-granin, pancreastatin, catestatin, chromostatin, vasostatin and parastatin[61].

CgA also possesses a cell adhesion sequence, Arg-Gly-Asp (RGD) near its amino-terminus in most species, except in the rat where it is located at the carboxy-terminus. Many proteins that are involved in attachment of anchorage-dependent mammalian cells to the extracellular matrix or in cell-cell adhesion, such as fibronectin, fibrinogen and integrins possess the RGD sequence. However, although the N-terminal fragments of CgA have been shown to have the pro-adhesive effect of fibroblasts under certain conditions, there is no evidence that this effect is related to the RGD sequence[47].

There are several oligoglutamic acid residue stretches present in the CgA molecule. These motifs are conserved and may play a role in determining the secondary and tertiary structures of the molecule. Bundles of negatively charged glutamic acid residues would oppose the formation of a stable α-helix and cause a disorganized "random coil" structure in these regions[131]. This has been proven by circular dichroism analysis of adrenal medulla CgA, which indicates predominately random coil structure (60-65%), with α-helix and β-sheet conformations representing about 25% and 10%

respectively. This random coil structure is the major reason for the high heat stability of the CgA molecule.

A unique feature of rodent CgA is the presence of a polyglutamine tract. A tract of 11 repeating glutamine residues, beginning at amino acid 84, can be found in the mouse sequence. In the rat, however, this polyglutamine tract displays a polymorphic structure. 16-20 repeating glutamine residues are found beginning at amino acid 74 in different rat CgA cDNA clone sequences. The structural and functional significance of the glutamine tract and its polymorphism in these rodent CgA molecules is presently unknown.

1.2.2 Post-translational modifications

Pulse-chase labeling of chromaffin cells from bovine adrenal medulla followed by two-dimensional electrophoresis revealed that the early-labeled CgA spots behaved identically to the *in vitro* translation products. However, at later intervals, significant changes in size and pI occurred. These modifications were obviously due to post-translational processes[158].

- **GLYCOSYLATION**

Bovine adrenal medulla CgA contains 5.4% carbohydrate, which is mainly present as O-glycosidically linked tri- and tetra- saccharides composed of N-acetylgalactosamine, galactose and sialic acid[43]. Recently, by electrospray and matrix-assisted laser desorption ionization time-of-flight technique, two O-linked glycosylation sites were located on Ser^{186} and Thr^{231} in the middle part of bovine CgA from chromaffin

granules[138]. Potential sites for N-glycosylation are present in human, rat and mouse sequences (Fig. 1.2), but it is unclear whether N-linked glycosylation actually occurs.

- **PHOSPHORYLATION**

CgA is a phosphoprotein whose phosphorylated residues are mainly serine and threonine. It has been proven that bovine adrenal CgA contains five phosphoserine residues per molecule[129]. In addition, phosphorylation of Tyr^{173} residue is responsible for the antibacterial activity of the CgA-derived fragment, chromacin (CgA 173-194)[139].

- **SULFATION**

Adrenal CgA incorporates [^{35}S] sulfate during biosynthesis. Most, if not all, of this sulfate appears to be bound to carbohydrates[120], especially oligosaccharides[53], and there is a proteoglycan form of CgA in the chromaffin granules. The relative amount of the proteoglycan form of CgA varies depending on species, tissue and developmental stage[156].

Other post-translational modifications of CgA include carboxymethylation[102], α-amidation of the CgA derived peptide pancreastatin[45], and the formation of a disulfide bond between the two N-terminal cysteines[10].

1.2.3 Proteolytic processing

As mentioned previously, the CgA molecule undergoes proteolytic processing by prohormone convertases to generate biologically active peptides during routing to and storage in secretory vesicles.(For a review of peptide hormone precursor processing, see ref. 19.) Both the pattern and rate of CgA processing vary in a tissue-specific manner. In adrenal medulla[132] and anterior pituitary[7], the rate and extent of processing are low. In contrast, CgA is processed extensively in the endocrine pancreas[68] and gastrointestinal tissues[152]. Proteolytic processing of CgA may also occur after its release from neuroendocrine cells[92]. (Table1.2)

Granin	*Peptide*	*Sequence*	*Cleavage*	*Origin*	*Effect*	*Target tissue*
CgA	Betagranin	1–128 (r)	NH_2-KR	Islet β-cell	Unknown	Unknown
	Betagranin-like peptide	1–113 (b)	NH_2-KR	Parathyroid	↓ PTH/CgA secretion	Parathyroid
	Vasostatin	1–76 (b)	NH_2-KK	Adrenal medulla	↓ vascular tone	Arteries, veins
	Chromostatin[a]	124–143 (b)	K-K	Adrenal medulla	↓ catecholamine secretion	Adrenal medulla
	Pancreastatin	240–288 (p)	R-KR	Islet β-cell	↓ insulin secretion	Islet β-cell
					↓ PTH secretion	Parathyroid
	Parastatin	347–419 (p)	K-K	Parathyroid*	↓ PTH secretion	Parathyroid

Table 1.2 CgA derived peptides and their biological effects
The species from which the sequences of CgA are derived are abbreviated as followings: r, rat; b, bovine. p, porcine (Adapted from ref.101) [a]Note that a recent correction [59] indicates that the inhibition of catecholamine secretion is not due to chromostatin but maybe to another CgA-derived peptide. The asterisk indicates the fact that parastatin was produced by exogenous protease.

Two-dimensional gel electrophoresis, HPLC, N-terminal sequencing and immunoblotting were used to analyze the proteolytic processing of CgA in bovine chromaffin granules[108]. The results demonstrate that CgA-processing occurs in both N- and C-terminal moieties of the protein. The major CgA fragments were obtained by

cleavage of peptide bonds localized in the C-terminal part of the protein (247-248 and 291-292). The cleavage sites observed include three dibasic sites: 77-78 (KK), 114-115 (KR) and 314-315 (KR); four single arginine residue sites 247-248 (RA), 353-354 (RG), 358-359 (RG) and 386-387 (RG); and the other four sites involved the peptide bond hydrolysis of Val, Leu or Phe residues which are 3-4 (VN), 64-65 (LA), 291-292 (LF) and 350-351 (FR). [92] When correlating the biologically active peptides derived from CgA to these processed fragments, it seems that the granin (1-113), chromostatin (124-143) and pancreastatin (248-294) are part of larger endogenous fragments: 1-115, 116-431 and 248-431 respectively. The occurrence of pancreastatin starting at position 248 and 292 suggests cleavage at the N-terminus in the secretory granule and maturation in the extracellular space. Chromostatin may represent the active sequence of a larger peptide.

It has been shown that the endoproteases PC1 and furin play a role in pituitary CgA processing. PC1 antisense studies demonstrate that endogenous PC1 is responsible for the generation of a ~50 kDa N-terminal fragment and a ~30 kDa C-terminal fragment from the intact CgA protein and an ~66 kDa C-terminal fragment in the mouse pituitary corticotrope cell line AtT-20[38]. Over-expression of PC2 in AtT-20 cells increased processing to and secretion of ~71 and ~27 kDa N-terminal CgA fragments. Moreover, antisense PC1 specifically abolished regulated secretion of CgA in response to the secretagogue, corticotropin-releasing hormone, demonstrating the cleavage-dependent trafficking of endogenous neuropeptides.

1.2.4 Putative functions

As for other members in the granin family, CgA is proposed to have extracellular functions and intracellular functions. (Table 1.3)

Table 1.3 The putative functions of chromogranin A

EXTRACELLULAR
1 Precursor of biologically active peptides regulating resident hormone secretion
2 Regulation of cell adhesion
3 Antibiotic functions
INTRACELLULAR
1 Packaging of peptide hormones and neuropeptides, granule condensation and exocytosis
2 Intracellular regulator of prohormone processing
3 IP_3-sensitive intracellular calcium-binding protein

1.2.4.1 Extracellular functions

1) Regulation of resident hormone secretion

The processed products of CgA are biologically active peptides that are proposed to possess multiple functions that ultimately regulate the secretory activity of the resident hormone in an autocrine or paracrine manner.

The first evidence that supports this notion came about through the characterization of a 49 amino-acid carboxyl-amidated peptide, pancreastatin, which shares identity with porcine CgA-(240-288). Pancreastatin inhibits glucose-stimulated insulin release from the perfused rat pancreas[144]. It was also shown to have activity in other cell types, for example, in inhibiting secretion of parathyroid chief cells[34,61], stomach parietal cells, and in affecting glucose metabolism in rat hepatocyte[86].

In addition to pancreastatin, β-granin, a 20 kDa peptide first isolated from rat insulinoma cells, is derived from the amino-terminus of CgA (amino acids1-113) and inhibits activity of parathyroid secretion[33]. The N-terminal fragment (amino acids 1-76) of β-granin has been shown to inhibit arterial smooth muscle contraction *in vitro*, and thus is called vasostatin[1]. The synthesized truncated fragment of vasostatin, CgA-(1-40), has been proven to reduce parathyroid hormone-related peptide and calcitonin secretion from a squamous-cell carcinoma line and parathyroid hormone release from primary cultures of parathyroid cells[31,61].

Parastatin is another peptide generated from CgA [porcine CgA-(347-419)] that inhibits of low calcium stimulated parathyroid cell secretion of PTH *in vitro*[40]. However, this peptide is of lower potency than pancreastatin or β-granin.

Chromogranin A was first identified in chromaffin cells of the adrenal medulla, but the functional peptide that regulates secretion from these cells is the most recently defined novel fragment of CgA [bovine CgA-(344-364)], named catestatin[147]. It has been shown to inhibit catecholamine release from chromaffin cells and noradrenergic neurons by acting as a non-competitive nicotinic cholinergic antagonist. The principle of its action is described as the following: Catestatin exists in nature as a β-strand/loop/β-strand structure with three arginine residues in the loop. The positive charge of the arginine residues "dock" the catestatin molecule to the negatively charged extracellular domain of the Torpedo nicotinic cholinergic receptor. The catestatin molecule interacts with the β and δ units of the receptor and the ligand-receptor complex occludes the cation pore, providing the structural basis for the non-competitive nicotinic cholinergic antagonist properties of the peptide[147].

CgA derived peptides appear to be inhibitors of endocrine and some exocrine secretions. There is a new model, named "zero steady-state error" (ZSSE) homeostasis control, that attempts to explain the crucial role of these CgA-derived peptides in counter-regulatory homeostatic hormone systems, such as the PTH/calcitonin system for plasma ionized calcium and glucagon/insulin system for plasma glucose[77]. The model indicates that the secretory activities of the counter-regulatory homeostatic hormones are controlled by plasma CgA levels and always return to an equilibrium point regardless of disturbances within physiological limits.

2) Modulation of cell adhesion

A new physiological role of CgA was addressed recently. It has been shown that mixtures of large molecular weight CgA fragments can inhibit fibroblast adhesion, and limited trypsin treatment can convert the anti-adhesive activity into pro-adhesive activity. The fragments corresponding to vasostatin I, vasostatin II (residues 1-78, 1-115) and a synthetic peptide encompassing the residues 7-57 exert pro-adhesive effects. These fragments induce adhesion and spreading of fibroblasts on plates coated with collagen I or IV, laminin and fetal calf serum. Since the peptides encompassing the residues 1-20, 25-46, 37-57, and 47-68 do not have this pro-adhesive effect, it suggests that conformational constraints in the N-terminal domain are necessary for activity[47].

The fact that anti-adhesive fragments can be converted to pro-adhesive fragments indicates that proteolytic processing of CgA could be critical for its adhesive functions. Thus, tissue specific patterns of CgA processing may regulate its adhesive effect. Since the change of adhesion of fibroblasts markedly changes their physiology, the adhesive activity of CgA may be important for the regulation of neuroendocrine tissue development and remodeling.

3) Antibacterial and antifungal activities

Bovine vasostatin-1 (amino acids 1-76 of bovine CgA), human recombinant VS-1, and rat synthetic CgA_{7-57} possess antimicrobial activity against Gram-positive bacteria and a large variety of filamentous fungi and yeast cells at micromolar concentrations, and they are also able to kill a large variety of filamentous fungi and yeast cells in the 1-10 μM range[87]. The rat synthetic CgA_{7-57} displays decreased antimicrobial activity.

The 19 residues at the C-terminal end (residues 58-76) of vasostatin are important for the formation of helical secondary structure, and probably play a role in its antimicrobial activity. Furthermore, opening of the disulfide bridge and *S*-pyridylethylation of CgA-derived peptides results in slightly decreased antibacterial activity. The disulfide bridge structure is not crucial for the antifungal activity. The molecular mechanisms responsible for antimicrobial activity are still under investigation.

1.2.4.2 Intracellular functions

1) Packaging of peptide hormones and neuropeptides to secretory granules

Besides its numerous extracellular functions, CgA is proposed to play an important role in the regulation of resident peptide hormone trafficking to the regulated secretory pathway.

Peptide hormones and neuropeptides destined for the regulated secretory pathway are stored at high concentrations in secretory granules, which are derived from the trans-Golgi network[20, 62], in endocrine and neuroendocrine cells until they are released in response to signals from various secretagogues. This storage may be required for the proteolytic processing of prohormones to active peptides. Other secretory proteins leave the cell through the constitutive secretory pathway, originating from the trans-Golgi network; or through the constitutive-like secretory pathway, originating from immature secretory granules. Thus, granule storage of secretory proteins requires two sorting steps, "sorting for entry" and "sorting by retention", which refine the final complement of stored proteins[3].

Several mechanisms for sorting of secretory proteins into secretory granules have been demonstrated. These include receptor-mediated transport of selected secretory proteins (the mechanism of the "sorting for entry" step), low pH and /or calcium-induced aggregation (the mechanism of "sorting by retention" step), and direct binding to specific lipid domains in granule membranes. It was recently proposed that sorting-for-entry dominates in neuroendocrine cells while sorting-by-retention dominates in endocrine cells[145].

It appears that different sorting mechanisms are responsible for the sorting of different proteins to the secretory granules. For example, pro-opiomelanocortin (POMC) is sorted by binding to carboxypeptidase E through the disulfide loop region at its N-terminal region[27]; CgB contains a similar N-terminal disulfide loop region that is both necessary and sufficient for sorting in PC12 cell, although carboxypeptidase E is not responsible for this sorting event[54]. Furthermore, while a narrowly defined aggregation domain of pro-atrial natriuretic factor is responsible for its sorting to the regulated secretory pathway in AtT-20 cells[22], the aggregation of chromogranins at low pH and the presence of calcium is neither necessary nor sufficient for their sorting in PC12 cells[49]. These findings indicate that individual secretory proteins use different sorting mechanisms in a cell type dependent manner.

CgA is a high-capacity, low-affinity, calcium-binding protein, which can aggregate at low pH and high calcium conditions[119]. These characteristics may contribute to CgA's ability to sort peptide hormones and neurotransmitters and package them into secretory granules. Evidence for this includes the fact that CgA can co-aggregate with a

peptide hormone such as parathyroid hormone (PTH), but excludes constitutively secreted proteins such as serum albumin[52].

The domains that are responsible for directing CgA to the regulated secretory pathway were investigated recently. Although CgA contains an N-terminal disulfide loop homologous to that of CgB which acts as the sorting signal of CgB in neuroendocrine PC12 cells [24,80], a CgA mutant with the deletion of the disulfide loop could still be directed to regulated secretory pathway in PC12 cells[54]. But this can be explained by the aggregation of CgA mutant to endogenous CgA molecules. However, in endocrine GH4C1 cells, which lack endogenous CgA, the CgA mutant without disulfide loop can still be sorted to granules. On the other hand, a 90 amino acid deletion at the C-terminal of bovine CgA impairs its aggregation properties and reroutes CgA secretion to the constitutive secretory pathway in GH4C1 cells[28]. This is the first evidence that low pH/calcium induced aggregation is necessary for sorting of CgA to the regulated secretory pathway of endocrine cells. Interestingly, such a mutant has no effect on CgA sorting in neuroendocrine PC12 cells. Thus, there must be a different mechanism and/or domains in the molecule responsible for CgA sorting in neuroendocrine and endocrine cells respectively. Sorting of CgA is therefore dependent on cell type.

2) Regulating the proteolytic processing of the resident prohormones

Chromogranin A has been shown to have an intracellular role in regulating prohormone processing. The multiple pairs of basic amino acids in CgA may serve as competitive substrates for serine protease activity in the Golgi apparatus and secretory granules. There is evidence that CgA can serve as a competitive inhibitor of the *in vitro*

processing of pro-enkephalin[128]. The relative amount of CgA within different cell types may influence the extent of processing in addition to the differential expression of prohormone convertase enzymes[61].

3) Regulating intracellular calcium flow as an IP_3-sensitive intracellular calcium-binding protein

Secretory granules have been identified as the major inositol 1,4,5-trisphosphate (IP_3)-sensitive intracellular calcium store of adrenal medullary chromaffin cells and pancreatic β-cells. CgA is the major component of granules with high capacity and low affinity for calcium binding. It has been shown to interact with several integral membrane proteins of the secretory granules, one of which is the IP_3 receptor/Ca^{2+} channel, at the intragranular pH of 5.5 and dissociate from them at a near physiological pH of 7.5. Recent investigation showed that tetrameric CgA binds to four molecules of the intraluminal loop region of the IP_3 receptor at the conserved N-terminal region[162]. This result is consistent with the proposed anchor role of the near N-terminal region, as well as the function of the free C-terminal region that contributes to dimerization and tetramerization of CgA molecules. From this information, one can imagine that the conformation change of the IP_3 receptor upon binding with IP_3 will be directly and rapidly transmitted to CgA, which in turn will undergo a conformational change that leads to dissociation of some Ca^{2+} for release into the cytoplasm.

1.2.5 CgA gene structure

The CgA gene is a single copy gene in the genome. The human, bovine and mouse genes are comprised of 8 exons with conserved exon-intron boundaries (Fig 1.2)[61] N-

terminal and C-terminal regions show conservation among the genes of different species[114]. Exon I and part of exon II encode the signal peptide. The remaining exon II to exon V collectively encode the highly conserved amino terminal domain (the β-granin sequence). Exon VI encodes a variable domain. Exon VII encodes another variable domain. which contains the pancreastatin and catestatin sequences. Exon VIII encodes the highly conserved carboxyl-terminal domain[61]. The human CgA gene is located on chromosome 14q32, near the immunoglobulin heavy-chain locus[99]. In rat, the CgA gene is located on chromosome 6 and in mouse on chromosome 12[133].

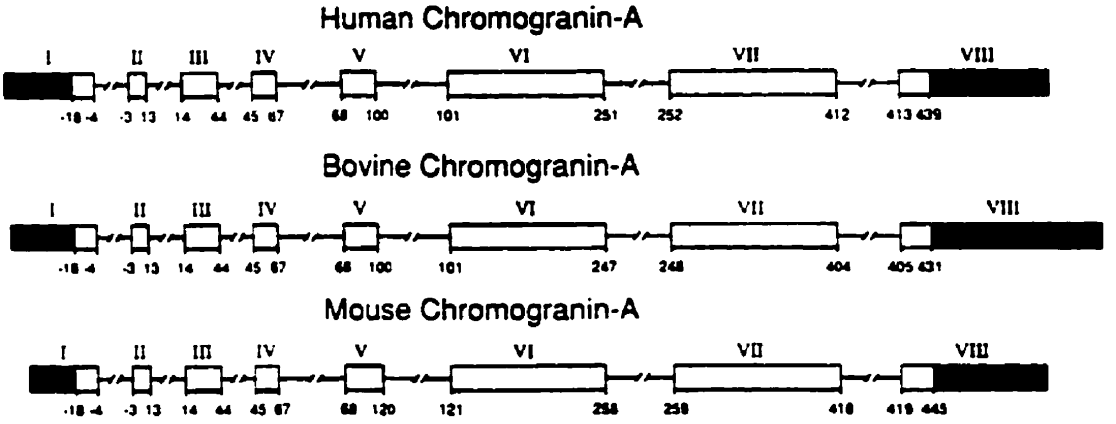

Fig 1.2 CgA gene structure (taken from ref.97)

The approximately 200bp proximal promoter of the CgA gene is well conserved[61]. Human and bovine CgA gene share around 80% identity whereas human and mouse gene have 50% similarity between each other[97]. In each of the promoters, there is a cyclic AMP response element (CRE) just upstream of the TATA box. The CRE was proven to be both necessary and sufficient for the neuroendocrine specific expression of the CgA gene[97]. In addition, consensus sites for the transcription factors, Sp-1 and AP-1 were found in all three genes. The importance of these elements remains to be clarified.

1.2.6 Distribution and ontogeny

- **Distribution**

The CgA molecule was present in the early stage of evolution of the animal kingdom[158]. CgA immunoreativity has been detected in *Drosophila*, fish and single-celled eukaryotic organisms such as the protozoan, *Paramecium tetraurelia*[61]. The subcellular localization of CgA is in the secretory granules of mammalian neuroendocrine and endocrine cells or their submammalian counterparts, such as the trichocysts of *Paramecium*. This conservation suggests a functional importance for CgA throughout evolution[61]. An effort has been make to search CgA sequence homology in yeast and C. elegans data base, but no significant homologous sequences found.

CgA was found originally in chromaffin granules of the adrenal medulla and dense-core vesicles of sympathetic nerves. Later, its expression was shown to be much more widespread and it is found in most neuroendocrine and endocrine tissues and some exocrine tissues.

The expression level of CgA varies widely from one type of neuroendocrine cell to another[143]. The largest amount is found in the adrenal medulla, with CgA representing almost half of the total soluble content of the secretory granule. The granules of the cells in pituitary contain 25% of the amount in adrenal medulla, the pancreas and stomach plus intestine contain 5% each, whereas all other endocrine glands represent less than 1%. In the pituitary, CgA immunostaining is mainly in LH and FSH-containing cells, and little is present in ACTH- or prolactin- containing cells. The thyroid C cell, insulin- and glucagon-containing cells and exocrine cells in the pancreas have CgA expression. CgA

is also present in scattered neuroendocrine cells in nonendocrine tissues, such as breast, lung, trachea, prostate, uterus, gastro-intestinal tract and the Merkel cells of the skin. CgA-positive cells are found in the spleen, thymus and lymph nodes of the immune system[2]. However, some endocrine cells that are not of neuroendocrine origin lack CgA. These cells include the follicular cells of the thyroid gland and the steroid hormone-secreting cells of the adrenal cortex and the gonads.

CgA is widely distributed in nervous tissue. In the central nervous system, CgA expression was found in the brain medulla and cerebral cortex, and the posterior hypophysis. In the peripheral nervous system, CgA is present in different neuron subpopulations and in the peripheral sympathetic neurons as well. Secretory vesicles of lumbar motor perikarya and cholinergic nerve terminals in skeletal muscle also contain CgA immunoreactivity[150].

- **Ontogeny**

The developmental expression of CgA has been examined in a small number of tissues in limited species. In the developing rat adrenal medulla, CgA shows a dramatic rise in expression in the secretory granules from fetal day 14[9]. The first ECL cells of rat appear at embryonic day 17[75]. Immunoreactive histamine and chromogranin A appear one day later. At 15 days of postnatal age, the level of CgA expression is similar to that of the adult[140]. In human, CgA mRNA can be detected as early as 6 to 8 weeks of gestation in chromaffin tissue and progressively increases during the first year of life[57]. In the endocrine cells of the human intestinal tract, CgA immunoreactivity appears together with the resident hormones at eight weeks of gestation. However, in the avian,

for example, the chicken, CgA first appears in the intestinal tract at various embryonic ages from day 10 in the cloaca to day 16 in the distal ileum and colon[123].

1.2.7 Regulation of biosynthesis and secretion

It has been observed that the biosynthesis of CgA is regulated by several steroid hormones and intracellular messengers.

Steroid hormones affect CgA synthesis in a cell type specific manner and opposite direction to the resident hormones. 1,25-dihydroxycholecalciferol [1,25-$(OH)_2D_3$] enhance CgA synthesis in the parathyroid but decrease that of PTH[88,95,137]; glucocorticoid stimulate CgA expression but suppress POMC[151]; However, in the case of estrogen, it decrease CgA expression in gonadotroph cells but increases LH and FSH secretion[44]. Thus, this phenomenon is complementary to the extracellular function of CgA in that the steroid hormones increase or decrease the level of CgA expression to affect the secretion of the biologically active CgA derived peptides, which lead to the decreased or increased secretion of the resident hormone respectively. GRE, ERE and VDRE like sequences were found at the 5'-flanking region of CgA gene[94].

CgA biosynthesis is also regulated by several intracellular messenger systems, including intracellular calcium and cAMP of the PKC and PKA signaling pathways[61]. In chromaffin cells, short-term treatment with phorbol esters elevates intracellular calcium level by the PKC pathway and stimulates CgA protein biosynthesis. There is a consensus cyclic AMP response element (CRE) in the promoter of the CgA gene. It was shown to be important for the neuroendocrine specific expression of the CgA gene. However, there has not been any evidence that increase intracellular cAMP levels can

obviously stimulate CgA biosynthesis, but has some modest increase of CgA mRNA level in chromaffin cells[69] and medullary thyroid carcinoma cells[98] by forskolin treatment.

1.2.8 CgA as a neuroendocrine tumor marker

Since CgA is widely distributed in neuroendocrine tissues and co-secreted with neuropeptide hormones, it is suitable as an immunohistochemical and serum marker for neuroendocrine tumors.

The serum level of CgA is elevated in patients with various neuroendocrine tumors, and the elevated level is correlated with tumor burden[106]. The highest levels recorded in cases of metastatic carcinoid tumor were with extreme elevations up to 1000 times the upper limit of normal[108]. However, serum concentrations of CgA are rarely elevated in subjects with small neuroendocrine tumors, such as insulinomas, paragangliomas and pituitary adenomas[105].

As many neuroendocrine tissues contribute to the circulating concentrations of CgA, its plasma pool is substantially greater than that of most peptide hormones. Thus, it is more difficult to acutely increase the level above the physiological background. Moreover, the correlation between serum levels of CgA and of the specific peptide hormones of the tumors is poor. Therefore, the usefulness of the serum CgA level as a marker for neuroendocrine tumors is limited in sensitivity and specificity. However, it is still useful in the following cases[107]: (1) 'non-functioning' neuroendocrine tumors, which do not secrete a detectable marker, such as non-functioning pituitary adenomas and silent neuroendocrine tumors of gastro-enteroendocrine system; (2) tumors with unstable markers or markers that are inconvenient for clinical use. For example, the

catecholamine levels and 24-h urine collections for catecholamines and their degradation products in patients with phaeochromocytomas. (3)Immunohistochemical measurement of CgA to differentiate the tumors of neural crest or other origin.

Other members of the granin family are also suitable for neuroendocrine tumor detection. With a combination of CgA and CgB antisera, the detection sensitivity rises to about ten times of that of the CgA antiserum only, both in immunohistochemical and serum detection[37]. Therefore, using a combination of antibodies against several chromogranins and their cleavage products might offer a more powerful tool for neuroendocrine tumor detection. Another future application of CgA as a tumor marker is its use in *in vivo* imaging techniques of neuroendocrine tumors. It has been shown that CgA and CgB molecules are tightly bound to the interior side of the secretory vesicle membrane. After exocytosis, the vesicle membrane becomes part of the plasma membrane, exposing the membrane bound granin molecules to the extracellular milieu. This provides the opportunity to detect the granin secreting neuroendocrine tumors by using the immunoscintigraphy with antibodies against CgA or CgB. Preliminary data in the use of a three-step pre-targeted immunoscintigraphy using anti-CgA monoclonal antibody in patients with several neuroendocrine tumors shows a higher diagnostic accuracy than conventional imaging techniques[89].

Chapter 2

Transgenic approaches for gene function and regulation study

2.1 Transgenic technology

2.1.1 Introduction

Recent advances in transgenic technology have provided an opportunity to explore the function and regulation of a single gene in a complex physiological system *in vivo*[59]. In particular, the mouse has been established as the model of choice for mammalian genetic and developmental analysis. Since all the studies of CgA gene regulation and function were obtained through *in vitro* approaches, the project outlined in this thesis is based mainly on the investigation of CgA gene function and regulation *in vivo* through transgenic approaches. In this chapter, I would like to introduce some recent progress in gene function and regulation analysis using transgenic technology in the mouse.

2.1.2 Development of transgenic technology

In the early 70's, Jaenisch & Mintz first demonstrated that SV40 viral DNA could be introduced into the mouse embryo and detected in various tissues of the animals that developed from these embryos[72], thus establishing the concept of a transgenic animal. Subsequently, the Moloney murine leukemia retrovirus was used as the vector to deliver the transgene into mouse embryos[73]. However, due to the limitation of the retroviral genome capacity, this method was replaced by direct pronuclear microinjection of purified DNA into the male pronuclei of fertilized one-cell stage mouse embryos[51].

Because of the efficiency and consistency of this method, it has become a standard procedure to produce transgenic animals.

However, the approach of pronuclear microinjection to create a transgenic animal suffers from two major problems: uncontrollable integration sites and copy numbers. Random integration and multiple copies of transgenes lead to unregulated expression and can cause side effects, such as variable expression levels, and disruption of other important genes. It has been reported that 5-10% of transgene integration events result in a mutant phenotype as a consequence of disruptions, deletions, or translocations.

The problem of random transgene integration was circumvented by the development of a homologous recombination-based gene-targeting procedure. The transgene or mutated DNA sequence can be delivered to a specific locus by using a recombinant DNA vector (the targeting vector) that allows homologous recombination to occur between the targeting vector and its endogenous counterpart at the homologous regions, thereby altering one of the normal alleles in the genome. This recombination event takes place in the embryonic stem cells (ES cells) in culture after transfection with the targeting vector. The ES cells are derived from the inner cell mass of the blastocyst[39], and maintain their pleuripotential characteristic in culture with the presence of a differentiation inhibiting factor such as leukemia inhibitory factor (LIF). The recombinant ES cells containing the desired genetic alteration can aggregate with eight-cell stage morulae or be injected into the blastocyst and contribute to the development of all kinds of tissues including germ cells of the chimeric mouse[11]. Therefore, the genetic composition of the ES cells could be passed on to the offspring. This results in germ-line chimerism, or heterozygous offspring. By crossing the

heterozygotes, homozygotes might be obtained. Thus the function and regulation of a single gene can be assessed in the complex *in vivo* situation.

2.2 Promoter activity study *in vivo*

2.2.1 Reporter systems

The transgenic approach enables the vigorous examination of gene regulation in different tissues and at different developmental times. The typical approach for this purpose is utilizing sequence cassettes that contain a regulatory sequence of a gene hooked up to a poly (A)-tailed reporter gene encoding a product whose biological or physiological activity can be detected easily, such as chloramphenicol acetyl transferase (CAT), β-galactosidase (lacZ), luciferase (Luc), or green fluorescent protein (GFP). Two common features are shared by all of these reporters: the biological activity is generally lacking in mammals, and the activity can be detected by a simple assay[155].

- **β-gal reporter system**

The most elegant result can be obtained by using the lacZ reporter to identify temporal and spatial regulatory DNA elements in the transgenic mice. LacZ activity can be determined by in situ or quantitative analysis[124]. Therefore, it is possible to follow the expression patterns of the promoter activity in the developing embryos in a three–dimensional manner under a stereoscope. Detailed expression patterns at a cellular level can be examined on selected tissue sections located in a specific region[154]. CAT and Luc reporters are more sensitive. However, due to technical limitations, these two systems are not as convenient as LacZ system in the in situ analysis assay. They are

preferred when the sensitivity of detection is a concern. Therefore, the choice of reporter system depends primarily on the purpose of the experiment.

- **GFP reporter system**

The recently developed GFP system contains a green fluorescent protein from the jellyfish *Aequorea victoria* that absorbs blue or UV light and emits detectable green light in a species-independent fashion, either in living or fixed tissues[70]. This process occurs without requirement for a substrate or cofactor. Thus, it allows direct imaging of fluorescent gene product in living cells and allows an examination of changes in gene expression in living tissue. Okabe et al. generated a transgenic mouse line with a mutated "enhanced" GFP cDNA under the control of a chicken ß- actin promoter and cytomegalovirus enhancer. Under excitation light, all tissues from these transgenic mice were green[104]. GFP is not toxic to the cells, and does not interfere with cell growth and/or function. So far it is reported that GFP does not alter the subcellular localization of proteins and is distributed both in the nucleus and cytoplasm[25]. Due to its small size, GFP diffuses throughout the cytoplasm of extensively branched cells like neurons and glia[12]. Thus, gene expression analysis using the GFP system offers distinct advantages over other reporter gene systems.

2.2.2 Conventional approach

The conventional approach to study promoter activity *in vivo* is to generate transgenic mice by pronuclear microinjection of the purified DNA construct containing the promoter-reporter cassette into male pronuclei of fertilized oocytes. This method has proven to be effective in achieving exogenous gene expression to address promoter activity. However, the transgene is integrated in a random manner and with unpredictable copy numbers, which may cause profound side effects that affect the accuracy of the detection of the promoter activity.

2.2.3 Introducing a single copy of transgene cassette at a chosen site

To avoid the disadvantages of the conventional approach, a gene targeting technique is used to introduce a single copy of the promoter-reporter cassette to a special chosen site in the mouse genome. This controls the copy number and eliminates the chromosomal location effect of the reporter gene expression. Therefore, the targeted transgenes should provide a more efficient and informative means of detecting the gene regulation elements. The chosen locus of transgene integration is usually a housekeeping gene locus, since at such a region, the chromosome construct is always in a open status and the inserted promoter activity is not likely to be restricted in its expression by unfavorable chromatin configuration. The most popular chosen site is the hypoxanthine phosphoribosyltransferase (Hprt) locus[15], because of the availability of a targeting vector capable of mediating a highly efficient, directly selectable homologous recombination event at this locus. In additional, a special karyotypically male ES cell line

containing a deletion at the X-linked Hprt gene site is used for generating the recombinant ES clones[64]. Therefore, the selection can be conveniently carried on in HAT medium directly. (Fig 2.1)

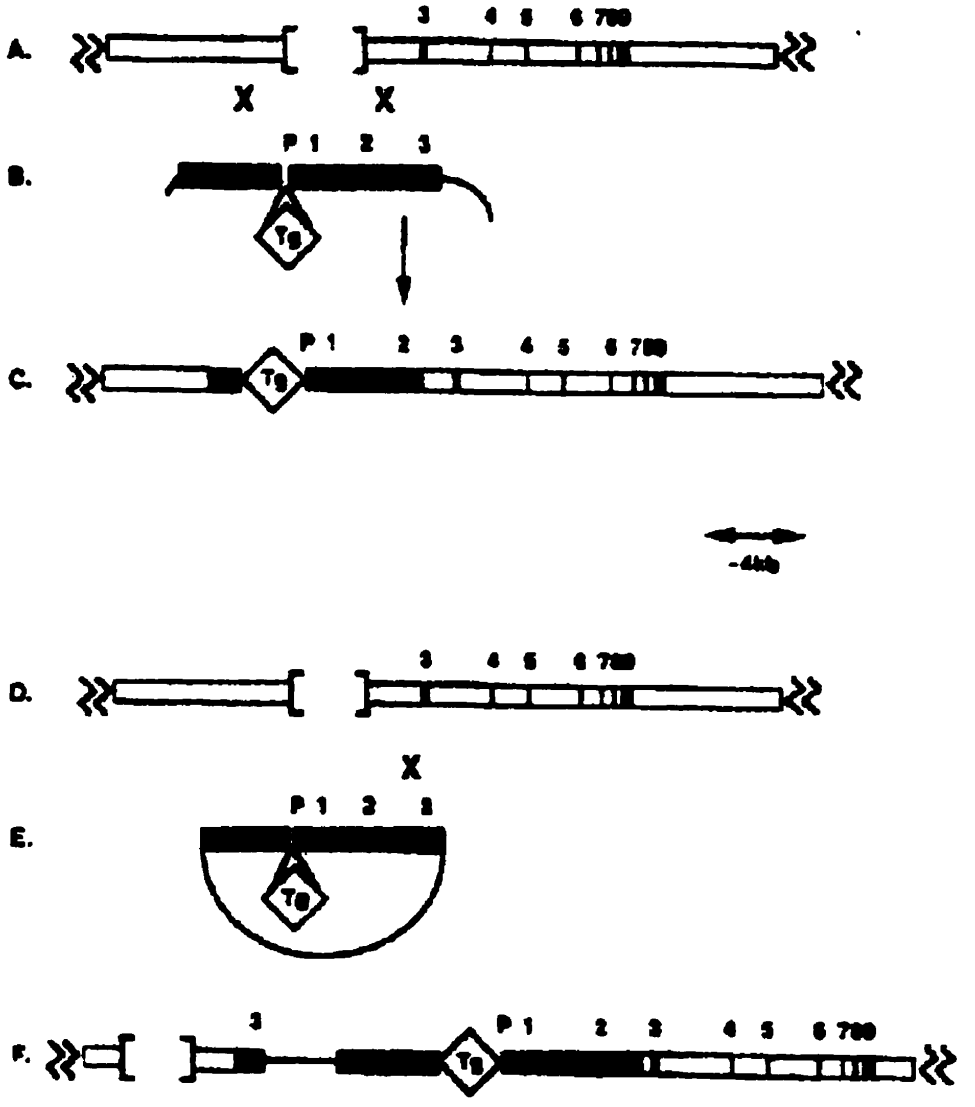

Fig 2.1 Mechanism of introducing the transgene into the Hprt gene locus (taken from ref.15) **A-C:** a replacement event mediated by two crossover(X and X); **D-F:** an insertion event mediated by a single crossover(X). **A&D:** the endogenous Hprt locus; **B&E:** targeting construct; **C&F:** resulting targeted loci modified by homologous recombination restoring Hprt function and introducing the transgenic sequences. **Unshaded box:** genomic locus; **Brackets:** deletion; **Shaded box:** targeting construct; **Tg:** transgenic sequences; **Thin solid line:** plasmid sequences; **Numbers:** representing exons.

2.3 Gene functional analysis *in vivo*

To study a single gene function, the gene can be over-expressed, disrupted or mutated in the genome of the transgenic mouse. Through the phenotype changes, the exact function of the gene can be obtained.

2.3.1 Gain-of-function analysis

Over-expression or ectopic expression of a gene of interest or its mutated form may shed light on its specific function. This can be obtained by pronuclear injection of the transgene construct into the nuclei of fertilized mouse embryos at the one-cell stage[155]. In transgenic mice, a transgene construct with its own promoter will be expressed in all tissues where the gene is normally expressed, while it can be targeted to selected tissues by an appropriate promoter. The normal function of a given gene can be altered by over-expressing a mutated gene encoding a dominant negative protein.

2.3.2 Loss-of-function analysis

Another approach to study gene function is to inactivate the target gene in the genome leading to a gene knockout.

- **Conventional gene knockout**

Conventional gene knockout strategy is to direct homologous recombination to a specific target gene in embryonic stem (ES) cells[22]. This procedure is mediated by a target vector containing two homologous sequences of the target gene flanking the target

region of the gene. After the recombination, the target region is substituted by the vector sequence containing the selectable marker.(Fig 2.2) The ES cells containing the modified genomic sequence are then injected into the morulae or blastocyte stage embryos, which will developed into the chimeric mouse. The offspring of the chimera may harbor the mutation in a heterozygous way. The homozygous mouse can be obtained by breeding the heterozygotes.

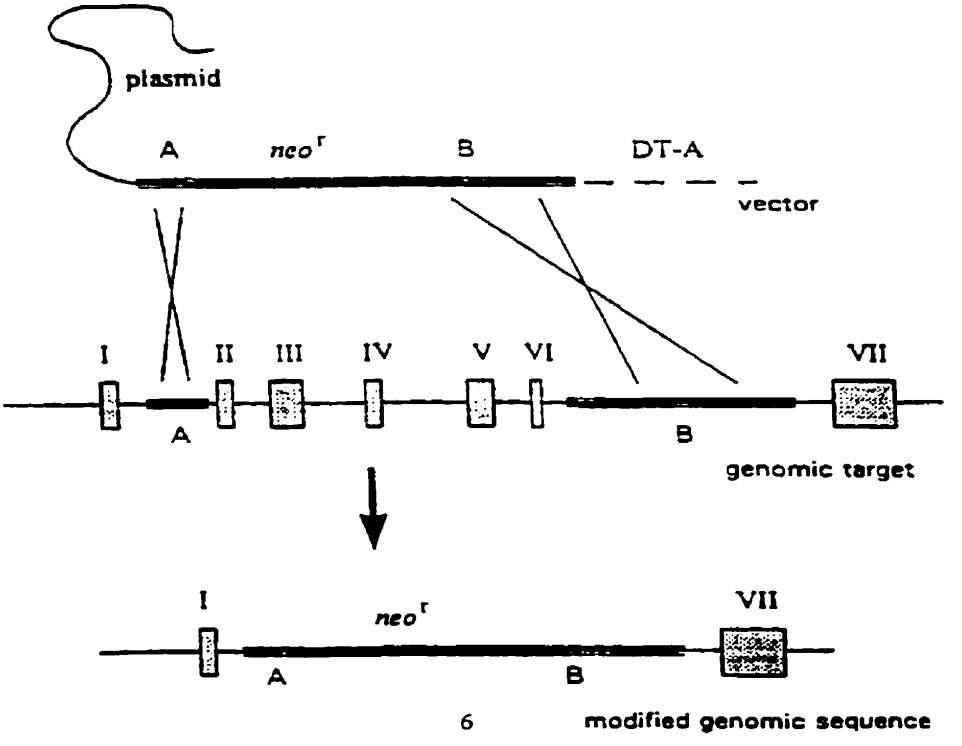

Fig 2.2 Targeted homologous recombination (taken from ref.118) Linearized targeting vector contains: **A&B:** sequences identical to the targeting sequences on genomic sequences; **neo^r**: positive seletable marker; **DT-A:** negative selectable marker; **I-VII:** exons of the targeted gene.

- **Tissue-specific inducible gene knockout**

However, there are a number of limitations of the conventional knockout strategy. The homozygous disruption of a gene involved in complex pleiotropic effects in individual development may be lethal and thus does not permit the development of an *in vivo* model system in which gene inactivation is restricted to a defined subset of cells. To overcome these limitations, gene knockout strategy was refined by targeted deletion of a gene from a specific cell type at a specific stage of development. This can be achieved by site-specific heterologous recombination systems. There are two such systems established so far. One is the Cre/lox system from bacteriophage[125]; the other is the Flp/Frt system from yeast[35]. The Cre/lox system appears to be more effective and widely used.

Cre is the name of the gene that encodes the bacteriophage P1 site-specific DNA recombinase. It recognizes a specific 34-base pair DNA sequence called lox. Cre is able to bring pairs of lox sites into proximity, and catalyzes breakage and religation between them. Where two lox sites occur in the same orientation, this results in looping out and deletion of the sequence between the lox sites. When lox sites occur in opposite directions, it causes inversion.(Fig.2.3) In the Flp/Frt system, Flp is the yeast version of Cre recombinase, and Frt is the Flp's version of lox.

For example, the gene knockout approach was applied to directly inactivate the insulin receptor gene in mice[26], as well as conditionally knockout the gene in the muscle[122]. IR deficiency in IR -/- mice led to a number of major metabolic alteration soon after suckling, including severe diabetes mellitus with ketoacidosis, hyperglycaemia, hyperinsulinaemia, reduced hepatic glycogen content and marked

postnatal growth retardation as well as skeletal-muscle hypotrophy. All of these disorders led to the death of IR-deficient pups within 1 week after birth. Due to the lethal nature of IR disruption, a conditional knockout model needed to be established to study insulin function in muscle. Therefore, the muscle-specific IR gene knockout was obtained by using the Cre/lox system. Mice harboring an altered IR gene with exon 4 flanked by lox sites were generated and bred with transgenic mice expressing the Cre recombinase gene driven by the muscle creatine kinase gene promoter/enhancer. The resultant muscle-specific IR knockout offspring showed impaired insulin-stimulated glucose uptake in skeletal muscle. Yet, these animals remained physiologically normal, normoglycaemic at birth and showed normal growth and development.

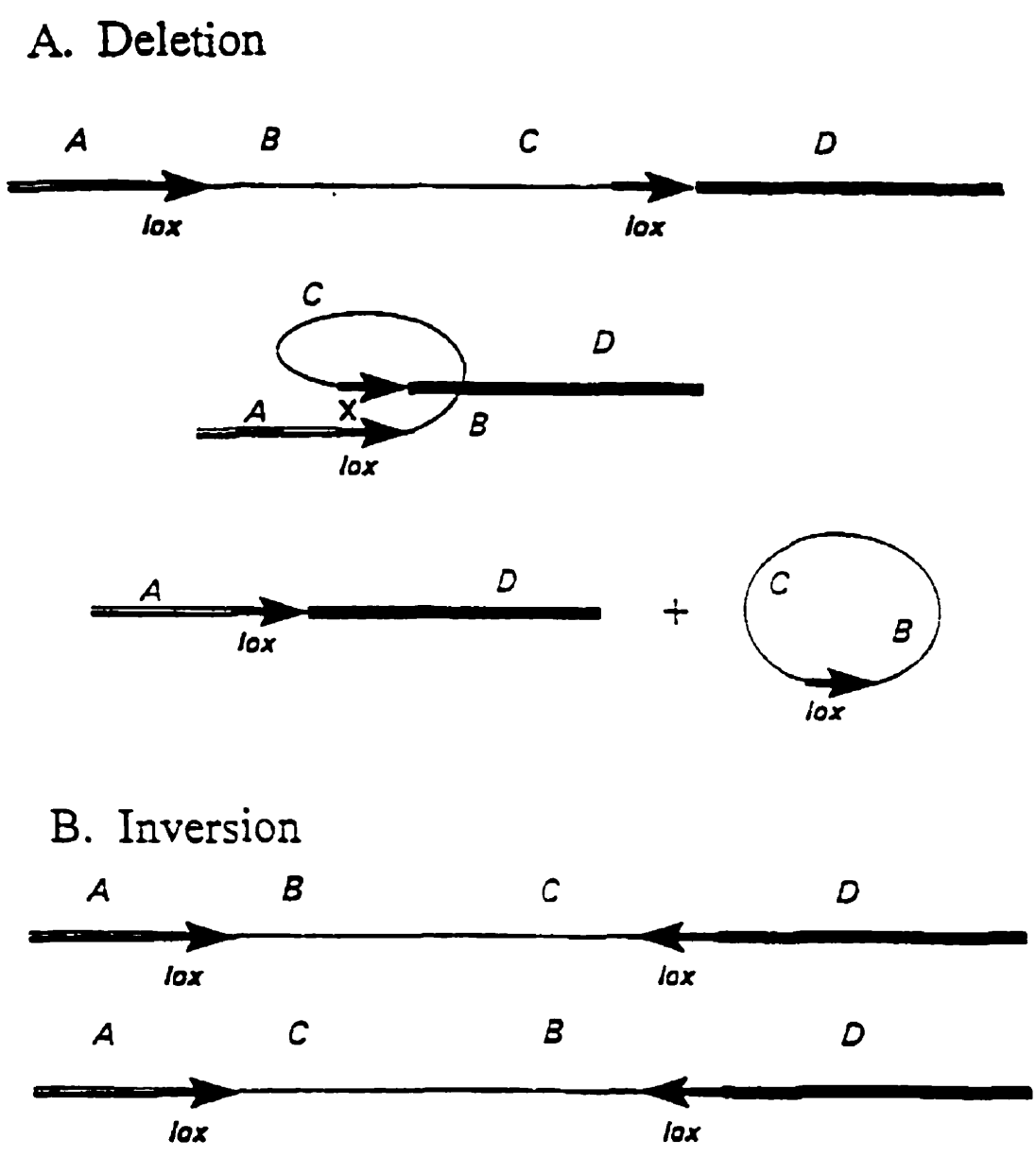

Fig 2.3 Mechanism of Cre/lox site-specific recombination (from reference 118) **A:** Cre recombinase catalyzes deletions between two 34-basepair lox site when they are integrated in the same orientation in a segment of DNA, producing both a deleted gene in the genome and a circle, each retaining a single lox site. **B:** targeted gene inversion is produced when lox sites are present in opposite orientations.

Besides the possibility of introducing somatic mutations in a given tissue, current mouse gene targeting technology is able to do this at a chosen time by inducible knockout strategies. When the Cre gene is under the control of inducible expression systems, such as the tetracycline-inducible system[46] and ecdysone control system[104], the target gene can be deleted upon administration of the exogenous inducible agents.

The temporal and spatial expression patterns of the Cre gene are dependent upon the nature of the promoters employed. Therefore, the control of timing and tissue-specific expression of the Cre gene can be achieved by application of specific and inducible promoters. Several inducible systems have been established, such as the heavy metal inducible promoter system[112] and the estrogen inducible system[13]. However, the physiologic (estrogen) or toxic (heavy metal ion) effects of the inducer and high basal transcriptional activity from the promoters limit their utility. Yet the development of a regulatory circuit based on the tetracycline-resistance operon *tet* from E.coli transposon Tn10 opened a new approach for controlling transgene expression. Figure 2.4 shows how the tet-system works. Transgene-1 utilizes a tissue-specific promoter-A to direct the expression of the transactivator TA for the tetracycline operon (tet-P). Transgene-2 encodes the cDNA of interest under the control of Tet-op. In the absence of Tet, TA produced in specific tissues is active, thereby activating expression of tg-2 in pr-specific tissue. In the presence of Tet, TA remains silent, and no tg-2 is expressed. The tetracycline regulatory system can provide temporal control of transgene expression under experimental conditions. It should be useful for experiments designed to address certain biological questions such as further definition of the roles played by growth modulators, oncoproteins and other proteins participating in developmental processes.

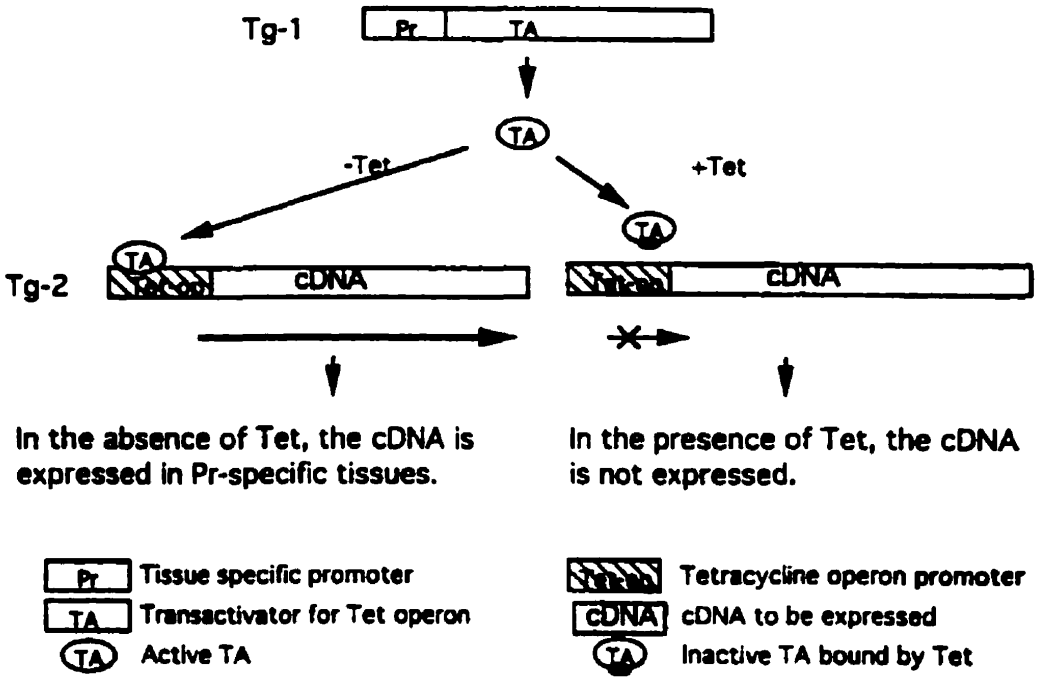

Fig 2.4 The mechanism of tetracycline-inducible binary system (taken from ref. 155) Transgene-1 (Tg-1) encodes the transactivator (TA) for the tetracycline operon. and under the control of a tissue-specific promoter-A (Pr). Transgene-2 (Tg-2) encodes the cDNA of interest under the control of Tet-op. In the absence of Tet, TA produced in Pr-specific tissues is active. thereby activating expression of Tg-2 in Pr-specific tissue. In the presence of Tet, TA remains silent. and no Tg-2 is expressed.

Another inducible gene targeting approach was achieved by the engineering of the Cre recombinase itself. Chambon's group fused the Cre gene to a mutated ligand-binding domain of the human estrogen receptor resulting in a tamoxifen-dependent chimeric Cre recombinase[41], which is activated by tamoxifen, but not by estradiol. Transgenic mice were generated expressing Cre-ER^T under the control of a cytomegalovirus promoter (Fig 2.5). PCR and southern blot analysis shows that the excision of a chromosomally integrated gene flanked by lox sites can be induced by administration of tamoxifen to these mice, whereas no excision could be detected in untreated animals.

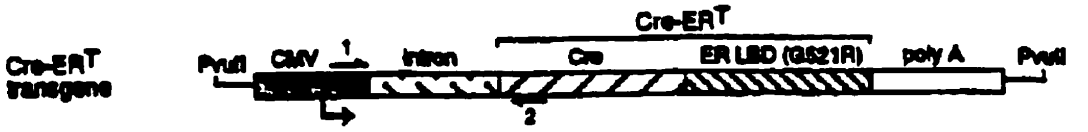

Fig 2.5. The structure of the Cre-ER^T transgene (taken from ref. 41)
The Cre-ER^T transgene, consisting of a Cre recombinase gene fused to the cDNA coding for the G521R mutant of the human ER ligand binding domain, is driven by the human CMV promoter, with a rabbit β-globin intron upstream and a simian virus 40 polyadenylylation signal downstream.

Such an inducible system could help in certain cases to distinguish between anomalies related to a mixed genetic background and those due to mutation of the target gene.

2.2.2.3 Introduction of mutations into the target gene

When it comes to study the effect of a specific variation of a gene structure on the gene function, the knockout strategy has proven to be invaluable. Therefore, instead of knocking genes out, a knock in method is needed to introduce desired sequence alterations into the genome[14]. The method consists of a two-step strategy. The first step is a homologous recombination with a single reciprocal recombination at the target gene resulting in a duplication of the homologous region. Successfully targeted cells are obtained by selection for neomycin resistance. In the second step, ES cells are selected that contain a spontaneous intrachromosomal recombination event leading to a loss of the endogenous locus and the thymidine kinase gene. Successfully reverted ES cells are used to generate transgenic mice carrying the mutations. An excellent example is the murine

model generated by introduction of the Arg-403 Gln point mutation into the α-myocyte heavy chain (α-MHC) gene in ES cells [59]. (Fig 2.6)

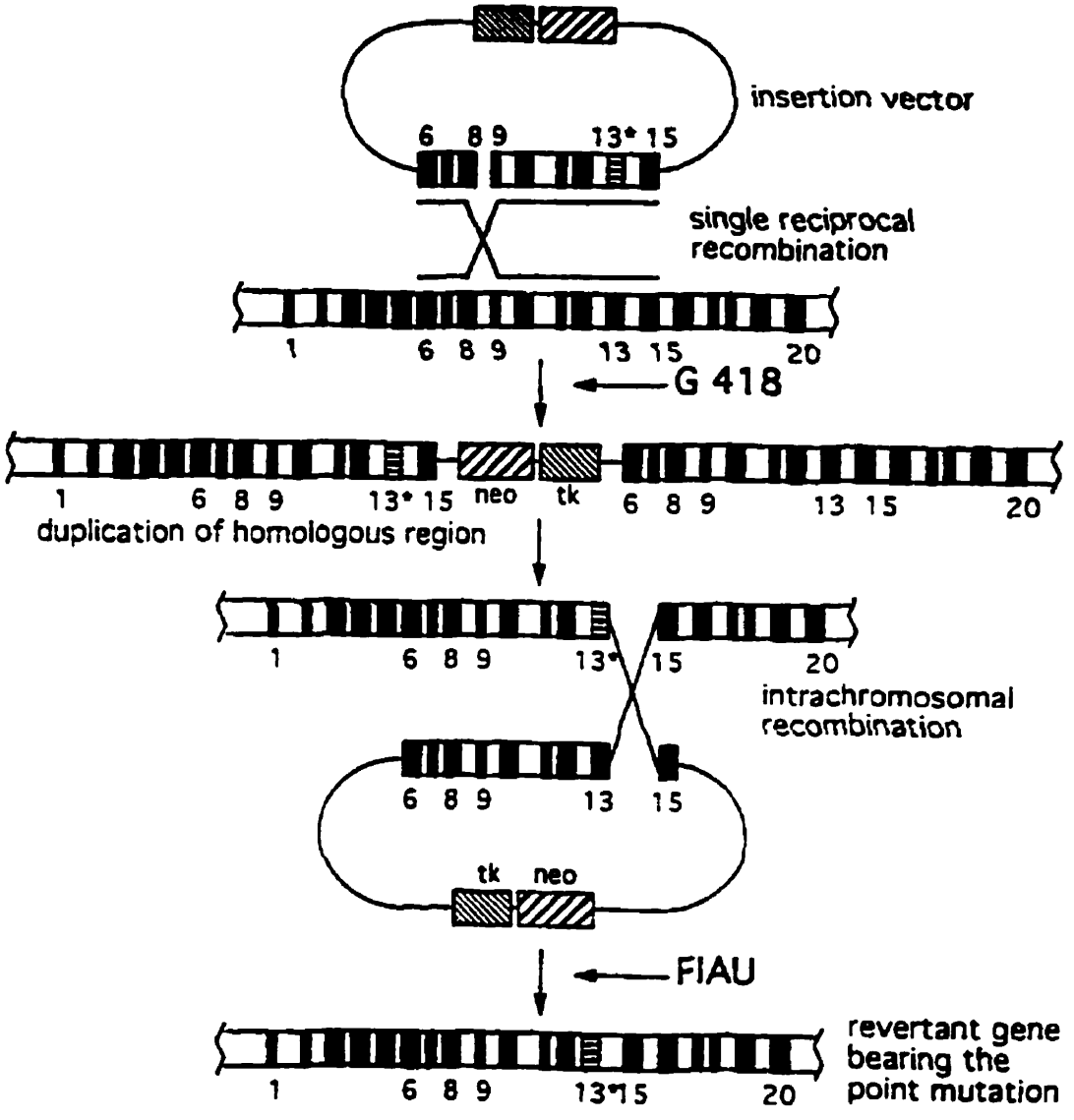

Fig 2.6 Introducing point mutation into the α-MHC gene (taken from ref. 59) The insertion vector contains exon 6 to 15 and two selectable marker genes, neo and tk. Exon 13 bears a single point mutation. The first step is a single reciprocal recombination leading to duplication of the homologous region. Successfully targeted cells survive selection for G418 resistance. The second step consists of a spontaneous intrachromosomal recombination event leading to loss of the endogenous locus. These cells can be selected by resistance to FIAU (fialuridine).

A similar scheme introducing defined genetic alterations can be envisioned using the Cre/lox system[56]. The targeting vector contains a region homologous to the target gene but with desired sequence alteration and interrupted by a selectable marker flanked by lox sites. Homologous recombination introduces the desired mutation with the selectable marker to the target site. The marker can then be removed by transient expression of the Cre gene in ES cells.

This approach was used to study the function of two structurally related genes: En-1 and En-2, which are involved in brain development[58]. Both genes become active early in the same regions of the developing brain, although En-1 is turned on 8 to 10 hours before En-2. En-1 null mice had serious abnormalities, including a deleted midbrain and cerebellum that caused them to die shortly after birth. In contrast, En-2 knockouts had only minor problems. There are two hypotheses that could explain the difference between the brain phenotypes of the two mutant genes. First, these two genes have different functions in spite of their similar structure and site of expression. Second, they have equivalent functions but the divergence in the temporal expression patterns results in the compensation for En-2 by En-1, which is expressed earlier.

In order to distinguish between these two hypotheses, Joyner and her colleagues spliced the En-2 gene into the DNA used to knockout En-1, As a result, En-2 was inserted into En-1 gene, simultaneously inactivating it, while En-2 itself was hooked up to En-1's regulatory sequences. (Fig 2.7) Fig 2.7 shows how this was achieved. The PGK/Neo selectable marker cassette flanked by lox sites was then removed by Cre recombinase. With En-2 now expressed with the exact pattern of En-1, the transgenic

mice exhibited a normal phenotype. Thus, it proves that from a biological point of view, En-1 and En-2 are functionally equivalent.

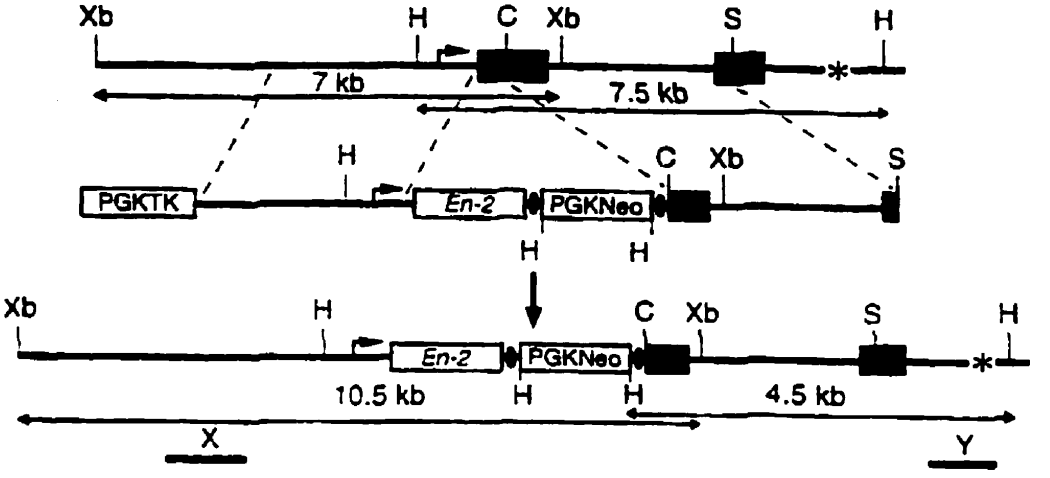

Fig 2.7 Targeting strategy for substitution of En-2 for En-1. (taken from ref. 58) **Top line**: the schematic of the En-1 genomic locus; **horizontal arrow**: En-1 promoter; **black boxes**: En-1 coding sequences. **Middle line**: linearized En-2 KI vector with the PGKNeo marker flanked by lox sites (represented by black ellipses). **Dashed lines**: the regions of homology between locus and targeting vector. **Bottom line**: structure of targeted locus. After homologous recombination, En-2 hooks up to En-1 promoter and inactivates En-1 by eliminating the of 5' coding region of En-1 gene. By Cre recombinase activity, the PGKNeo resistant marker can be deleted to generate the perfect replacement of En-1 by En-2.

Part II

Experimental section

Chapter 3

Construction of constructs for studying neuroendocrine specific regulation of CgA gene *in vivo*

3.1 *In vitro* study of neuroendocrine specific expression of CgA gene

3.1.1 Introduction

Chromogranin A is expressed throughout the scattered neuroendocrine system. However, it is still unknown what factors govern the activity of the CgA gene to yield such a widespread yet neuroendocrine-selective pattern of expression. At present, the study of CgA regulation is focused on addressing the neuroendocrine specificity of its expression.

The luciferase reporter system was used to analyze a series of mouse CgA gene promoter deletions in neuroendocrine and nonendocrine cell lines[160]. The results established both positive and negative transcription regulation domains, which included a distal positive (-4.8/2.2 kb) domain as well as a proximal negative domain (-258/-181 bp) and a positive domain (-147/-61bp). There is a minimal neuroendocrine-specific element between –77 and –61, which contains a cyclic AMP response element (CRE, -66/-61).

An elaborate investigation of mutations in the CRE of the human CgA promoter in our laboratory demonstrated that the CRE is both necessary and sufficient to direct neuroendocrine-specific expression of CgA gene in many cell lines[21]. In the human endocrine BEN cell, CgA activation can be induced by dibutyryl cAMP and inhibited by the PKA-inhibitor H89, thus the signaling pathway was shown to be mediated by PKA. However, the studies also showed that the expression of CRE-binding protein (CREB) and CREB-binding protein (CBP) are not different in many neuroendocrine and nonendocrine cells. In addition, immunoblot analysis indicated that the phosphorylation levels of CREB in neuroendocrine and nonendocrine cell lines are not different.

Therefore, the neuroendocrine-specific transcription of the hCgA gene mediated by CRE does not function at the level of total CREB or phosphorylated CREB or its cointegrator CBP. The specificity may be achieved by a PKA-responsive CRE-binding protein other than CREB expressed specifically in neuroendocrine cells, expression of a repressor protein that binds to CREB in nonendocrine cells, or may lie downstream of CREB which couples transactivators to the basal transcriptional machinery.

The signaling pathway of activation of CgA gene expression in the rat chromaffin cell (PC12) by nerve growth factor (NGF) was studied[91]. NGF can induce a 7.6 fold increase in CgA expression, and CRE is both necessary and sufficient for NGF-induced CgA expression. It was indicated that TrkA and mitogen-activated protein (MAP) kinases transduce the signal from the NGF receptor to the nucleus and that CREB is involved in the signaling pathway. The expression of the dominant negative mutants of Sos, Ha-Ras, Raf1, MAP kinase, ribosomal protein S6 serine kinase II (CREB kinase) or CREB (KCREB) each inhibited the NGF-induced increase in CgA promoter activity. This makes the role of CREB in CgA gene expression complicated. There may be other factors that form complexes with CREB to affect neuroendocrine CgA expression. Further studies on the transcription cofactors and signaling molecules are needed to address this question.

3.1.2 **DNA response elements in the CgA 5'flanking region**

There are several consensus hormone response elements in the 5' flanking region of the CgA gene. Table 3.1 shows these consensus sequences in human CgA 5' flanking region and compares them to that of some well-defined transcription factor elements. For

example, the CRE is at –51, TATA box is at –28, and some Sp-1 and AP-1, AP-2 elements are in the proximal promoter region, supporting the results that PKC and PKA are involved in the regulation of CgA gene expression *in vitro*.

CgA synthesis can be modulated by several steroid hormones, whose genomic actions are mediated by nuclear receptors. There are both estrogen- and glucocorticoid-like response elements located in the 5' flanking region of the bovine and human CgA gene. $1,25(OH)_2D_3$ can enhance CgA gene transcription in cultured parathyroid cells[95,96]. There are two VDRE-like motifs localized at –663 and –126 relative to the transcription start site.

Table 3.1 Putative Cis-Acting Response Elements* in the 5' Flanking Region of the Human Chromogranin A Gene

[Adapted from Mouland (1993)]

Element (position[1])	Consensus Sequence[2]	Sequence in hCgA Gene
TATA (-28)	TATATAA	TATATAA
CRE (-51)	TGACGTCA	TGACGTCA
Sp-1 (-63)	GGCGGG	GGCGGG
VDRE-like (-126)	$CGGTGAN_6CCCTGA$	$GGGTGAN_6AGGTGG$
(-663)	$TCACCCN_6TCACCC$	$TCACCCN_6TGAACC$
VDRE		
1/2 sites(-891)	$^{A}_{G}GGTCA$	AGGTGA
(-846)	$^{A}_{G}GTCA$	AGGTGA
ERE(-1547)	GGTCACAGTGACC	GGTCAGGCTGGTC
GRE (-893)	$GGT^{T}_{A}CA^{C}_{A}{}^{T}_{A}NTGTTCT$	GGTGAGAGcTGTTCT
AP-1 (-421)	$TGA^{G}_{C}TCA$	TGAGTGA
AP-2 (-376)	$G^{G}_{C}{}^{G}_{C}{}^{T}_{A}GCGCC$	CCCCAGGC
IR[3] (-1738)		GTTTGAGAcAGAGTTTC
(-1919)		GCATTAAgtccAATTATG
PR[4] (-79)		$[(GGG)AGN]_3$[5]

• To be confirmed by functional studies.

1 Number indicates beginning of element relative to transcriptional initiation site.

2 When alternate basepairs are indicated (Locker and Buzard., 1990) the best fit to the hCgA sequence is chosen.

~ Inverted repeats.

~ Purine-Rich (PR) sequence.

~ "N" denotes any basepair.

3.2 Strategies for studying the neuroendocrine specific expression of CgA gene in vivo

3.2.1 Promoter portions studied

Although elaborate work has been done on CgA gene regulation *in vitro*, the sequence requirements for the neuroendocrine specific expression of CgA *in vivo* are still unknown, and this we intended to elucidate in this study. Such knowledge would be important for a variety of studies such as those attempting to express therapeutic toxic genes in neuroendocrine tumor cells using CgA gene regulatory elements.

According to *in vitro* CgA regulation studies, there is an active domain downstream of –147bp relative to the transcription start site, including the CRE, in the proximal promoter region. Also, there is a proximal negative domain at (-258/-181 bp) and a distal positive domain around –4.8 kb upstream of the cap site. Therefore, we planned to study two portions of the 5'-flanking region of CgA gene in this project. The short portion is terminated at the 184th bp upstream of the transcription start site. We

chose this region to address the necessity and sufficiency of the proximal promoter region for neuroendocrine specific expression of CgA gene. The other is the longer portion, which consists of a 6 kb sequences upstream of the cap site. This portion is chosen for addressing the entire sequence requirement for the neuroendocrine specificity of CgA gene regulation.

3.2.2 Reporter systems used

The reporter systems we used are the typical β-galactosidase system, which has been proven to give elegant results, and the green fluorescent protein system. Since the detection of GFP activity does not require any substrate and there is no need for tissue fixation. the GFP system provides real time observation of the developing embryos[50]. We planned to make four constructs. each of which contains one of the two reporter genes hooked up to one of the two promoter portions of the CgA gene separately.

3.2.3 Introducing a single copy of the promoter-reporter cassette to the mouse genome

Since the expression of the reporter gene may be altered by the effect of the integration location, the copy number and the insert disruption of other genes in the genome, the most efficient way to obtain the reliable results is to introduce a single copy of the transgene cassette to the targeted locus in the genome. The targeted vector we proposed to use is the Hprt gene targeting vector--pMP8SKB[15], which contains the homologous sequence of the Hprt gene and can deliver the transgene cassette to the Hprt gene locus by homologous recombination.

3.3 Material and methods

3.3.1 β-gal reporter system construction

For the β-galactosidase reporter system, the pKS(Sal)^2-SDKlacZpA plasmid[163], which was kindly provided by Dr. Mark Featherstone in the Cancer Research Center of McGill University, was used as the vector. It was derived from the pBluescript KS vector by inserting a Kozak consensus sequence containing *E.Coli* lacZ gene followed by the SV40 polyadenylation signal.

The proximal promoter (-184bp upstream of the cap site) of the CgA gene was amplified by polymerase chain reaction (PCR) with primers that have SphI and HindIII-KpnI sites added at the 3' and 5' ends respectively. The PCR product was inserted between the SphI and HindIII sites of the vector upstream of the β-gal gene to generate the short promoter-reporter construct, β-gal construct I. Then, the 6109 bp HindIII-KpnI portion of the 5'-flanking fragment, which is cut from the phage DNA clone that contains the CgA genomic sequence of the mouse 129 sv/J strain, was inserted upstream of the proximal promoter sequence to construct the β-gal construct II. The phage clone is from a mouse kidney genomic DNA library constructed in Lambda Dash II phage. (Fig 3.1)

The primers used for the amplification are: (1) Forward primer: 5'-GACC**AAGCTT**GCCG**GGTACC**CTTTACAG-3' The HindIII site and KpnI site are indicated in bold letters. (2) Reverse primer: 5'- TGGAGC**GCATGC**GAGCCG-3' The SphI site is indicated in bold letters. Although there is a ATG codon (as CAT) in the sphI site, the purine G at –3 (three nucleotides upstream from the ATG translation start codon) position, the conserved CgA gene Kozak sequence, was shifted one base pair

downstream due to the deletion of a T at the −1 position in the primer. Therefore, the ATG may not serve as a translation start site. According to the literature [4], 95% of the published eukaryotic mRNA sequences have a purine, A or G, in position −3 and position +4. Mutation studies of a cloned preproinsulin gene indicated that translation declines 5 to 10-fold when either of the purines in position −3 or +4 is replaced by a pyrimidine, and translation decreases 20-fold when pyrimidines are substituted in both of those positions [164]. Here is the comparison of the original CgA gene Kozak sequence and the mutated sequence in the reporter construct:

```
CgA:      ---GCTCGCTATGCG---
                  -3  -1  -4
Reporter: ---GCTCGCATGCC---
```

CgA gene Kozak sequence is not very conserved. It contains only one purine in position −3 but a pyrimidine in position +4. In the reporter sequence, both −3 and +4 positions are occupied by pyrimidines. Therefore, this ATG in the reporter sequence may not serve as an efficient start codon, especially compare with the strictly conserved Kozak sequence of the lacZ gene. So, the translation will still mainly start at the lacZ gene. There was a previous study , which retained the SphI site upstream of the lacZ gene, conducting the Hoxd4 gene promoter study very successfully[163]. Our result shown on the following pages, which indicate the expression of lacZ gene in AtT20 cells, confirms the feasibility of this strategy.

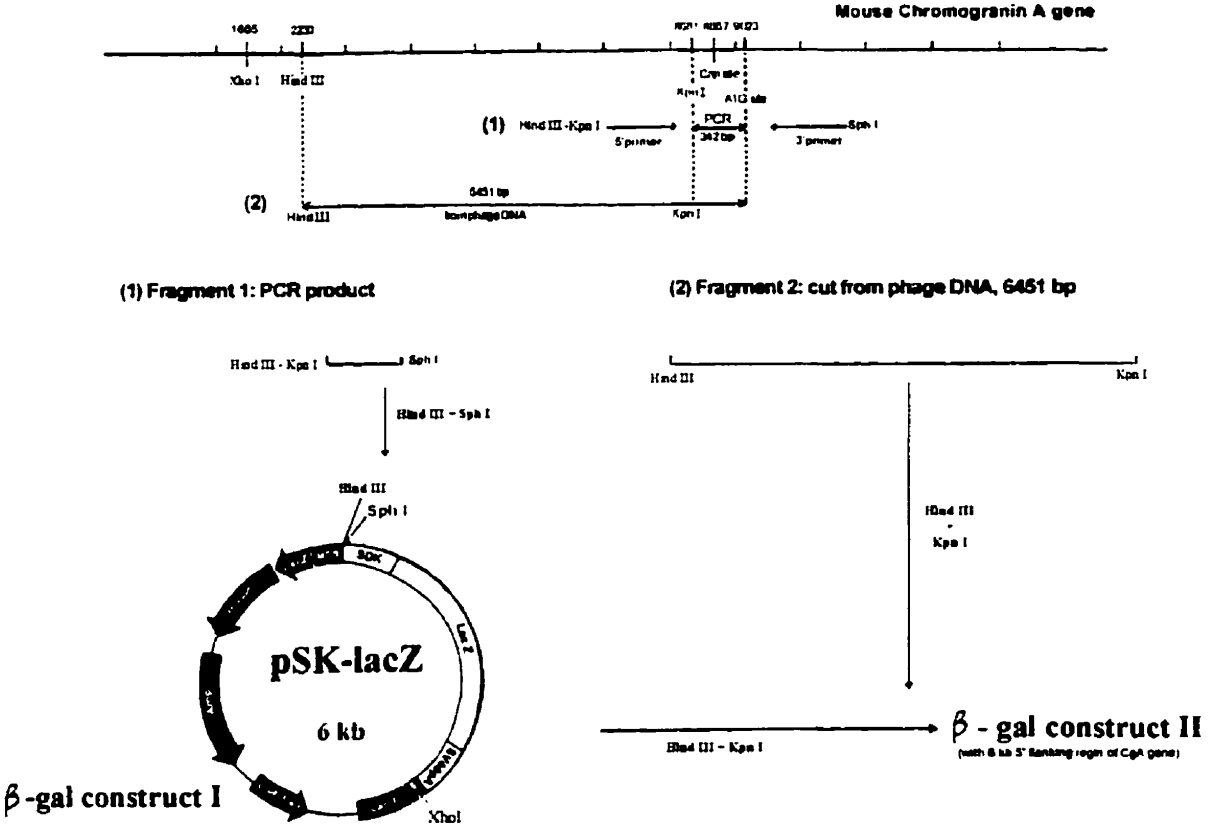

Fig 3.1 β-gal reporter system construction A 342bp PCR fragment amplified from the proximal promoter region containing the sequence from −181 to ATG site with HindIII-KpnI and SphI sites introduced at both ends was inserted into the pSK-lacZ vector at HindIII-SphI sites to construct the β-gal construct I. The second large distal promoter fragment is a 6451bp sequence cut from phage DNA clone with HindIII-KpnI sites, and it was inserted upstream of fragment 1 of the β-gal construct I to form the β-gal construct II.

3.3.2 **GFP reporter system construction**

In the GFP reporter system, we used the pGlow-TOPO plasmid (purchased from Invitrogen), which contains the superGFP gene[23.29] downstream of a TA cloning site[130]. The 342 bp PCR product of the CgA gene proximal promoter portion, which contains the 184 bp sequence upstream of the cap site and a KpnI site at the 5' end, was inserted in the TA cloning site to construct the short promoter-reporter construct, GFP construct I. Then, a 6109 bp HindIII-KpnI fragment of the CgA 5'-flanking region from the phage clone was inserted upstream of the proximal promoter sequence to generate the longer promoter-reporter construct, GFP construct II. Since there is no Hind III site in the pGlow-TOPO plasmid, the AatII site was used to cut the vector and filled in by T4 DNA polymerase to generate a blunt end, and the HindIII cut site on the phage fragment was also filled in to generate a blunt end. Then the two ends were able to be ligated at the 5' of the long promoter fragment to form GFP construct II. (Fig. 3.2)

The sequences of the primers used to amplify the CgA short promoter are: (1) Forward primer: 5'-AGCCG**GGTACC**CTTTACAGCCA-3': (The KpnI site is indicated in bold letters.) (2) Reverse primer: 5'-TGGAGCG**TCG**ACCGAGCCGGAC-3'. The ATG start codon in the CgA gene is substituted with TCG in the primer:

CgA: . GTCCGGCTCGCT**ATG**CGCTCCA
Reporter: GTCCGGCTCGGT**CGA**CGCTCCA

Therefore, the translation will not start here but at the reporter gene start codon downstream.

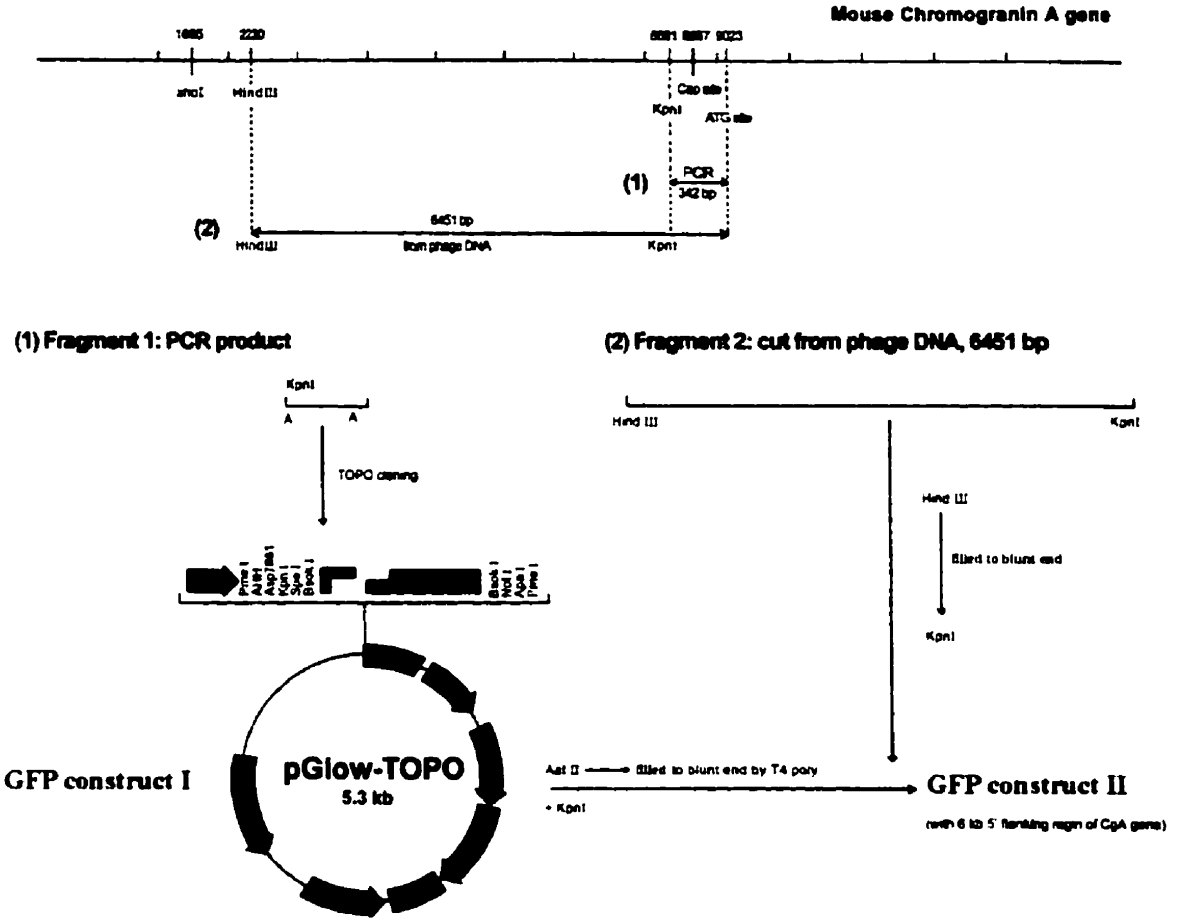

Fig 3.2 GFP reporter system construction A 342bp PCR fragment amplified from the proximal promoter region containing the sequence from –181 to ATG site was T-A cloned into the pGlow-TOPO GFP reporter vector to construct the GFP construct I. The second large distal promoter fragment is a 6451bp sequence cut from phage DNA clone with HindIII-KpnI sites, and it was inserted upstream of fragment 1 of the GFP construct I to form the GFP construct II.

3.3.3 Testing the functionality of the promoter reporter cassettes

In order to test the feasibility of the strategies used to make the promoter-reporter constructs, the four constructs were transfected into the neuroendocrine cell line, AtT20 cells, which is derived from mouse pituitary corticotrophs, and non-endocrine fibroblast NIH-3T3 cells. Reporter activity was then detected by the β-gal assay or fluorescent microscopy.

3.3.4 Preparing to generate the targeting vector for introducing a single copy of the promoter-reporter cassette into the Hprt gene locus

The pMP8SKB vector, which contains homologous regions of the Hprt gene, is to be used as the targeting vector for introducing the promoter-reporter cassettes into the Hprt locus of the mouse genome. This vector was kindly provided by Dr.Alan Peterson in the Molecular Oncology Group at the Royal Victoria Hospital. The promoter-reporter cassettes are to be excised and inserted into the polylinker site of the vector (Fig 3.3). Since there are just 3 restriction endonuclease sites in the polylinker, the choice of enzyme to use is limited. It has been shown that EcoRI and MluI sites are present in the insert sequence, so the only site that can be used is the NotI site. Therefore, we have to generate NotI sites flanking the promoter-reporter cassette and use them to excise it from the original vector (either pKS(Sal)^2-SDKlacZpA or pGlow-TOPO) for insertion into the targeting vector.

The oligo-mediated site-specific mutagenesis was used for the NotI site generation at the appropriate location in the β-gal construct I. Since there is already a

NotI site upstream of the promoter-reporter cassette. we need to generate the other one at the downstream

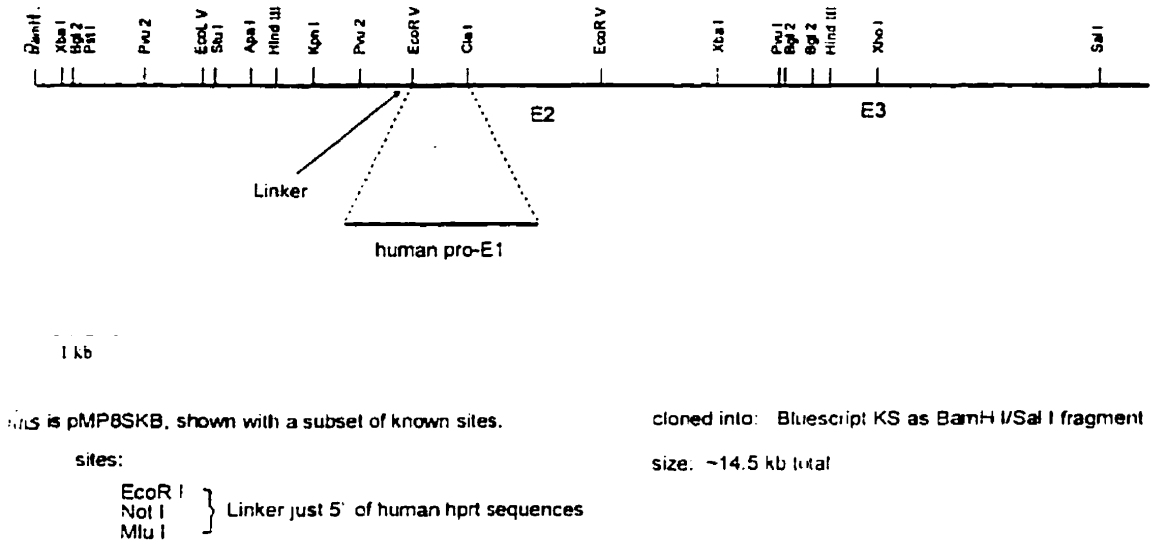

Fig 3.3 polyclonal site of the vector pMP8SKB

E1. E2 and E3 are exons 1. 2. 3 of human Hprt gene

site. The sequence of the oligonucleotide is (TCGAG**AGCGGCCGC**TC). This sequence. containing a NotI site in the middle (bold letters) and flanked by 2 incomplete XhoI sites (underlined letters). is a palindrome. When two oligonucleotides anneal together. there is a XhoI protruding end at both ends. which help to insert the small fragment into the XhoI site downstream of the promoter-reporter cassette.

TCGAG**AGCGGCCGC**TC

CT**CGCCGGCG**AGAGCT

The construct then contains two NotI sites that can be used for further cloning of the targeting vector.

3.4 Results

3.4.1 β-galactosidase reporter constructs

The short portion of the proximal promoter is from ATG to −184 bp relative to the cap site. which is 342bp. There is a KpnI site at the 5' end of the sequence. We chose this site for the further cloning of the long portion promoter which terminates about 6kb upstream of the cap site at Hind III site. Therefore. the KpnI and HindIII sites are the ones to be used upstream of the Lac Z gene. However. since there is a KpnI site downstream of the LacZ gene at the pKS(Sal)^2-SDKlacZpA vector. we have to eliminate it first. The T4 DNA polymerase was used to fill in the sticky end and the blunt ends were re-ligated by T4 DNA ligase.

The short promoter portion was amplified by PCR. with the HindIII and SphI sites added at both ends for further cloning. The fragment. which is 350bp. was inserted into the HindIII and SphI sites upstream of the LacZ gene of the pKS(Sal)^2-SDKlacZpA vector to form β-gal construct I(Fig 3.4 Lane 2 shows the 350bp promoter fragment released by SphI and KpnI digestion).The long portion of the 5' flanking region was cut from the phage DNA by HindIII and KphI (6109bp). and inserted just upstream of the short promoter portion to form β-gal construct II. Since the vector is also around 6kb. an additional enzyme site for EcoRV was used to give the 2.1kb and 4.4kb bands of the insert promoter fragment (Fig 3.5 Lane 3).

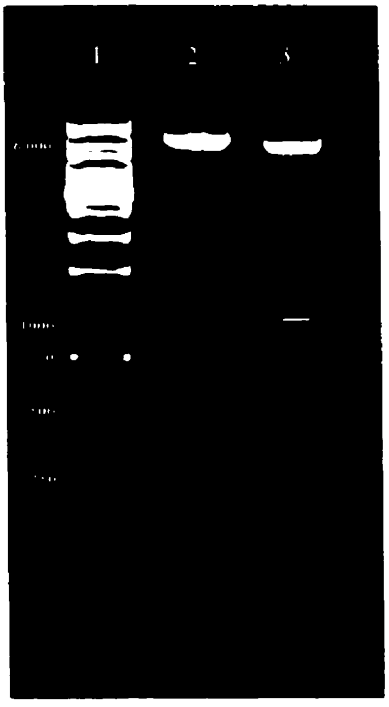

Fig 3.4 β-gal construct I and GFP construct I

Lane 1: marker

Lane 2: β-gal construct I digested with SphI & KpnI

Lane 3: GFP construct I digested with KpnI & ApaI

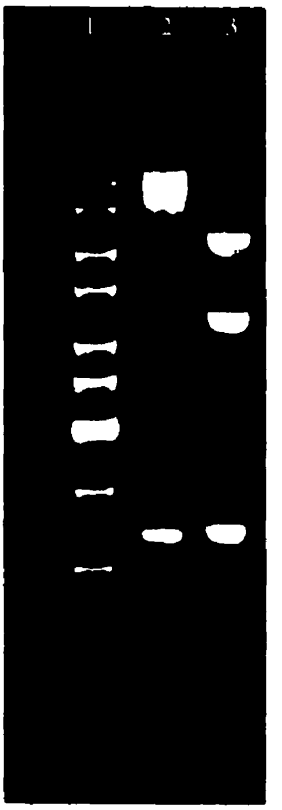

Fig 3.5 β-gal construct II and GFP construct II

Lane 1: marker

Lane 2: GFPconstruct II digested with EcoRV & KpnI

Lane 3: β-gal construct II digested with EcoRV & KpnI

3.4.2 GFP reporter constructs

For the GFP system, the short promoter portion was also amplified by PCR and inserted into pGlow-TOPO to form the GFP construct I by TA cloning. In this case, we had to check the orientation of insertion. At the 5'end of the PCR product, there is a KpnI site; there is another KpnI site in the polylinker of the vector. Downstream of the GFP gene, there is an ApaI site. KpnI and ApaI were used to digest the mini-prep DNA. When the insert is in the right orientation, there will be two bands: one is 1.1kb, which consists of the short promoter and GFP gene; the other is 4.5 kb, the vector backbone. (Fig 3.4 Lane 3) The long portion of the CgA 5' flanking region was cut from phage DNA and inserted upstream of the short portion. Since there is no HindIII site in the pGlow-TOPO, AaTII was used and the sticky end was filled in and the blunt end was ligated to the blunt end of the filled in HindIII site at the 5' end of the insert fragment. KpnI and EcoRV were used to confirm the correct orientation of the insertion. The positive clone should give a 2.1 kb and a 9.7 kb band. (Fig3.5 Lane 2)

3.4.3 β-gal assay of the β-gal construct transfected cells

To test the feasibility of the strategy, the two reporter constructs as well as the empty pKS(Sal)^2-SDKlacZpA vector and a β-gal constitutive expression plasmid pCH110 were transfected into both the neuroendocrine cell line AtT20 and nonendocrine cell line 3T3 cells. The negative and positive control showed expected results and the two promoter portions drive the neuroendocrine specific expression of the reporter gene in the AtT20 cells. (Fig 3.6)

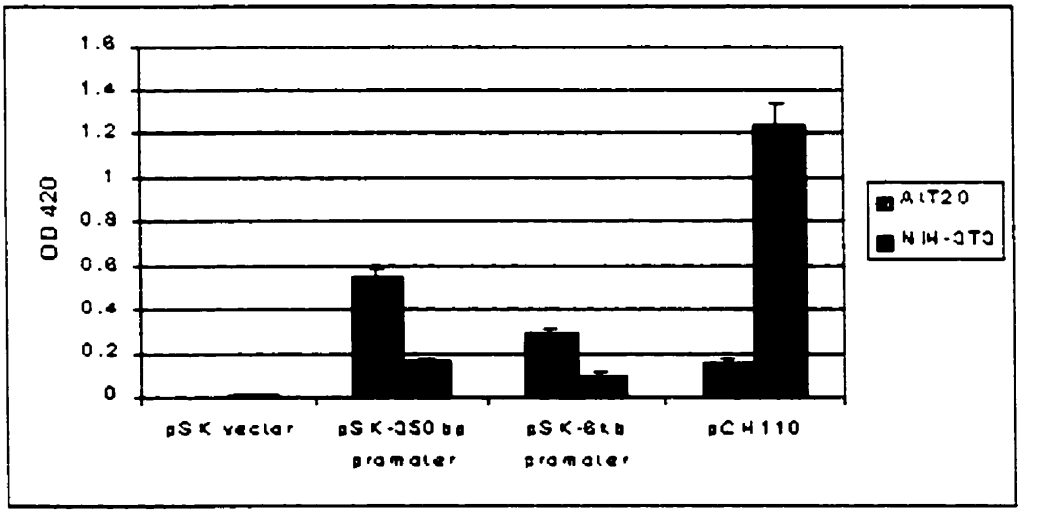

Fig 3.6 β-gal assay result of the β-gal constructs β-galactosidase activity of the cell lysate of both AtT20 cells and NIH-3T3 cells transfected with pKS(Sal)^2-SDKlacZpA reporter vector alone, reporter vector with 350bp short promoter, reporter vector with 6kb long promoter, and pCH110 positive control construct is shown. Reporter vector alone transfected cells show very low or absent β-gal activity. The positive control shows high activity especially in NIH-3T3 cells, which indicates better transfection of NIH-3T3 cells than AtT20 cells. The cells transfected with β-gal reporter constructs containing short and long promoter portion show β-gal activity, which is stronger in AtT20 cell than in NIH-3T3 cells. With the fact that AtT20 cell is more difficult to transfect and the β-gal activity is more intensive in AtT20 cells than in NIH-3T3 cells, we may conclude that CgA promoter is more active in neuroendocrine cell line AtT20 than in non endocrine cell line NIH-3T3. In the figure, the value is the OD of the light with the wavelength of 420nm. The error bars come from the measurement deviation.

3.4.4 Green fluorescent microscopy of the GFP constructs transfected cells.

The two GFP constructs were transfected into AtT20 and NIH-3T3 cells by lipofectamine-mediated transfection on chamber slides. The transient transfection results showed that the GFP constructs of CgA gene promoters did direct neuroendocrine expression in AtT20 cells. (Fig. 3.7)

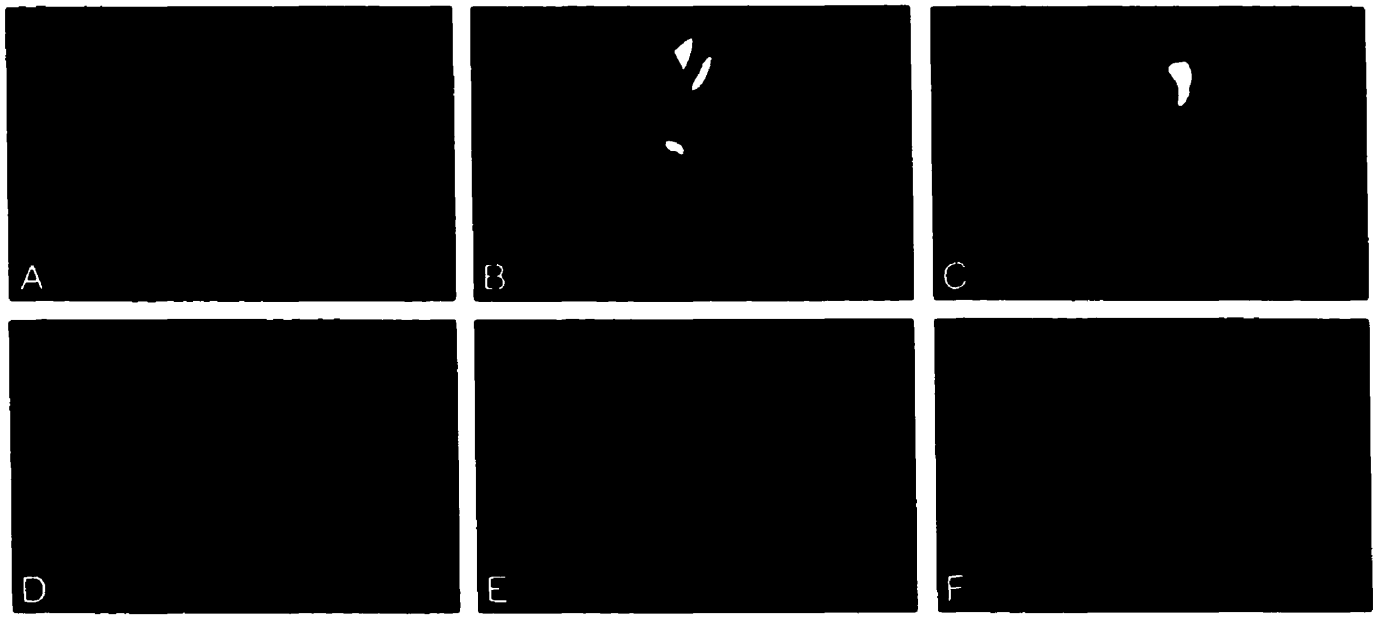

Fig 3.7 Transient transfection result of the GFP constructs

AtT20 cells transfected with A) empty vector;
B) GFP construct I; C) GFP construct II;
NIH 3T3 cells transfected with D) empty vector;
E) GFP construct I; F) GFP construct II

3.4.5 **Introducing a NotI site in the β-gal construct I**

In order to excise the promoter-reporter cassette and insert it into the Hprt targeting vector, a NotI site should be created downstream of the LacZ gene. An oligonucleotide-mediated site-specific mutagenesis was performed successfully (Fig 3.8) The insert is 4kb and the vector is 3kb. Therefore, the insert can now be excised out of the plasmid and inserted into the targeting vector.

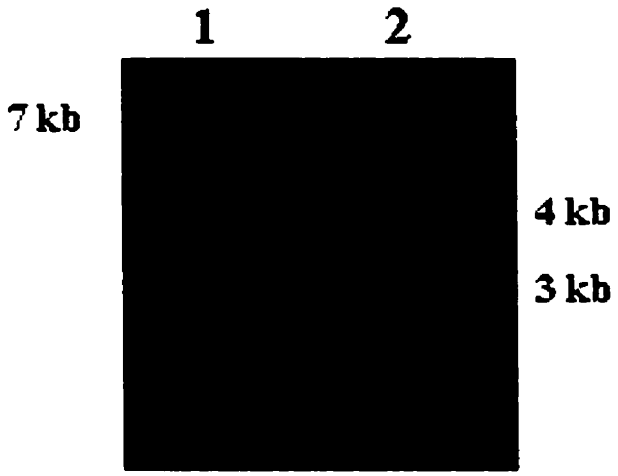

Fig. 3.8 Introducing a Not I site in the β-gal construct I

Lane 1: β-gal construct I digested with NotI

Lane 2: Oligo-mutated β-gal construct I digested with NotI

3.5 Discussion

The aim of the study is to discover the *in vivo* sequence requirements for the neuroendocrine specific expression of the CgA gene. This knowledge would be important for a variety of studies, for example, CgA gene regulatory elements could be used to express therapeutic toxic genes in neuroendocrine tumor cells; to express exogenous genes in neuroendocrine tissue for transgenic studies; and to express the Cre gene in neuroendocrine tissue for tissue specific knockout studies.

We plan to study two portions of the CgA 5' flanking region. One is a short proximal promoter region (184 bp upstream of the cap site), which contains the crucial CRE element that has been proven to be both necessary and sufficient for the neuroendocrine specific expression of the CgA gene in many neuroendocrine cell lines. Previous studies in our lab indicated that the cAMP and PKA levels affect the neuroendocrine specific expression of CgA gene[21]. Other studies also demonstrated that NGF and IGF-I stimulate the CgA expression in neuroendocrine PC12 cells by the MAP kinase pathway through the CREB phosphorylation by the CREB kinase---RSKII[91]. However, the studies in our lab suggested that the expression level of CREB and CBP, as well as the phosphorylation level of CREB is not different in neuroendocrine and nonendocrine cells. Thus, the phosphorylation of CREB is necessary but not sufficient for the neuroendocrine expression of the CgA gene. There may be neuroendocrine specific factors or nonendocrine suppressors that cooperate with CREB to direct CgA expression. To identify these factors and their underlying mechanisms, protein/protein interaction studies should be undertaken.

Besides CRE, other *cis*-elements also participate in CgA gene regulation. For example, in the enterochromaffin (ECL) cell line AGS-B, CgA expression is regulated by gastrin[63]. 5'deletion analysis and scanning mutagenesis of mouse CgA 5'-flanking sequences showed that a Sp1 site spanning –88 to –77 bp and a CRE at –71 to –64 bp are essential for gastrin-dependent mCgA transactivation. Mutation of either the Sp1 site or the CRE site completely abrogated gastrin responsiveness, suggesting the possibility of some sort of cooperative interaction between Sp1 and CREB. However, the exact nature of this interaction is unclear at present. In addition, there is a proximal positive domain downstream of –181bp which is active in the neuroendocrine cell lines AtT20 and PC12[160]. Therefore, we chose the proximal promoter region, which contains the sequence downstream of -184 bp relative to the transcription start site, and includes these critical regulation elements to address the minimal requirements of the *cis*-elements for the neuroendocrine specific expression of the CgA gene *in vivo*. If this small fragment directs the neuroendocrine specific expression pattern well enough, it will be used for many studies such as those mentioned above.

Other elements upstream of the proximal promoter region also take part in CgA gene regulation. A negative domain at –258/-181 as well as a positive domain at –4.8/-2.2 kb was detected in AtT20 cells[160], and a sequence located between –726 and –455 was shown to stimulate CgA expression approximately 5-fold in the human lung neuroendocrine cell line, BEN cells[21]. These results indicate that distal 5'-flanking region contains *cis*-elements that play a role in fine tuning CgA gene regulation. The long portion of the 5'-flanking region we are studying consists of the 6 kb sequence upstream of the cap site. Thus, it may contain the complete sequence requirements for CgA

regulation. It may help to answer the question whether the 5'-flanking region of the CgA gene is sufficient to control CgA gene expression.

The results of the transient transfection of the constructs show that the short and long portions of the mouse CgA gene 5'-flanking region do drive the neuroendocrine specific expression of the reporter gene in mouse pituitary adenoma cell line AtT20 cells. The short portion of the proximal promoter has higher activity than that of the long portion promoter. It confirms that the upstream distal region contains negative *cis*-elements. However, the difference may also due to different transfection rate of the short and long promoter containing plasmids. In either case, the *in vivo* study will be carried on later to show the exact expression pattern of the two promoter regions.

The crucial point of studies on gene regulation is to identify the precise localization of transgene expression in different mouse tissues. An appropriate reporter system facilitates this procedure. Because the method should localize reporter proteins to the site of their production, reporter assays that employ either precipitating substrates or substrates that generate a luminescent or fluorescent signal are most effective. So far, there are three reporter systems for *in vivo* gene regulation study: β-galatosidase, firefly luciferase, and green fluorescent protein (GFP). The most commonly used approach is the β-galactosidase system. The reaction with X-Gal produces a rich blue color that can easily be scored against background in most applications. The detection of lacZ expression in whole embryos has been particularly useful for characterizing tissue-specific gene expression during early development. For instance, the activity of the regulation elements of Hoxd4 gene in developing embryos was successfully detected by the β-galactosidase system[163].

The GFP system is a newly developed reporter system. The GFP molecule absorbs blue light and emits green light with no exogenous substrates are required. It is of small size (26.9kDa) and nontoxic to the transgenic animal with ubiquitous high level expression in all of the tissues[110]. Moreover, it can be examined both early in mammalian embryogenesis when the embryo is relatively clear and in organ preparations late in embryogenesis when the embryo is opaque. In particular, laser scanning confocal microscopy has been used to obtain optical sectioning images of GFP-expressing embryos without histological sectioning. This technique was applied successfully in the detection of targeted GFP-Hox gene fusions during mouse embryogenesis[44]. In addition. It has been shown that the GFP reporter system is more sensitive than the β-galatosidase system. The study of mouse β1 integrin promoter activity by both of the GFP and β-gal reporter systems showed almost eight times more fluorescent cells than β-galactosidase staining in transgenic liver with the same level of reporter protein expression[25]. Due to the many advantages of the GFP system, we decided to apply it to our CgA regulation study as well as using the conventional β-galatosidase system. The GFP gene we used is an enhanced form of GFP protein that gives bright green fluorescence[23,29]. The result of the transient transfection of the promoter-GFP constructs to the AtT20 cells shows that the bright fluorescence distributes throughout the cell including the attenuated processes. We expect to obtain satisfying results with these constructs in the *in vivo* CgA regulation study.

To refine our strategy, we plan to apply the gene targeting approach to introduce a single copy of the promoter-reporter cassette to the house keeping gene--- Hprt gene locus. To fulfil this task, the promoter-reporter cassette should be inserted between the

two homologous sequences of the Hprt gene. Since there is a size limitation, the insert must be less than 6 kb, otherwise the homologous recombination rate will drop. Therefore, although the short portion of the proximal promoter can be used directly, the long portion 5'-flanking sequence should be cut to about 3 kb.

Chapter 4

Functional analysis of CgA gene *in vivo*

4.1 Introduction and Strategy

At present, CgA functional analysis is restricted to *in vitro* studies. It remains unclear whether the intracellular functions or the extracellular functions are predominant and relevant *in vivo*. In addition, there are different secretory mechanisms and different factors affecting the secretory functions of the endocrine organs. Therefore, the question of what exact role CgA plays *in vivo* can only be addressed by transgenic technology.

Nowadays, transgenic technology provides an opportunity to explore the function and regulation of a single gene on complex physiological system *in vivo*. I will now discuss how it could be used in functional analysis of the CgA gene.

To knockout a gene in the mouse genome, the targeting vector should be constructed first. In our laboratory, the CgA gene targeting construct had been made before I started my work. The conventional knockout approach is used in this case. The targeting vector, which is constructed from the pPNT plasmid[74], contains two sequences which are homologous to the 5' flanking region (6 kb) and exon 5 to exon 6 (2.9 kb) of the endogenous CgA gene of the mouse129sv/J line, from which the embryonic stem cells used for gene targeting are derived. (Fig.4.1) By homologous recombination, the exon1 to exon 4 of the CgA gene was substituted by the sequence in the targeting vector, which

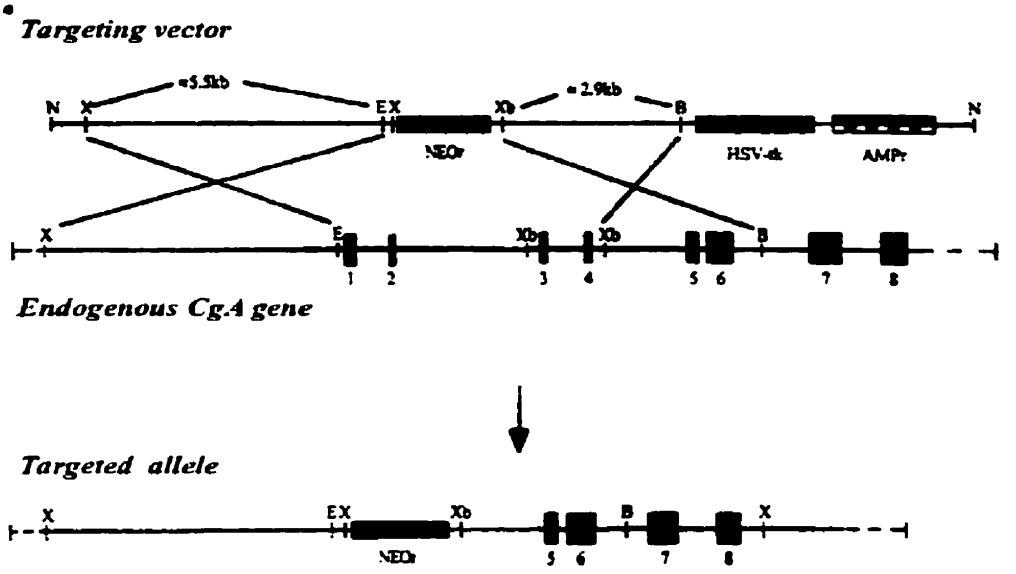

Fig. 4.1 Targeting vector and targeting event for CgA gene Top line: targeting vector of CgA targeting, which contains two homologous regions (a 6 kb 5'-flanking region and a 2.9kb sequence at exon 5 and exon 6 region), neomycin resistant marker, HSV-tk marker and ampicilin resistant marker. **Middle line**: endogenous CgA gene region, shows 5'-flanking region and exon 1 to exon 8. **Bottom line**: targeted allele with exon 1 to exon 4 substituted with the neomycin resistant gene. (adapted from ref. 41)

contains a neomycin selective marker to form the targeted allele in the mouse genome. Another marker, HSV-TK (herpes simplex virus thymidine kinase) gene, was located outside the region of homology to the target gene and used as a negative selection marker. In the presence of the TK gene, the cells are sensitive to acyclovir and its analog, gancyclovir. The HSV-TK enzyme activates these drugs, allowing them to incorporate into the growing DNA, causing chain termination and cell death. Homologous recombination results in the loss of the sequence outside the regions of homology to the target gene. Therefore, ES cells with the targeted recombination become resistant to both neomycin and gancyclovir, but the ES cells with random targeting vector integration are

still sensitive to gancyclovir. Using both neomycin and gancyclovir helps to enrich the targeted homologous recombinants.

The targeting vector is transfected into the embryonic stem cells by electroporation. The cells are selected by neomycin and gancyclovir selective media and the surviving colonies are screened by Southern blot analysis. and the recombinants are then confirmed by an alternative Southern blot strategy. The recombinant embryonic stem cells that contain the targeted allele are injected into the blastocyst stage embryo and transferred to the pseudo-pregnant female mouse. The chimeric embryos are developed to the chimeric mice. some of whose germ cells are derived from the recombinant ES cells. Therefore. the offspring of the chimera may be heterozygous CgA gene knockout mice. By breeding the heterozygous mice. homozygotes may be obtained. (Fig. 4.2)

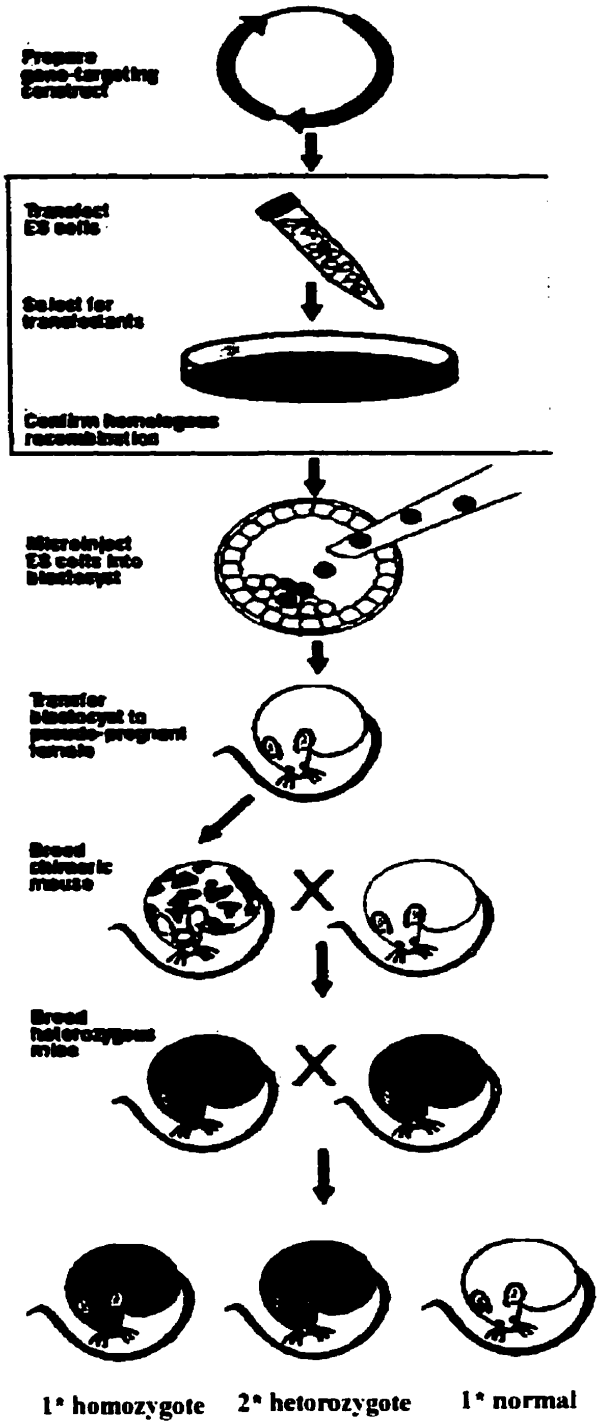

Fig. 4.2 Generation of gene knockout mice (Taken from ref. 41)The targeting vector is tansfected into ES cells by electroporation, and the cells are selected by neomycin and gancyclovir. The surviving clones are screened and confirmed by Southern blot. One of the positive clones is injected into host blastocysts and transfered to pseudo-pregnant female mice to develop into chimeric mice. Some of the germ cells of the chimera derived from the targeted ES cells give rise to heterozygous mice. By breeding the heterozygotes, homozygotes can be obtained.

4.2 Material and Methods

4.2.1 Generation of CgA gene ablation embryonic stem cell lines

Since the targeting vector was ready, my work started with the transfection of the vector into the embryonic stem cells by electroporation[146] to generate recombinant cell lines with one allele being targeted and substituted with the sequence on the targeting vector.

4.2.1.1 ES cell culture

Embryonic stem (ES) cell lines are derived from the inner cell mass of a blastocyst-stage embryo. They are very small round cells that grow in tightly packed smooth, round colonies in culture, which retain the ability to differentiate into every cell type present in the mouse[109]. The ES cells grow very fast and are very easy to differentiate. The differentiated ES cells become flatter with distinct intercellular boundaries. Such cells are not suitable for chimeric mouse generation. To prevent the ES cell from differentiating, inactivated (unable to proliferate) embryonic fibroblast cells are used as feeder cells to provide appropriate conditions for maintaining the ES in an undifferentiated state. In addition, leukemia inhibitory factor (LIF) is also used for the same purpose.

4.2.1.2 Transfection and colonies selection

In our protocol, ES cells were cultured in ES/LIF medium. Medium was changed every day and cells were passaged every 2-3 days by seeding a 100-mm gelatin-coated and feeder cell covered ($5X10^6$cells/plate) tissue culture plate with 1-$2X10^6$ cells/plate. The cells were harvested at $1X10^7$ by adding trypsin/EDTA. Since ES cells package very

tightly in colonies, it is very important to dissociate them to single cells by pipetting up and down. Then, the cells were suspended in 1ml electroporation buffer in the electroporation cuvette. 10 μg targeting vector, which had been linearized by NotI and sterilized by ethanol precipitation, was added to the cuvette. The electroporation was done at 240V and 500μF. The cells were then kept on ice for 10 min and plated at $2X10^6$ cells/100-mm plate with feeder cells. Selection started on the second day with 0.3mg/ml G418 and 2μM Gancycline. After four days Gancycline was stopped and only G418 was used for the selection. Eight days after electroporation, individual colonies grow large enough to be picked up from the plate. By using pipette tips, individual colonies were picked up, dissociated to single cells and transferred to 24-well microtiter plates containing feeder cells and incubated until colonies were visible, but the cells are not differentiating (typically 3 to 4 days). Half of the cells were passaged to a clean 24-well plate for continued growth and the remaining cells were frozen at -70°C.

4.2.1.3 Screening of the colonies for recombinant cell lines

- **ES cell genomic DNA extraction**

The growing cells in the 24-well plate were used for DNA extraction for Southern blot analysis. The cells were allowed to grow till near confluence (3 days). Then, 400 μl digestion buffer (10mM Tris-HCl, PH8.0, 1mM EDTA, 0.1M NaCl, 1% SDS, 0.5 mg/ml proteinase K) were added to the wells and the cell lysates were transferred to 1.5-ml microcentrifuge tubes and incubated overnight at 55°C. The next morning, 500 μl phenol were used for the extraction. The genomic DNA was precipitated with ethanol: 3 M NaAc (25:1), and washed in 70% ethanol. Finally, it was dissolved in 30 ul TE.

- **DNA digestion with XhoI**

In this study, since a relatively high concentration of G418 was used (300μg/ml), only 17 colonies survived, which were picked and extracted for DNA to be used in Southern blot analysis. 15μl of the genomic DNA was used for the restriction enzyme digestion. In our screening strategy, XhoI was chosen for the digestion. Since it has been shown that XhoI is a rather difficult enzyme for genomic DNA digestion, a relatively large digestion volume (60 μl) and extra amounts of enzyme (60 units) was used for each reaction. The enzyme was added every two hours in three boosts, with a final overnight incubation. In addition, we added spermidine to the digestion reaction to assist the digestion.

- **Southern blot analysis**

The digested ES genomic DNA samples were loaded on 1% agarose gel with ethidium bromide, and electrophoresed for 16 hours at 50 volts. After the picture was taken, the gel was washed with 0.25N HCl solution for 10 minutes, and 0.5N NaOH/1.5MNaCl solution for 20min, then neutralized in 1.0M Tris HCl pH8.0/1.5 M NaCl solution for 2X15 minutes.

DNA on the gel was capillary transferred onto a supported nitrocellulose membrane (Hybond-C 0.2 micron/Amersham) overnight with 20XSSC as the transfer buffer. The membrane was baked at 80°C for 2 hours to immobilize the DNA bound to it.

Prehybridization was performed in 10ml hybridization buffer (1%BSA, 7% SDS, 0.5M sodium phosphate pH6.8 and 1mM 0.5M EDTA pH8.0) at 60°C for 3 hours. Then, the ^{32}P labeled probe (labeling procedure see page 78) was denatured and added to the hybridization buffer at a concentration of $1X10^6$ cpm/ml. The hybridization was allowed to go overnight at 62°C. the membrane was then washed with wash buffer A (0.5% BSA,

5% SDS, 40mM $NaPO_4$/pH6.8 and 1mM EDTA/pH8.0) for 2X 15 minutes and then with wash buffer B (1% SDS, 40mM $NaPO_4$/pH6.8 and 1mM EDTA/pH8.0) for 2X 15 minutes at 65°C. The membrane was then removed and sealed in a plastic bag. Autoradiography was performed in an autoradiography cassette with a sheet of Kodak Scientific Imaging MS film and two intensifying screens at -80°C for 3 days, and the film was developed using a Kodak RP X-OMAT developer.

- **Probe generation and labeling**

There are three probes we used for the screening and confirmation of the recombinant ES cells: **(1) 3'-end probe:** a sequence of about 350bp in the exon 7 coding region, which is localized downstream of the homologous recombination region; **(2) Neomycin gene probe:** 450bp sequence in the neomycin coding sequence; **(3) 5'-end probe:** a 1.2 kb sequence in the 5' flanking region of the CgA coding region, which is located upstream of the homologous recombination region. The first two probes were used for screening and the third one was for the confirmation of the recombination.

The 3'-end probe and the neomycin gene probe were amplified by PCR from the mouse CgA cDNA and the neomycin resistant gene on the pPNT plasmid respectively. The primers for 3'-end probe are: (5'-GAAGCTCTGCCGTCTGAA) and (5'-TCTGGAAGATGGCCTCCA); The primers for the neomycin probe are: (5-CAAGATGGATTGCACGCAGG) and (5'-TCCAGATCATCCTGATCGCA). By using the TOPO-TA cloning kit (from Invitrogen), the PCR products were subcloned into PCR2.1 vector. The sequences were confirmed by sequencing. Then, the inserts were released by appropriate restriction enzymes and purified from agarose gel with a Qiagen II gel purification kit and stored at -20°C for future labeling use. The 5'-end probe was

generated by restriction enzyme (HindIII & BamHI) digestion of phage clone DNA that contains the mouse CgA gene genomic sequence followed by gel purification of the appropriate fragment.

Purified probe DNA fragments were randomly labeled with α-^{32}P-dCTP using the Ready-to-go Reaction Labeling Beads (Pharmacia). Forty ng of the DNA fragment was added to sterile water in a total volume of 45 µl, and heat denatured. Five µl of α-^{32}P-dCTP (50 µCi) and labeling bead were added, then incubated in 37°C for 15 minutes. Spin column was applied for eliminating the unincorporated radioisotope. 2 µl of the labeled DNA sample was used for scintillation counting (1219 RACKBETA Liquid Scintillation Counter/LKB Wallac). The labeled probes were denatured before adding to the hybridization reaction at a concentration no less than $1X10^6$ cpm/ml.

4.2.1.4 Culture of selected ES cell lines

Two positive ES cell lines were obtained by the first screening. These two lines were thawed and grown up for more storage and large amount of DNA extraction. Since they might be used for the microinjection, they were cultured on feeder cells with LIF added to the medium.

The 5`-end probe was used for the confirmation of the recombinant ES cell lines. Fifteen µg genomic DNA of the ES cells was digested with BamHI and KpnI. The BamHI was added first and the DNA was digested for 6 hours and purified by ethanol precipitation. Then, the KpnI was applied and digested overnight at 37°C overnight. The digested DNA was then analyzed by Southern blot.

4.2.2 CgA gene knockout mice generation

The cells of clone number 5 were injected into the blastocyst of the Balb/c mouse and developed in the pseudopregnant Balb/c female mouse. Several chimeric mice were obtained, which were crossed with Balb/c mice to give rise to heterozygotes. By breeding the heterozygotes, homozygotes may be obtained.

4.3 Results

4.3.1 Screening results

From the first round of electroporation and selection. 17 colonies of ES cells survived. DNA from these ES cell clones was extracted and analyzed by Southern blot after digestion with XhoI. There are two XhoI sites flanking the CgA gene coding region. (Fig. 4.3) When genomic DNA is digested with XhoI. the wild type allele gives a 17kb fragment. and the targeted allele gives a 7 kb fragment. Part of the sequence of exon 7 (350 bp) was used as the probe. A neomycin probe (450 bp) was also used to confirm the short fragment.

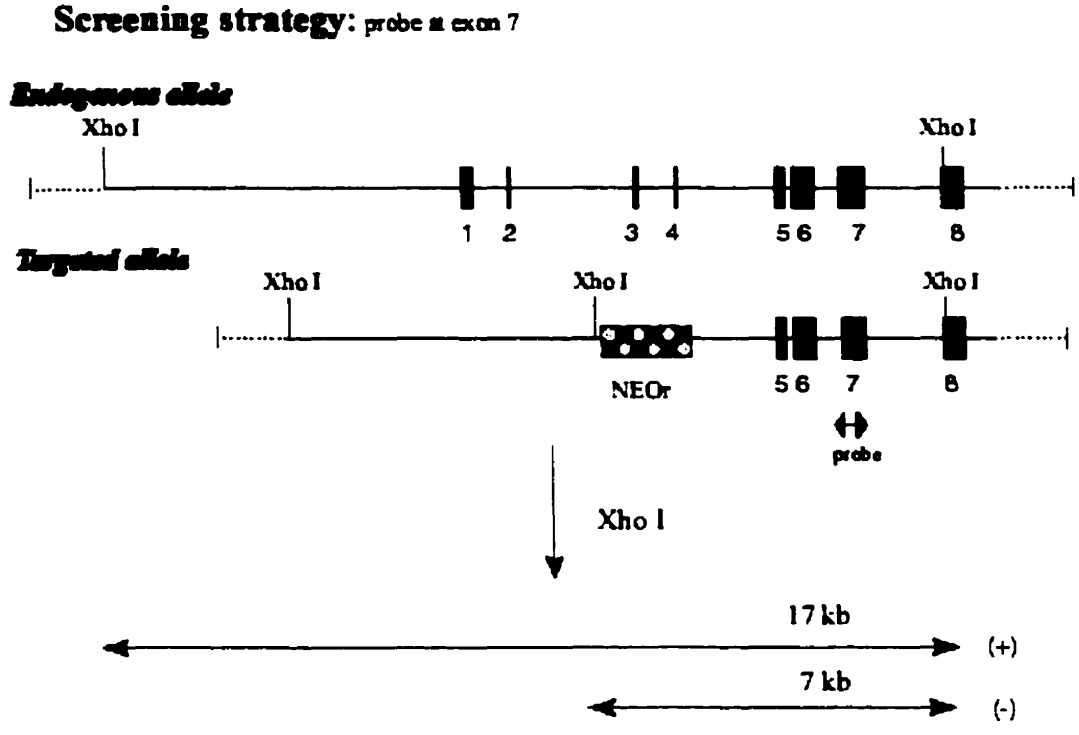

Fig 4.3 Screening strategy of CgA gene knockout in ES cells
Upper line: When digested with XhoI, endogenous CgA gene locus gives a 17 kb fragment. **Lower line**: Targeted allele gives a 7 kb fragment with XhoI digestion. While probed a fragment of exon 7 sequence, the endogenous allele and targeted allele give a 17 and a 7 kb band respectively.

Fig 4.4 Southern blot screening of the surviving ES cell colonies

Fig 4.5 Neomycin gene reprobing of the blot

Fig 4.4 & Fig 4.5 : By southern blot, clone #5 and #9 give two bands, which are17kb and 7kb, corresponding to endogenous allele and targeted allele, when probed with exon 7 probe and neomycin resistant gene probe.

The Southern blot result (Fig 4.4) shows that when probed with the exon 7 probe, clone number 5 and 9 give the expected bands, indicating they may be ES clones in which the desired recombination events have taken place.

Theoretically, the smaller band (7 kb) of the targeted allele contains the neomycin resistant gene. To confirm this, the blot was stripped and reprobed with the neomycin probe. The result (Fig 4.5) shows that neomycin positive fragments in sample #5 and #9 are located at the same places as those of the smaller bands in the exon 7 probed blot.

4.3.2 Recombination conformation

To confirm the result, an alternative strategy was used. In this strategy, the genomic DNA of clone #5 and #9 was digested by BamHI and KpnI. The probed region is at the 5' flanking region of the CgA gene. The wild type allele gives a 8.3 kb band, and the targeted allele gives a 12kb band. (Fig 4.6) The Southern blot shows that these two clones are the recombinant cells. (Fig 4.7)

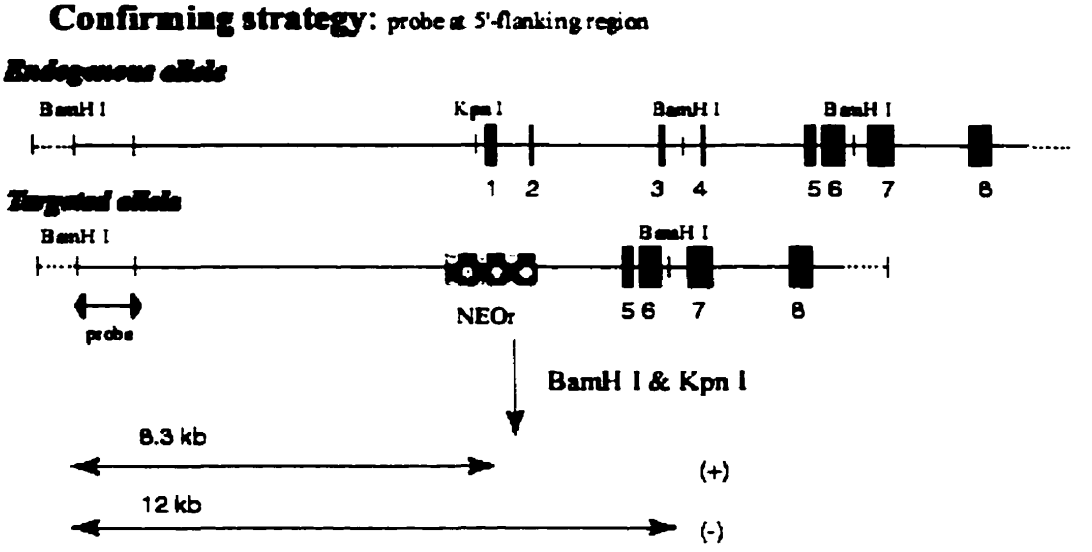

Fig 4.7 Southern blot confirmation of the ES cell lines containing targeted allele

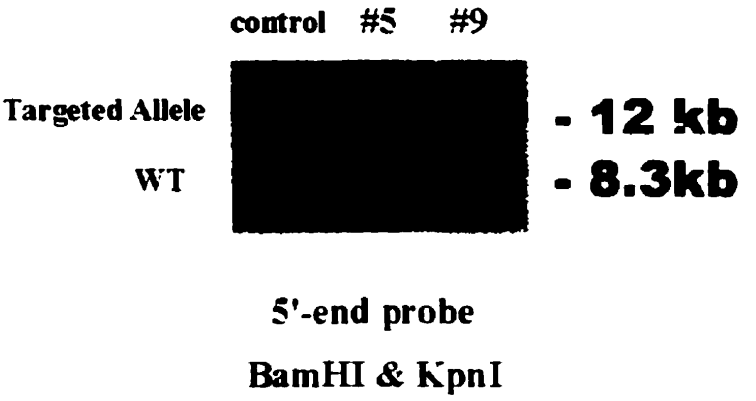

Fig 4.6 & Fig. 4.7: While digested with BamHI and KpnI and probed with the sequence at the 5' flanking region of CgA gene locus, the endogenous allele gives a band of 8.3 kb and the targeted allele gives a band of 12 kb. The result of Southern blot confirms that clone #5 and #9 are cell lines containing targeted CgA gene.

4.3.3. Transgenic mice generation

Since the R1 ES cells were derived from brown colored 129sv/J mice, the chimeric mice, generated by the injection of clone number 5 ES cells into the Balb/C mouse blastocyst, were brown/white. Two months after the injection, 2 chimeras were obtained. The chimeras were bred with Balb/C mice. Three brown heterozygote mice were obtained three months later. Southern blot analysis was used to confirm the genotype of the mice with the tail DNA extract. The heterozygotes display no obvious distinguishable phenotype.

4.4 Discussion

4.4.1 ES cell manipulation

Embryonic stem cells are pluripotent cells of the inner cell mass of preimplantation blastocysts that can give rise to all cellular lineages of mature tissues including germ cells, and are termed 'totipotent'[109]. They provide an important system for site-specific introduction of genetic alterations by gene targeting technology. Only ES cells from the mouse give rise to a successful germ line chimera. The barrier to establishing totipotent ES cells from non-mouse embryos is not understood. Additionally, the most successful and commonly used ES cell line is derived from 129sv/J, a mouse strain that tends to develop spontaneous teratocarcinomas.

The ES cells can be cultured *in vitro* and can accept transfected DNA to change their genotype. When transferred back to the blastocyst, they are incorporated into the developing embryo and contribute to all kinds of tissues of the chimera. However, if they are not treated appropriately in culture, they will undergo differentiation and lose their

totipotency, especially the ability to form germ cells in the chimeric mice. Therefore, the crucial point of ES cell culture is to prevent their differentiation.

According to ES cell signaling studies, maintenance of the pluripotency of the ES cell depends on the balance between the signaling pathways that control self-renewal and differentiation[16]. Leukemia inhibitory factor (LIF) has been proven to play a crucial role in this event. LIF binds directly to a heterodimeric receptor complex containing two transmembrane glycoproteins, gp130 and LIFR[141]. The major downstream signaling pathways related to gp130 in ES cells are the JAK-STAT3 and SHP-2/ERK pathways[17]. Overexpression of the dominant negative form of STAT3 in ES cells block their self-renewal, and the inner cell mass of the STAT3 knockout (STAT3-/-) blastocysts does not thrive in culture[142]. These results indicate that STAT3 may contribute to maintaining the pluripotency of ES cells, which would provide a rationale for its role in promoting ES cell self-renewal *in vitro*. On the other hand, although SHP-2/ERK signaling plays an important role in somatic cell proliferation by pushing cells through the G1/S phase, it does not contribute directly to stem cell self-renewal[17]. These stem cells have an unconventional cell cycle with a shortened G1 phase that does not require dephosphorylation of the retinoblastoma protein, and lack the control mechanisms that regulate the G1/S transition in somatic cells. Evidence indicates that ERK activation is an important effector of differentiation in ES cells[82]. Therefore, STAT3 may act by repressing transcription of differentiation genes, whereas ERK-dependent factors may induce their expression. The balance between these two pathways may determine the choice between self-renewal and differentiation. However, there is a report of a gp140-independent factor secreted by extraembryonic cells that can suppress differentiation of

ES cells but does not activate STAT3, thus supporting the existence of an alternative self-renewal pathway[30]. Therefore, LIF alone is not sufficient to sustain the full pluripotency of ES cells. In our experiment, ES cells were cultured in medium with LIF and on top of a layer of inactivated embryonic fibroblasts, which give additional factors that help to maintain totipotency of the ES cells.

Another trick to maintain the best health of ES cells is the growth time. Most ES cell lines exhibit an inverse correlation between the number of passages and the percentage of germline contribution, which may be due to abnormal variants that accumulate in the cell population. Therefore, extra passages and prolonged culture time should be avoided during recombinant ES clone selection. It is necessary to freeze down half of the cells before passage and try to use the batch of cells that are frozen as early as possible for injection.

The manner in which ES cells are treated is also very important for successful ES cell culture and transfection. ES cells are very subtle and fast-growing cells (doubling time is 8-12 hours). They need to be cultured in high quality reagents and fed every day. Also, since the ES cells grow in tightly packed colonies, thorough dispersion is needed by adequate trypsin incubation and appropriate pipetting at every passage, especially the one before electroporation, to insure that each colony is derived from a single cell.

4.4.2 Screening strategy

Southern blot is used for screening recombinant ES cell clones. An appropriate strategy must include the following: first, choose a suitable restriction enzyme to digest the genomic DNA, which will result in obvious different size fragments between the wild

type allele and targeted allele. The selected enzyme should be an efficient enzyme for the genomic DNA. In our first screening, we used XhoI, which turned out to be a rather difficult enzyme for genomic DNA digestion. Addition of both boosts of enzyme and spermidine, which helps to loosen the DNA structure, was necessary to facilitate the digestion. For the confirmation strategy, BamHI and KpnI were chosen. They are enzymes that were proven to digest genomic DNA well. Second, choose an appropriate probed region, which is outside the homologous recombination region (the targeting vector sequence). Therefore, only the targeted recombination can be detected, but not the random integration event. In our case, it should be in the 5'-flanking region or downstream of exon VI. Third, determine probe sequence and length. Usually, the length of the probe is from 100 bp to 1 kb, with unique sequence that avoid repeats. The exons are generally a good choice. The 3'-end probe in this work is 350bp of exon 7, and the 5'-end probe is a 1.2 kb restriction fragment without repeat sequence at the 5'-flanking region.

4.4.3 CgA gene knockout outcome expectation

Although we have not had the CgA gene knockout mice yet, I would like to predict the potential outcome of the knockout mice. There are three possibilities: (1) early lethality; (2) defined phenotype; (3) no obvious phenotype.

Since CgA expression is apparent in some endocrine cells in the fetal rat pancreas, stomach, intestine, adrenal gland and thyroid at E13.5, E14.5, E15.5, E15.5 and E18.5 respectively[75], it may be important for fetal development. Therefore, the possibility of fetal lethal may exist. However, since we have three heterozygotes that do not exhibit

obvious phenotypes, the first possibility seems unlikely. However, if it does turn out to be an early lethal, we will think of using the Cre-lox system to develop tissue- specific and/or inducible knockout mice, whose CgA gene will be ablated in a specific tissue and/or at a desired stage of development to diminish the harmful effect of a CgA complete knockout.

The second possibility is that the CgA gene knockout mice display a defined phenotype. According to the previous discussion in Chapter 1, the major function of CgA is its potential role as a regulator of the secretory function of endocrine organs, either extracellularly or intracellularly. In the CgA gene ablation mice, the secretory function of some endocrine organs may be altered. So, there may be changes in the viability, metabolism, locomotor function and fertility in these mice. If the extracellular functions of CgA are its major functions *in vivo*, we may expect that the regulation functions of the CgA-derived peptides, such as pancreastatin, catestatin, parastatin, β-granin and vasostatin, will be lost. As a result, the insulin secretion of the pancreas islet β cells, the catecholamine secretion of the adrenal medulla and the parathyroid hormone secretion of the parathyroid gland will not be suppressed by the CgA derived peptides. Therefore, the CgA gene ablation mice may suffer from high PTH induced hypercalcemia and skeletal abnormalities; excessive catecholamines such as epinephrine and norepinephrine induced hypertension; and alterations in the thyroid hormone affecting the basal metabolic rate.

If the intracellular functions of CgA are more important, the CgA gene knockout mice may have trouble sorting some of the prohormones into the regulated secretory pathway, and lose control over regulated secretion. For example, parathyroid hormone is secreted through the regulated secretory pathway. If CgA plays a role in directing its

sorting or retention in the secretory granules, the ablation of CgA will result in mis-sorting of PTH into the constitutive secretory pathway, and the excessive amount of PTH being secreted will induce hypercalcemia. In addition, since the prohormone convertases are mainly found in the secretory granules (e.g. PC1, PC2), CgA knockout mice may also lose the ability to process some prohormones, such as proinsulin and procalcitonin, to their active forms. Proinsulin was shown to be sorted into the regulated secretory granules by carboxypeptidase E (CPE) through the sorting-for-entry mechanism. The CPE mutant mice display a metabolic disorder that results in obesity. We know that one of the putative functions of CgA is to play a role in protein sorting through a sorting-by-retention mechanism. The sorting-for-entry and sorting-by-retention mechanisms exhibit cell type and hormone specificity. Therefore, these two mechanisms may affect different hormones' sorting events. There has not yet been any direct *in vivo* evidence that CgA directs the sorting of a particular hormone. If we can detect any elevation in the level of a specific prohormone in the CgA knockout mouse plasma, we will shed light on the intracellular function of CgA.

To define the phenotype, macroscopic and microscopic examination of endocrine organs will be made; provocative endocrine testing will be performed to compare adrenal, parathyroid, pituitary, pancreas and thyroid secretory response in CgA gene knockout mice with those of the normals. Also, the analysis of viability, changes in locomotor function and fertility will be performed in the CgA gene ablated mice.

The third possibility is no obvious phenotype. Since CgA belongs to the granin family, other members in the granin family may compensate for some functions of CgA, especially the closely related CgB. It was shown that exons II, III and VIII of CgA, which

encodes the most conserved N- and C- terminus of CgA molecule including functional peptides such as β-granin, vasostatin and pancreastatin, are homologous to exons II, III and V of CgB. Therefore, it is possible that we will see no phenotypic changes at all. In this case, the CgB knockout mice will be generated and mated with CgA null mice to generate double mutants for further investigation of the granins' function.

Chapter 5

CgA gene study

Preface

Besides regulational and functional studies of the CgA gene, we also began to analyze the complete sequence of the gene. Due to previous work in our lab, including cloning of the human CgA gene and isolation of the mouse CgA gene for targeting vector construction, the complete sequences of human and mouse CgA gene could be obtained. We then compared the two sequences and obtained some information encoded within the sequences. In addition, since zebrafish genetic approaches are currently being widely applied to gene functional analysis, we attempted to map the CgA gene on the zebrafish genome.

5.1 Zebrafish CgA gene mapping

5.1.1 Introduction

The zebrafish is an excellent genetic system for the study of vertebrate development and disease. It has become a powerful approach for gene functional analysis to randomly mutate the DNA of thousands of zebrafish and then screen their offspring for embryonic defects and relate these defects to specific genes. Such a method was originally developed in fruit flies in the late 1970s. The particular value of the transparent, externally developing zebrafish embryo is the ability to screen for vertebrate-specific attributes, which are not present in fruit flies and more difficult to visualize in the intra-uterine mouse.

In order to relate the specific phenotype of the mutated embryo to genes, a detailed gene map has to be constructed. The haploid zebrafish genome has 25 chromosomes, containing 1.7×10^9 base pairs of DNA. The first genetic map of the zebrafish included about 400 genetic markers, mostly random amplified polymorphic DNAs (RAPDs), along with a few genes and mutations. RAPD markers were rapid and cost-effective, but unable to be used across different strains. Recently, a microsatellite genetic linkage map was constructed with about 700 simple sequence length polymorphisms (SSLPs) markers that were more consistent from laboratory to laboratory and from strain to strain. Since the localization of cDNAs and expressed sequence tags (ESTs) on these meiotic maps requires the identification of polymorphisms, the alternative technology of somatic cell fusion or radiation hybrid mapping were widely applied to gene mapping.

The fusion of cells from two different species results in a hybrid cell which is usually unstable and tends to lose chromosomes from one of the two parental species[76]. In the case of zebrafish and mouse somatic cell fusion[36], the hybrid cell maintains at least one zebrafish chromosome in the entire mouse genomic background. The reduction in complexity of the donor DNA in such hybrids enables assignment of genes to chromosomes. Although whole cell fusion has been used for a few decades, this technique can only provide chromosomal assignments of specific markers or very low resolution maps since only a few large donor DNA fragments (up to entire chromosomes) are present in each hybrid. Recently, production of radiation hybrids (RHs) has allowed a greater degree of precision in locating genes on chromosomes[81]. RHs are generated by irradiating cells from a donor species (e.g. the zebrafish), causing random chromosomal breaks, and fusing these to a cell line from a different species, such as the mouse. Donor-cell chromosome fragments are retained to different extents in the ensuing hybrid cells. Typing a panel of RHs with PCR-based sequence-tagged sites creates an RH map in which the frequency of breakpoints between two markers is proportional to the distance between them.

5.1.2 Mapping the CgA gene in zebrafish genome

- **LN54 RH panel**

The aim of mapping the CgA gene in zebrafish is to link a CgA gene mutation to a phenotype of the mutated zebrafish and try to obtain some clues to CgA gene function in this organism. Mapping is based on PCR amplification of the appropriate CgA gene fragment in a panel of RHs. In our study, the panel of RHs used was a newly developed

panel LN 54[66]. This panel of hybrid cells was produced by fusion of irradiated zebrafish AB9 cells (derived from fin amputation of adult zebrafish AB strain) with mouse B78 melanoma cells. There are 93 RH cell lines in the panel with 22% overall retention of zebrafish sequences. Characterized with 235 markers, which consists of SSLP markers, ESTs and cloned genes, the total genome coverage of the LN54 panel was found to be 88%. Comparison of marker positions in the LN54 RH map and the genetic meiotic maps indicated a 96% concordance. Therefore, the LN54 RH panel should be an effective tool for the mapping of a gene in the zebrafish genome. However, there are still four gaps found on the framework map of the panel, which may indicate the incomplete retention of sequences from the zebrafish.

- **Primer design & PCR amplification**

The procedure of mapping a gene by RH panel includes PCR amplification of an appropriate length of the target gene in each of the cell lines of the panel, comparison of the pattern with that of the markers, finding the related markers and locating the gene on the map. Since the entire gene sequence is usually unknown, the ESTs are always used for the primer design. Therefore, amplified sequence should be within an exon or across two short introns.

To map the CgA gene in the zebrafish genome, a pair of primers was designed according to zebrafish CgA-like EST sequences from the gene bank. There are 5 ESTs homologous to the CgA gene in the gene bank, two of which are similar to the 5' region of the gene and three of which are similar to that of the 3' region. By comparing the amino acid sequences of the ESTs to that of the mouse CgA, the potential exon junctions

were found. The 5' ESTs contain exon1 to 4 of the CgA gene and the 3' ESTs contain piece of exon7 and exon8. According to the mouse CgA gene structure, exons 2, 3, 4 and the coding region of exon1 are of small size, with a maximum of 96bp; and the portions of exon7 and exon8 in the ESTs are also not large, the maximum of which is 158bp in exon7. Therefore, we decided to use a cross-intron strategy. The mouse CgA gene structure indicates a short distance between exon1 and exon2, therefore the primer sequences were designed adjacent to the potential boundary of exon1 and 2. The sequences are : 5'-end primer: 5'-**GCTCTCCTGCTCCTGGTTAATATT**-3'; 3'-end primer: 5'-**CATGTGACCTGGAGAAACTGG**-3'.

To verify the feasibility of the strategy, the primers were used to amplify zebrafish genomic DNA and the PCR product was cloned into the PCR2.1 vector using the TA cloning kit (Promega). The PCR conditions is: 94°C 30 sec, 58°C 30, 72°C 40sec. The insert was sequenced to confirm its identity. The primers then were sent to Dr.Marc Ekker's lab at the Loeb Research Institute at University of Ottawa and used for RH panel mapping with the LN54 panel. The PCR reaction contains 2 ul 10x reaction buffer, 2 µl 2mM dNTP, 0.25 µl oligo (20µM), 0.1 µl Taq (10U/µl), 5 µl DNA (20 ng/µl), and the final volume is 20 µl. The thermal cycler used was Tetrad DNA engine (PTC 225) by MJ Research. The program is 94°C 4 min 1 cycle, 94°C 30 sec, 30°C sec, 72°C 30 sec, 30 cycles. Then 72°C 7 min. The PCR products were electrophoresed on 1% agarose gel, (Fig 5.2) and the results were analyzed with the software called "RH MAPPER" through the internet at the web site: zfish.wustl.edu

5.1.3 Results and Discussion

Based on the sequence of the mouse CgA gene, a pair of primers was designed to amplify intron 1 of zebrafish CgA gene. The size of the PCR product was 1,164 bp (Fig. 5.1). The underlined letters represent the primer sequences and the bold letters represent the exon sequences in the EST. When designing the primers, 3' ends of the primers were located a few base pairs away from the exon/intron boundaries of the CgA EST sequences, so that the conserved exon sequences can be used for the confirmation of the correct amplification. We can see that the correct EST sequences appear downstream of the primers in the amplified product. Thus, our strategy amplified intron 1 sequence of the zebrafish CgA gene. The length of zebrafish intron 1, which is 1.1 kb, is greater than that of mouse CgA intron 1, which is 0.72 kb.

After confirming the success of intron 1 amplification, the primers were sent to Ottawa for CgA gene mapping. The mapping result is shown in Table 5.1 and Fig 5.2.

Zebrafish CgA In1

```
   1 GCTCTCCTGC TCCTGGTTAA TATTGGTAAG TGTCTAAGAT ATATTAAAGA TCATGCTGTT
  61 GGAATACTTT GCGTTAATTT GGATGACTTA TTGACTGAGT TTACTTGCAC ATTTTATATC
 121 ATCCATGTGA GCTGTGCAAT GTCTCCGTCA GCACCAAGGA CAGCGAAGAA ACACTTGAGA
 181 GGAACACAAA CGGGCTTTTC ACACTTGACA TTTGTTTCCG CAGTGATTTC ACACTTCTCA
 241 TTGTTAATGC GTTATTTTGC GTTAGTAATT AGTTCTTCTT TTTCGAGATG TCATACTGTT
 301 ATTCTTCTTT ACTTCACATA TAATGTCAGT CAGGAACATT TATCAGCCAG AAATCTTTGT
 361 CTTTCTTTGT AAATCTGCCT TCTGGTTTAC TTAAACTAAT AATCTGTCCG TTAATGGCTT
 421 CCCTTTGTTG TTCGACATGT TTGACATTAT AAGCAGCCCG CGAAAGTTTC ACTACAAAGC
 481 ACATGAATAT TTAATGAGAT GACCGCTGAA CAGTTTATGA TTGGTTGTTT GTTGCTAAAC
 541 GCTCACCATA CCTTTGTTAC TTAACCTGAA AGCTAAAAGA CAGAAGTTAA AGGGCAGTTC
 601 ACTCAAAAAT GAAAATGTAT TCATTCACCC TCAAGTGACT CCAAACCTTT ATAAGTTTTT
 661 TTCTCTGCTG ACAATAAAAA AAAGATATAT TGAAGAGAGC TGGAAACGTG TAAACATTGA
 721 CATCCATTTG TTGTTCATAC TATGGAAGTA AATGGTTAGA AAATTTGTTT AAAATATCTT
 781 CTCTAGTGTT CAGCAGAAGG AAAGAAGTCA AACTGGTTTG GAACAAGTGG AGGTTGAGTA
 841 AATAATGACT AAAATTTTAG TTTTGGGTGA ACTATCCCTT TAAGTCAATG GTCTCAAACT
 901 CAATTCCAGG AGGGCTACAG CTCTGCACAG TTTAGCTCCA GCCACCTCTA ACTCACACCT
 961 GCTTAATAGT CTCTTAAAGA CCTTGAGTAG GGTTGGAGCA AAACTGTGCA GAGCTGCGGC
1021 CCTCTAGGAA TCCATTTGA GACCTGTGTT AACCCAAGTT AACCCATGTG AAGGTAATTG
1081 CATTATAACA CCAGGCTTCT GTTAATCAAA GCCTTTACTC TTTTTTTTCC AGTCATCTCC
1141 GTGCCAGTTT CTCCAGGTCA CATG
```

Bold letters: exons (exon 1 and exon 2 fragments)
Underline letters: primer sequence

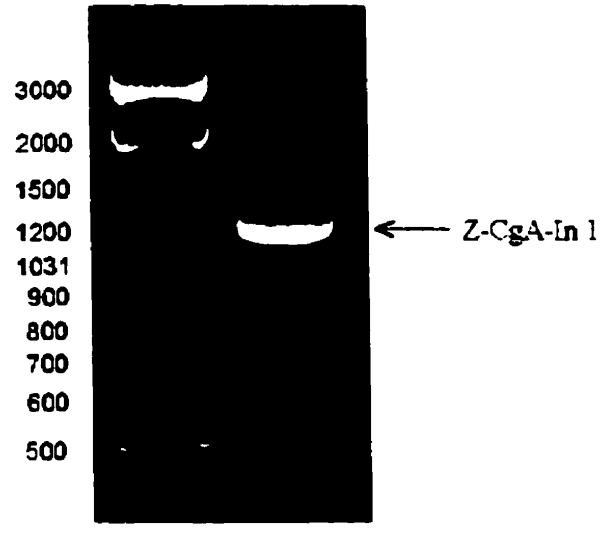

Fig 5.1 Sequence and electrophoresis of Zebrafish CgA gene intron 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	2	4	5	(6)	8	9	**11**	13	16	18	19
B	23	25	(27)	29	30	33	**36**	39	**40**	41	47	48
C	(49)	50	52	54	57	59	63	65	66	70	**73**	74
D	79	80	83	84	85	86	87	88	89	91	92	96
E	97	98	101	**104**	105	(106)	(108)	**109**	114	117	119	121
F	123	125	132	135	136	137	138	150	151	152	153	154
G	169	174	175	176	178	182	183	184	190	300	**301**	**302**
H	303	304	305	306	308	**309**	310	311	**312**	**mix**	**AB9**	B78

Table 5.1 mapping result of Zebrafish CgA gene: positive hybrids are bold & underlined

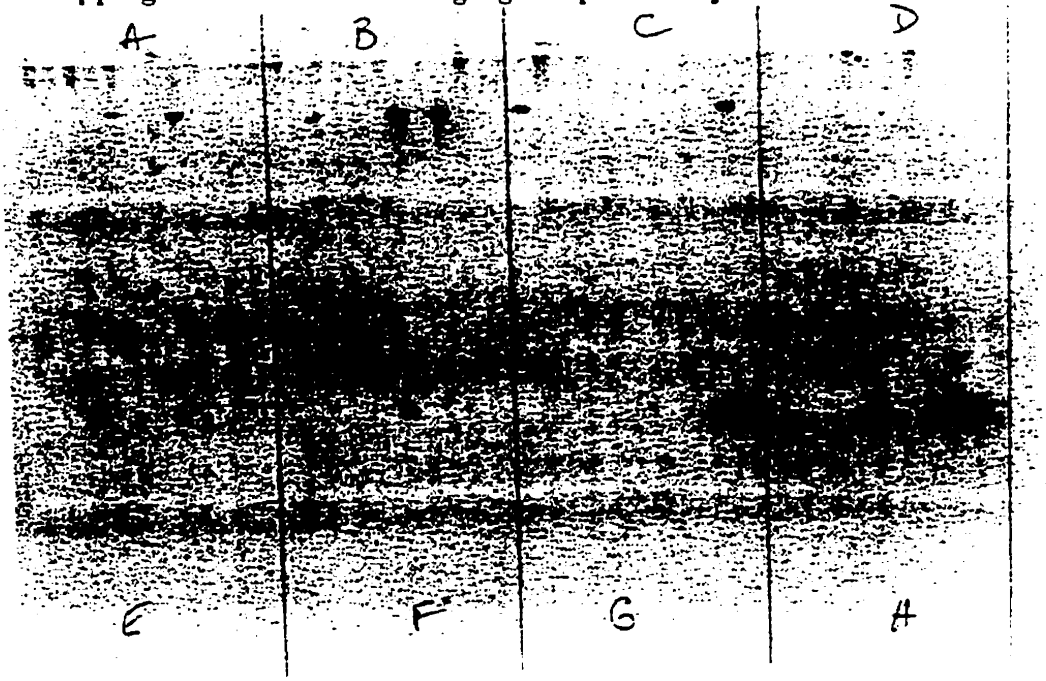

Fig 5.2 PCR of Z-CgA-In1 in LN 54 RH panel

Table 5.1 lists the LN54 hybrid clones used in the mapping. **Fig 5.2** shows the corresponding PCR results. The first 93 clones are the hybrid clones. H10 is the mixture of zibrafish AB9 cells and mouse B78 cells; H11 is AB9; H12 is B78. The genomic DNA of these 96 cell cultures was used as the template to amplify Zebrafish CgA intron 1 fragment. H10 and H11 were used as the positive controls and gave a strong band with expected size. H12 is the negative control. The hybrid clones that gave positive signals on the gel are indicated as bold, underline numbers in the table. The clones that gave weak signals are indicated as numbers in parentheses.

Table 5.1 shows the number of the RH clones. The positive clones of CgA mapping are indicated in bold and underlined fonts, and the uncertain semi-positive clones are shown in parentheses. Fig 5.2 shows the PCR results of the mapping. The last three cell lines are the controls, including parental cell lines Zebrafish AB9 cells, mouse B78 cells and the two cells' mixture DNA. The results show the specific amplification of the band in Zebrafish AB9 cells and mixed DNA but not in the mouse B78 cells. Therefore, the amplification is specific and reliable.

The PCR results were analyzed with the software RH MAPPER. We were disappointed to receive information that the mapping signal is not related to any known markers. It means that the CgA gene can not be mapped in this panel of RH cell lines and another strategy should be applied.

The reason that the mapping failed may be that there are too few markers mapped at the region where the CgA gene is located. This situation may be due to sequence features in this region. Other mapping methods should be applied instead.

The other reason for the failure of the mapping maybe due to incomplete coverage of the zebrafish genome by the LN54 panel. First, the overall retention of the LN54 panel is 22%, which is barely within the parameters suggested for full coverage of a genome (20-50% retention for 100 hybrid lines). Therefore, some linkage groups have even lower retention. If the CgA gene is located in such a linkage group, the nearby markers may not be retained in the panel and therefore there is not enough information for the mapping. In this case, other RH panels should be used.

5.2 Comparison of Human and Mouse CgA gene

The entire human and mouse CgA genes have been sequenced in our lab. The human CgA gene sequence was obtained by screening a human fetal liver genomic DNA library[97]. It contains the entire coding region plus 2.3kb upstream of the exon1 and 1.3kb downstream of exon8 (App. 1). The mouse CgA gene sequence was obtained from mouse strain 129sv/J female kidney genomic DNA library. The entire sequence that we have consists of the sequence including 8.8kb upstream of exon1 and 4.5kb downstream of exon8 (App. 2).

Most of the genetic information is encoded in DNA sequences. Besides coding sequences, there are numerous regulatory elements and a variety of repetitive sequences in the genome that are responsible for gene regulation and polymorphism, which are critical for our understanding of the development and evolution of the CgA gene. On the present study, we intend to investigate the sequences of introns and flanking regions of the CgA gene to obtain more information on gene regulation and evolution.

5.2.1 Methods of comparing the hCgA and mCgA genes

The hCgA and mCgA gene in λ phage clones were digested and the resulting fragments were subcloned in the Bluescript KS vector for sequencing. (The quality of the sequencing data for flanking regions and introns is such that the sequence is presently 95% accurate. Further analysis will be require to provide a definitive result.) The subclones of human and mouse CgA gene were mapped and reunited as the entire gene

sequence. The sequences of these two genes were compared by using NCBI blast, and regions with more than 80% identities are shown (Fig 5.3). The result shows that the homologous regions are mostly in the exons, especially exons 2,3,4,7 and part of exon 8, which correspond to the conserved N- and C- terminal important functional regions of the CgA protein molecule. Interestingly, there are several highly homologous regions scattered thoughout the introns and the flanking regions. Usually, the important regions are highly conserved. However, there are abundant repetitive sequences in the mammalian genome, which may also be highly conserved. Therefore, to eliminate the effect of the repeat sequences, it is better to identify them first. In addition, the repetitive elements may affect gene function and regulation, and contain information of evolution, which is also very interesting and worthwhile to study. Thus, the interspersed elements in hCgA and mCgA were found by software Repeatmasker, and the results are shown in Fig 5.4.

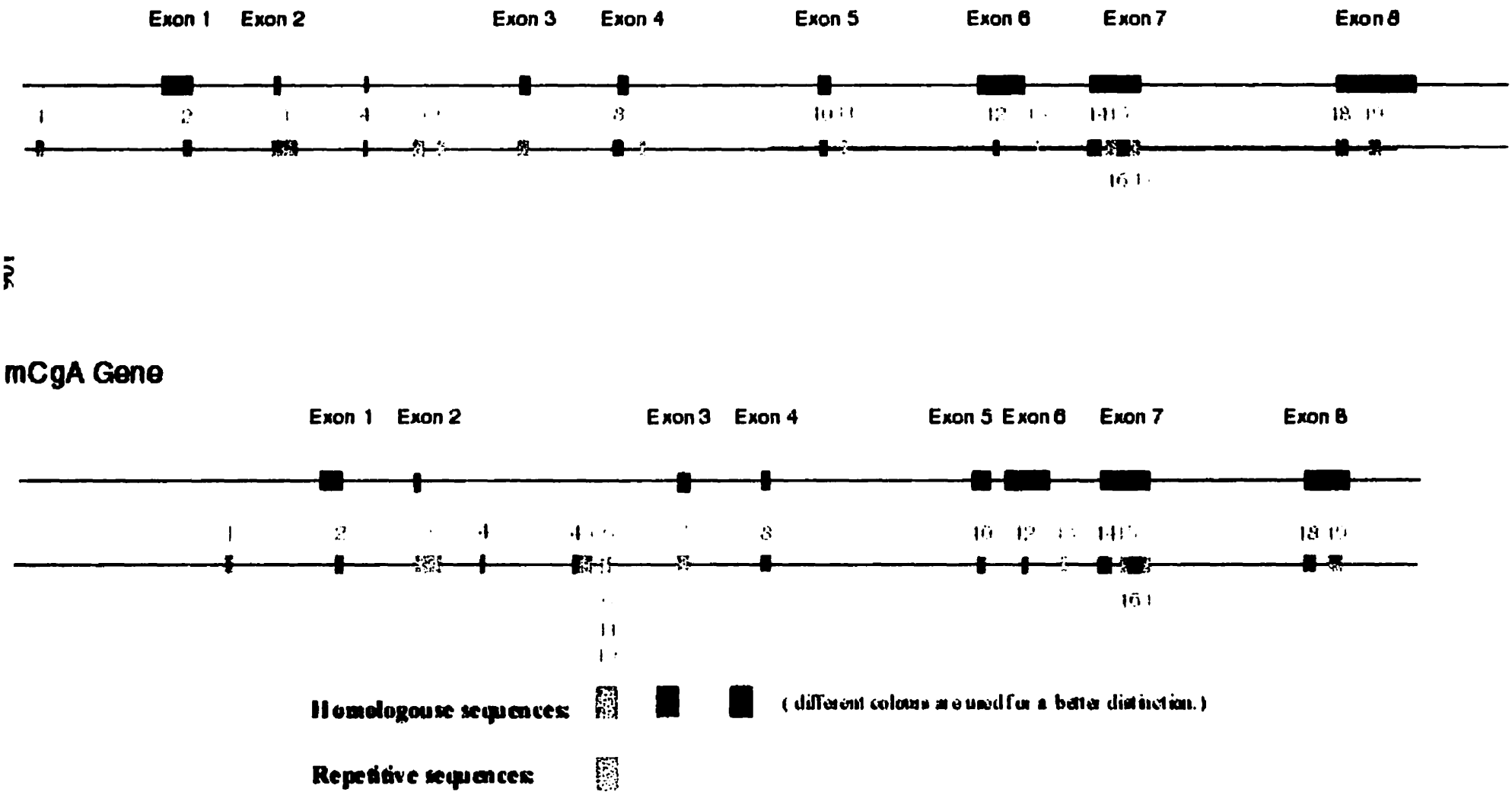

Fig 5.3 High conserved regions between hCgA and mCgA hCgA: upper line: schematic diagram of hCgA, indicating the exon-intron arrangement. **lower line:** conserved regions (>80% homology) with mCgA, marked with numbers corresponding to those marked on mCgA. **mCgA: upper line:** schematic diagram of mCgA, indicating the exon-intron arrangement. **lower lower line:** conserved regions (>80% homology) with hCgA, marked with numbers corresponding to those marked on hCgA.

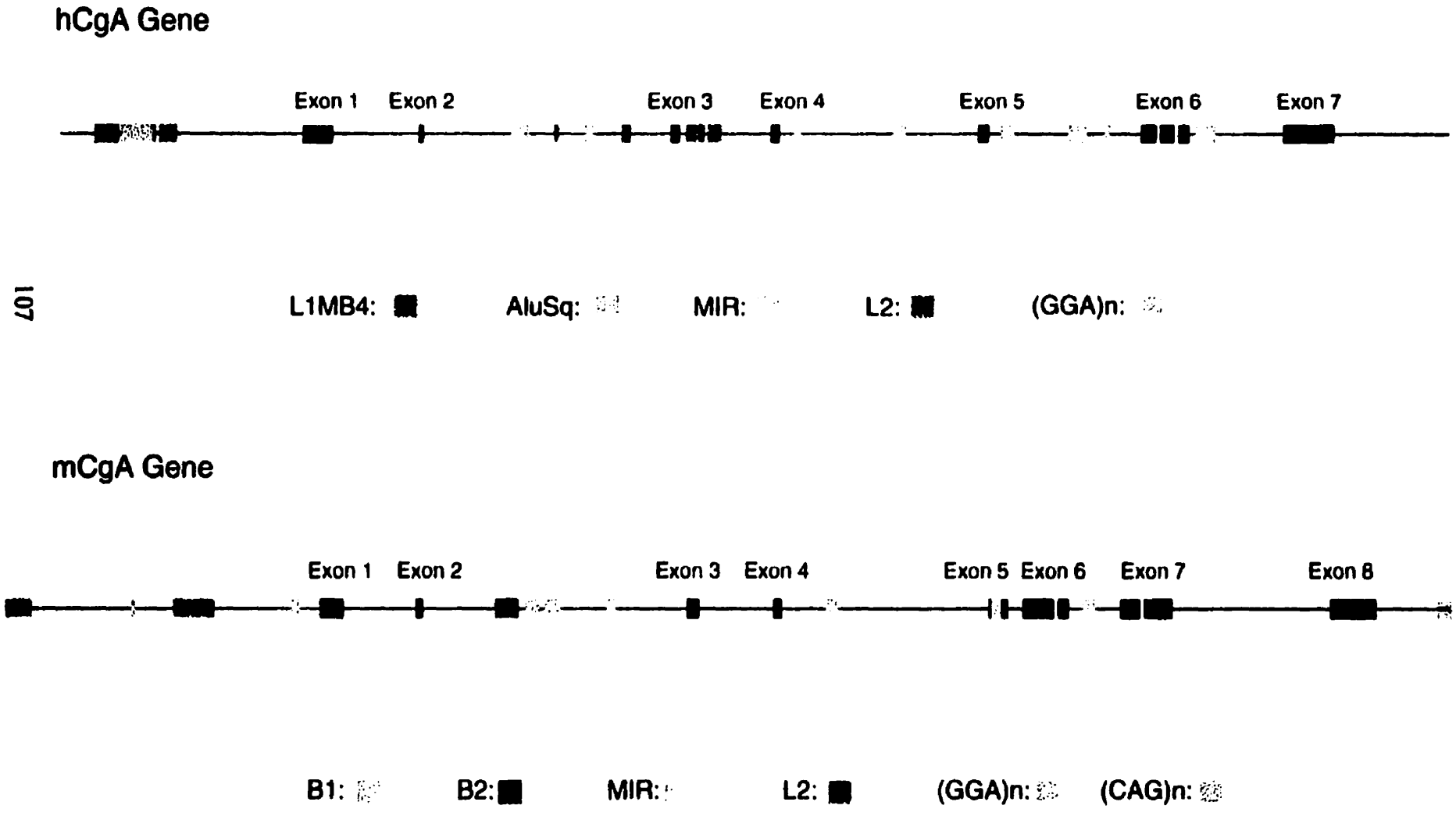

Fig 5.4 Interspersed repetitive sequences in the hCgA &mCgA gene coding region
Black box: exons; colour box: different interspersed repetitive sequences in the region

5.2.2 Discussion

5.2.2.1 Interspersed repetitive elements in the hCgA gene and mCgA gene

In mammals, Only 10% of the genome at the most represents the coding regions. Much of the remainder consists of a variety of repetitive sequences including tandemly repeated sequences (simple repeats, satellites, etc.) and interspersed repeats (IRs)[136]. The tandemly repeated sequences arise primarily through slippage of the replication machinery and/or unequal exchange during genetic recombination[32, 113], and the interspersed repeats are the fossils of transposed elements, which are the sequences that can create copies of themselves at other sites in the genome[136]. The fossilized progeny of transposable elements may well comprise more than half of our genome, although many are too old and divergent to be recognized as such. The mammalian genome could be compared, somewhat poetically, with a coral reef, in which the transposable elements are the coral, the reef is built of the fossils of their ancestors, and the genes are the inhabiting fish, anemones, seastars and so on[134].

In mammals, DNA transposition can occur via reverse transcription of an RNA intermediate, retroposition, or via excision and reintegration of the DNA itself (DNA transposition).

Retroposed repeat elements fall into three categories: short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), or retroviruse-like elements[136]. SINEs are usually 100-400 bp long and harbor an internal (mobile) polymerase III promoter. The best examples of SINEs are MIR (mammalian-wide

interspersed repeat[135], Alu element in primate and B1 in rodent[149]. The LINEs are 6-8 kb and encode a reverse transcriptase and other proteins necessary for retrotransposition. The most abundant LINEs in the mammalian genome are L1 and L2. It has been proposed that reverse transcriptase recognizes the 3' end of the LINE transcript and initiates reverse transcription simultaneously with integration by using nicked genomic DNA as a primer. The fact that SINEs and LINEs share a variable length insertion duplication site and a poly A or simple repeat tail suggests that SINE mobility is dependent on proteins provided by LINEs. The most typical example is that MIR, a mammalian specific SINE element shares a 50bp sequence at the 3'-end of L2, a LINE-like element. The equal distribution of MIR and L2 in human genome indicates their related transposition[135]. Retrovirus-like elements (RLEs) have long terminal repeats (LTRs), which carry the transcriptional regulatory sequences and maintain the full length reproduction[136]. Besides retroposons, DNA transposon fossils make up more than 1.5% of human genome[5]. These elements, flanked by short inverted terminal repeats, move by excision and reintegration. The 'cut' and 'paste' mechanism of DNA transposition can lead to element duplication when the gap at the site of excision is repaired using the sister chromatid.

In mammals, the most abundant SINEs are the Alu/B1 family of SINEs derived from 7SL RNA, which is the small cytoplasmic RNA in the signal recognition particle (SRP) that promotes translocation of newly synthesized secretory protein across the endoplasmic reticulum. There is a 2 bp mutation in the 7SL RNA which creates a Pol III promoter upstream of the element[116]. Alu elements are members of family in human genome. There are around a million copies of Alu elements in the genome that can be

classified into several subfamilies. They are usually heterodimers around 300bp in size. However, the most ancient subfamily of Alu elements is a monomer that is homologous to the left Alu sequence, and named Free Left Alu or FLA[117]. The other dimeric subfamilies are classified and the evolution of these subfamilies is shown in Fig 5.5. Since Alu elements are the most widespread IRs in the human genome, the homologous recombination between these sequences and the continuous insertion of these elements in various portion of the genome may cause mutations in genes and lead to genetic disease. For example, the insertion of an Alu element at codon 876 of calcium sensing receptor gene leads to a truncated mutant form of the receptor, resulting in an inactivation of the calcium sensing receptor and causes familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism[6]. In addition, it has been reported that the recombination of Alu family members causes deletions in the genes for the LDL receptor[84], α-globin[103], β-hexosaminidase[100] and α-galactosidase[78], as well as duplications in the genes of LDL receptor[85], Duchene muscular dystrophy[65], hypoxanthine phosphoribosyltransferase[146] and ALL-1[126]. Therefore, Alu elements are the most intensively studied IRs in human genome. However, ironically, the transpositional mechanism of the Alu elements is still unknown. By using the Repeat Masker software developed by Dr. A.F.Smit, human and mouse CgA genes were analyzed to reveal the interspersed repeat elements within the entire sequences. The result (Fig.5.4) shows that in the human CgA gene, there is just a single Alu/Sq sequence at the 5'-flanking region of the gene. On average, there should be a Alu element every 3 kb of the genome sequence. This human CgA coding region is 13kb, and it is expected to have 4 Alu elements. Therefore, it is considered to be a low Alu element region. Since Alu elements,

like other SINEs, tend to be distributed in clusters, the hCgA gene is not in a cluster region of Alu integration. Interestingly, the Alu element that is present, is inserted in a L1 3'-end sequence (Fig 5.4). Although we don't know the mechanism of Alu retrotransposition, in this case, this special Alu element can use the L1 retrotransposition machinery to complete its self-reproduction together with this L1 sequence. In mCgA, however, there are 5 B1 elements in the 23kb coding region, much more than in hCgA.

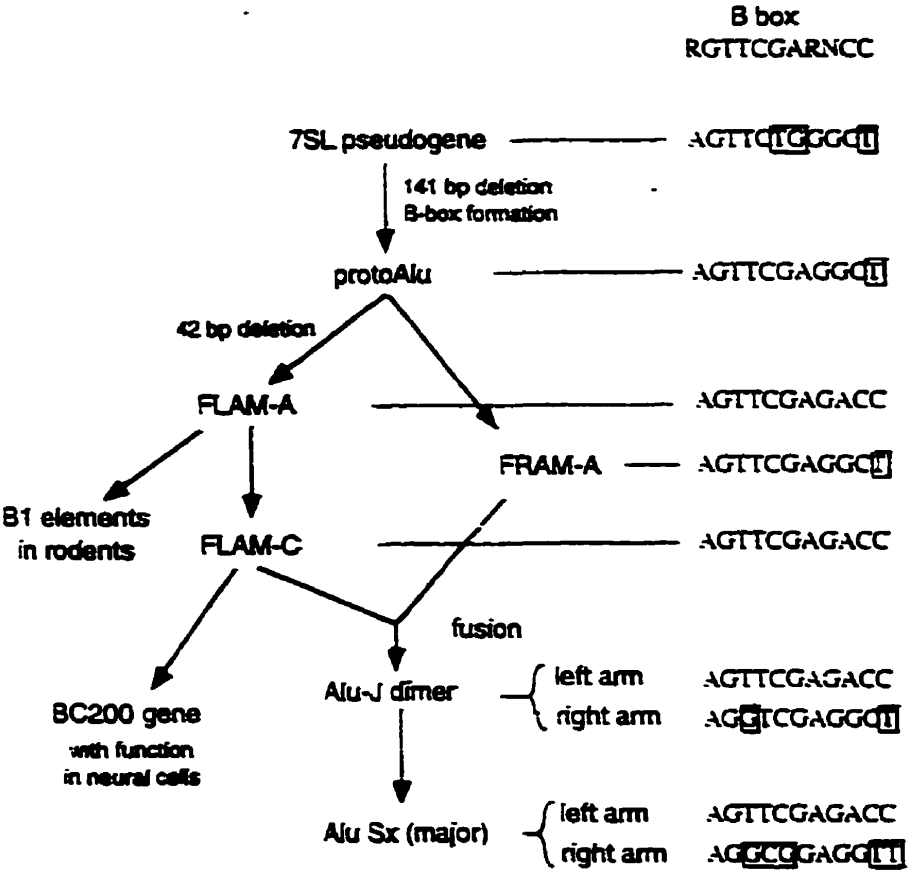

Fig 5.5 The evolution of Alu family members (taken from ref.134) Shows the possible origin of Alu repetitive sequences. The Pol III B box region of each subfamily is shown at the right. The highly conserved sequences are boxed. The Alu-Sx or Major subfamily is the most abundant Alu in the human genome.

On the other hand, the second most widespread SINE, MIR, appears to be more abundant in this area. MIRs are classic tRNA-derived SINEs that were amplified before the mammalian radiation. The entire sequence of MIR is 260bp, with a conserved 70bp long consensus sequence at the middle[135]. There are 12 MIR fragments in hCgA and 5 in mCgA gene coding region. Usually in primates, an estimated 300,000 copies of MIR are still discernible, which account for 1 to 2% of the genome. However, in hCgA, 8% of the sequence consists of MIR fragments. It means that hCgA is in a MIR rich region. Comparison of the hCgA with mCgA (Fig 5.3) shows that some of the highly conserved regions of these two genes are MIR fragment sequences, such as the fragment number 6 upstream of the exon3 and fragment number 13 downstream of exon 6 of both genes. The integration of these fragments may be earlier than the divergence of the rodent and the primate. As mentioned before, the replication of MIRs is probably dependent on the machinery coded within L2. We can see some L2 elements littered in the CgA gene coding region. As for other IRs, although the retrovirus-like elements account for 2% of the genome, there is none in either hCgA or mCgA gene. In addition, no DNA transposon sequence occurs in this region.

5.2.2.2 Transcription regulation elements in the conserved regions

Besides the IRs, there are two highly conserved regions between the hCgA and the mCgA gene. One is a 62bp sequence about 1 kb upstream of the human and mouse CgA exon1 and the other is a ~200bp sequence in intron2. We expect there may be some important regulation elements in these conserved regions. Therefore, a software named MatInspector [based on TRANSFAC database] search was applied to reveal any clues to

this expectation. TRANSFAC is a database of transcription factors, their genomic binding sites and DNA-binding profiles. It contains information about *cis*-regulatory DNA sequences and *trans*-acting factors in a computer-readable format and is released for public use.

The results show the potential transcription binding sites in these two regions. Both of the 62bp sequences at the 5'-flanking region in hCgA and mCgA contain a AP1_Q2 sequence (consensus sequence: TGACG; in hCgA is TGACG at –1071 from the transcription start site; in mCgA is TGACA at –904 from the transcription start site), which is a sequence motif critical for binding of a group of cellular transcription factors (ATF, CREB, E4F, and EivF) and for activation of certain E1a-inducible and cyclic AMP (cAMP)-inducible promoters[83]. It is shown to be an enhancer of several cAMP-dependent genes, such as human vasoactive intestinal polypeptide (VIP) gene [148] and human CYP21A2 gene.[152] Since CgA is also a cAMP-inducible gene and its neuroendocrine-specific expression is also dependant upon the presence of a CRE, this conserved element may play a role in such a regulation. Further studies focussed on this region may answer this question.

The other conserved region is a 200bp sequence just downstream of exon 2. In these regions, two general transcription factor AP4 sites occur at +1179 and +1242 in the hCgA gene and +1028 and +1095 in the mCgA gene. In addition, there is a NF1 (nuclear factor 1) site at +1205 in hCgA and +1054 in mCgA. Whether these sequences are really functional and how they affect the CgA gene expression are not known. Further studies on the deletion or mutation of these sequences may address their function.

Part III

References

1. Aardal, S., and Helle, K. B. (1992). The vasoinhibitory activity of bovine chromogranin A fragment (vasostatin) and its independence of extracellular calcium in isolated segments of human blood vessels. Regul Pept *41*, 9-18.

2. Angeletti, R. H., and Hickey, W. F. (1985). A neuroendocrine marker in tissues of the immune system. Science *230*, 89-90.

3. Arvan, P., and Castle, D. (1998). Sorting and storage during secretory granule biogenesis: looking backward and looking forward. Biochem J *332*, 593-610.

4. Kozak, M., (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol *196*, 947-950

5. Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M. H., Meunier-Rotival, M., and Periquet, G. (1995). Human and other mammalian genomes contain transposons of the mariner family. FEBS Lett *368*, 541-6.

6. Bai, M., Janicic, N., Trivedi, S., Quinn, S. J., Cole, D. E. C., Brown, E. M., Hendy, G. N. (1997). Markedly reduced activity of mutant calcium-sensing receptor with an inserted Alu element from a kindred with familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. J. Clin. Invest. *99*. 1917-25

7. Barbosa, J. A., Gill, B. M., Takiyyuddin, M. A., and O'Connor, D. T. (1991). Chromogranin A: posttranslational modifications in secretory granules. Endocrinology *128*, 174-90.

8. Barinaga, M. (1994). Knockout mice: round two [news; comment] [published erratum appears in Science 1994 Aug 12;265(5174):855]. Science *265*, 26-8.

9. Barkatullah, S., Curry, W., Johnston, C. (1997). Ontogenetic oexpression of chromogranin A and its derived peptides, WE-14 and pancreastatin, in the rat neuroendocrine system. Histochem Cell Biol 107, 251-5.

10. Benedum, U. M., Lamouroux, A., Konecki, D. S., Rosa, P., Hille, A., Baeuerle, P. A., Frank, R., Lottspeich, F., Mallet, J., and Huttner, W. B. (1987). The primary structure of human secretogranin I (chromogranin B): comparison with chromogranin A reveals homologous terminal domains and a large intervening variable region. EMBO J *6*, 1203-11.

11. Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-6.

12. Brand, A. (1995). GFP in Drosophila. Trends Genet *11*, 324-5.

13. Braselmann, S., Graninger, P., and Busslinger, M. (1993). A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. Proc Natl Acad Sci U S A *90*, 1657-61.

14. Bronson, S. K., and Smithies, O. (1994). Altering mice by homologous recombination using embryonic stem cells. J Biol Chem *269*, 27155-8.

15. Bronson, S. K., Plaehn, E. G., Kluckman, K. D., Hagaman, J. R., Maeda, N., and Smithies, O. (1996). Single-copy transgenic mice with chosen-site integration [see comments]. Proc Natl Acad Sci U S A *93*, 9067-72.

16. Burdon, T., Chambers, I., Stracey, C., Niwa, H., and Smith, A. (1999). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. Cells Tissues Organs *165*, 131-43.

17. Burdon, T., Stracey, C., Chambers, I., Nichols, J., Smit, A. (1999) Suppression of SHP-2 and ERK signaling promotes self-renewal of mouse embryonic stem cells. Dev Biol 210, 30-43

18. Canaff, L., Brechler, V., Reudelhuber, T. L., and Thibault, G. (1996). Secretory granule targeting of atrial natriuretic peptide correlates with its calcium-mediated aggregation. Proc Natl Acad Sci U S A *93*, 9483-7.

19. Canaff, L., Bennett, H.P.J., and Hendy, G.N. (1999) At the cutting edge. Peptide hormone precursor processing: Getting Sorted? Mol Cell Endocrinol 156, 1-6

20. Canaff, L., Bennett, H.P.J., Hou, Y., Seidah, N.G., and Hendy, G.N. (1999) Proparathyroid hormone processing by the proprotein convertase PC7: comparison with furin and assessment of modulation of parathyroid convertase mRNA level by calcium and 1,25-dihydroxyvitamin D_3. Endocrinology 140, 3633-41

21. Canaff, L., Bevan, S., Wheeler, D. G., Mouland, A. J., Rehfuss, R. P., White, J. H., and Hendy, G. N. (1998). Analysis of molecular mechanisms controlling neuroendocrine cell specific transcription of the chromogranin A gene. Endocrinology *139*, 1184-96.

22. Capecchi, M. R. (1989). Altering the genome by homologous recombination. Science *244*, 1288-92.

23. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-5.

24. Chanat, E., Weiss, U., Huttner, W. B., and Tooze, S. A. (1993). Reduction of the disulfide bond of chromogranin B (secretogranin I) in the trans-Golgi network causes its missorting to the constitutive secretory pathways. Embo J *12*, 2159-68.

25. Chiocchetti, A., Tolosano, E., Hirsch, E., Silengo, L., and Altruda, F. (1997). Green fluorescent protein as a reporter of gene expression in transgenic mice. Biochim Biophys Acta *1352*, 193-202.

26. Joshi, R. L., Lamothe, B., Cordonnier, N., Mesbah, K., Monthioux, E., Jami, J., Bucchini, D. (1996). Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality. EMBO Journal. *15*, 1542-7

27. Cool, D. R., Normant, E., Shen, F., Chen, H. C., Pannell, L., Zhang, Y., and Loh, Y. P. (1997). Carboxypeptidase E is a regulated secretory pathway sorting receptor: genetic obliteration leads to endocrine disorders in Cpe(fat) mice. Cell *88*, 73-83.

28. Cowley, D. J., Moore, Y. R., Darling, D. S., Joyce, P. B. M., and Gorr, S. U. (2000). N- and C- terminal domains direct cell type-specific sorting of chromogranin A to secretory granules. J Biol Chem *275*, 7743-8.

29. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol *14*, 315-9.

30. Dani, C., Chambers, I., Johnson, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J., Smith, A. (1998) Paracrine induction of stem cell renewal by LIF-deficient cells: A new ES cell regulatory pathway. Dev Biol 203. 149-62

31. Deftos, L. J., Hogue-Angeletti, R., Chalberg, C., and Tu, S. (1990). A chromogranin A-derived peptide differentially regulates the secretion of calcitonin gene products. J Bone Miner Res *5*, 989-91.

32. Dover, G. A. (1993). Evolution of genetic redundancy for advanced players. Curr Opin Genet Dev *3*, 902-10.

33. Drees, B. M., Rouse, J., Johnson, J., and Hamilton, J. W. (1991). Bovine parathyroid glands secrete a 26-kDa N-terminal fragment of chromogranin-A which inhibits parathyroid cell secretion. Endocrinology *129*, 3381-7.

34. Drees, B. M., and Hamilton, J. W. (1992). Pancreastatin and bovine parathyroid cell secretion. Bone Miner *17*, 335-46.

35. Dymecki, S. M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. Proc Natl Acad Sci U S A *93*, 6191-6.

36. Ekker, M., Speevak, M. D., Martin, C. C., Joly, L., Giroux, G., and Chevrette, M. (1996). Stable transfer of zebrafish chromosome segments into mouse cells. Genomics *33*, 57-64.

37. Eriksson, B., Arnberg, H., Oberg, K., Hellman, U., Lundqvist, G., Wernstedt, C., and Wilander, E. (1990). A polyclonal antiserum against chromogranin A and B--a new sensitive marker for neuroendocrine tumours. Acta Endocrinol (Copenh) *122*, 145-55.

38. Eskeland, N. L., Zhou, A., Dinh, T. Q., Wu, H., Parmer, R. J., Mains, R. E., and O'Connor, D. T. (1996). Chromogranin A processing and secretion: specific role of endogenous and exogenous prohormone convertases in the regulated secretory pathway. J Clin Invest *98*, 148-56.

39. Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-6.

40. Fasciotto, B. H., Trauss, C. A., Greeley, G. H., and Cohn, D. V. (1993). Parastatin (porcine chromogranin A347-419), a novel chromogranin A-derived peptide, inhibits parathyroid cell secretion. Endocrinology *133*, 461-6.

41. Feldstein, R. (1998) Construction of a targeting vector for analysis of the neuroendocrine function of chromogranin A. M.Sc. thesis. McGill University

42. Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proc Natl Acad Sci U S A *93*, 10887-90.

43. Fischer-Colbrie, R., Lassmann, H., Hagn, C., and Winkler, H. (1985). Immunological studies on the distribution of chromogranin A and B in endocrine and nervous tissues. Neuroscience *16*, 547-55.

44. Fischer-Colbrie, R., Schmid, K. W., Mahata, S. K., Mahata, A., Laslop, A., and Bauer, J. W. (1991). Sex-related differences in chromogranin A, chromogranin B and secretogranin II gene expression in rat pituitary. J Neuroendocrinol *4*, 126-30

45. Funakoshi, A., Tateishi, K., Shinozaki, H., Matsumoto, M., and Wakasugi, H. (1990). Elevated plasma levels of pancreastatin (PST) in patients with non-insulin-dependent diabetes mellitus (NIDDM). Regul Pept *30*, 159-64.

46. Furth, P. A., St Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., and Hennighausen, L. (1994). Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc Natl Acad Sci U S A *91*, 9302-6.

47. Gasparri, A., Sidoli, A., Sanchez, L. P., Longhi, R., Siccardi, A. G., Marchisio, P. C., and Corti, A. (1997). Chromogranin A fragments modulate cell adhesion. Identification and characterization of a pro-adhesive domain. J Biol Chem *272*, 20835-43.

48. Gerdes, H. H., Rosa, P., Phillips, E., Baeuerle, P. A., Frank, R., Argos, P., and Huttner, W. B. (1989). The primary structure of human secretogranin II, a widespread

tyrosine-sulfated secretory granule protein that exhibits low pH- and calcium-induced aggregation. J Biol Chem *264*, 12009-15.

49. Glombik, M. M., Kromer, A., Salm, T., Huttner, W. B., and Gerdes, H. H. (1999). The disulfide-bonded loop of chromogranin B mediates membrane binding and directs sorting from the trans-Golgi network to secretory granules. Embo J *18*, 1059-70.

50. Godwin, A. R., Stadler, H. S., Nakamura, K., and Capecchi, M. R. (1998). Detection of targeted GFP-Hox gene fusions during mouse embryogenesis. Proc Natl Acad Sci U S A *95*, 13042-7.

51. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A., and Ruddle, F. H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A *77*, 7380-4.

52. Gorr, S. U., Shioi, J., and Cohn, D. V. (1989). Interaction of calcium with porcine adrenal chromogranin A (secretory protein-I) and chromogranin B (secretogranin I). Am J Physiol *257*, E247-54.

53. Gorr, S. U., Hamilton, J. W., and Cohn, D. V. (1991). Sulfated secreted forms of bovine and porcine parathyroid chromogranin A (secretory protein-I). J Biol Chem *266*, 5780-4.

54. Gorr, S. U., Huang, X. F., Cowley, D. J., Kuliawat, R., and Arvan, P. (1999). Disruption of disulfide bonds exhibits differential effects on trafficking of regulated secretory proteins. Am J Physiol *277*, C121-31.

55. Gridley, T., Soriano, P., and Jaenisch, R. (1987) Trends Genet *3*, 162-6

56. Gu, H., Zou, Y. R., and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell *73*, 1155-64.

57. Haley, K. J., Drazen, J. M., Osathanondh, R., and Sunday, M. E. (1997). Comparison of the ontogeny of protein gene product 9.5, chromogranin A and proliferating cell nuclear antigen in developing human lung. Microsc Res Tech *37*, 62-8.

58. Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A. B., and Joyner, A. L. (1995). Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. Science *269*, 679-82.

59. Galindo, E., Zwiller, J., Bader, M. F., Aunis, D. (1994). Correction: Chromostatin Inhibits Catecholamine Secretion in Adrenal Chromaffin Cells by Activating a Protein Phosphatase. Proc Natl Acad Sci U S A *91*, 832.

60. Hasty, P., Ramirez-Solis, R., Krumlauf, R., and Bradley, A. (1991). Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells [published erratum appears in Nature 1991 Sep 5;353(6339):94]. Nature *350*, 243-6.

61. Hendy, G. N., Bevan, S., Mattei, M. G., and Mouland, A. J. (1995). Chromogranin A. Clin Invest Med *18*, 47-65.

62. Hendy, G.N., Bennett, H.P.J., Gibbs, B. F., Lazure, C., Day, R., and Seidah, N. G. (1995). Proparathyroid hormone (ProPTH) is preferentially cleaved to parathyroid hormone (PTH) by the prohormone convertase furin: a mass spectrometric analysis.

63. Hocker, M., Raychowdhury, R., Plath, T., Wu, H., O'Connor, D. T., Wiedenmann, B., Rosewicz, S., and Wang, T. C. (1998). Sp1 and CREB mediate gastrin-dependent regulation of chromogranin A promoter activity in gastric carcinoma cells. J Biol Chem *273*, 34000-7.

64. Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature *326*, 292-5.

65. Hu, Y., and Wang, Z. (1991). Effects of hyperlipidemic sera and lipoproteins on prostacyclin synthesis by rabbit aortic smooth muscle cells. Chin Med Sci J *6*, 48-51.

66. Hukriede, N. A., Joly, L., Tsang, M., Miles, J., Tellis, P., Epstein, J. A., Barbazuk, W. B., Li, F. N., Paw, B., Postlethwait, J. H., Hudson, T. J., Zon, L. I., McPherson, J. D., Chevrette, M., Dawid, I. B., Johnson, S. L., and Ekker, M. (1999). Radiation hybrid mapping of the zebrafish genome. Proc Natl Acad Sci U S A *96*, 9745-50.

67. Huttner, W. B., Gerdes, H. H., and Rosa, P. (1991). The granin (chromogranin/secretogranin) family. Trends Biochem Sci *16*, 27-30.

68. Hutton, J. C., Davidson, H. W., Grimaldi, K. A., and Peshavaria, M. (1987). Biosynthesis of betagranin in pancreatic beta-cells. Identification of a chromogranin A-like precursor and its parallel processing with proinsulin. Biochem J *244*, 449-56.

69. Iacangleo, A. L., Grimes, M., Eiden, L. E. (1991) The bovine chromogranin A gene: structural basis for hormone regulation and generation of biologically active peptides. Mol Endocrinol 5, 1651-60.

70. Inouye, S., and Tsuji, F. I. (1994). Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. FEBS Lett *341*, 277-80.

71. Ischia, R., Lovisetti-Scamihorn, P., Hogue-Angeletti, R., Wolkersdorfer, M., Winkler, H., and Fischer-Colbrie, R. (1997). Molecular cloning and characterization of

NESP55, a novel chromogranin-like precursor of a peptide with 5-HT1B receptor antagonist activity. J Biol Chem *272*, 11657-62.

72. Jaenisch, R., and Mintz, B. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. Proc Natl Acad Sci U S A *71*, 1250-4.

73. Jaenisch, R., Fan, H., and Croker, B. (1975). Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. Proc Natl Acad Sci U S A *72*, 4008-12.

74. Karaplis, A. C. Gene targeting in, Priciples of Bone Biology, Eds. Bilezikian, J. P., Raisz, L. G., Rodon, G. A. Academic Press, San Diego. pp. 1189-201

75. Kent, C., and Coupland, R. E. (1989). Localisation of chromogranin A and B, met-enkephalin-arg6-gly7-leu8 and PGP9.5-like immunoreactivity in the developing and adult rat adrenal medulla and extra-adrenal chromaffin tissue. J Anat *166*, 213-25.

76. Keydar, J., Gilead, Z., Hartman, J. R., and Ben-Shaul, Y. (1973). In vitro production of mouse mammary tumor virus in a mouse mammary tumor ascites line. Proc Natl Acad Sci U S A *70*, 2983-7.

77. Koeslag, J. H., Saunders, P. T., and Wessels, J. A. (1999). The chromogranins and the counter-regulatory hormones: do they make homeostatic sense? J Physiol (Lond) *517*, 643-9.

78. Kornreich, R., Bishop, D. F., and Desnick, R. J. (1990). Alpha-galactosidase A gene rearrangements causing Fabry disease. Identification of short direct repeats at breakpoints in an Alu-rich gene. J Biol Chem *265*, 9319-26.

79. Krisch, K., Buxbaum, P., Horvat, G., Krisch, I., Neuhold, N., Ulrich, W., and Srikanta, S. (1986). Monoclonal antibody HISL-19 as an immunocytochemical probe for neuroendocrine differentiation. Its application in diagnostic pathology. Am J Pathol *123*, 100-8.

80. Kromer, A., Glombik, M. M., Huttner, W. B., Gerdes, H. H., (1998) Essential role of the disulfide-bonded loop of chromogranin B for sorting to secretory granules is revealed by expression of a deletion mutant in the absence of endogenous granin systhesis. J Cell Biol 140, 1331-46

81. Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, R., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N., and Schmitt, K. (1998). Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. Nucleic Acids Res *26*, 3562-6.

82. LaBonne, C., Burke, B., Whiteman, M. (1995) Role of MAP kinase in mesoderm induction and axial patterning during xenopus development. Development 121, 1475-86

83. Lee, K.A., Fink, J.S., Goodman, R.H., Green, M.R. (1989) Distinguishable promoter elements are involved in transcriptional activation by E1a and cyclic AMP. Mol Cell Biol *9*, 4390-7

84. Lehrman, M. A., Schneider, W. J., Sudhof, T. C., Brown, M. S., Goldstein, J. L., and Russell, D. W. (1985). Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science *227*, 140-6.

85. Lehrman, M. A., Goldstein, J. L., Russell, D. W., and Brown, M. S. (1987). Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell *48*, 827-35.

86. Lewis, J. J., Zdon, M. J., Adrian, T. E., and Modlin, I. M. (1988). Pancreastatin: a novel peptide inhibitor of parietal cell secretion. Surgery *104*, 1031-6.

87. Lugardon, K., Raffner, R., Goumont, Y., Corti, A., Delmas, A., Bulet, P., Aunis, D., Metz-Boutigue, M. (2000) Antibacterial and antifungal activities of vasostatin-1, the N-terminal fragment of chromogranin A. J Biol Chem 275, 10745-53

88. MacGregor, R. R., Hinton, D. A., and Ridgeway, R. D. (1988). Effects of calcium on synthesis and secretion of parathyroid hormone and secretory protein I. Am J Physiol *255*, E299-305.

89. Magnani, P., Paganelli, G., Songini, C., Samuel, A., Sudati, F., Siccardi, A. G., and Fazio, F. (1996). Pretargeted immunoscintigraphy in patients with medullary thyroid carcinoma. Br J Cancer *74*, 825-31.

90. Mahata, S. K., O'Connor, D. T., Mahata, M., Yoo, S. H., Taupenot, L., Wu, H., Gill, B. M., and Parmer, R. J. (1997). Novel autocrine feedback control of catecholamine release. A discrete chromogranin A fragment is a noncompetitive nicotinic cholinergic antagonist. J Clin Invest *100*, 1623-33.

91. Mahata, S. K., Mahata, M., Wu, H., Parmer, R. J., and O'Connor, D. T. (1999). Neurotrophin activation of catecholamine storage vesicle protein gene expression: signaling to chromogranin A biosynthesis. Neuroscience *88*, 405-24.

92. Metz-Boutigue, M. H., Garcia-Sablone, P., Hogue-Angeletti, R., and Aunis, D. (1993). Intracellular and extracellular processing of chromogranin A. Determination of cleavage sites. Eur J Biochem *217*, 247-57.

93. Monnat, R. J., Jr., Hackmann, A. F., and Chiaverotti, T. A. (1992). Nucleotide sequence analysis of human hypoxanthine phosphoribosyltransferase (HPRT) gene deletions. Genomics *13*, 777-87.

94. Mouland, A. J. (1993) Structure and functional analysis of the chromogranin A gene and study of the regulation of expression in the parathyroid. PhD. Thesis. McGill University

95. Mouland, A. J., and Hendy, G. N. (1991) Regulation of synthesis and secretion of chromogranin A by calcium and 1,25-dihydroxy cholecalciferol in cultured bovine parathyroid cells. Endocrinology *128*, 441-9

96. Mouland, A. J., and Hendy, G. N. (1992). 1,25-Dihydroxycholecalciferol regulates chromogranin-A translatability in bovine parathyroid cells. Mol Endocrinol *6*, 1781-8.

97. Mouland, A. J., Bevan, S., White, J. H., and Hendy, G. N. (1994). Human chromogranin A gene. Molecular cloning, structural analysis, and neuroendocrine cell-specific expression. J Biol Chem *269*, 6918-26.

98. Murray, S. S., Burton, D. W., and Deftos, L. J. (1988). The effects of forskolin and calcium ionophore A23187 on secretion and cytoplasmic mRNA levels of chromogranin A and calcitonin. J Bone Min Res *3*, 447-52

99. Murray, S. S., Deaven, L. L., Burton, D. W., O'Connor, D. I., Mellon, P. L., and Deftos, L. J. (1987). The gene for human chromogranin A (CgA) is located on chromosome 14. Biochem Biophys Res Commun *142*, 141-6.

100. Myerowitz, R., and Hogikyan, N. D. (1987). A deletion involving Alu sequences in the beta-hexosaminidase alpha-chain gene of French Canadians with Tay-Sachs disease. J Biol Chem *262*, 15396-9.

101. Natori, S., Huttner, W. B. (1994) Peptides derived from the granins (chromogranins, secretogranins) Biochimie *76*, 277-82

102. Nguyen, M. H., Harbour, D., and Gagnon, C. (1987). Secretory proteins from adrenal medullary cells are carboxyl-methylated in vivo and released under their methylated form by acetylcholine. J Neurochem *49*, 38-44.

103. Nicholls, R. D., Fischel-Ghodsian, N., and Higgs, D. R. (1987). Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. Cell *49*, 369-78.

104. No, D., Yao, T. P., and Evans, R. M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A *93*, 3346-51.

105. Nobels, F. R., Kwekkeboom, D. J., Coopmans, W. (1997). Chromogranin A as serum marker for neuroendocrine neoplasia: comparison with neuro-specific enolase and the α-subunit of glycoprotein hormones. J Clin Endocrinol Metab *82*, 2622-8

106. Nobels, F. R., Kwekkeboom, D. J., Coopmans, W., Hoekstra, R., De Herder, W. W., Bouillon, R., and Lamberts, S. W. (1993). A comparison between the diagnostic value of gonadotropins, alpha-subunit, and chromogranin-A and their response to thyrotropin-releasing hormone in clinically nonfunctioning, alpha-subunit-secreting, and gonadotroph pituitary adenomas. J Clin Endocrinol Metab *77*, 784-9.

107. Nobels, F. R., Kwekkeboom, D. J., Bouillon, R., and Lamberts, S. W. (1998). Chromogranin A: its clinical value as marker of neuroendocrine tumours. Eur J Clin Invest *28*, 431-40.

108. O'Connor, D. T., and Deftos, L. J. (1986). Secretion of chromogranin A by peptide-producing endocrine neoplasms. N Engl J Med *314*, 1145-51.

109. O'Shea, K. S. (1999). Embryonic stem cell models of development. Anat Rec *257*, 32-41.

110. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T., and Nishimune, Y. (1997). 'Green mice' as a source of ubiquitous green cells. FEBS Lett *407*, 313-9.

111. Ottiger, H. P., Battenberg, E. F., Tsou, A. P., Bloom, F. E., and Sutcliffe, J. G. (1990). 1B1075: a brain- and pituitary-specific mRNA that encodes a novel chromogranin/secretogranin-like component of intracellular vesicles. J Neurosci *10*, 3135-47.

112. Palmiter, R. D., Norstedt, G., Gelinas, R. E., Hammer, R. E., and Brinster, R. L. (1983). Metallothionein-human GH fusion genes stimulate growth of mice. Science *222*, 809-14.

113. Pausova, Z., Morgan, K., Fujiwara, T. M., and Hendy, G. N. (1995). Evolution of a repeat sequence in the parathyroid hormone-related peptide gene in promates. Mammalian Genome *6*, 408-14

114. Pohl, T. M., Phillips, E., Song, K. Y., Gerdes, H. H., Huttner, W. B., and Ruther, U. (1990). The organisation of the mouse chromogranin B (secretogranin I) gene. FEBS Lett *262*, 219-24.

115. Quandt, K., Frech, K., Karas, H., Wingender, E., Werner, T. (1995). MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data Nucleic Acids Research 23, pp. 4878-4884

116. Quentin, Y. (1992). Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. Nucleic Acids Res *20*, 3397-401.

117. Quentin, Y. (1994). A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. Nucleic Acids Res *22*, 2222-7.

118. Reeves, R. H. (1998). Exploring development and disease through germ-line genetic engineering in the mouse. Anat Rec *253*, 19-23.

119. Reiffen, F. U., and Gratzl, M. (1986). Chromogranins, widespread in endocrine and nervous tissue, bind Ca2+. FEBS Lett *195*, 327-30.

120. Rosa, P., Hille, A., Lee, R. W., Zanini, A., De Camilli, P., and Huttner, W. B. (1985). Secretogranins I and II: two tyrosine-sulfated secretory proteins common to a variety of cells secreting peptides by the regulated pathway. J Cell Biol *101*, 1999-2011.

121. Roush, W. (1996). A new embryo zoo [news]. Science *274*, 1608-9.

122. Bruning, J. C., Michael, M. D., Winnay, J. N., Hayashi, T., Horsch, D., Accili, D., Goodyear, L. J., Kahn, C. R. (1998). A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. Molecular Cell. *2*, 559-69

123. Salvi, E., Vaccaro, R., and Renda, T. G. (1998). An immunohistochemical study of the ontogeny of the neuroendocrine system in the chicken oesophagus. Anat Embryol (Berl) *197*, 283-91.

124. Sanes, J. R., Rubenstein, J. L., and Nicolas, J. F. (1986). Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. Embo J *5*, 3133-42.

125. Sauer, B. (1994). Site-specific recombination: developments and applications. Curr Opin Biotechnol *5*, 521-7.

126. Schichman, S. A., Caligiuri, M. A., Strout, M. P., Carter, S. L., Gu, Y., Canaani, E., Bloomfield, C. D., and Croce, C. M. (1994). ALL-1 tandem duplication in acute myeloid leukemia with a normal karyotype involves homologous recombination between Alu elements. Cancer Res *54*, 4277-80.

127. Seidah, N. G., Hsi, K. L., De Serres, G., Rochemont, J., Hamelin, J., Antakly, T., Cantin, M., and Chretien, M. (1983). Isolation and NH2-terminal sequence of a highly conserved human and porcine pituitary protein belonging to a new superfamily. Immunocytochemical localization in pars distalis and pars nervosa of the

pituitary and in the supraoptic nucleus of the hypothalamus. Arch Biochem Biophys *225*, 525-34.

128. Seidah, N. G., Hendy, G. N., Hamelin, J., Paquin, J., Lazure, C., Metters, K. M., Rossier, J., and Chretien, M. (1987). Chromogranin A can act as a reversible processing enzyme inhibitor. Evidence from the inhibition of the IRCM-serine protease 1 cleavage of pro-enkephalin and ACTH at pairs of basic amino acids. FEBS Lett *211*, 144-50.

129. Settleman, J., Fonseca, R., Nolan, J., and Angeletti, R. H. (1985). Relationship of multiple forms of chromogranin. J Biol Chem *260*, 1645-51.

130. Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J Biol Chem *269*, 32678-84.

131. Simon, J. P., and Aunis, D. (1989). Biochemistry of the chromogranin A protein family. Biochem J *262*, 1-13.

132. Simon, J. P., Bader, M. F., Aunis, D. (1989) Proteolytic processing of chromogranin A in cultured chromaffin cells. Biochim. Biophy. Acta *1051*, 123-30

133. Simon-Chazottes, D., Wu, H., Parmer, R. J., Rozansky, D. J., Szpirer, J., Levan, G., Kurtz, T. W., Szpirer, C., Guenet, J. L., and O'Connor, D. T. (1993). Assignment of the chromogranin A (Chga) locus to homologous regions on mouse chromosome 12 and rat chromosome 6. Genomics *17*, 252-5.

134. Smit, A.F. (1996) Structure and evolution of mammalian interspersed repeats. PhD dissertation, University of Southern California.

135. Smit, A. F., and Riggs, A. D. (1995). MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res *23*, 98-102.

136. Smit, A. F. (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev *6*, 743-8.

137. Soliman, E., Canaff, L., Fox, J., and Hendy, G. N. (1997). In vivo regulation of chromogranin A messenger ribonucleic acid in the parathyroid by 1,25-dihydroxyvitamin D: Studies in normal rats and in chronic renal insufficiency. Endocrinology *138*, 2596-600

138. Strub, J. M., Sorokine, O., Van Dorsselaer, A., Aunis, D., and Metz-Boutigue, M. H. (1997). Phosphorylation and O-glycosylation sites of bovine chromogranin A from adrenal medullary chromaffin granules and their relationship with biological activities. J Biol Chem *272*, 11928-36.

139. Strub, J. M., Goumon, Y., Lugardon, K., Capon, C., Lopez, M., Moniatte, M., Dorsselaer, A. V., Aunis, D., and Metz-Boutigue, M. H. (1996) Antibacterial activity of glycosylated and phosphorylated chromogranin A-derived peptide 173-194 from bovine adrenal medullary chromaffin granules. J Biol Chem *271*, 28533-40.

140. Sundler, F. (1998). Ontogeny of ECL cells in the rat. Yale J Biol Med *71*, 155-61.

141. Taga, T., Kishimoto, T. (1997) gp130 and the interleukin-6 family of cytokines. Annu Rev Immunol *15*, 797-819

142. Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., Akira, S. (1997) Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proc Natl Acad Sci USA *94*, 3801-4

143. Takiyyuddin, M. A., Cervenka, J. H., Pandian, M. R., Stuenkel, C. A., Neumann, H. P., and O'Connor, D. T. (1990). Neuroendocrine sources of chromogranin-A in normal man: clues from selective stimulation of endocrine glands. J Clin Endocrinol Metab *71*, 360-9.

144. Tatemoto, K., Efendic, S., Mutt, V., Makk, G., Feistner, G. J., and Barchas, J. D. (1986). Pancreastatin, a novel pancreatic peptide that inhibits insulin secretion. Nature *324*, 476-8.

145. Thiele, C., Gerdes, H. H., and Huttner, W. B. (1997) Protein secretion: puzzling receptors. Curr. Biol. *7*, R496-500

146. Thomas, K. R., and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-12.

147. Tsigelny, I., Mahata, S. K., Taupenot, L., Preece, N. E., Mahata, M., Khan, I., Parmer, R. J., and O'Connor, D. T. (1998). Mechanism of action of chromogranin A on catecholamine release: molecular modeling of the catestatin region reveals a beta-strand/loop/beta-strand structure secured by hydrophobic interactions and predictive of activity. Regul Pept *77*, 43-53.

148. Tsukada, T., Fink, J.S., Mandel, G., Goodman, R. H. (1987) Identification of a region in the human vasoactive intestinal polypeptide gene responsible for regulation by cyclic AMP. J Biol Chem *262*, 8743-7

149. Ullu, E., and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. Nature *312*, 171-2.

150. Volknandt, W., Schober, M., Fischer-Colbrie, R., Zimmermann, H., and Winkler, H. (1987). Cholinergic nerve terminals in the rat diaphragm are chromogranin A immunoreactive. Neurosci Lett *81*, 241-4.

151. Wand, G. S., Takiyyuddin, M., O'Connor, D. T., and Levine, M. A. (1991). A proposed role for chromogranin A as a glucocorticoid-responsive autocrine inhibitor of proopiomelanocortin secretion. Endocrinology *128*, 1345-51.

152. Watanabe, N., Kitazume, M., Fujisawa, J., Yoshida, M., Fujii-Kuriyama, Y. (1993). A novel cAMP-dependent regulatory region including a sequence like the cAMP-responsive element, far upstream of the human CYP21A2 gene. Eur J Biochem *214(2)*, 521-31

153. Watkinson, A., Jonsson, A. C., Davison, M., Young, J., Lee, C. M., Moore, S., and Dockray, G. J. (1991). Heterogeneity of chromogranin A-derived peptides in bovine gut, pancreas and adrenal medulla. Biochem J *276*, 471-9.

154. Wei, L. N., and Chang, L. (1996). Promoter and upstream regulatory activities of the mouse cellular retinoic acid-binding protein-I gene. J Biol Chem *271*, 5073-8.

155. Wei, L. N. (1997). Transgenic animals as new approaches in pharmacological studies. Annu Rev Pharmacol Toxicol *37*, 119-41.

156. Weiler, R., Marksteiner, J., Bellmann, R., Wohlfarter, T., Schober, M., Fischer-Colbrie, R., Sperk, G., and Winkler, H. (1990). Chromogranins in rat brain: characterization, topographical distribution and regulation of synthesis. Brain Res *532*, 87-94.

157. Wiedenmann, B., and Huttner, W. B. (1989). Synaptophysin and chromogranins/secretogranins--widespread constituents of distinct types of neuroendocrine vesicles and new tools in tumor diagnosis. Virchows Arch B Cell Pathol Incl Mol Pathol *58*, 95-121.

158. Winkler, H., and Fischer-Colbrie, R. (1992). The chromogranins A and B: the first 25 years and future perspectives. Neuroscience *49*, 497-528.

159. Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, Meinhardt, T., Prüß, M., Reuter, I. and Schacherer, F. (2000). TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28, 316-319

160. Wu, H., Rozansky, D. J., Webster, N. J., and O'Connor, D. T. (1994). Cell type-specific gene expression in the neuroendocrine system. A neuroendocrine-specific regulatory element in the promoter of chromogranin A, a ubiquitous secretory granule core protein. J Clin Invest *94*, 118-29.

161. Yoo, S. H., and Albanesi, J. P. (1990). Ca2(+)-induced conformational change and aggregation of chromogranin A. J Biol Chem *265*, 14414-21.

162. Yoo, S. H., and Lewis, M. S. (1998). Interaction between an intraluminal loop peptide of the inositol 1,4,5-trisphosphate receptor and the near N-terminal peptide of chromogranin A. FEBS Lett *427*, 55-8.

163. Zhang, F., Popperl, H., Morrison, A., Kovacs, E. N., Prideaux, V., Schwarz, L., Krumlauf, R., Rossant, J., and Featherstone, M. S. (1997). Elements both 5' and 3' to the murine Hoxd4 gene establish anterior borders of expression in mesoderm and neurectoderm. Mech Dev *67*, 49-58.

164. Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.

App.1

Human CgA entire sequence

```
   1 GAATTCATGC CAGCTCTGAG CCTACATCTC CCCATGGGGA AGAGAGGGCT ACTTTTGGGG
  61 CCTCTCTCGA AATCAGCATT CTGGAGAGGA AAACATCTGT GAGAATCCTC CCTGGATCCC
 121 AGTAGACCAA GCCCGGGCAG GGCTCTGGTG CCCTCCCCTT CTGTTCACTA GTTCCAGTGC
 181 CTGTGGTGCT CCCAGAGATG CTGGCTGGAC AGCTTCTGTC CTCCAGGGAG GCCAGGACAC
 241 CTGCTCTGCC CCCTGCTCTC TATCCTCCAG GCTAGAACCA GCCACCTTCT TGCAGTCAAC
 301 ATCCCCACAA GATACTGAGA GTGATTTTTT TTATTGTGAT AAAATATCCC TAACATAACA
 361 TTTACCATCT TAACCATTTT TAGGGGCACA ATTCAGTGGC ATTAAGTCCA ATTATGTTGC
 421 TGTGCAGCCA TCATCACTAT CCATTTCCAG AACTTTTTCA TCTGCCCAAA CTCTGTACCC
 481 ATGAAACACT AACTCTCGTT TCCTCCTCCC CCTCAGCCCC TGGAAACCAG CATCCTATTT
 541 CCCATTTCTG TTGTTGTTGT TGTTGTTTGT TTGTTTGTTT GAGACAGAGT TTCACTCTTG
 601 TTGCCCAGGC TGGAGTGCAG TGGCATGATC TTGGCTCACT GCACTCTGCC TCCACCTCTT
 661 GGGTTCAAGC GATTCCCCTG CCTCAGCCTT CCAAGTAGCT GGGATTACAG ATGCGCCACC
 721 ATGCCCAGCT AATTTTGTAT TTTTAGTAGA GATGGGGTTT CACCATGTTG GTCAGGCTGG
 781 TCTCGAACTC CTGACCTCAG GTGATCTGCC CGCCTCAGCC TCCCAAACTG CTGGGATTAC
 841 AGGTGTAAAC CACTGCGCCC AGCTGACCCA CCCTATTTAA GTATAACCCC ACCCTGTCTC
 901 AGCACAGCCG TCCCCTTCCG TGCTGACAAA CTGTGGATGT TACTTCTTAT TCATTGTCTG
 961 TCTCCACCAT CTAGAATGTA GGCTTCTGTC TGTTCTGTTT AATGCTGTAC CCTCAACTGT
1021 GTACTGTTCT AGGATGTTCT AAGCACATAG TAGGTGCTTA GCAAAGCAAA TATGTAAATT
1081 CATGAATGTG GCCATTTCAG TCTCCTTTGA GCTCCCACCA GGTTCTCATT TAGGGACAGG
1141 CGTGAGCACA GGTGTGGGAA AGGAAGTATT AGTAGAAAAC AGCAGCCGCA GATGGAGAGG
1201 GGTGGGCAGG AAGGAGGGGC TCTGGACGAT GCCAGCCAAT GACAAACCGT CACCCAGATC
1261 CCAGGTTCCC GTGACAGTGG GGTCAGGTGG CAAGAGCTTC CTACATGGCC CCCAGAGCTG
1321 GAGCCCAAGA TCTGGAGCAT CAGGGGCCCC AGAAATAAGG GAAGCCAACC AGGAAACTGA
1381 CCTTTTACAC AAAATATCCA GGTTTTGAGA ATGCTCAGGA AGAGGTGAGA GCTGTTCTCA
1441 AATTCAGATG GGCTGCCAGT GGCAAGAAGG TGATGCTAAG GGGTTGCACT AGGGCTAATG
1501 GGTAGGTGTT GCAAGGACTC CTATTTTGGC GGATGTAATA AGAACTGGCC CCCACTCCTC
1561 CCTCTACCTA ACAAGAGGTG TCCAACAGTG GCTTAAGCTT CCTGGTGAAA GTGAGCCCCT
1621 GGTTCCCGTG TGTATGCAGA GCCCAGGATT CACTCACCCA TGGCTTGAAC CTTTAATGGA
1681 TACATTCATC AGTTACCTGT CAAGTGCGTT TCCTCTGTGC CAGATTGTGT GGCGCTCAGC
1741 ACTGTGAAAG TAATGATGAC TAAACATTAT TCAAATAAAT ACACATTCAA CAACAAAGTG
1801 TGAAGGCACA AAACTAGCTT GCCAGAGAGA GAGAAGCCTC ACTCAGACAG GGGATCGAAG
1861 GCCCCTCTAG AGCCTCCTAG ATATTGGAGA GAGCCATGAG TGACGGGCAT GCCTCCAGCC
1921 TGTGAGACCG GGGGAGCATA CCCCAGGCAG AGGGCACTTA GGAAAAATAC GACGGGGTCG
1981 TAGCGCTGCC AGGGACCTCA GTACACATGG AAGAAGGCAA GGTCCAGAGA TCCGAGTGAC
2041 TTGCCCAGGG ACACAAGGCA AATCGGTGGA ATCGTCGAGG GGTGGAGGAT CAGCCACAGC
2101 CCCCCGCCTC CCACGCCACC TCTTGGAAAC CAGATACCCG TCGCGGCCAA GACCCACCAG
2161 CTCCAAGCGG CGGAGGCCAG AGGTAGCGAG GGGTGAGGTT AGAGGTGGGG GCGAGCGGGG
2221 ACTGGACCCC TGGGGAGTGG GGAAAGGGGA AGGGGCGGGT ACCGCTGACG TCATTTCCGG
2281 GGTCGGGGTA TATAAGCGGG GCGCGAGGGC GCTGCT***GCTG CCACCGCTCC TGCCACTGCA***
2341 ***GTGCTCGATC CCCGTGCAGG GGAGCTTGCG GGAGGATCGA CCGACAGACG GACGCACGCC***
2401 ***GAGGCACTGC GCCCCCAGCC CCGCGCCGGT GCCACCGCAG CCCGACCCCG GCCGCCAGTT***
2461 ***CAGCCGCCCC TCGCCCGGTG CCTAGGTGCC CGGCCCCACA CCGCCAGCTG CTCGGTGCCC***
2521 ***GGGTCCGCCA TGCGCTCCGC CGCTGTCCTG GCTCTTCTGC TCTGCGCCGG GCAAG***GTGAG
2581 CGAGCGCGGG GAGCTCGCGG GGAGAGGGTT CCGGGCGCCC CTGCCCACCT TGAGGTCCGG
2641 GCACCGCGCG GCGCCCCGCA CCCCTCCACA CTTCCCTTCG GGCGCGGCGA GTTTTCAGCA
2701 CCGCGGACAG CGCCTCCGCC TCCCGCCTGA CCCTGCAGTG TGGCCTCCGC CTGGGACCCA
2761 CAAGACACTT GGGCCTGACT CCCAACCCCG CGGGGCAACT TCCCCTTGTG CCCACCCCTT
2821 GTCTGGGCTT GGCGCCTCGG TTGCATCAAG GCCAGTGCCG GTTGCGCGGT GGCCGGCCAA
2881 GGTCACTGCG GGGAGGTGGG AGCATGGCTG CTCGCAACCC TCCACCCCTC TTCCCCTTTC
2941 ATCTCAAGCT CCCCGCCTTT CTGCCCTTGA CTTCTACTTC TCCATATCAT CCCTCACAAT
3001 CCCCTTCCTC CATCTCTTCC CCCCGCCCCC CCCACCTCTG GGGTCTCTCT CCCCTTCGTC
3061 TGTCTCCCAC CCGTCCCACT GTGCAGCTCT TGGCCTTGTT TGATGGCGCT CCCCGCTCCA
```

3121 AAGCACACCC GGTCCAGAGC AAGAGTTTCC AAAGGGGGAG GGGGAGGACC TCTCAGTGCA
3181 GCCTCTGGGG GACCACCCCA TCTGCTGGGG AGTAAGCCAG GGGCCAGCCG AAATCGTGGT
3241 CAGCCCTCGG CCTCCTGTTG AATGGTGGGG AGGAGTTTGG AGGGGTGGAA GGTGGTGAGC
3301 GGGCCCCAGG TGAGCCAGGT CAGAAACGAC TGGCCTTGAG TCCGAGCAGA GAAGTGTCCC
3361 CAGCACTCTG TGCAGATTAC TAG***TCACTGC GCTCCCTGTG AACAGCCCTA TGAATAAAGG***
3421 ***GGATACCGAG*** GTAAGAAGGG GTGCTGGGGA TGAGGGGTAG GAGGCTCCAG TGGACACTTG
3481 ACAGCCAGTT CTTGGGCAGC TGCAGGAGTA AGTTCGGCAT AAAGCCAAGA GCCAGCACTG
3541 CGGCTGCTGA GCCTGGGGAC AGCTTGCAAA AGAAGATATC AAAGATTGCT TGCTTTCCCT
3601 GCTGTCCTGC TGTGGGGACT TCCCCAGCTT CCTCTGAGCT GCACGTGCAG CGGAAGAGGC
3661 CTGCATCTTA CCTGCTGGGC CGAGGTGCCT GGCTTCAGGC TGTCTTGAAC ACTCTGAATC
3721 CAGTAGGCTC CCCAGGGGCT GGTGGGGGTG GCTTCTCTGT AGTTGTAGTT GTCCACAGAC
3781 CTGTAAAAAT GGCCTTTGCC TGCCTGGCTC TGGCCTCCCC GCTTTGCTCT TTCTCCCTGA
3841 AAGGCTCTGC CAGTAACAAC CCACACCCAG CAGGCTCTCC AGCCCTGGCT GTCCCAAAGG
3901 AACACAAGCA AAACATCTTG ATCATGGCTC CTGCACCCAC GGTCCCACCA ATAGGGGCTA
3961 GTTGTGGAGG TCACCCAGGA AAGGGCCTTA GGAGACACTA GTAGGGGTTT CTCTGGCAGC
4021 AAGAGACACA CCACGCTGGG TGTGTACTTG GAATTCATTC ATGGCTGAGC TGGGGCTGTT
4081 TTTTCCAGGC TCTAGTGCTC ATCCAAATCT CTGACTTGAG GGTCAGAGAA ATTTCTACTC
4141 CCCACCTTCC CAGAAACTAT TTTCTTCCAC TAGACACCAC AGTTAAAATC CAAACCATCC
4201 AAGAGATTCC AACCTCAGTT TGCTCAATTA TATTTCCTAT GTTTTAGTTG CAGATAGGAC
4261 AGAATTTAAA CCTCTACCAC TCATTCTAGC TGTGTGACCT TGGGAAAATT CCTTAATTTC
4321 TCTGAGCTCT GCCTTTTTCT CCTTTTCCTT GAGGGGAGGT GTAAAGAGTA GGAATCATGG
4381 ATGTGAATTG CCTAGCACAT AATAGGTCCT CAGTAAATGG GCACTTTTAT TCCTAGAGCT
4441 GCCTTCTTGA TCTCCCCATC CCTCTCCAGA CCCTCTACCC AAATTCCTGC CTCCCAGAAA
4501 ATGCCCTCTC GAACAGGTGG TTATCTGATA TTCCCAATGC AAGGTCCGGG GTTGCAGGGA
4561 GATACAGGGT ACAGGCTGCT GTGCTCTCCT TGCTAGATCC CTTATGGGAG ACACCCTGGT
4621 CACCCTATTA TACCCAGCAG CCTGTGGTCA GGGGCACCAG AGCAGCTGGC TAAGACCTGA
4681 AGATTCTATA CGAGGTTGTT GTGACCACAT GCCTTATACA GCACTTGCCA GTTTGCCATC
4741 TTGGATCCTT GGACCGGTTG CCTTGGCATC CCGTAGACAC AGCCTCCCAC CCCCTCACTC
4801 CCTGCCGGTG TTGCCTGATG CCTGTCTGAC ACCTCCTGCT CTCTACTCAG CTCTAGGCTG
4861 TGGAAGCAGG GCTGGAGTTG TGAACCCCTC CCTGGGCGTG CTGACACTCC ATGCTTGAAA
4921 GAGTGTTTTT CCATCTGTAC CTCATTACCC TGGGAGAGCA GTTGGGAAGG TATTCTCACT
4981 CCCGTTTTAC AAATGAGGAC ACTGAGGCTC AGAGAGGTTA AGTCACTTCT CCAAGATCAC
5041 ACAGCTAGGA TACCACATTA CGGTGGGCAC CATCATCAAA GATCACTCAT TTATCAGATG
5101 CCTAATTATC CACAGCACCT TTGCAGGCAC AGGATAAGAC TTTCAGAGGC GTGGCTGTGG
5161 TTGACCTGGG ATATGCTCCC AGGTGTAGTC AATGTGTGCT CACATGCCGC CCAGGGATTC
5221 ACCCATCCCA TGATGAACAT GGAGTGAGAA TATGAAGCTG TGACTCCTTG CATGCGACAA
5281 CACCATGGGT GTTTGGAGAC AGTAACCTAC ATCCTATTTC CAAACAGGGT AGAGCCAGGG
5341 AAGGTCTTTG TCTCCTCTGG GCAGTCCAGA AGGACTTCCT GGAGGACCTG ACCTTTGGGC
5401 ATCTTCACGA TCAGAGCACT TTTGTAAGAC ATCATTTTAG TGCCCTGAGG CAGTTTTCTC
5461 ATGTTTAGAG AAATAGCACT TGGGGATGAG AAGGTACCAT TCCCTACACC CCATTCTCAC
5521 TATGATCCAC CTTGGAGGGG AACCTGCTGG CCATGTGGTT GGAGTGGCTG ACACATGGTA
5581 AGCACTGTGT TAAGGGTTTT CTATTATCAT TGCTATCACT GTCTGGCTAA GAAAGCAGAG
5641 CTGTGTCCCT GAGCCAGGGA TCATGGAATC CCAAAGACAA AAGCAAATAC CCTCTGTCCT
5701 CACTAGAAAT TCACCCTAGA GCCTCCAGGG ACTGAGCCCC ATGCTCAAAC CAGCCCCAAC
5761 CTGCCCCTGC ACTGTGTTCC CAG***GTGATGA AATGCATCGT TGAGGTCATC TCCGACACAC***
5821 ***TTTCCAAGCC CAGCCCCATG CCTGTCAGCC AGGGATGTTT TGAGACACTC CGAGGAG***GTA
5881 TGAGCTGGAG GCTAGGGGTG AGGGCTGCTG CCTGCTGGGC TGGGAGGCTA GGACATGGGT
5941 GTGTGGCTTT TAGTGGAATT AATGATTTAA CTCAACAGTC ACTGACTGAG TGCCCGTCAA
6001 TGTCCAGGCA CTGCAACCAG GGGCCAAGGT GTACAGACAG TTGAAGGGGT TGTAGCCCAG
6061 GTGCTTACAG ACTGGCAGAG GGGATGCATA TGAACACAGT TAACTAGAAC TCTGGGCAAA
6121 ATAAAACCAG GGCTGGAACA GAAGCCCCCA GACTGTGAGC TCCCTATGGA CAAAGATTTT
6181 GTGTGCTGTA TCCCAGTGCC TAGAACAGCC TCTGGCCTCT GCAGACACCC TCTAGATGTG
6241 TGAGGAATAA AGGAATCTTA TGGGTGGACA GGAGAAGGAC GAAATCTTTC CAGGCAACCA
6301 GTCAAGTTTT TACTAACTGC TGGATAGAGC CAGCAGCATT TCGGTTGGGC TTTAAAAGAA
6361 ACATTGGAAA CCGATGAGTG GAAAAATGAA GACAGGGTTG TCCAGGTGTG GTAGACAGTG
6421 TGAACACAGG CATGGAGATG GGAACCTGTT GTATTTAAAG AGACCCCACT CATCCTCAGG
6481 GGTTGCCACC CTTTGGCTAA GGTCTGGCTG CTGGGGCAGT GGCTGACTTC TCTGATGTAG

```
6541 TAGGAAACTG ACACATACCT TGGAAAGGAG GTAATTCCAA AAGTGTTATC TGAAGACCCA
6601 ATCTTGCTGC CACCACCCTG GTTTTGAGTT TCCAGAGTAA ATTCCTTGTG CTCAGCTGAA
6661 AATATTGATG GTAATCTCTT TCTCGGCTGT TCTCTGGAAA CCACCCATGA TGACTCTTCT
6721 TCATTGCAGA TGAACGGATC CTCTCCATTC TGAGACATCA GAATTTACTG AAGGAGCTCC
6781 AAGACCTCGC TCTCCAAGGT ATTTTCCAGC CACTGCACTT GACTCTGGGT AAATTCCAGT
6841 AACTGAGAGT CAGAAAATCT GGTGGAAGAT TCAGCAAGTT CCTCTCTGGG ACTGTCATTT
6901 CTTTTAATTA TAAAGCGGAA AGCAGGATCC GCCGTGCACC TGGCTCAATC TGTTATTGAT
6961 ATTGTTAAAG AAATCAAGGC CCAGAGAGGT TGAGTGACTT GTCCAAGGTC ACACAGTGAA
7021 TAGAGACATG CCAGGTGCAG CTTCCCTAAA ACCCACGGCC TCTGCAGTTC AGTCATTTTA
7081 GGAAATACCT TTATGAAGTA GTGAGCCACT CATCACTGAT GATATTCAAG CCAAGCCTGG
7141 ACAGGATACT GGAGTGGAGG GTTAAGCTGC CATCTTCCCA CAGTTAGCCT CTATGCCCAG
7201 CCCCCCAGCT CCTTCCAGGC TTGATGCAAA GCACGGCCCT AGCCTTGCAA GTCAGGGAAC
7261 CTGAGTTTGG CCCTGCACCT GCCCCGGCTG ACTTCCTGTG GGAAGCGGCC GTTTCCACCA
7321 GGGCAGGACC ACTGGTGGTG GTGTGTCTCT CCATGAGAAC AGGGCATTAC AGCATGAATC
7381 CAGCATCTGC TTCCCAACGA AGCGAGAGTG TTTTTCAGGG CTGGGCTTGT GTCTGGGGTG
7441 GGGTGGGGGT ACGAAGGCTG CTCTGGCTTC AGGATGGGCC CCACTCTGAG GCCTTTGGGA
7501 CAGACCAGGC CACGTGTGAG AACAGAGAGG GCTTTGGGGA GGAGGCGCGT AAACACAGCC
7561 ACGTCAGTAG GATTCCTGTG GCCCTTTCCA AGCTAACACA CTGTTCTTCC TTTACAATTG
7621 ATAACTTTAA AATTAAATAT TGAATGACTA ACTGATTACA TTTAAATAAT TAAATAGACT
7681 GTCAACAGAG GCTATTATAG GAGAATGCGT CCTGAAACAT CTCTATTCAC ATGCACGTGT
7741 GTATCCCGTA CAAGCAGTAT CCAATATATG TATATATGAC TATTTTCCAT TAACAGATTC
7801 CTGGATGTAT TGCTGTCCAT CCGTCTATCC GTCCACTCCC TCTTGGGCTC TCATAACATC
7861 TTCTGTGTGC CTCTGTCTTA GCATTTACCA CACAGCAACC CTATTTTCTA GACAAGCACC
7921 CAACTCTTGG AGAGCCCAAG TTGCCCATGG AGCTAGGAAA TAGCCAGAAG GGGAGTGGAA
7981 CCCACGTTGC CTGAGACCAG GGCTGGGCAT GCTCCCTGGC CCACAGCTGC CCAGGAAAGC
8041 CCATTGTTGG GGAAGAAATG AGCTGGGCTA GCTGAGGGGC CAACATCAGG AATAAGGCAC
8101 AGGTGTGGGG ACAGTCAGAG GAGGCAGCGG CGGCGGGAGG AGAGGGGTCT CTGGGTCATG
8161 GCTGCAGGCG CCAACCCTGG GCTCAGATGT AGGCACAGGG CTTGGCCTTC TCCCCGGGCC
8221 TGGACTCAGC GTCCAAGGTT AGCAGAGCTT GCCTGAATCA AGAGCTGTCA CCATTACTGC
8281 CTGCTGCAGG GTTCCTGGAG CTACCAGCCA GGTCCTGAGG GGTTCCCCAC CCCCAACACC
8341 TGCCCAGCCA GGTCCTGAGG GTTCCCCACC CCCAACTCCT GCCCAGCCAG GTCCTGAGGG
8401 GTTCCCCACC CCCAACTCCT GCCCCTGGGA TCCTTGCACC ATTTGAGCCT TTCTTGCTGT
8461 GTGGCCACAG CACGCTTGCT CTGGCCTCCC AGGGAGAAGC TGAGTCTGGC GAGACAGGGT
8521 GGCTGCCTGA GGGCCCAGCT GAGAGAGGTG AGGGACGCCA ACCCTTCACC GCTCCCCAAC
8581 TCACGCCCCA TCAGTGTCTC ATTGGGCAGG CCCTGAACAT GATGTGCCCA GCTTACAGAT
8641 GGGGAAATGG AGGTCTTCAC TCTTATAGAC AAATATGCCA CAGCTGCACC CAATGGGACT
8701 TTTCCCTCCT TTTCTCATAC CAGGCGCCAA AGTAGAGGGC ACATCAGCAG TAGAAACACA
8761 GCGGTTTTGA AGATGAACTC TCAGAGGTTC TTGAGAACTA GAGCAGCCAG GCCGAGCTGA
8821 AAGGTCTGTC CCAGGCGGTC TGGCTGGAGG TGGGGAAGGG AGGGTGGCAG TGGGAACAGT
8881 GGCTATGATG GACCCAGGGT TCCAGTGAGT CGGGAGCCCC ACCCATAATC CCTTATTTCC
8941 CTCTCTCTAC AAATGGGGAA ACTGCTGCTC AGAGAGACTA GGTCCCTTGC CCAAGGTGAG
9001 TGGCATGACC AGGGGTTTCG ATCTGGTCCT GCCTCACACT AAAGCCCTCG CTCGGTCGCG
9061 TCTGCCCAGA AGAGGGGCTG GCACTCTAGG GCAGGAGATC TCTGGCATGG CAGGTGGACA
9121 GCTTCAACTC TGGGTCTACC GAGTGGCCAC TGTAGCAGAG CCCCCACCCC CAGTGCCTGT
9181 CCTCCTGGGA AGTGGTGGGG CCTCAGAATA GCCAGGATGC AGGAGGACTA CAGACAGAGT
9241 AGGTGTGAAA CGTGAAATGA ACCCAGATCA TCTTTCTTGA AACTAAAGAG CTTCTAGGAG
9301 TCTAAGAGGT CATGCCTTAG GTCAGTGCGA CACTCTGAGC CATCCTCCCT TCCCGGGGAG
9361 AGCCACAAAC TGGCCCCGTG GCGCTTCTGC TCTCCATGTA GAAGATGGGC CGGTAAGGCC
9421 GGTGTCCAGG AGCTGGATCC TGGTGTCTCC CATCTGGGGC TGAAAGGGAA AGCCCGCAAG
9481 TGAAGCCCAG GTGCCCCCAC TTTGGTTGAA GCTTTAATGA GGCAGGCAGC CATGTGTCAG
9541 TGCAAAGAGA TCCCGAGCTT GTCATTTGGA GGCTTGGGTT CCAGGCTCAG CTCTGCTGCC
9601 CGCAAACTGT GAGATGGACA GGTCATCTTA CCTTTCTGCA GATCATTTTG GGTTTCTGTG
9661 AAATAGGCAT AGGAGCACCT TTCACCAGAG TCCTGGTGAA GCTCAGATGG TGGAATGTAC
9721 TTGGGAGCAC TTTGTTCAGT GGTTATCATT CTAATGACGG GAGGAATTTT CACTTGTTCA
9781 AAGGCTGTCT TTAAGGCAGC CACAATAAAA TACCTTTTAA TGAAACCCTA AAATATTGCA
9841 GAATTGAGCC CTTTCACAGG ATGCTGAAGA CAGAGAAGCC CCACATGGAG GGGAGAATGA
9901 AGTTGCAGTC ATGGTGCCTC TTACACAGAG AGGTTAAGCA ACTTGCCCAA GATCACACAG
```

9961 CTGACATTAA TTTTCAGTTC AGGCTTTATC TCCTGGATTA ATCTTAAGGT CTGGTTTTTT
10021 GTTTTTGTTC TTTGAATAAT AACCTGTTTT AAAAGTCAGG ATTAAAAAGA AAGAAACAAA
10081 TCCAGCTGTG CTCAGATCAA GCCTGGACAT GGACAGAGGG GTAACCCTAA TCGTTGTCCT
10141 GGGCTGGGCT CGCTGGAAGC CAAGAAACAT AATGAGTAAC ATCTGAAATT AGCCTGTGGG
10201 GAGGCCCCAC ACGTGGCCCA GTCCTCAGGG CCTGACTTGG CTGTGCTGTG TCTGCAG***AGG***
10261 ***CGGTGGAAGA GCCATCATCC AAGGATGTAA TGGAGAAAAG AGAGGATTCC AAGGAGGCAG***
10321 ***AGAAAAGTGG TGAAGCACAG ACGGAGCCAG GCCCCAGGCC CTCCCGGAGC CCATGCAGGA***
10381 ***GTCCAAGGCT GAGGGGAACA ATCAAGGCCC CTGGAGAGGA AGAGGAGGAG GAGGAGGAGG***
10441 ***CCACCAACAC CCACCCTCCA GCCAGCCTCC CCAGCCAGAA ATACCCAGGC CCACAGGCCG***
10501 ***AGGGGGACAG TGAGGGCCTC TCTCAGGGTC TGGTGGACAG AGAGAAGGGC CTGAGTGCAG***
10561 ***AGCCAGGGTG GCAGGCAAAG AGAGAAGAGG AGGAGGAGGA GGAGGAGGAG GCTGAGGCTG***
10621 ***GAGAGGAGGC TGTCCCCGAG GAAGAAGGCC CCACTGTAGT GCTGAACCCC CACCCGAGCC***
10681 ***TTGGCTACAA GGAGATCCGG AAAGGCGAGA*** GTACGTATGA TGGCGAAGAC CTCAACGAAC
10741 GTGTCTGGGA GAGGGGGATG GGTGGGAGGA GAGCCTTCTC CATAACCTCA TTCCACCTTC
10801 ATAACAACCC TGAAAGGTAG GTATTAGCTC CATTTCCCAG GTGGACAAAT GAGGCTCAGA
10861 GAGGTTAAGT AATTTGCCCA AGGTCACACA GCTAATTTGT AGCAAAAGCT GAGGTTCTAC
10921 CCCAGATTTA TCTGACTCCA AAACCATGCA TTTCCCACTA TTCCACATAG CCTCCCCCCA
10981 CCAACCCCAG AAGCTAGGGG TGGGTGATGG GTGGGCTGGC TTTGGGAAGA GAGACCATGG
11041 CAGGAGCGCA CAGGCTGATC TGGGAACAGT GTCAGCTTCA ACTACGGTTT AGGAGAGGCC
11101 CTGGCTCCCG CTGCGGGCCT GTCAGGGTCT TTCCTCACTC TCCAGCTGCC GGGCTTCTGG
11161 GGTGAGGATG AGGGGAAGAG GCAGGCTCCA GCTAACCCAC CCCTCTGAGC CGAGAGGGTC
11221 TTGCAAAGCC TGGCATCAAG AAGGTTTTTC CCGCTAAGCG TCATCACTGT GGAGAGGCTG
11281 GGCTGTGGCC GCAGCAGAGG CCCCCAGGGA GTGGCAGAGA CTGGGAAAAT GGTGGTCCCC
11341 CACCCATTCT CCTGCTCTTG CCCACCACCT GCTCCAG***GTC GGTCGGAGGC TCTGGCTGTG***
11401 ***GATGGAGCTG GGAAGCCTGG GGCTGAGGAG GCTCAGGACC CCGAAGGGAA GGGAGAACAG***
11461 ***GAGCACTCCC AGCAGAAAGA GGAGGAGGAG GAGATGGCAG TGGTCCCGCA AGGCCTCTTC***
11521 ***CGGGGTGGGA AGAGCGGAGA GCTGGAGCAG GAGGAGGAGC GGCTCTCCAA GGAGTGGGAG***
11581 ***GACTCCAAAC GCTGGAGCAA GATGGACCAG CTGGCCAAGG AGCTGACGGC TGAGAAGCGG***
11641 ***CTGGAGGGGC AGGAGGAGGA GGAGGACAAC CGGGACAGTT CCATGAAGCT CTCCTTCCGG***
11701 ***GCCCGGCGTA CGGCTTCAGG GGCCCTGGGC CGCAGCTGCG ACGAGGCTGG AGGCCATCCT***
11761 ***CCTGGGAGGA CAGCCTTGAG GCGGGCCTGC CCCTCCAGGT CCGAGGCTAC CCCGAGGAGA***
11821 ***AGAAAGAGGA GGAGGGCAGC GCAAACCGCA GACCAGAG***GT TGGTATGGGG CGGGAGCCAG
11881 CTCTGTGCCA GGCCACGGAG CAGCAGGGGG CAGCCGCACC CAGACACACT GCCCCTGCCC
11941 CACTGAGGGG ACAGGGCCCC CCCGCCGAAG TCTGGGGATG GAGAGATGCT CAGACCGGGG
12001 GCTCTCAGGG TGGGAGAACA CCCCAGCTCA CAGGGGGCAC CCAGCAAAGC TGGCTGGGAG
12061 ATGGGAGGAG CTCAGGTCAG CCTGTGGCAG CGGGGGAGGC GCTGAGTGGG GACGTTGTCC
12121 AGAAGGCAGT AGGGTGCCCT GGAAAAACCA TGCGGTGGGT CTGACCGTGT CCCCTCCTGG
12181 CCCCCAGGCA GCCCAGGCAC AGCCTCCTCC CAGGCTCAGG CCCTGCCTCT GGCCCATGAG
12241 GGTAACGGGG CAGGGCAGGG GCAGGGCCAA GGGATCTCTT GCCTGTGGGG AGTGGGTGGG
12301 AGCCACACCG GATGTGAGCT GGAAGACCTG GGCTCTGCCT CCTCTTGGCT GCGTGGACTT
12361 GGGAGGGTCA CTCCCCTCCC CGAGCCTCAT CAGTTCCATT TATGAAGTAG GGGTAAGGAG
12421 TCTCCTCTGC CCACGTCATT ACCATGATTG GAACACACAA AGGCTTTGAA AACCAGGGCC
12481 CCTGGGGAAG ACAAATGGGG CGTGGGGACA CAGCCCACAG GAGCAGGGGG GAGTTGGGGT
12541 CCCTGGCAGC TAGGCGGCTG GGATACCTGG TTCAAGCCTA GTGTCCAGGT AGGAACCCAG
12601 GCTCTGTGGG AAGGCTGACA CCTGTCCGCA GCCCTGACCC CAAGGCTCTT CTCTGCCCAC
12661 CTAGAAGAGG GTGTGGCCCC TTCACTGGGC CCCCGACCCC CTCTCTGGGC CTGGCAGTCT
12721 GAGAGGAAGA AGGCTCCATG TTGGCAGCTG TAGCGGGAGA GGGCATCTGC CCACGTGCCT
12781 GGGAGCCCAG GGCGAGGGGC TGAGGACAAA GCTGGAGAGA ATGGGGGCCC AGAGGTGCCA
12841 CCAGCTACCT ATCTCCACAA GCAGAGGGAG CTCTGGACCG CAGGGCAGAG AAACTGGACC
12901 CCTGAGGCAT ATGGAGAGGA TGGGGTGGGG GCCTCCTCAA CACAGCCCAA GGGTCATCTT
12961 AGAAACCCGA GCCTCGCAGA GCACCTGCCC CAGGGCCTCC TCCAAGACTA AGGGGCTGTA
13021 GAGTCTATTT CCACGGCCTG CAATGTCGGG GTCCCTGGGC CAGGATGCCC TCGGCAGCCT
13081 CAAATCTGGC AGTGACCCCG GGAAAATGAC CCCTCCCAGG TCAAAGGAGG CAGCCTTCCC
13141 TGCGCTCCTG CTCTAAGTCT TGAGGTTGTG CTCTTGTGGG TGTCCAAGTT CTGAGGCCTG
13201 AAAGCTTCTG CAGACGGGCA TAGAACAGGG CTCAGAGGAG GGGTCCCTGC TTCCCTAAAC
13261 TTGGGTAGCC AGAGTTCGGA TCCTCTCACA ATCAGGACCA GGGATCAAAG CCCAGCCCTA
13321 GGCAGGCTGA TGTGGGGCAG CTCCCCAGCA TGTAGCCCTG TCTCAGGCCC CACATTCCCC

13381 TCAGGTGGGA AGGCTGCAGA AGACAGACCC GGAAGTCCCA GGCTGGCCTC ACCCTGGGTC
13441 CTGGCAACCA CGGGCTAGGC CGAGGGAAGG GGGCTCTTCA GGGCACTGTA AGCTTTGTTT
13501 TCCATGTGCC CTAAATTTTC AGACTTCCAC CTGTGCGGAA CCATTTGATA GGAAGTTGGG
13561 CCTGGATAAG AGGCCACCAA GAGCAGTGTG GGGACAGGGC CCCGGCATCA GCCACCTTCT
13621 TGTGCATAAT GATCCCGAGA GCAGTACAGT AAGCTCATTA TCACTGTCTC CATGATGGAT
13681 GAGGGTACAG AGAGACAAAG CCACCTGTCC GAGGCCACAC AGCTAACCCA GCGCCTTTCT
13741 GGCTTACTGT GCCTTCCATG CCACTGGCCA GGCTGGCCCG AGTGAACCCC AATGTCTCTC
13801 CTAG***GACCAG GAGCTGGAGA GCCTGTTGGC CATTGAGGCA GAGCTGGAGA TAGTGGCCCA***
13861 ***CCAGCTGCAG GCACTACGGC GGGGCTGAGA CACCGGCTGG CAGGGCTGGG CCCAGGGCAC***
13921 ***CCTGTTGGCC CTGGCTCTGC TGTCCCCTTG GCAGGTCCTG GCCAGATGGC CCGGACGCTG***
13981 ***CTTCCGGTAG GTAGGCAGCC TCCAGCCTGC CCAAGCCCAG GGCCACCCTA TCGCCCCTAC***
14041 ***GCGCCTTGTC TCCTACTCCT GACTCCTACC TGCCCTGGAA CATCCTTTGC AGGCAGCCCA***
14101 ***CAACTTTAAA CATTGACGAT TCCTTCTCTG AACACAGGCA GCTTTCTAGA AGTTTCCCTT***
14161 ***CCTCCATCCT ATCCACTGGG CACAACTGCA ATAACTTCTG ACCTTTTGGT GAAAGCTGAG***
14221 ***AACTCCTGAC TGTAACATAT TCTGTATGAA CTTTATCTAA AGAAAAATAA ATCTGTTCTG***
14281 ***GGCTCTTTCC TCTGA***ATTTC TTCGCCTCTC AAGAGTGGAG TGAGATGCAC TGGATTTGGG
14341 CGAAGGGACA AAGGGAGGTT CAGTCCTCCT GGGGACCTCC TTTCTGGGCC ACTTCCTGGC
14401 TGTGGGATCT TGAGCAAGTG ACTTTGTTAG CTCTGAGCCT CAGTTTCCTC ATCTGTAAAA
14461 CGGGAATGCT GACCTGCCTC AAAGAGTGTG TGCAGGGATT AAAGCACACT TGCTGAGGGC
14521 CTCCAGGTTC CTTCCTCTCC CCGTCCCCAG CCACACCTGA GCTGGGCCAT GCAGAGGCAG
14581 CGCCAGGTGG GGACACGGGC TCTGTTTCCT GAGATCAGGT CGCCCCTCCC TGATGGTCCC
14641 CGCCATCCAG TCACCTGTGG TGCCGGGTGG AGGCCCTGAC AGCCCCAGTC TAGTGCCTGT
14701 CGCAGGACAC AGGCATCAGA CTTGAGCCTC TGAGCTGCAG GGACCCCTTG GCATAACCAC
14761 TTATTAGGCC CTGCTATGAC TTCTGGCCTA GGAAGAAGGC ATAGGAATGT GGGGCGTCTG
14821 AAGATCTCGT GTATGTAACG CTTCACTAGA GCACGGGCTG CTCTTAGTCC CAGTTTTGTG
14881 AATGAGGAAA CTGAGGCAGA GAAGGTGAAA TGACTCGACC CAAGTCACAG GCGGGGATGA
14941 TGCTCAGAAC ACTGAAGCGT GGGGACTGTG AGGATTACAC AGTCAGGACC CATGTGTACT
15001 GGCTGAATAT CAGGGCCCAT CATTCAGTCA CTCGTCAAAC ATTTGTGCAT GGCTGCTCTG
15061 GCCTGGTAGG CCTGAGCATG GCTGAGACAA GCCCTGGCTT TAAGAAGCTC ACTGTCCAGC
15121 GGGGACAGCT GCACCCTTCA ACAGCCAAAG CAACCCATGG GACAAAAGTA CGCTCAGCAG
15181 GAGGAGCACA CGCCCACGCT GGAGGGCGCT CTTGCAGGCA TGAATTCcag ctgagcgccg
15241 gtcgctacca ttaccagttg ggctggtgtc attaataata ataaccgggc aggccatgtc
15301 tgcccgtatt tcgcgtaagt aaatccattg tactgccgga ccaccgactg tgagccactc
15361 cggccatggc gtacgcactg acctgcttac tgatttgtaa aaccggtccg gccatcacgc
15421 tcacataacg tccacgcagg ctctcatagt gaaacgtatc ctccccggtc atcactgtgc
15481 tgctcttttt cgacggcgcg aaccccaggg aagccatcac cccacactg tccgtcagct
15541 cataacggta cttcacgtta atcccttca gatgactcac accggtatcc ccgcccgaca
15601 acgacggcaa tgtacccggt ttcacttgaa aatagcccac cgtaaacgta ccatgtccac
15661 cttccgcacg ggccggagtg actgtcaccg caagtgcggc aaagacagca acggcaatac
15721 acacattacg catcgttcac ctctcactgt ttataataaa acgcccgttc ccgggacgaa
15781 acctctgtaa cacactcaga ccacggctgat gcccagcgcc tgtttcttaa tcaccataac
15841 ctgcacatcg ctggcaaacg tatacggcgg aatatctgcc gaatgccgtg tggacgtaag
15901 cgtgaacgtc aggatccact agttctagag cggc

App.2

Mouse CgA gene sequence

```
   1 GAATTCGGAT CCTGCAGGAG TGGTGGGTGC TGTGCCAAAG GCCTTGGGCT TGGGAGTCAG
  61 ATGGCCCAAG GCAAAGGCCT GTCTTGCTTC CTGACCTTGG ACCAATGCCA GTGAGCTGAG
 121 GCCTCGAGTT CCTGGTGATG TTTCGGATAG TGGGTAGTTG TAGGGACTGG AAGAAAGCCT
 181 GGAAAGGCTT GGCTTGTACC TGGAATACAG TACGTGTCCC ATACATGTAG CTGGCACCCT
 241 TGTTCCAACC TCATTCTCCA TCCACCCTCA GGGCTCAAGG TCTCCTGCCT CAGGTCACCT
 301 TGTCCCATCC TTGGTTTGTA AAGTGTAAGA AGAGAATGAG CCAGCAAGGA ATCAAGGTGG
 361 TGATAGAAGT GGAGGTTTGT GTTAGCTGAG CCTTGTGCGG GTCATTAAGG CAACTAACCA
 421 TTTCACAGCA GGATCCAGAC CCACAGAGCC TGCTGTGTCA GGAGTAGAGC CCGCACTTGT
 481 TTGTCTTCTG TTATCTGATC TGCTTTGACA TCTGCTCCTG TCTGGATGTC CAGAGCTGAG
 541 GCAGGGCACA AGGGACAGAA AGCTAGTGGG TGCTTGGACC AGCCTGAAAA TGTACTCTAA
 601 CCACTACTAT CTTCCAACTC CTCTTCCTTC TACTGTGACC TCTTGTCCTC TTGTCCTAGT
 661 GTTCTTCAGG GATGGTGGCC TGAGAAGGCT CCTCAGAACA CTGTGAGCAG GGGCTGGAAT
 721 CCCTATCGAG GGTGCCCCTC CCTTGTCAGA TAAGAGCATC ATTAGGCAAT TGTATTTGCC
 781 CTTATCTCAC TCTGCTGTGA ATGGGACTTC TCAGCTATTA AGCTACAAGG CAGGCTGAAG
 841 GTCTGCAGCT AGGGGAGAGG AGGCAGGGAG GTGAGATTGC ATCTTCAGAT GAAAGGACAC
 901 CCAGTCCTTC AGCTCAGGGT GTGTTGAAAG GCCCTGGTGG CTCCTGAGAG CACAGCCAGT
 961 ACCTTGTCAG AAGGCCAAAA AGAGCATGGA CAGTTTGCTA CAGGGTCTGC ACTAAGTGGC
1021 TATTTAATCT CAGAACTCGT GCTCAACCTC TCTGGATCTC AGGTTCTGCA TGGGTCAGAT
1081 AGAAACACCA GGTTTCTCCT ACAATGATCA AACCTACCTG AAATTTAAGA GACTCTCTAT
1141 ACCTAATCCC TGTGTTTGTA CAGTGGCTTT AAAAAAAAAA GACAACTTTA TTGACATATA
1201 TTTGTACATG GCTTAAATTG TACAATTCGG TAAGTTTTTA CTTGCGTGAC TATCTGGGCA
1261 GTCGTTGTCA CAATCAAAGC TGTGCCTAGA GGCATTGCCT TGTTTGTTTG CTTTTTTTTT
1321 TTTTTTTTTT TTGATGTTGC TGTTTTAGTT GTTACTCTGA GACAAGCCTT TGCTATTTAA
1381 CCCAAGCTGG CTTTAAAATT CAGTCTTCCT AGTTGGCCTC CTGAGTGCTG TGATCACAGG
1441 GACAATCCTG CTTTTTCTGA TTGCTAAGCA CTGCCTTATT CACAGCAACG CCTGGGTCCC
1501 AAGCCTTTCC CCATCCATAT GTCCTTAGCT GGGATGGAGG GCAGGGCTCT CCAAGCCGTA
1561 GAGTTCTCAG TGTGTGCCCC TATCCTTTCT TCCTCATAGG AGAGCTATAA TGGACAGGTG
1621 ACCTCCCAAG ACCTTCCCAG GCCTGTGATT CCAGGACAAA TGCACTCGAG ACCCAGCTTT
1681 CCTTTTGTAT AGAAAGTCTT TGAGGGCTGG CAAGATGGCT CAGTAGGTAA GAGCACTGAC
1741 TGCTCTTCCA GAGGTCCTGA GTTCAAATCC CAGCAACCAC ATGGTGGCTC ACAACTACCT
1801 GTAATGAGAT CTGATGCCCT CTTATGGTGT CTGAAGACAG CAACAGTGTA CTTATTTATA
1861 TTAATAAATA AATCTATGGA CCTGAGCGAG TGGGGTTGAC CAGAGTAGAG GTCCTAAATT
1921 CAATTCCCAA CAACCCCATT AAAAAAAAAA AGAATTAAAA AGNAAGAAAG AAAGNAAGAA
1981 AGAAAGAAAG AAAGAAAGGA AGAAAGAAAG AAAGAAAGAA AGAAAGTCTT TGAGCCCTCA
2041 ATCCCGCTGT TCCGGGTGGG CTCACTTCAG TATAGTGTTA TTTATTAGTA TTCAATAGTC
2101 TTGGGAGCCG GGTGCAGGAT TAGGGTTCTG AGCCCCCTAT CTTACCACAA GACCCAAACC
2161 ACTTGTCCTA AAGCTGAGGG AAAGGACGAC TTGATTTCTT CTGGGGTCCC TCATCAAGGG
2221 TGACCACAGC TCAAGCTTTC TGGGGCTGTC CAAGGTGGCA GAATCCCCAG GTCTCCGTGT
2281 AACTCAGTAG ACTTTGTTAT TCCAGGCTTC CCACTTGTGA GAGGAACATC TGCATGGCCT
2341 CTCTGTGGAA GCACAGAAGA GTGAGTGTAT GGTGGTCTTG TATGATGACT GTAGGTATAA
2401 GCTCTCTGAA CATTCTTAGT AACCACAAGT AAGGCCATCC ATAACAAACT ATCTTTTGAT
2461 ATGATGTGGG ATTGTTTTTT TTTCTTTATA AAAATTAAAT TTAATTTTTT TATTTATTCA
2521 TTTTACATCC TGCTCACTGC CCTGCCCCCT GTCCTGGTCA CTCCATCACA ATCCTTCCTC
2581 CATCGCCCTT CCCTTCTCCT CTGAGTGAGT GCCCCCCCCC CTCTCGGTAT CCCCTTGCCC
2641 TGGCACTTCA AGACTCTGAA GCTAGGCCCT TGCTCTCCCA CTGAGGCCAA ATAAGGCAGC
2701 CCAATGAGAA GAATATAGAC GTGGGATTCA TTTACTAATT TTACTTACAT GTTTGTGCCT
2761 GTGCCAGTGT GAGTGTATGT GTACTATATG TACATGTGAA GGAGCCCATG GAGGACTGGG
2821 ATTACAGGTG ATTGTGAGCC TAATGTGGAT GCTGGGAACT GAACCAGTTC CTGCAAGAGC
2881 AGTAAGTGTT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT TTTCTCTGTC
2941 TCTCTCCCAT CTTTTTCCTT CCTTCCTTCC TTCCTTCCTT CCTTCCTTCC TTCCTTCCTT
3001 CCTTCCTTTC TTCTTCTTCT TTCTTTCTCT CTCTCTATGA AATAACTTCT ATTTAATACT
3061 GATTCAACCT GGAACTAATT GTAGACAGGG TGACCTGGAC CTAATAGTGA TCCCCCTGCC
```

```
3121 TCTGCCTCCC CAGTGCTGGG ATTAAAGGCG TGCACCTCCA TGACCAGCCT CAGTAAACTC
3181 TTCTAACTGC TGAGCCCCCC AGATATTCTC ACTTTATTGC AGAGCTTAGG AGGACGACCG
3241 TGGAGGTGTG GGCAGTGGGC CAGGGAGCCA GAGGACACAT TTCACCGCAG AAAGCAATCT
3301 GCAGATAATC CAGTTCAAAC TTCTTCCATT GCTTCTACCA TTTTATCAAT TAATCAATCA
3361 GTCAATTTAA TTAATTAGTC ACTAATCTAG TTAATCATCT AATTAATTAC CCAATTAGCC
3421 TAATTAATTT AGACAATCTA ATTAATCACC ATGTAGCCTA AGATGGAGTC AAACTCAAGA
3481 TATTTTTTTT CCTCAGCATC ACAGGTGCAG TAATTATAAA CCTGGTTGTT TCCACCGTCA
3541 TAAAGGGGAA AAACCGTAAT GATGAATTGA CTTGTCAAAA GTCATGTAGC CAGCGTGAGC
3601 TAGAGCTGGG GCAATTTTTT GTTTTGTTTT GTTTTTTTTT TAATGATGCA AAGGTCCTAA
3661 TAATTTAGGT GTGGTGGTGT GTGCGTGTGA TCCCTGCTCT TGCACAGGTG GAGCGGGTCA
3721 CAGTCTTAAT CTAAAGCCAG CCTGAACAAC ATGGGAGGAT TCTGTCTGAG CAGCAGCAGC
3781 AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAACA ACAACAACAA CAACAACAAC
3841 AGCCCACAGA GCCACTGTAA CTGAAATTAC ATGGTACTCA CATCAAAATG GACTCACAGA
3901 CAAACAGAAT AAAAGAGAGC CCAGACAGAC GCATACACAC CCATGAGCAT CTGTTGGAGA
3961 ATAAAACAGC CTCCTTTAAC AAACAGTGTC AGGAAGGCTG AACACCCATA TATAGAAGAT
4021 TGAAACCAGA ACCCTACCTC TCACCTCGTA TAAAAGCCCA GCTCAAAATA GATCAAAGAT
4081 CTTAATAAAA GGAATAAAAA GCCGTGAGCT TTGACTACTT GCATGAACGC AGGGGAAGGT
4141 GGGCATGGCG ACCCACATCT GTAATTCCAG CATTTGGCAT GGAGCTAGGA GGATCAGGAG
4201 TTCAAGGGCA GCTTCTGCTA TGCAGCAAGT CTGAAACCAG CCTGGACTAC GTGGGGCCCC
4261 ATCTCCCAAA AGTGCCAATA ACTTTCCTAT AGAGAAAGTA GGGCCTATAG ACCACAAAAG
4321 TGTCATTGTG TGGGGTATTA CGAGGGAAGA ATCGATCCTC TAGGCTTCAG TTTCTCAGAG
4381 GTGAAAAATA TACAACTGTC GGAGACTTGA CAAAACTTCA TAAGTTGTTT GCAAAAGAAC
4441 ATAGCCCTAA TTATTCTGTC ATGGTTATTT ATTTACTGTC TGTCTGTTTG TCTGCCTGCC
4501 TGCCTATTAA GTTTAAGTCA TTAGAAGGAC GTAGTATCTT AATACAATTT GTTATTTAAG
4561 TCCCCAAGCT AGGAGTCCCA GACAGGATGA GTGTTTCAGG ACCACGGACA GCGACACGGC
4621 CACACACTAA AGCTGGGCCT GAGACTCTGA GCCTAGTAGC GGTCTCTACC CCGCGTGCTC
4681 CAAAGCGAAG CTGCAGCACT ACCGCGGGGC ATGCATGTCA GAGTTCAAGC CTACCCACAC
4741 GACCCTCCTA TGGAGCCCCC GAGGAGCTGC CATCAGGAGG CTCTACATGG CTTCAGAAGT
4801 TCATAACTTA TTTCCAGAGA GAGATACTTG CTGGGGAACC TATGGCACAC AGACCAGAAG
4861 ATTGAAAAGG AAGCCGTGTG GAATCTGCAG CCTCGAAGCT CTCCCAGGTT TGGATCTTCT
4921 TGCCAGCTTG CTTGTTATGC TCGGTGTCAT TAATGAGCCA ACCTAAGGGT CTAGGGTTTG
4981 GTGTATGTGG CACTTCCCTC TGCTACCCCC ACCTAGGCTG GGGCCCACTA AAGCGGTCAT
5041 CTGCCTGCTG AGGAGGCCTG GGATCTGCTT TATTAGATGA AAGCAAACAT TTAACAAAGA
5101 AAGAAGCCAA GTCCTTAATT GGAGGCCCCC AGGGCTGTTT ACAGAGACAG ATGGGAGGAG
5161 ATCTAATGAA ACCACCTCAC ACAGTCCTTG GTGCGCACCT CACAGATAAA GCCTCATCGC
5221 CAGGCAACAG AGCTTTTCCA TCCAGTGGCC TGATAAGGAG CTTGGTAGGA CACTATGTCT
5281 GGAAAATATG ATTTTGTAAG CAGGGCCCAG AGTGTGTTGA CTCTGCCCAA TGCCAGAACT
5341 TACTGCCAGA TAAGCAGCTG CCACGTTCCC TGGGGGCACT GGCAGGGTTG TTTTTCCCAG
5401 GACCTCTGAC AGAGCCTGTC AGCAAGAAGA CACCCTGAGC CACAGCCAGA GAGGCCCATC
5461 ACTACCTAAG ATCAGATGCT ACTCAAGGGA CTGCTTTAGC CACCGAGGGC AGGGTCCAGG
5521 CCTGGGGTTG AGGGTGCAGA GACTGGCTAC ATACCTCAGC CCCTGCCGTC TTCTGCTGCC
5581 ACTCCAATGA CTTCCTCTCC CTCAGACCAG AAGCCAAAGT CCATGCAGAT GCTCATGGGT
5641 GTCTGGATGA CCCTGGCTCA ACACTGTCTC CCCTCTACAG ACCAGTGCTG GGCTCCTTAT
5701 CTTCTAGTCA CAGTTAATTT ATTTTTCTTC TCTACCTCAC ATGGGCTGGG TGAGTGGGCT
5761 GCCCCAGGGT TTTTGTATAA GCTGCTCATT CTTCATACAG TTAATGTTCT TCCATATCCC
5821 CCTCTCTCTT CTCTCTGTCT GTGTGTGTCT CTCTCAGGAG CCCTGTTATA CCACCCCCCT
5881 GCTATTCCAA GGCTCCTTAC CCTGCTCAGT GCTTTCTTTT CTCCCTAGCC ACCCCAGCCT
5941 TCTAACATAC TGTCTAATGT GCCTTATGGC TCATTGCTCA CTGCCTGTGT CTTCATTTTA
6001 GCAAAGCTCC AAGGAGGCAG AACTCTGGCT TTCCTTCCGC ATTAGCACGT CCCAGACATC
6061 AGACACAGAG CGCACAATAG GTGCTCAATA AGTGTTTGAT GACTAATGAA ATAATTTGGC
6121 TCGGTTCCAA AGTATTCTAG TTGAACTAAT GAACAGGAGT TCATGCGCAT AACAAGCCCC
6181 TGCTGGGTCA CTGAGCCTCT GGGAGGGGAT TGGCAACAAC ATTTAGGAAG GAGAAGCATC
6241 CTAAGACATT TCCCCAAACA ATCACCAAGG AACAGTCAAG CCACAGAGCC AACGTGGAGG
6301 GGGAAACACC TGCATCCTGG AGAAAGACCC AGACTCTGTG CAGTTACCAA GGACCAAAGA
6361 CCATCAGAGG AGCAGGGTAC TGTCAGCCCA CCCACACCCC TGAATCAAAT GGTGTTGACC
6421 AGCATTTGAG CATTTGAACT TGGGTTCCCT GAGATAATGA TTCTGTTTAC CAGAGACCCC
6481 TATGAGAGAG AGGGGGGGGG GGAATGGGAG AGGGAGGGGA GTAGACAAAG GATTGATGGA
```

6541 GTGTGCACTA TTGGCAGTGT GCTATTGATA AAGTAAGATA TCAGCGAAGT GTCTGTTATT
6601 AATATAAAAC TCTACTGATA AAGTACATTA CTGGTGGTGT CCTGTCAATA GTGTCTGCCA
6661 TTGATAAATC ACACTATTGA AAAGGTGCCG TCACTTAAAT ATATGCTATT GATAAAAGTG
6721 TGCTGTTGAT AAAGTATGTA CTCTTGGAGA AGATGGCAAA GTGAGAATCG GGGTTTAATT
6781 CTTACGGGTG ATAGTGATGT TACTGGAAGT CCTCTTCAAT GCATCCCACC CAAGGGGAGG
6841 AAGCCCCTTG TGAACTCTGA AACTAGAGGT GGCCTTTGGC TCCCAGTAAA TCTCTGTTAC
6901 ACACTATGCT AAAATAAGTA CATATTTATA TTTTTACATA ATATGAAAAT AAATATAAAA
6961 ATAAAACAGC CAGCACATCT AAACTGTGAC CCACATAAGG GTGTCTGGGA TGATGATGCC
7021 CAGAGTGCAA GGAGGGTTTA TTTGTTTATT TATTTATTTA GAAACAGGGT CTCACTATGT
7081 AGCTCTGGCT CGTCCTGGAA CTCGTCACAT GGACCAGGCT GGCATTGAAC TCAGAGATCT
7141 GCTTGATTCT GACCCCCGGG TTCAGGGATT AGACGTGTGT GCCACCATGC CTTGCTGAGG
7201 CTACTATTTA CGAGACGATG CTAGGAGCTT GCACATGATC ACAGCGACCT CCCATCCTAG
7261 GACCTGGAAC AACCGTTCTC CCAGTGCTGA GCTGGAGTCC TTCTCTAGGA ACTAATATTA
7321 AATGATAGGA GATGCCTTAA ATAAATGCCT TAAAAGAAAG TTGTCGGCCC TTGTCCCTGG
7381 GTGGAGCCAC TGTGAGGGAT CTCCTATATT TCTCAAGATT GAGATGAGAT GTGTGCTATT
7441 CATCTCTGAA AGCCTTGTAT GACCCACCTG CTTGCTGTTC CTTAGAATGG ACCAGACACC
7501 TTCTCCCTTA GGGGGCGTTT GGCTGGCTGC TCCCTCCCCT GGAGAGAGAG CTCCCCCTCT
7561 ACACAGGTTC TGGCTCACTT TAAATGCCAC CTCCTGGTAA GGCCTCCTCA GAACCAGCCT
7621 CTTCAATGTA TTCCTGTCCC TCACTAGAAC TCCGCCCGCT TCTCTGCTTT ATGCTCGTAG
7681 CATGGCCACT ACCCAGCACA GTGTACATCT TGTTTCTTTC TCTGGAAAGG GAATTCTATA
7741 AGGGTTGGGT TTGCTGTTTT GTTTACAGCT GCGTCTTTGG CACCCAGCAC AGCTGAGTGG
7801 TTCTAAGCCC ACGTCGATGC TTAACACATG GTTGTTGAAT GAATACACGC GAAGCGGTTC
7861 TCATTTAGGG GCATGAGTAG GCAGAGGTGT GGGCAGGAAG CAGGAAAGAG CGGAAACAGG
7921 TGCGGACAGA AAGGAGGGGC TCTGAAGGAT GCCAGTCAGT GCCAAACTGT CATCCAGATA
7981 CCAGGTTCAC TGTGGCCCTA GGCCAGGCTG CACGGGGCTT CCCATGTGGT CTGCCCAGGG
8041 TGAGAGCAGA ACTGCGGTTG GGCGGGGCAG AAGGAAACCA ACCAGGAAGC AGGGTTGCAC
8101 CCAAATTATC CAGGTTTTAA GTACATTTAA GAGACAAGGC TGGGCTGTTG AAGGCAGAGG
8161 TGTCCCTGGG GTGCTGGACT AGGACTGACC ACTTCTGTTT TAGTTTAATG GTGAGAACTG
8221 CCTCACACTG CTACCTGCCT TACTTGCCCC TTGAGAGCTG TGAGCCTAGG ACCCACCCAT
8281 GTGTGGGTTG GACCTTCAGT CCACACACTG AACGTGTGTG AAGCCACTGG TTGTCAGCGC
8341 AGGGCTCTCG GCACTGAGGA AGCAGTGACC ACTATGCCCT ATCAAATAAC AATTAAATAC
8401 ACACAGAATG CGAGGCACAC AACTGAGTTT CAGGAGAGGC CTCGCTCAGG CAAGGGGTTC
8461 AAGAGGCTTC TGTGGGGACC CGCTGGATGT TCCAGGGAGT TCTTAAAGAT GGGCGTGCCT
8521 CCAGCCAAGT GAAATCAAGA GAAAAGTACG CGAAGTATAG GAAAACTCAG CAGTCTGGAG
8581 AGGTAAATAG GGGAGGAATC CGAGGCTCAG AGACAGGAGT GACTTGCCCA CGGACGCACA
8641 GCAAGTTGGC AGGTGGAGTT TCAGCTGTGC CACCTTCTGA AGCCGGGTAC CCTTTACAGC
8701 CACCAGATAC AAGCGGGATA GAGACAGCTG ATGGAGAAGC TGGAGGTGGG GGGCGGGACC
8761 CCGAAGTTGG GGAAAGGGCG CGGGGGGGCG GTCCTATGAC GTAATTTCCT GGGTGTGTGC
8821 GCGCGTGTGC GTGCGTGTGC GTGTATATAA AAGCCGGCAT AGCATTGC***TG CTGCTGCCGC***
8881 ***CGCCACCGCC ACCATCACCG CTGTTACCAC CACCGCTACT GCAGTGTTCC CGCTGGTGCA***
8941 ***GAGCTTTGGT AGCCAGACTA CAGACCCACT CCCGCCATCC TCCTGCAGCA GCTCGTCCAC***
9001 ***TCTTTCCGCA CCGTCCGGCT CGCTATGCGC TCCACCGCGG TTCTGGCGCT TCTGCTGTGC***
9061 ***GCCGGGCAAG*** GTGAGCGGCT CCGGGCTGGG CAGCACGATC CCCTCTCTTG CCCCAGGTCC
9121 CTAGAGGAGC CGGAGGATTC AGCACCTCGG ACAGCACCAG GTCTCCTGCT TGAGTCCAGG
9181 CGCAAGACCC ATAATGTTGG CTATGGTCGC CCACCACACC CGAAGAAACT ATAAAGCCTG
9241 ACCCAACCCC ACTGGGCAGA CCATCCCACC CATGTCCGCA CACCACCCAC TGGTGCTCTG
9301 GCTCCACGCC GACCAAGGTC ACTGGGGGGA GATAGGGGAC TGGCGCGGGC TTCTCAACCC
9361 CTCTCCTGTC CTTCCACCCT CCAGCCTGTC CTTGATCTCT TCCCTCACCC CTGCTCTTTC
9421 CCGCCACGTC TGGGGTTTCT CCCTTGCCTG CCTCTGTTCC TATCTCAGTC CCACTACATG
9481 GCTGGGACTG GGCTTGTTTT CATTGAACTT GATTTCAGCT CTGTATTCCC ATGGGGAAGG
9541 GTTGAGGCTC TTCAGTGCAG CCTCTGAGCC CAGCCCCCCC CCGGGGTGCA GCATGGAGGG
9601 GGGGATAGCA GAATCAGTGC TGCTAACTTA TTTGGGTTTT GAGGGTGTGT TGGTTGGTGG
9661 GTCCCTTGAG CCTCCCGCAG GAAGGACTGG GGTTCCCTCG TTGCAGAGAG ATAGCCCCAA
9721 TTAGCAATCT CAAGTGCTCT CTCTCTCGCT CTCTCTCTCT CTCTCTCTCT CTCTCTCCAG
9781 ***TTTTTGCCCT TCCTGTGAAC AGCCCCATGA CAAAAGGGGA CACCAAG***GTA AGAAGGAATG
9841 TTGGGGTACC CTGGGAGAGG AAGGGTTCCT GTGGACTCCT CAGGGCCAAT CCTTGGGCAG
9901 CTGCAGGAAA GAGTCTGGCA TAGAGCCAAA AGGCAGCACT GCAGCGGCTG AGCCTGCACA

```
 9961 GGGGACAGCT TGCACAAGAA GACATCAAAG ATCTCTTGCT TTTGTTGCTG CCCAGCTACA
10021 GTGACTTCCC AAGGGGCCCA TGTCTCACCC GATGAGTCCT GGGACCAGGG ACACTGGTCT
10081 GAGCCATGAC CATAGCTTCT ACTGCCCTAG GTGCACAGAC CTGCCTGGCT GTCACCTCTC
10141 TTTGTCATTC CTCTTCTCCT ACCCTCACTC TGCCCAGAGC TTCCTCTGGC AGTAAGAAAG
10201 AAGCCACATC AGATGCTAGG ACCACATTCA CATCAGAGCC AGGGGTGTTC TTACTCCAGA
10261 GAATCGGAAA GTTCCGTTCT CTATGTTCGG AGGGCTGGCT TCTTATGCAA GAAGCCACAG
10321 TTTAAAAATT GTGCCGTTAA AGGGACTCAC CGTTTACTTG TTGAATTGTA TTCACTTGTT
10381 TATTTGTTGT CACTGAACAC AGGGCAGTGT TTAAGTTTAC ACTACTCATT CTAGCTGTGT
10441 AACCTTGGAA AATTCATCTC TCTATATCTC TATCTCTATC TCTATCTCTA TCTCTATCTC
10501 TATCTCTATC TTCCTTCCTT CCTTCCTTCC TTCCTCTCTG TCTATCTCTC TTTTTTAAAG
10561 ATTTATTTAT TGTTATATGT AAGTACACTG TAGCTGACTT AAGATACACC AGAATAGGGC
10621 ATCAGACCTC ATTACGGATG GTTGTGAGCC ACCATGTGGT TGCTGGGATT TGAACTCAGG
10681 AAATTTGGAA GAGCAGTCAG TGCTCTTACC TGCTGAGCCA TCTCTCCATC CCTCTTTCTT
10741 CCTTCCTTCC TTCCTTTCTT TCTTTCTTTC TTTCTTTCTT TCATTTCTTT CTTTCTTTCT
10801 TTCTTTCTTT CTTTCTTATT TTGAGACAGG GTCACTCTAT GCAGCCATGG CTGACCTGGA
10861 ACATCAGTTA TATAGACGAG GCTGGCTTGA ACTCAGAGAA ATCTACCAGC TTCTCTCTCC
10921 CAAGTGCTGG GATTAAAAAT ATGCACCACA TTATACTGCA CCCCAGCATT CTTTTTTGTA
10981 TGATTGTTTA GGTTTTAGGT TTTGTTCTAG ACAGGGTTTC TCTGTGTAGC CCTGGCTGAC
11041 CTGGAACTCA CTTTGTAGAC CAGGCTGGCC TCGAACTCAG AAATCCACCT GCCTCTGCCT
11101 CCGGAGTGCT GGGATTAAAG GCATGCACCA CCACGCCCGG CTGCACCCCA GCATTCTTAA
11161 GGCCCTTACC CAACTCCTGT CTCTTGGTGT CTGCCCTCAA AGGCTGGTGG TTGCCAAGTA
11221 TTCTCCATGT AAAGTCTAGT GTTAGGGAGA CAGTACATGC TTCTAGGAGA TACTTACCTG
11281 GTCCACTCCA CACCATAGTA TGGGAGCAGG GGTCAGGGGT GCCAGCAGGT AAGGAAGACC
11341 TGTATGCTCC CTTTGATGCT GCTATGACTA CATACCCTGA CAACCTTTGC CAGCTTGCTA
11401 TGGCATCCCA TAGACACAGC CTCCCACACC TTCACTCCGT GCTAGTGTTG CCTGACGCAT
11461 GTCTGGCACC TCCTGCTTCT CCACCCAGCT GCAGGCTGTC GAAGCAGGTC TGATGTGATT
11521 CCCTCCCTGA GTGAGCAAGC GAGCCTGCGT GGCACTTCAT AGTGGGCAGA GTGTTTTGCA
11581 TTGGGGAGGA GTTCTCATCC TGCTTTACAT GAGAGGAATC CAAGGCTCAG AGAGACTAAG
11641 TCACTTGCTC AAGGTCACAC AGCTAGCATA CCACACAACA GTGGACACCA ACTTCAGAAA
11701 TTATCCATGG TGCTTTGTGA TTTGCTGCTC ACCTCACAGG GTCGGATAAA ATATTCAGAG
11761 GTGTGTCGTG TTGGTCTGGA CCAGACATCC AGCTGTAACT AGTTAGTATC CACATCCAAC
11821 CAGGGGATTC TTCCCGTGAA AACGTGGAGC CACACACTGT GGCTGTAGAA TGGGAGGTGG
11881 TCCTGTTTCT GGCAAGTTCA AATCCAGGTA GGTTCGTGTT CTTCAATGGG CAGTCTGGAG
11941 GAACTGACCC TATGGACTCT GCTAGACCCC AGTCCTTTGC AAGAGGTCCT TTTAGGTCTC
12001 AGAGGCAGCG CCATCCTTGG AGAACTACCT TACGAGTGAG AACATGTTAG GCACCTGCTC
12061 CCACCCCTCC TTTAGCAGTG GTTCCCTCAA GACACATGCT GGAATGCCCA TGTGATTTGA
12121 CTGGCCCACA GTAAGCAGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGGT
12181 GTGTGTGTAG TGTCCTGTTT CTGTGGCTCT ATGGTCTTAT TAAAGAAGCA GAGCTGGGCC
12241 CCATGGCCCC TTGAAGACAA CACAAATGCT CTTCTGCCCT TCTGGAAAAG TTGCCCTGGC
12301 ACCTCAGGGG CTGAGCTCCA TGTTCACACC GGCCACAGCC CACCTTGCTG TCTCCCAG***GT***
12361 ***GATGAAGTGC GTCCTGGAAG TCATCTCCGA CCTCGCTGTC CAAACCCAGC CCCATGCCTG***
12421 ***TCAGCCCTGA GTGTCTGGAG ACCCTCCAAG GAGG***TAGGAG TCAGAGGCTG GAGGAGGAAT
12481 TGGGAGGCCA GGGTGTGGGA GGGTAGCTCT TAATCATCAC ATCCAGTACT CATTGCCCCA
12541 GTGCCCTGGT ATGTGCTAGA ATCTGGGGAA GACACAGGAA TGGGTGACAG CTCCTGGGAC
12601 CCTGAGGTTT GTAGACACTC AGAGGAACAT ATGGCAGTTA ACTAGAAGCC TGGGCAGAAT
12661 GAAGGAGCTG GAACAGCAAG GTGCAGACTG GGTGCCCTGC GTACTAGGGT TCTGTGTTCT
12721 GTACTCTCAG TGTTTGGAGT CTCGCTGACA CCAGTAGCCG TTCTCTGGAT CTGGGATGAA
12781 TTCAGTGCTG AGTGGGCCTA AGGAGGACCA CATCTTTCCA AACAGACAGC CCAGTTTGTC
12841 TAGCGCATGA TAGCGCTGGG ACAGGTAGAA AATGAGGAAG GATGTCAGGA GGAACAGACA
12901 ACAGGCAGCA TGGAAAAGGG AAAGGAAGAA TTTACACACT CACTGGAGTC TGAGGTCCAC
12961 AGTCCCTTCC TTGGAGGCTG GGACCAGAGT CAGCAGGTGA CCTTTCTGCT GCAGTAGGAA
13021 GGTGATTGAC ATCTTGGAAA TGAGGAATTT CCAATGCGTT CTACCAAGAT CCAGGCTTGC
13081 CCCTGGTGCC AAGTTTCCAG AATAAATGCT TGAGCTCAGC TGAAAATGTC CACACTAATC
13141 TCTTCTGAGA AGCCACCAGT AGCCATGCTT TCTCCCTGCA G***ACGAGAGGA TCCTCTCTAT***
13201 ***CCTGCGACAC CAGAATCTGC TGAAGGAACT TCAAGACCTG GCTCTCCAAG*** GTATTTCAGC
13261 AGCATACGCA GTTCCACACA GAAAAGCCAG CAACAGAGCC TGAGGGGACC TGGGTGGAGG
13321 CTGAGTGTAC TCTTCCAGGT GCTCAGTTCC AGAGCAGAAA GGCGGGGTCC TCTGTGTCAC
```

13381 CTGGCTCCAT TGTCTAGAGA CTAAGAGCTC AAGGCTCAGT GAGGCTCAGG GATTCCNTCA
13441 GGGTCACAGT GAGTAGTCTT GTTAAGGCAG TCCAGGTCTG CTGATGTTCC TCCGCCATCC
13501 AGGGAGGCCG AGAGAGCTCT TCTCTGCAGA CCCAGAATTG AACTTACACA GAGAGAAGAC
13561 TTTCATTTCC TGCGACTCTT TGTTCTATGG GTAATGCTTG CCGATAAACA CAGAGCAGAT
13621 GAGTGTGTTT CGGACTGAGA GGTAACTTGC CCCCAAGAGA TTAAGAAAAC AGTAGACAGG
13681 CAGTGGTGGC ACATACCTTT AATCCGAGTT AAGCCAGCTG ATCTACAGAA TGAGTTCTGG
13741 GACAGCCAGA GCTACACAGA GAAACCCTGT TGCAAAAAAA CAAAGAAAGA TAGTAAAAAT
13801 CAAAACTAAT GAACAAACAA ACAATCAATA CAAAAAAGAA AGAAAGGAAG GAAGGAAGGA
13861 AGGAAGAAAG AAAGAAAGTC CACAGATACT TTTCAGACAG GAGTGTGAGG CTGGGGTCTC
13921 ATAGCATTGA AGATATACAA GCTAGACAGC GGCTAGGATG GAGATGTTCT TAGGCTGCAG
13981 TCTCCCACCC TTCCTGTTTG AGGTCGTGCT ATAGATTGAT CCCAGGGCTT CATGCGTGCT
14041 TGGCAAGCAC CCCACCAAAT GAAGCCCAGC TCCAGTTCTA TGGCATATTC ATGGTATCTC
14101 CTGGCACACG GAGGGTGGTG GGTGGTATAG GGGTTGTTAT AATCAATTTT GTCCTGGCTT
14161 TGTAAACACA AACAAGGTCT CATTGTGTAG CCACACTGTC CTGACTCTGC AGTCCTGCCT
14221 CAGCCTTTTG AGTGCTGGGA TTGCAGGCAT GAGCCACCAC ACCCACCATG CCAGGTCCAG
14281 GTACAAGACA CTGTGGGTCT GCGCCCTTAA ACATGCCCAC TTTCCTCAGG AGGCATCCAG
14341 GGTAGAAGCA GGGCTGGATG CCAGGCGATC CGGTCTCCGG GGTGGCACTT TTTAAACAAA
14401 ACATTAGATA TTTTAATGAT AAATTTAAAA TTAACTATTG GGCAGATATT CTATTTAAAT
14461 AAATAAGTAA TTCAATGAGA CCATTAGAGA GGAGTGTGCT CTATTCTACG TTCTCACAAG
14521 TCTGGATCTA ACCATGCACA GCATCCAGTG CACACAGGGT TGGTTTCCGT TATGAGGATA
14581 CTGGGGTGTG TCTTCAGCTA CATCATCATC ACCCCATTCC CTCCTGGTTT CTATATAATA
14641 GCCTTGCATG GAGACCTCCC TCCACTTTAA CATCTACCAC AAGTCTGTTT ATAACCTTGG
14701 GGAGCCCATG TTACCCAAGG TCACAGAGCT ATAGGCATGG CCAGGGCAGG CCTGGAGCTC
14761 ACTGTCGCTG GGTTGGGCTT ATTTCCTAGC CTACAGTGAC CCCACAGCTG CAGAGCTGTG
14821 TCCTGATTCA GACACGGCTG CAGGATACTC AGGGCAACAT CTGCAGCAAT AGCTGCAGGG
14881 GTTAGGTTGG CAGAGGGACT CTGGGCAATG TAGGTATAAC CCTGGGCTCA AAGGTAGACA
14941 TAGCACCCAC CTTTTCCTGG GGCCAGACCT GGAGTTCCAG AACAGTGGAG GCTGTTTAAC
15001 AAAAGAGCAC CCATGAGGGC GTGCTGCCCT GGAGATCTGT CTGTCCCATT TGGGCCTTTA
15061 TTGCTCTCTA AGGTACAAGT AGAACTTTCT CACATTTCCC ACTAGGGACA CTGAGTCAGA
15121 GGGGTGGTGT GGCCTGATGG GTATAGCTGA GAGCCTGTAC CCTCAAGGGA GGCCTTCAGT
15181 GTTACGGACA GGTATGTCAG GTCCACATGG AACCTTCCCT TGTCCCCTAC TACAAG***GTGC***
15241 ***CAAGGAGCGG GCCCAGCAGC CGCTGAAGCA GCAGCAGCCG CGAAGCAGCA GCAGCAGCAG***
15301 ***CAGCAGCAGC AGCAGCAGGA GCAGCAACAC AGCAGCTTTG AGGATGAACT CTCAGAAGTG***
15361 ***TTTGAGAACC AGAGCCCTGA TGCCAAGCAC AGAG***GTTAGT GTGGCCAGCT AGGGGGGAAG
15421 GGAGCTACAA GGAGAACAGT AGTCCCCAGG TGATGCTGCT TCTTCACAGC AGGAAGTCAG
15481 CTACCATTGA ACTTGGCCTC AGGAATCCCC TAGGGCATCC GGGTCCTCTG CTTGACCCTA
15541 TCCATTTGTG TTTGCAG***ACG CAGCAGCAGA AGTGCCCTCC AGGGACACTA TGGAGAAGAG***
15601 ***AAAGGATTCT GACAAGGGAC AACAGGATGG CTTTGAGGCA ACCACAGAGG GACCTAGGCC***
15661 ***TCAGGCCTTT CCAGAGCCTA ACCAAGAGTC CCCCATGATG GGAGACAGTG AGTCCCCTGG***
15721 ***GGAGGACACA GCCACCAATA CCCAATCACC AACCAGCCTC CCCAGCCAGG AGCATGTGGA***
15781 ***CCCACAGGCC ACGAGAGACA GTGAGAGAGG CCTGAGTGCC CAACAGCAAG CCAGAAAAGC***
15841 ***CAAGCAAGAG GAGAAAGATG AGGAGGAGGA AGAGGAGGCT GTGGCTAGAG AGAAGGCTGG***
15901 ***GCCTGAATGA AGTCCCCACT GCAGCATCCA GTTCCCACTT CCATGCAGGC TACAAAGCGA***
15961 ***TCCAGAAAGA TGATG***GTATG TATGGGGACA GGGACCTCAA CTGATGTGTC TGGGAGATGG
16021 GTGGGTAGGG ACTTCATGCC AACCTCATAA TAATCCTGAT AGATGAATAG CATCTCCACT
16081 GTGTGTGTGT GTGTGGGGGG GGGGGGTAGG GAGCAAGAGG CTCAGCGAGG TTATGTAATT
16141 TGCCTAAGGC TACACAGCTA ATTAAAGACA AGTGTAGTTT CAAGCCCAGT TTATCTGACT
16201 CCTAAACCAA GTGTCAACCA CTCCTCTACA CCCGGGAGCA AGAGGAGCTT AGGGGGCAGG
16261 TGTTCCTCTT AGAGAAGGGA GAGACAACAG GGTGAGCTTC AGTGAGGGGC TCTGCCAACT
16321 AGCTCCTTTG AGGTTTGCAG GGTCTTGCTT CATGCTCCCC AGTGGAGCTG GCTGGATCCT
16381 GACGGTGAGC TTGGCTTTAG AGGTGGCTGC ACAAAAGACC TTGAGGAGTG GCAGGGGCTG
16441 GGGAAATGGC ACTCCTTCCT CACTTCCTGT TCCTGCCACA CCTGCTCCAG ***GTCAGTCGGA***
16501 ***CTCTCAGGCA GTGGATGGAG ATGGAAAGAC AGAGGCTTCC GAAGCTCTGC CGTCTGAAGG***
16561 ***GAAGGTGGAG CTGGAGCACT CTCAGCAGGA GGAAGATGGG GAAGAGGCCA TGGTAGGGAC***
16621 ***CCCCCAAGGT CTCTTCCCCC AAGGCGGGAA GGGCCGGGAG CTGGAACATA AGCAGGAGGA***
16681 ***AGAGGAAGAG GAGGAGGAGC GTCTGTCCAG AGAATGGGAG GACAAGCGAT GGAGCAGGAT***
16741 ***GGACCAGCTG GCCAAGAGCT GACAGCAGAG AAGCGGCTGG AGGGCGAGGA TGACCCTGAC***

16801 ***CGCTCCATGA AGCTCTCCTT CCGGACCCGC GCCTATGGCT TCAGAGATCC TGGGCCTCAG***
16861 ***CTACGCAGGG GCTGGAGGCC ATCTTCCAGA GAAGACAGTG TGGAGGCCCG AAGTGACTTT***
16921 ***GAGGAAAAGA AGGAAGAGGA GGGCAGCGCC AACCGCAGAG CAGAG***GTTGG TATAGGCAAG
16981 gGCAGCCACC CCCAGAAGTC CCTCTCCCAC TGAGGTGACG TGGTTCCCTT AGGTGGGTGC
17041 TGGACTTGGG ATAGGGGGAA CAACTTGCTT ATAGCAGGGG CTCAACAGAG TTACTGGGTC
17101 AGGTGGACAG TGGGAAATGG CTGGAACTGG AGGCCTCAGA CCTGCCTgAG GGACCGGAGt
17161 GGGGGCATTg TCTCTGCTGG TATGTGAGCA GGACACGGGA GTGATAGGGT CAGGCCTGGT
17221 GCTTTGGATG GAGAGCTTgG CAGGACTTGC CTGGGCCTCT CAGTACAGAC ACCTCTTGTC
17281 CCAGGAACTG GAACGGATGC TAGGGGGACC AGTGGTGACC GAGGAAGACA GCCTGGTTGA
17341 TGGCCATGTC TTCCTGGTCC CCAGGGAACT CAATTGTGCC TCCTCCTGGG TTTAAACTCA
17401 CAGAGGCCAC AATGGCATAG CAGGGCGAGC AGATAGGGAG TCATGGAAGT GCTGGCAGAT
17461 TTGGGTTCCA GCTGGCCCTG GCTGCGTGAC TTTGGGCAGG CCATTACCCC TCTCTAGGTC
17521 TCACTTGTTC CCCTTTGTGA AGTGGAAGTA TGGGAGTACC TCTAACTTAC CTTATCATAA
17581 TCAGAACATG ACCTTgAAAA CCCTACTATG CCCAGGGTCA CTTGGGCAGC TGCAAGGGTG
17641 GTATAGGGAC ACAGAGATGT GACCTCATGG GAGGTGAGAG CTGGGCTTCT TGGCAGGACA
17701 GGATgcCTCC TGAGCTGGTG GGACCTGGCT AGGCCAGTGC TCcGGAGGGA GTTGGGCTCC
17761 TATAGGGAGC CTGATCTATC cACAGCCCTG ACTTCAAGGC TCTTCCAATC AGGCACAGaA
17821 GAGGACCTTC TTTAGGACCT GCCCCTGTCT TTCCACCTGA AACAGGGTCC CGTTGATACC
17881 TGAGAGCAAT AGGCTCTTCT TAGGAGCGGG CGCTGGCAGG AAACATCTGT TCCTGGAGGT
17941 GGGGGAAGAG TGAAAGCATT GTGCTCAAGA TGGGACAGCC TCAAGAACCC CCTTACCCAC
18001 CCATAAAAGC TTCTTTCTCA TGGACAAAGG GAACCGGTAG TGAAGAAGTG GTAGTGAAGA
18061 CCCTGGGTGT TATGGACATG GTGAGGCTTG GGAGAGTCTC GTTAGAAACT TTAGCAACAC
18121 AGAACTCATA CTCCAGAGGC TGCTCTAAGA TCAAGAGATC TAGAATCCCT AAGTCAAAGG
18181 CGGCCACTGT CAACATTCTG TGGCCATCGC ATTGGCAGGA CAGTGTAGAA CAGGCTCCTA
18241 GTGGTCCAGA GTCCCATAGA CAAAATGGGG GCTGGATAAT CAGCCATGGG GGCCTTGGAC
18301 ATAATGGTCT CCAGAGAGTC CTAGCATCAC TGGGTGGCAT CGTTGTGCCT ATAATGCGGA
18361 CAAATCTTCT GGGGCTCAGA GGCAAAGTCA CTTGTCTGAG CCTAGCAGCT AACAAACAGC
18421 AGGGTTAGTC TGGCACACTG TGCCATCGAC ACTTCTAACA CAGGCCAGTA CCAAGTGACC
18481 CTGTCTGTTC CCTCCTAG***GA CCAGGAGCTA GAGAGCCTGT CAGCCATCGA GGCAGAGCTG***
18541 ***GAGAAGGTGG CCCACCAGCT GCAGGCTTTG CGGCGGGGAT GAGGCCCTGG CTGGTGGGGC***
18601 ***CAGCCATGGC TTCAAGGCCA GACTGCTGCT CGAGTAGGGC TGCTTCCAGC CACAGAGCCC***
18661 ***AGGTTACCCC CTTTGCCCCT CACCTTCTCT TGTTCTCAGC CCCTGCCCTG GACACTTCTG***
18721 ***CAGGGCAGCC CTGAATGTCA GCACAGATTC CTTCTCTGAA CACAGGCAGC TTTCTAGAAG***
18781 ***TTTCTCTTCC ACCTTCTTAT CCATGGGGCA CAACTGCAAT AACTTCTGAC CTTTGGGCGA***
18841 ***AAGCGAGAAC TCCTGACTAT TAAGATATTC CAGATAAAAT ATATTGAAGG AAGAATAAAG***
18901 ***TGCTTTGGGC TCTTT***GCTTT CTTTGATTTA TTTTCGTTCC CTTTCCGGTC TGAGCTGGGT
18961 AGGGGACGAG GGACAGGGTA GTTTTTACAC TGGGCTGGGT GTGTGGTTTG GGAGACGACT
19021 CCTTTCTGGC CCATTTCCTG GGCAAGCGAC TTTGCACTGT TGAACCTCAG TTTCATCTGT
19081 AGATGGAGGT GCTCAACTGC CCAGGGTGGG GGTAGGGTGG GGATGCGCAC CACCACAGGG
19141 GACTTCATCA CACCCTGCTT GCTTATCTCT CAGCCACGCC TGAGCCAGGC ATTTTCTGCT
19201 GAGACCGGCC TCCTGGGATG TCCCCTAGTT ATCTTGGTGC TGGGTGGAGG TTGACAGGCC
19261 CCAGGATGAT ACTTTCCATT GGCCATAGGT TGGCCATGCT GGACTCCCTA GCATGCTTCT
19321 TGTTGCTGAA GTCCACCGTG CCCATGACCA CAGCTTGTGG TATGCATGGG ACCACAGGAT
19381 GTCTGAAGAG TAAGGCAGGG ACAATTTACT AGAAAATAGC AGGTCTTAGG TCCAATTTTT
19441 AAAGAGGGAA ACAGGCAGGG AAGATAAAAC GGCACATTGA AGTCACAGGT AGAGCTGACC
19501 CCCCCCCCCC TTTATGTTTT TCCAAGACAG GGTTTCTTTg TGTAGACTTG GCTGTCCAGG
19561 AACTAGCTGG CCTAGAACTC AGgGATCCAC CTACCTGCCT cTGCAAGTGC TGAGATTAAA
19621 GGCGTGCACT ACCAcAGCCC TACTGACATA CCACCAAACT TTTTTTTTCT TTTTTAGATT
19681 TATTTATTTT ATGGGTGTTT TACCTGCATG CCTGTGCACC ATGTGCTGTC TGGTGCCTTT
19741 GGGTCCCGTG TAACTGAAGT TATGGATGGT TGTGAACCAC CATGTGGATG CTGGGAACTG
19801 AACCTAGGTC CTtTGCAGGA GCAATAAGTG CTCTCAACCG CTGAGCCTCA ACACACCAAA
19861 CTTAAGACTT TTGTAGGTTA TTCAATCAGG CCAGGCTGGG ATGTGGTACT GCACCTATGG
19921 ATGCTGGCTA CCAGGACCAA TCATTCAGTC ACTCACCAAG CACGCCTGCA CATGGAACCC
19981 ATGATGAAGC CTGGCTTTGG AAACCCATGG CCTCAGGGAT AAGTAACATC CTTCTATGGG
20041 CAAAGCAGCC TCTCGTGTGA CGGGCTTGTG TGTGCCTCTG CTGGGGTCAG CTTCAGACAG
20101 GAACGTTACC TCATCCCGTT CTCTGTGTAC CTCAACTGAA AACCAGGTGG TGATGATGAT
20161 GCTAACCATG TGGTCTTAGC AGTCAGCCTA TGTGAGGCTC TCCTGCCAAC GAAAGCCTGC

```
20221 CCCTGAGGTG GAGACATCTG TATGATGTCA CTGCTGAGGC GGGGCTTTCT TCATGTTATG
20281 CCTCCTGATA GTTAATTACA CAAACTCCCA CAAGGCCACC AGAGACCAAG AATATCCACC
20341 AACTAAGACT TTAAGTGGCA ACGGACACCA TGTCACCAAA GGACCCCATC CCAAAGGCAA
20401 GCCACCTGCA TTGTATCATG GGAAGCAAAG ATGCAGCCCA GGCAGGCACT GACAGCCAGC
20461 TTTGTAGGCA GTACAGGGCC AAGTGTACGT GATATACTGA TGTGCACACG ATTAATGGGA
20521 GAGGAAGCTG CAGCCATCAT GGACACAACT CAGCTTCTGG GGTAGGAAAG CTCTGTATGG
20581 ACCCTGGATG AGCTGTCAAT TAGGGAGCCT CAGGGAACTA TTATCTCCTA TGTCCTTCAG
20641 TAAGTAGCCA GCCCTTCTAA GATGCCATTA AACCCAGGAA TCAGAAAACT CAATCATTCC
20701 TGCAACGGAC TCTTGAGAAC TGGCTAGTAC TTCCTAGTCA CAAGCAGAAG GAAGAGGACC
20761 CAAGCAGGAG GGTCATGCTG CTATGGAGGT CAGTCCTTCT CTGCACAGAG CACTGAAAGG
20821 GGTCATCCCC AGAGTGTACT GGCATTGGCT TGGGAGAGGA AAGGAAGTGA GTCATAGGTC
20881 TTTACCATGA GGGGAGAATC TGTGGAGCCG CCTTACAGCC ATCATCTGAG CGGGAGCCAG
20941 CTGAGCAGCA GCAGTCACAA CAGAGCTATC CAAACTGTTG GCTCCTGGAC CCCAGGGGGT
21001 TCCCACAGAG GGCCATGGCA CCTGTTCATC CACACTGGAG GAGCAGGCTC TTTCAGGCTA
21061 GCTTCAGGGC TCTGAAGTGA CAGGTGCAGC AGGCCCAGCA TGCTGGGGTG GGGAATGTTG
21121 CTGGAATCCC ACTGGCTAAG TGACTGATTG AGAAGAAATA TAAAAAGGCA GGGAACCCTT
21181 TCCAGAACAC CAGTAAAGCC ATCTGGTGGC TCTACAGCCA GCAACCAGTG TGACTGCTCT
21241 CGACTACAGA CAAGGACTAG TGGTCATGTC AGATGGATGA CATTGTCCTT TTGTCTCTGA
21301 ACGCCTGGCC AGCAGACACC ACTTGCCCCA GCTTCAGGGC TGGAGAGATG GCTCAGCAGT
21361 TAAGAGCACT GACTGCTCTT TTGAAGGTCC CGAGTTCAAA TTCCAGGAAC CACATAGTGG
21421 CTCACAACCA TCCGTTAGAT CTGATGCCTC GTTCTGGGGT GTCTGAAGAC AGCTACAGTG
21481 TACTACTACT ACTACTACCA CTACCACTAC CACTACCACT ACAACTACCA CTACTACTAC
21541 TACTGATAAT AATAATGAAA AAAAGAACCC CAGGTTCTTG CTCTGGCATT TGGGGCTATC
21601 TATAAATGCT CCAAGACAAA AAAAGGGCTG GATTACTCTC TGCCCACTGC TTTAATATAG
21661 GGTTTCTCCC TGAACTTGGG GTTCATATTT TCCTACGTAC GCTGGAAGCC AGCAAGCACC
21721 AGTGATTCCT CCGTCTCTGT CCCCTACACA CCTGAGCTGG AGTTAGTGTG CGGATGCCTG
21781 GCTTGTTATG CAGCTGCTGG GATCTCAAAA CCTCACCACT ACATAGCAAG CGCTCTTAAC
21841 CACTGAGCCA TCTCTCCAGC ACCCCCACCA ACTCCCTGTG GAGCATGCTC AGCAGCCTCA
21901 GGGCAATGCG CACTCACGGT GGTTTCTTCC TCCAGTGATG GAGCTGTAGG GGATGGTCCC
21961 CTTAGCCTTG GGGATTCGGA GGGTCCAATT TTCCAGGCTA GCTTACCCAT GGTGTGTGGC
22021 AGGCCCCACC CACTCTTCCA TGACACCCAC AACGACCCAG TCCCAGGGAA GATCCCTCTC
22081 GGCTCCTTTG GTGATGACAG ACACTAGCAA CTAGTGCTGT TCAGGGTCAA TGTTTACAAA
22141 GGTCCATTTC TGTATCATTA AAGTGTCTGC AAGAATGCAG GTGGCCCAGG TTTCTGTTCC
22201 TGTTTCCAGT CCTGGAAAGT TCACCTGCCC ACCCTCCTCT ATTCCTGCAC GGGAGACCTG
22261 ACCCGTGGAG AATGCTGCCT GGCCAGGAGG CAGGTACTTG ATAACCGGGG CTGGTTTACC
22321 ATGTGTGACC CTGTAGAGGA AGGAGCTCCA GGGGACGCAG CACTGCCCAC TCTGCTTGTT
22381 CTCAAGGCAG AGTCTGGTGA TTCTACTCTG AAAGGGCAAA GCTGTCATTA AATGTCTTCT
22441 CTTTAAAGAC TGTGACCACG TACAGAATGC GGTTTGATCT CAGTAGTTTA TTCCGGAGAC
22501 AAAGCAGTGC AAGAGGCGGC CACGCTTCCC ACCCCCTGTT TTCAGGGTCA GCCTGGGAAT
22561 TTTACTTACA GACATCAATC AGTTCAACAC TCAAGTCCTT TAAACTTGTT TAGTCCTGGG
22621 GCTCTGCTTC CGAGGAAGCC AATCGGCTTC TCCTTGCCCA AAGAATGCCC CTGCATGGCC
22681 CTTATTGGCC TTCCCAGGGC TGCTCCTTCT AGTGTACTCT ATCTTATTAG ACCTAAGGAC
22741 AACCCAGAGC CCTCATTCTC TTTACATAGT AAGCTGATGT CTTAAGACAC AAGATCACCC
22801 TTGGGGGTCC TGGCCTGCTG ACCTGGGGCT CCACAGGGGT CCTTTCCTGT GGGCTTTGGG
22861 GCCAGTGGGT GGTCTGCCAG CTGCTGTGGG GTCAGAGTGA GTCTGAAATG GGGCACAACT
22921 CCCTTGGGGG CTTGCCTGGT GCTCAGTGCA AGGGACTCGC TATGAACACA GCAGGAAGGA
22981 GCTGGGCCTT GCTGTTCCAG GAGGGCTGTC GCGTGTGAAC AGGAGCAGCC CTGTGGGTGG
23041 TCTCCTCTCA CCTGAAAAGA GGGACCGGAT GATGACATGA GGAAGCAGAG ATCCTCGTGG
23101 GCTAGACCAC TAGACCAGGC CTTCAGTTCT CTGTTTCAGG AGCCACCCTG AAGAGATGAG
23161 GTGGTGGAGG CCTGAATACT CACGGAATGA TCTGTTGGAA GTCCCAGGGT AAATCCTGGG
23221 TCCTGGCAGA CTCCTGGACT CCTAGTCAGA TGGCTCCTTC CTGGATCTCA GTGAGCAATA
23281 GGGTCAGGGT AGGGTGGATA GTGGCCTGCT GTATTCCACG AGCCAGTGAA GACTAGCAAT
23341 AGCCAGTGGG CTTCGGCTAC ACCCACAGGA AGCTCCTTTC CCTTCCTACA ACCCAGATCC
23401 GAATTC
```