MODIFIERS OF TRANSGENE METHYLATION AND EXPRESSION, AND GENOMIC IMPRINTING IN THE ZEBRAFISH, *Danio rerio*

BY

C. CRISTOFRE MARTIN

27

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Zoology
University of Manitoba
Winnipeg, Manitoba (c) September, 1994

ISBN 0-612-13343-5

Name Cristofre Martin

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Molecular Biology
SUBJECT TERM

0307
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology ... 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology .. 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics ... 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

MODIFIERS OF TRANSGENE METHYLATION AND EXPRESSION, AND GENOMIC IMPRINTING IN THE ZEBRAFISH, *Danio rerio*

BY

C. CRISTOFRE MARTIN

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Table of Contents

Abstract

Previous reports involving mammalian systems, particularly mice, have demonstrated the existence of *cis*-and *trans*-acting modifiers of transgene methylation and expression. These modifiers are thought to be important in dominance modification, genome imprinting and cellular expression mosaicism. The role that they may play in the penetrance and severity of many complex human diseases could be of even greater significance. In the present investigation I demonstrate that modifiers exist in a non-mammalian vertebrate, the zebrafish *Danio rerio* that act in a similar fashion to those identified in mice. I also provide evidence that the transgene methylation pattern may be influenced by genetic background, the sex of the individual and environmental modulators such as temperature and sodium butyrate. These data support the theory that this type of dominance modification is mechanistically analogous to *Drosophila* position effect variegation. I provide evidence that genomic imprinting is occurring in the zebrafish, and that it can be monitored at the level of transgene methylation as observed in imprinted murine transgenes and endogenous genes. Furthermore, these data suggest evolutionary conservation of the genomic imprinting phenomenon and the modifiers involved, at least within vertebrates, and implies that they and their actions are important in normal vertebrate development. The presence of genomic imprinting in the fish forces us to re-evaluate the proposed theories for the evolution of imprinting which are based primarily on mammalian data .

Acknowledgements

There are so many people that aided me in getting to this point in my career.

I would like to thank Dr. Richard Gordon for being the first to have faith in me as a scientist, for teaching me the wonders of developmental biology, and for being a close friend. I thank the members of the Department of Botany: Margerat Smith, Dr. John Stewart, Dr. Larry Van Caessalle, Dr. Sam Badour, and Dr. Isobel Waters for providing me with employment during my time as a poor starving student.

I thank my co-workers Rene Harrison, Katrin Stedronsky, Karen Yeow, Anindo Choudhury, Joe Carnie, Ming Chung Lu, and Tom Pratt for their support, friendship and good times.

I would like to thank Dr. Terry Dick for providing the indispensable role of the "petty tyrannt" in my quest to be a Man of Knowledge.

I thank Dr. Erwin Huebner for being on my graduate committee, providing technical expertise, being a perpetual teacher and a friend. I thank Dr. Murphy for being on my graduate committee and for the use of equipment.

I must extend my greatest gratitude to Dr. Ross McGowan for being a wonderful advisor, and providing a friendly and productive work environment. Most importantly I thank Ross for being a treasured friend.

Chapter 1: Genotype-specific Modifiers of Transgene Methylation and Expression in the Zebrafish, *Danio rerio*.

INTRODUCTION

Mammalian DNA modification involves among other things, the addition of a methyl group to position 5 of cytosine and in vertebrates occurs most frequently at CpG dinucleotides. This epigenetic modification of DNA by methylation is associated with transcriptional silencing of genetic loci with the result that inactive DNA sequences tend to be more methylated than active sequences. This relationship has led numerous investigators to propose that DNA methylation may play a role in developmental regulation (Monk *et al.*, 1987), genomic imprinting (Monk, 1990) and even, possibly, the speciation process (Varmuza, 1993).

The advent of transgenic technology has made it possible to analyse the inheritance of methylation within an individual allele. This is because, by selective breeding, it is possible to maintain a locus in a hemizygous state where only one allele is present. Using this method to analyse a transgenic line of mice McGowan *et al.* (1989) observed variable levels of methylation at the transgene locus in different individuals. This transgenic line also displayed mosaic expression of the *lacZ* reporter transgene between cells sharing the same developmental and phenotypic lineage, *i.e.*, not all apparently identical cells within a tissue expressed the transgene. The degree of mosaicism in transgene expression was inversely correlated with the level of the transgene's methylation. Mice with hypermethylated transgenes had only a few expressing cells whereas many expressing cells were evident in mice with a hypomethylated transgene locus. The alterations in the methylation (and expression) of the transgene locus was

shown to be affected by strain-specific modifiers of methylation consistent with what has been found for other transgene loci (Sapienza *et al.*, 1989; Allen *et al.*, 1990; Forejt and Gregorova, 1992). Transcriptional silencing, mosaic expression, strain-specific modifiers and positional specificity associated with transgene loci is very reminiscent of position effect variegation as observed in *Drosophila* in which there is mosaic expression of alleles that can be enhanced or suppressed by unlinked dosage sensitive modifiers. This similarity lead McGowan *et al.* (1989) and Sapienza (1990a) to propose that we are looking at a dominance modification phenomenon analogous to position effect variegation. Genomic imprinting would result when one or more of these modifiers is sex-linked and would therefore produce a dosage difference. Cytosine methylation itself is rare in *Drosophila*. However, the inactivation of genetic loci by the spreading of heterochromatin into adjacent euchromatic regions shares many characteristics with the expression of transgenes previously described in mice. Two-classes of dosage dependent modifiers of position effect variegation are thought to exist in *Drosophila* (Tartoff and Bremer, 1990). Class I modifiers are loci that enhance variegation when duplicated and suppress when mutated or deleted. Class II modifiers enhance variegation when mutated or deleted and suppress when duplicated. Class I modifiers appear to be the proteins which bind DNA and cause heterochromatin condensation. Locke et al. (1988) suggests that Class II modifier genes code for proteins that: 1) interact with, or inhibit indirectly, the assembly of Class I products into heterochromatin; 2) may bind to hypothetical termination sites that define euchromatic-heterochromatic boundaries, limiting the spread of heterochromatin; or 3) promote the formation of euchromatin. If any of these modifiers genes is located on the X

chromosome then males which are the heterogametic sex in mice, would have half the modifier gene dose as compared to a female during early development. The effect of the differential dose of modifier genes in the sexes would manifest itself as genomic imprinting.

Position effect variegation of the *white* locus as well as many other loci in *Drosophila* has been shown to be affected by a variety of factors including environmental temperature, the presence of sodium butyrate, the sex of the individual displaying the variegated phenotype, and the sex of the parent contributing the variegating rearrangement (reviewed in Spofford, 1976). Lower temperatures generally enhance variegation while higher temperatures suppress the effect. The presence of sodium butyrate in the culture media causes an increase in mutant cells presumably due to its effects on histone acetylation (Mottus *et al.*, 1980). In *Drosophila* the mutant cells are frequently more extensive in normal females as compared to normal males. Probably the most interesting modulator of PEV, in relation to vertebrates, is the effect of the parental source of the variegating rearrangement where variegation is enhanced when passage is through the egg and suppressed after passage through the sperm. This gamete-of-origin phenomenon is analogous to the genomic imprinting phenomenon observed in mice (and zebrafish, see chapter 2). The similarities between these processes suggest that the modifier genes *(Su(var)* genes) involved in position effect variegation (Henikoff, 1990) may be functionally analogous to the modifiers identified in association with the imprinting of transgenes in mice. These types of modifier genes in *Drosophila* have been shown to exert their effect in *cis-* or *trans-*, and behave in a dosage dependent manner. Many of these genes are known to encode proteins such as zinc fingers, heterochromatin associated proteins,

and histone proteins that interact directly in DNA binding and heterochromatin formation (reviewed in Reuter and Spierer, 1992).

Some recent observations on the expression of a transgene in fish suggested the possibility of similar modifiers being present in a non-mammalian vertebrate. Stuart *et al.* (1990) produced several stable transgenic lines of zebrafish containing the plasmid pUSVCAT (Appendix 1,2,3), the reporter gene chloramphenicol acetyltransferase (CAT) positioned downstream of the Rous sarcoma virus LTR. With the use of immunohistochemical staining, they observed a mosaic pattern of CAT expression in the skin epithelium of the caudal fin. The similarity between the mosaic transgene expression in zebrafish and the previously described mosaic transgene expression in mice (McGowan *et al.*, 1989) suggested the possibility of a similar modification system operating in fish.

In the present report I will demonstrate the presence of strain-specific modifiers of methylation in a line of transgenic zebrafish, *Danio rerio*, containing the plasmid pUSVCAT. The identification of genetic modifiers in zebrafish that are similar to those identified in mice, suggests evolutionary conservation of this modification system among vertebrates. I also present evidence in support of the PEV-like models of McGowan *et al.* (1989) and Sapienza (1990a,b) for the mosaic expression of mouse and zebrafish genomically imprinted transgenes. I further suggest that, based on the results of my work, the zebrafish provides a model system for analyzing the modification phenomenon that may be superior to that provided by the mouse, particularly for studies of early developmental events.

Materials and Methods

Zebrafish

Transgenic zebrafish, *Danio rerio,* containing the plasmid pUSVCAT and derived from the AB wild-type line of the University of Oregon (CATfish IV) were generously supplied by Monte Westerfield (Stuart *et al.*, 1990). A lab stock of wildtype fish (small inbreeding population) were also generously supplied by Hans Laale (Dept. Zoology, University of Manitoba). Wildtype, longtail zebrafish, and leopard danios (*Brachydanio rerio frankei*) were obtained from a local pet supply. Zebrafish were maintained in 10-20 gallon aquaria at 28°C on a 14 h light/10 h dark cycle. Prior to breeding, individuals were placed in 1 gallon aquaria for 1 week. On the morning of egg collection a single male and female were placed in a spawning tank containing a floor of glass bars which served to protect the eggs from being eaten by the adults. Eggs were collected 2-3 hours after the initiation of spawning, which occurs spontaneously at the onset of the light cycle. Embryos and larvae were maintained in plexiglass holding tubes with fine mesh bottoms. Tubes were suspended in a 20 gallon aquarium. Larval fish were fed live *Paramecia*, finely ground flake food, and baby brine shrimp. General zebrafish care was according to Westerfield (1989).

DNA Probe Purification

E. coli (strain JM109) were made competent and transformed with plasmids containing inserts used as probes for Southern and northern hybridization using the $CaCl_2$ method (Sambrook *et al.*, 1989). Transformed bacteria were plated on LB media agar plates containing 50 ug/ml. ampicillin as a selective agent. A single bacterial colony from these plates was used to inoculate a 500-1000 ml. culture of L-broth liquid media

containing 50 ug/ml. ampicillin. This culture was grown overnight at 37°C with rapid agitation in a shaker waterbath. The plasmid DNA was purified from the confluent bacterial culture using a Magic Megaprep DNA Purification System (Promega). The purified plasmid was digested with the appropriate restriction endonuclease to release the DNA fragment to be used as a probe. The resulting DNA fragments were separated by electrophoresis in a 1.0% agarose gel. The separated fragments were visualized using a UV transilluminator following ethidium bromide staining. A scalpel was used to cut a line in the gel just ahead and immediately behind the DNA fragment of interest. Pre-wetted NA45 DEAE ion exchange membrane (Schleicher & Schuell) was placed into the gel incisions. The gel was replaced into the electrophoresis box and the voltage re-applied for 10-30 minutes. The fragment migration was monitored periodically using a hand-held UV lamp. When the fragment of interest had completely migrated into the collection membrane, electrophoresis was discontinued. The collection membrane was removed and rinsed in Tris-acetate running buffer, and placed in a high salt elution buffer (1.0M NaCl, 50 mM arginine) and incubated at 68°C for 1-2 hours. The fluid from the tube containing the DEAE membrane was extracted with PCI (25 phenol: 24 chloroform: 1 isoamyl-alcohol) and CIA (24 chloroform: 1 isoamyl-alcohol), and precipitated with 0.3 M sodium acetate and isopropanol. The resulting DNA pellet was washed in ethanol, dried, and re dissolved in TE^{-4} (100 mM Tris pH 8.0, 1 mM EDTA) and stored at -20°C until later use.

Southern Hybridization

DNA samples were prepared as previously described by Sapienza *et*

al.(1987) from whole fish and pooled samples of small fish from a single breeding. DNA was extracted by placing minced zebrafish tissue or whole fry in a 1.5 ml. eppendorf tube containing 500 ul. of tail buffer (1% SDS, 100 mM NaCl, 100mM EDTA, 50 mM Tris pH 8.0) and 0.5 ug/ml proteinase K (Sigma). The samples were incubated overnight at 56°C in a water bath. The DNA was extracted from the digested tissues with PCI and CIA. The extracted DNA was precipitated with sodium acetate and isopropanol. The resulting DNA pellet was washed in 70% ethanol, 100% ethanol, and dried in a 65°C oven. The DNA pellet was redissolved in TE-4. The DNA concentration was determined spectrophotometrically. DNA was cleaved with the restriction endonuclease Hpa II, or Hpa II/BamH I as suggested by the manufacturer (BRL). Complete digestion was tested by the addition of control plasmid DNA (typically pGEM) to a sample of the experimental digest. Restriction digestion was considered complete when the internal control digest gave a digest pattern identical to the pattern obtained with test plasmid and the same enzyme alone. DNA fragments were separated by electrophoresis in a 0.7% agarose gel. Hind III cut lambda DNA was added to a single well to act as a DNA fragment size marker. Following electrophoresis, the DNA within the gel was visualized by ethidium bromide staining and a photograph was taken using a transilluminator and Polaroid camera. Prior to Southern transfer of DNA to nylon membrane (Hybond-N+, Amersham), the DNA fragments were denatured in 0.6 M NaCl and 0.2 M NaOH. The DNA was immobilized on the nylon membrane by baking for 2 hours at 80°C. The membrane was washed in 0.1X SCC (20X stock=3.0 M NaCl, 0.3 M citric acid) and 0.1% SDS at 65°C, followed by overnight prehybridization at 43°C in a solution containing 6X SSC, 10X Denhardt's solution (100X stock=2% ficoll, 2%

BSA, 2% polyvinylpyrrolidone), 500 ug/ml. denatured salmon sperm DNA and 0.5% SDS. ^{32}P-labeled probes were prepared by random primer labelling (Sambrook *et al.*, 1989). Blots were hydridized overnight at 43°C in a solution containing 50% deionized formamide, 4X SET (1X=0.15 M NaCl, 0.03 M Tris, 2.0 mM EDTA, pH 8.0), 1X Denhardts solution, 100 ug/ml. denatured salmon sperm DNA, 0.5% SDS and 4 X 10^6 cpm/ml. of ^{32}P-labeled probe. Blots were then washed at 65°C in several changes of 0.1X SCC and 0.1% SDS. Blots were patted dry with paper towels, wrapped in cellophane, and exposed to X-ray film for 1-10 days with an intensifying screen.

Blots were probed with a 1.4 Kbp and 1.6 Kbp Hind III fragment of pUSVCAT containing the Rous Sarcoma virus LTR and the gene encoding the enzyme chloramphenicol acetyltransferase, respectively.

Immunohistochemistry and Immunofluorescence

The chloramphenicol acetyltransferase antigen was visualized in the epithelial cells of individual scales, whole mount caudal fins, and whole 10-12 day fry and transgenic fish using a rabbit anti-CAT primary antibody (5'-3' Inc.) with a goat-anti-rabbit fluorescein conjugated secondary antibody (5'-3' Inc.) or a VECTASTAIN ABC- alkaline phosphatase kit (Vector Laboratories Inc.). Tissues for wildtype, non-transgenic zebrafish were also stained at the same time to act as a negative control. Fish were anaesthetized with 2-phenoxyethanol. Individual scales dorsal to the lateral line and immediately posterior to the dorsal fin were removed with fine forceps. Scales, fins and fry were rinsed in phosphate buffered saline pH 7.3 (PBS) (0.07 M Na_2HPO_4, 0.03 M NaH_2PO_4, 0.5 M NaCl) for 15 minutes and fixed for 30-45 minutes in 4% paraformaldehyde in PBS pH

7.3 at room temperature. Following fixation, samples were washed in three changes of PBS for 30 minutes and incubated overnight in blocking buffer containing 10% normal goat serum (v/v), 0.1% crystalline BSA (w/v), and 0.01% Triton X-100 (v/v) in PBS pH 7.3. Overnight steps were performed at 4°C. Endogenous biotin was blocked by applying avidin (3 hours) and d-biotin (3 hours) (Zymed Laboratories). Washes in buffer containing 0.1% BSA and 0.05% Triton X-100 in PBS pH 7.3 were performed for 30 minutes between all steps. The primary antibody was diluted 1/200 in blocking buffer and applied overnight. A secondary biotinylated goat anti-rabbit antibody was diluted 1/200 in washing buffer and applied overnight. The avidin-alkaline phosphatase complex was produced according to the manufacturer and applied for 1.5 hours. The substrates and a buffer containing 1mM levamisol were prepared and applied for 15-30 minutes. The staining reaction was stopped by washing specimens in distilled water for 30 minutes. Samples were temporarily mounted in glycerol and viewed using a compound microscope.

Similarly, the CAT antigen was visualized using immunofluorescence. Tissues were fixed, blocked, and treated with the primary rabbit-anti-CAT as previously described. After washings, a secondary goat-anti-rabbit fluorescein conjugated secondary antibody was applied to the samples. Samples were washed, mounted in 90% glycerol in PBS pH 9.5 and viewed using a Zeiss Photo-2 epifluorescence microscope using the appropriate filter set for fluorescein. Images were digitized and enhanced using a Hamamatsu CCD camera, PC computer, and Image 1 frame grabber and software (Universal Imaging Corp.). 3-D intensity profiles were produced from these images to aid in visualizing the mosaic expression of the transgene and relative levels of the staining signal.

RESULTS

To identify *trans*-acting modifiers similar to those found by McGowan *et al.* (1989) for the Tg4 transgenic mouse line and by Sapienza *et al.* (1987) for 379 tropinin I transgenic mice, I bred individual homozygous transgenic zebrafish to a variety of non-transgenic mates including two different wildtype populations, a single population displaying a longtail phenotype, and the leopard danio *Brachydanio rerio frankei*. These particular breedings were used on the assumption that different strains represent different genotypes. The homozygous, transgenic parent shown in Figure 1.1 displayed a relatively high degree of transgene methylation, *i.e.*, was hypermethylated. However, this fish carried two alleles at this locus and we do not know if they were methylated differently. Methylation sensitive, Hpa II restriction digests of F1 progeny DNA from these crosses were hydridized with CAT-specific sequences and revealed that methylation differences do exist that are, presumably, due to the genotype of the non-transgenic parent. Breeding the same transgenic fish to four different non-transgenic mates produced different degrees of transgene methylation in individual offspring (Figure 1.1) and samples of pooled offspring (not shown). These results strongly suggest the presence of *trans*-acting modifiers in the zebrafish that behave similarly to those previously identified in the mouse, *i.e.*, that produce changes in the transgene methylation pattern.

In some transgenic mice previously studied (Swain *et al.*, 1987; McGowan *et al.*, 1989) changes in transgene methylation were associated with changes in the numbers of transgene expressing cells producing

different degrees of mosaicism. Immunohistochemical and immunofluorescent analysis of CAT transgene expression in epithelial cells of our transgenic zebrafish clearly demonstrates a mosaic expression phenotype (Figure 1.2) as was reported by Stuart *et al.* (1990). Cellular expression of the transgene was detected as early as the time of hatching and in young fry in the primitive fin folds (Figure 1.3). It is also apparent from these stained scale epithelia that expression of the transgene is cell autonomous. Clusters and individual expressing cells are often entirely surrounded by non-expressing cells.

A comparison of the methylation of the transgene in whole fish DNA and transgene expression in the scale epithelia indicates that the mosaicism can be inversely correlated with the degree of transgene methylation (Figure 1.4). Epithelial cells from female fish consistently produced more non-specific staining (which appeared to be morphologically distinct) than epithelial cells from males. The reason for the sex-specific difference in the background is not yet clear but because of it only males were used in the analyses. Precise quantification of the number of staining cells was difficult, but relative comparisons between scales could be made. Like the mice, fish with a relatively high level of transgene methylation in somatic DNA displayed a relatively low number of CAT expressing cells in their epithelial layer and the opposite was also true. Low methylation correlated with high numbers of expressing cells (Figure 1.4). Most importantly, the proportion of expressing cells is variable between individuals and is inversely related to the transgene methylation pattern. By extrapolation, the differences in methylation observed when the transgene is passed onto different genetic backgrounds or through different parents is the result of changes in the relative proportion of hypo- and hyper-methylated cells in

the population.

The mosaic expression and dominance modification observed in zebrafish is virtually identical to that seen in mice. Because of mammalian internal development it has been difficult to modify the developmental environment of individuals during the time that this epigenetic modification is occurring. Zebrafish embryos, because of external fertilization, can be directly manipulated prior to first cleavage and throughout early development. The zebrafish provides a marvelous opportunity to test predictions of the position effect variegation model. Figure 1.5 shows the methylation pattern of two different spawns, of which half of each spawn were raised at 20°C and the other half raised at 30°C. The cold temperature fish display a higher transgene methylation pattern than the higher temperature fish. Similarly, fish raised in 6mM sodium butyrate displayed a lower methylation pattern than their siblings raised in normal aquarium water (Figure 1.6). Because fry were analyzed shortly after hatching, the effects of these environmental factors on the expression of the transgene could not be determined. However, the link between transgene methylation and CAT expression suggests that such a relationship may well exist (Figure 1.4). The similar effects of environmental modulators on transgene methylation in the zebrafish as compared to PEV in *Drosophila* supports the idea that the genetic modifiers involved in both processes are acting through a similar mechanism.

Most crosses between transgenic and non-transgenic parents resulted in all offspring within a spawn having virtually identical transgene methylation phenotypes (tested on individual and pooled samples). Certain genotypic combinations, however, produced a distinct difference in the methylation pattern between male and female siblings (Figure 1.7). In this

cross, the males showed a lower degree of methylation than the females. There is evidence, therefore, of sex differences in the degree of mosaicism.

Figure 1.1: *(Top)* A pedigree is presented and shows the mating of a single homozygous pUSVCAT transgenic male (circle in square) to different wildtype strain females (circles): (A) commercial wildtype, (B) longtail phenotype, (C) lab stock wildtype, (D) *Brachydanio rerio frankei.* HpaII-cleaved DNAs from individual offspring hybridized with a 1.6 Kb transgene-specific sequence are shown beneath the pedigree and demonstrate that the same male can give rise to progeny with different transgene methylation phenotypes, depending on the genotype of the non-transgenic female. A HpaII digest of the homozygous, transgenic male used in the above breedings is shown under the symbol representing that individual.

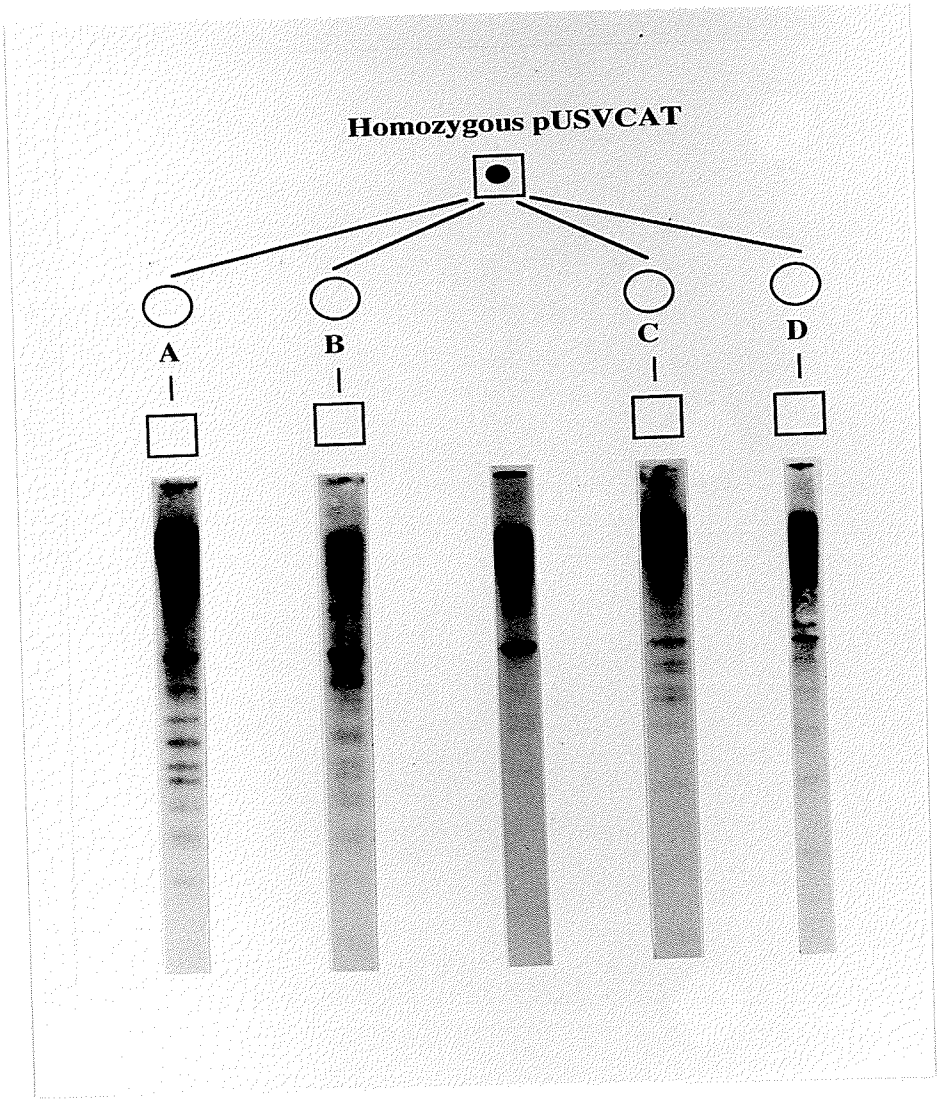
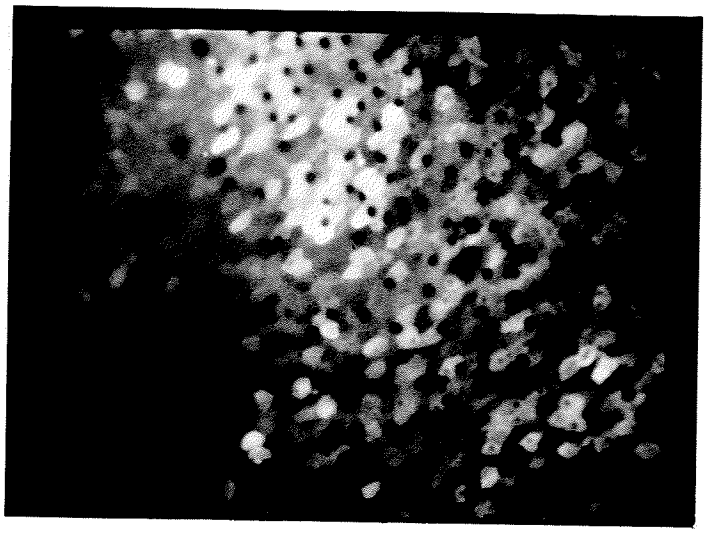
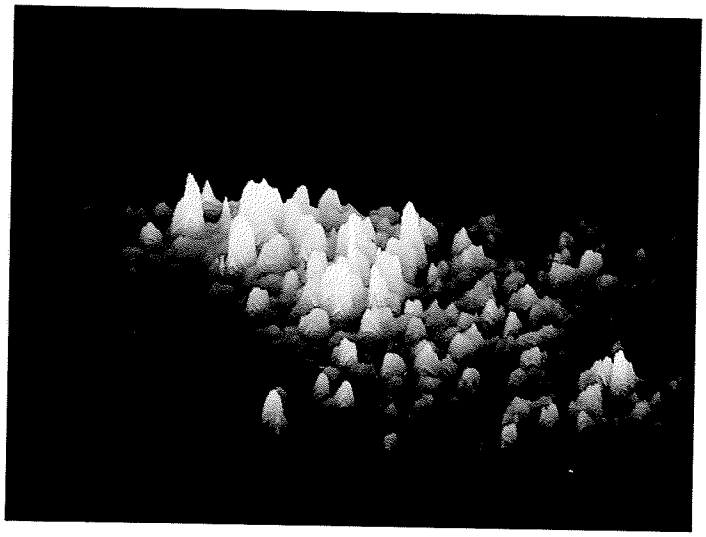

Homozygous pUSVCAT
A
B
C
D

Figure 1.2: Immunofluorescent staining for the mosaic transgene expression, using anti-CAT antibody, in the epithelial layer of the caudal fin of a homozygous transgenic zebrafish (A,C). The digitized images were recorded on black & white film using a Polaroid Freeze-Frame recorder, and converted to a 3-D intensity profile using Image-1 software (B,D).

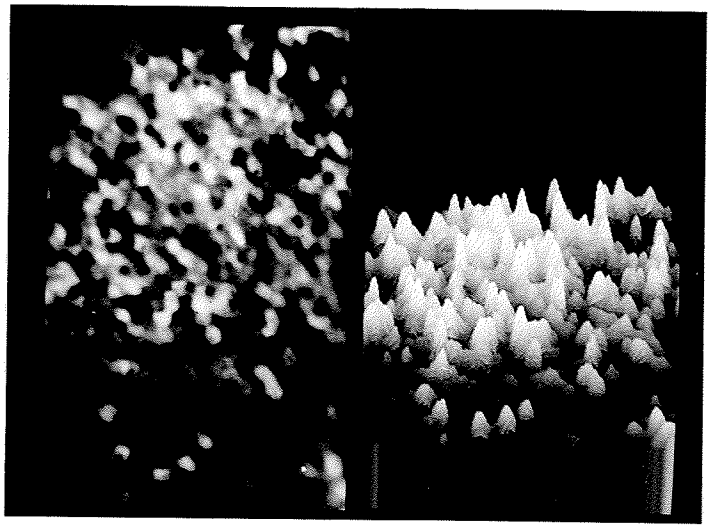
A



B



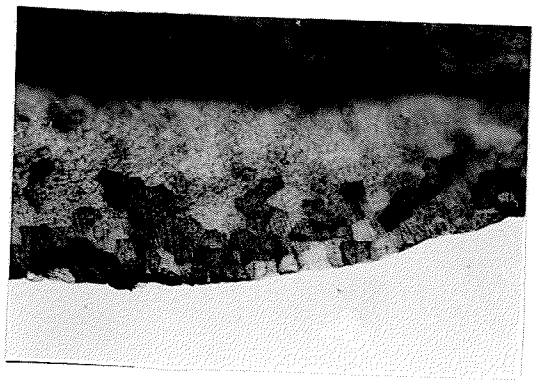
C



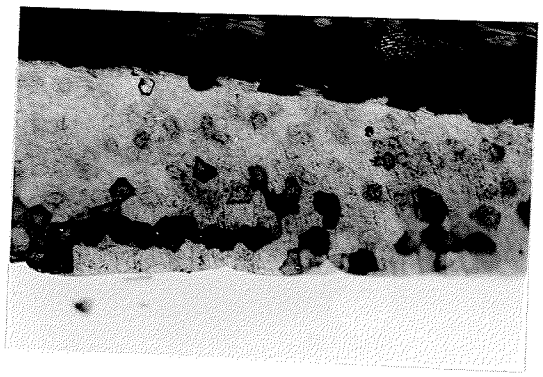
D

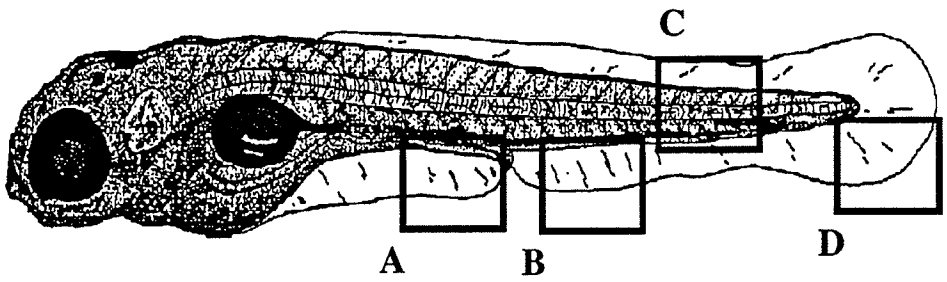
Figure 1.3: Immunohistochemical staining for transgene expression in a 7-10 day old transgenic zebrafish fry along the anal (A), ventral (B), and dorsal (C) finfolds and caudal fin (D). Black or dark cells are alkaline phosphatase positive.

A



B





C

D

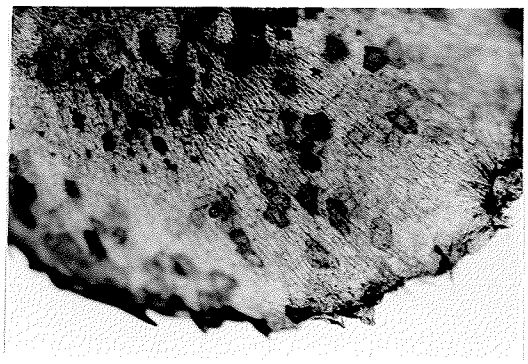
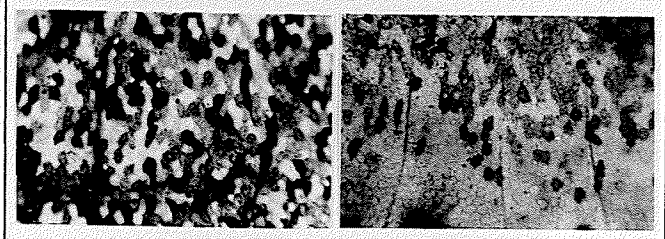


Figure 1.4: Chloroamphenical acetyltransferase expression in scale epithelial tissues and restriction endonuclease digestions of DNAs from male F1 hemizygous pUSVCAT transgenic zebrafish from different breedings. At *top* is shown the variation in the number of CAT expressing cells visualized by immunohistochemical staining. Beneath each stained tissue is HpaII-cleaved DNAs from the same individual. Restriction digests were probed with a 1.6 Kb sequence representing the CAT gene. Individuals having fewer expressing cells carry a hypermethylated transgene, whereas individuals with more expressing cells carry a hypomethylated transgene.

CAT Expression in Scale Epithelial Layer



Hpa II Digestion of Whole Fish DNA

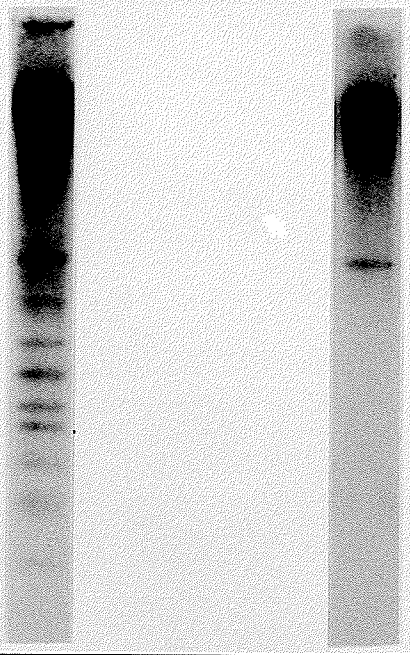


Figure 1.5: The effect of developmental temperature on zebrafish transgene methylation phenotype. Hpa II cleaved DNAs from transgenic sibling offspring raised at 20°C and 30°C till hatching. The level of methylation is increased in the 20°C fish as compared to their siblings raised at a higher temperature.

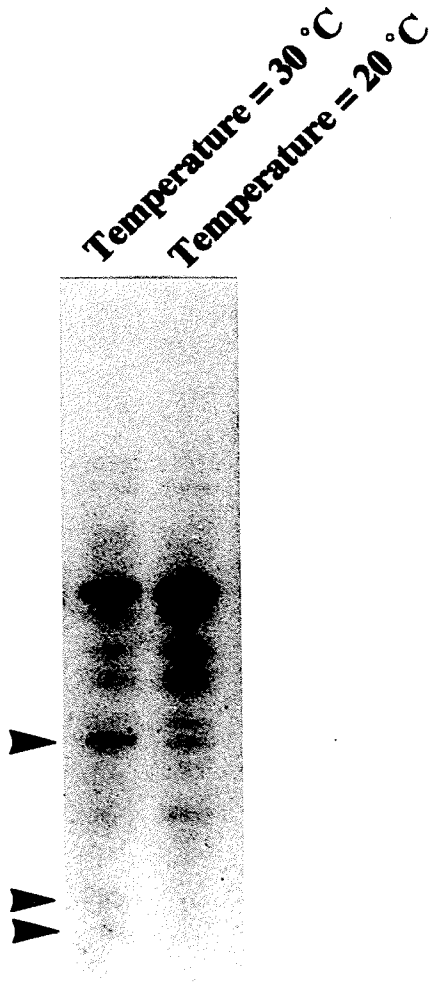

Temperature = 30 °C
Temperature = 20 °C

Figure 1.6: The effect of sodium butyrate on the zebrafish transgene methylation. Hpa II (A) and Hpa II/BamH I (B) cleaved DNAs from transgenic sibling offspring raised in a 6 mM sodium butyrate solution and the control group raised in normal aquarium water. The treated group shows a lower transgene methylation pattern than the control group. The directionality of the transgene methylation differences between treatment and control groups is best illustrated by comparing the relative intensity of the indicated fragments (arrows).

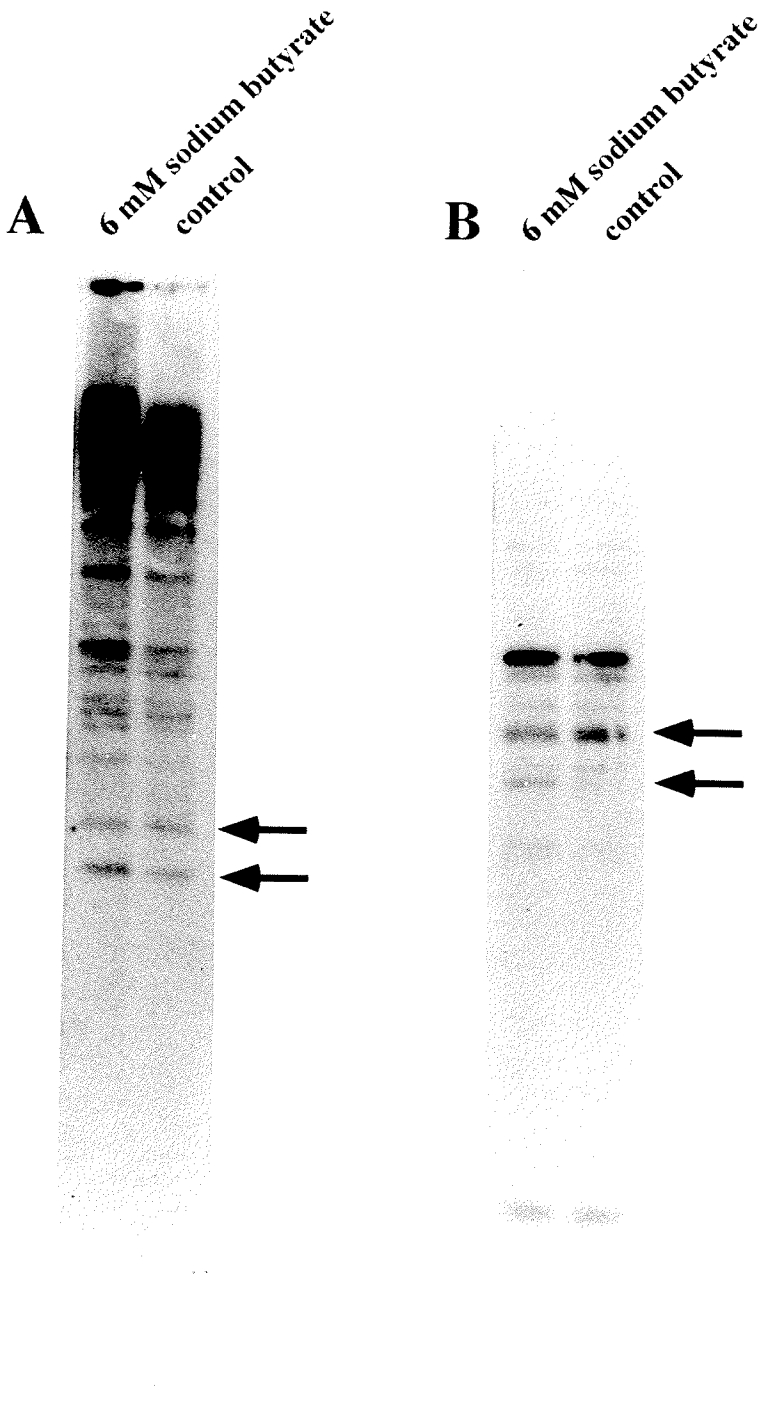
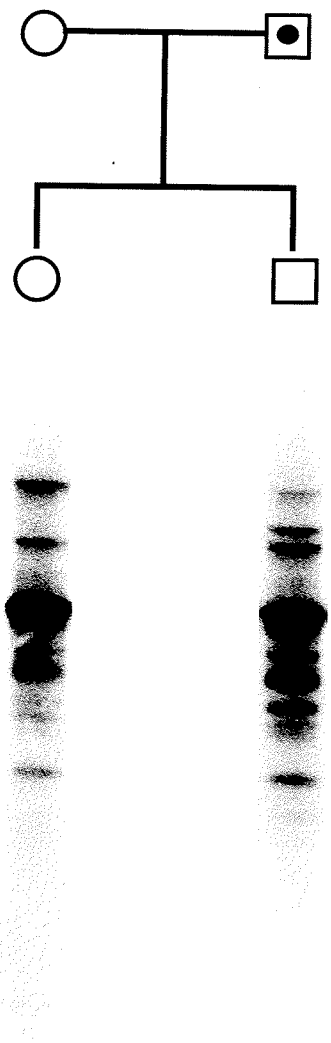

A
6 mM sodium butyrate
control
B
6 mM sodium butyrate
control

Figure 1.7: Hpa II/BamH I cleaved DNAs from female and male transgenic sibling offspring. In this particular cross, there is an apparent sex effect difference in the transgene methylation pattern where females have a higher degree of transgene methylation than sibling males.



DISCUSSION

The existence of genotype-specific loci that influence the expression of other loci was proposed by Fisher (1928) and his contemporaries in the early part of this century. This phenomenon was called dominance modification. The modification of dominance which produces incompletely penetrant genetic traits is now recognized as a common genetic phenomenon and results from allelic variation at loci that interact with other genetic loci. Dominance modifiers that are presumably of this type have been identified in mice at the level of the methylation and expression of transgene loci (McGowan *et al.*, 1989; Sapienza *et al.*, 1989; Allen *et al.*, 1990).

Genetic modifiers apparently analogous to those found in mice can be demonstrated in zebrafish by passage of a transgene onto different genetic backgrounds. Breeding the same transgenic fish to different non-transgenic mates resulted in different transgene methylation phenotypes. Clearly, loci must exist in the non-transgenic mates that affect the methylation of the transgene locus and that can: a) act in *trans* and; b) show genetic variability. The presence of modifier loci, similar to those identified in mice, in a non-mammalian vertebrate suggests evolutionary conservation of these factors. It stands to reason, therefore, that the action of these modifiers may be important in the biology and/or development of these animals and, potentially, other or possibly all sexually reproducing organisms. Whether these actions simply maintain the sexual process, ensure production of gametes or are more complex and fundamental to animal development remains to be determined.

In several of the murine transgenic lines examined, the expression of the transgene loci was observed to be inversely correlated with the extent of their methylation (Swain *et al.*, 1987; McGowan *et al.*, 1989; Allen *et al.*, 1990). Highly methylated transgene loci exhibited low to non-existent expression, whereas less methylated loci exhibited high levels of expression. This is consistent with previous findings on the relationship between methylation and expression (Cedar, 1988). This correlation between methylation and expression was not absolutely necessary to prove tthe existence of this modification system, because other factors besides methylation can affect expression. However, its presence allowed for the demonstration of a correlation with mosaic expression patterns (*ie.* relative proportions of expressing and non-expressing cells) within individual tissues in the transgenic mice (McGowan *et al.*, 1989) and the subsequent generation of a model to explain methylation modification and imprinting based on position-effect variegation in *Drosophila* (Sapienza, 1990a).

In zebrafish, the transgene expression was also inversely correlated with its level of methylation and was found in a similar mosaic pattern. The effect of changing the genotypic background on both the methylation and, by extrapolation, the expression of the transgene suggests that the modifiers that we have identified in these fish are acting (at least partially) by changing the number of affected cells present in a similar manner to that seen previously in transgenic mice. Therefore, at least some of the genotype-specific modifiers appear to affect the transgene indirectly by altering the proportion of differently methylated cell types present rather than altering the extent of transgene methylation itself.

The similarities between variegated expression observed with transgenes in mice (McGowan *et al*, 1989) and position effect variegation

in *Drosophila* have led to the creation of models to explain transgene modification based on the action of the *Drosophila Su(Var)* genes on heterochromatin formation (Sapienza, 1990a). *Drosophila* actually have little or no DNA methylation. However, this does not mean the *Drosophila* position effect variegation cannot be used as a model. In mammals the centromeric heterochromatin is found to be enriched in 5-methyl-cytosine (Miller *et al.*, 1974), and its role in the production of heterochromatin can be demonstrated by the use of methylation inhibitors which leads to the undercondensation of the same centromeric heterochromatin (Schmid *et al.*, 1984). The methylation being monitored in mice therefore is probably just reflecting the formation of heterochromatin rather than being the inactivator itself.

These models have been applied to the study of several diseases including Huntington's disease (Sapienza, 1990a; Laird, 1990) and heritable cancers (Scable *et al.*, 1989; Sapienza, 1990b) and are capable of successfully explaining all of the disparate information available about these diseases. They have also allowed for the formation of testable predictions which have lent further support to the model (Scrable *et al.*, 1989; Sapienza, 1990b).

Using the mouse, it has not been possible to test the PEV model for dominance modification by direct manipulation of the developing organism as is possible with *Drosophila*. In the zebrafish, however, external fertilization has allowed us to test predictions based on PEV in *Drosophila*. The increase in transgene methylation as a result of lower developmental temperature is analogous to the effect observed in *Drosophila* PEV where lower temperatures enhance variegation (increase in mutant cells) (Gowen and Gay, 1934). Very little is known about the mechanism for this

temperature effect but it has been suggested that it affects protein-protein interaction and the formation of protein multimeric complexes and protein-DNA interactions (Spofford, 1976).

Similarly, the effect of sodium butyrate on transgene methylation in the zebrafish was also shown to fulfill predictions based on its effects on *Drosophila* PEV. In variegating mutants of *Drosophila*, sodium butyrate suppresses the variegation (decreasing the number of mutant cells). The spreading of heterochromatin, by histone-mediated chromatin condensation, into euchromatic regions of the chromosome is thought to be responsible for the PEV phenomena. Transcriptionally active chromatin is associated with hyperacetylated histones, particularly H4 (Levy-Wilson *et .al.*, 1979). Sodium butyrate has been demonstrated to prevent histone-deacetylation in both *Drosophila* and mammalian cells (Mottus *et al.*, 1980; Peter *et al.*, 1978). It, therefore, suppressed variegation by blocking the formation of heterochromatin. In our transgenic zebrafish, sodium butyrate also caused a decrease in the transgene's methylation level suggesting that here also, sodium butyrate may be blocking heterochromatin formation.

Sex-specific suppression of PEV is thought to be due to the presence of the highly heterochromatic Y chromosome. The additional heterochromatic regions on that chromosome are thought to sequester heterochromatin forming proteins and therefore act to suppress PEV. This effect of the Y-chromosome has been well documented in *Drosophila*. The basal heterochromatin of the X-chromosome centromeric regions have also been shown to suppress PEV to a nearly equal extent as the Y-chromosome (reviewed in Spofford, 1976). Together the sequestering of DNA binding proteins by the sex chromosomes produces a distinct sex effect in the extent

of allelic variegation. Typically, PEV suppression is greatest in males. An identical effect which is probably occurring via the same mechanism is observed in the zebrafish where in some crosses the male offspring have a distinctly lower transgene methylation phenotype than females. This sex-effect is complicated, however, by the possible presence of dosage-sensitive Class I and/or Class II modifiers on the sex chromosomes. Further, there exists no information on the type of sex determination in zebrafish (*ie.* heterogametic sex in zebrafish) or the extent of heterochromatic regions in the zebrafish sex chromosomes. Potentially, the inactivation of a sex chromosome in either sex to achieve dosage compensation (which does not occur in *Drosophila*) may soak up considerable heterochromatin factors. The types and origins of the modifiers producing this effect in fish, therefore, are difficult to interpret.

The inactivation of loci by heterochromatin formation may not be just a genetic anomaly of *Drosophila*. The *Drosophila HP1 gene* (which is allelic to the suppresser of variegation *Suvar (2)5*) encodes a product shown to be a structural component of heterochromatin (James and Elgin, 1986). A region of the *HP1* gene termed the chromobox has been shown to be highly conserved across a wide range of phyla (Singh *et al.,* 1991) and has been used to isolated the murine *M31* and *M32* and the human *HSM1* counterparts which have now been shown to be involved in the packaging of mammalian chromosomal DNA into constitutive heterochromatin (Wreggett *et al.*, 1994). Further, the similar homology between *HP1* and the repressor of homeotic genes *polycomb* (Paro and Hogness, 1991) and the relationship between heterochromatin and yeast mating-type silencing (*SIR* genes) (Braunstein *et al.*, 1993; Lee and Gross, 1993) argues that this type of gene control may be widespread and developmentally very

significant.

An understanding of dominance modification in mammals is important, particularly in terms of human disease. However, the identification of a similar phenomenon operating in a non-mammalian vertebrate is equally significant. It suggests that the process is evolutionarily conserved and, therefore, is probably important in the biology of animals. Furthermore, there is considerably more potential using the zebrafish as a model system rather than the mouse, particularly in a developmental sense, because of the obvious experimental benefits of external fertilization, large reproductive potential and optically clear embryos. These fish also represent a considerably more natural population then inbred mice which have been selected for unusual traits (such as coat colour or cancer susceptibility, for example) and, therefore, should be more useful in trying to determine the biological relevance of this phenomenon.

Chapter 2 - Genomic Imprinting in the Zebrafish, *Danio rerio.*

Introduction

Dominance modification involves the production of phenotypic differences in the degree of penetrance of a particular locus. This epigenetic modification is produced by the action of other, distinct modifier loci, most commonly acting in *trans*. Genomic imprinting, thought to be a particular sub-type of dominance modification (Sapienza, 1989), involves the production of epigenetic and heritable differences through cell division between the paternal and maternal genomes. It is logical to assume that the epigenetic differences observed in the paternal and maternal genetic contributions would be established during the only time the two genomes are segregated, during gametogenesis.

One type of epigenetic modification is the addition of a methyl group to bases in the DNA. This type of DNA modification is significant because, in general, methylated DNA sequences are transcriptionally inactive. Furthermore, the methylation of DNA residues can be stably preserved through the replication process by the action of maintenance methylases which use the replicated hemimethylated DNA as a template (Razin and Riggs, 1980). It has, therefore, been suggested that DNA methylation may be the epigenetic marking responsible for the imprinting phenomenon, or at least reflect the imprinted state of a locus produced during gametogenesis (reviewed in Sasaki *et al.*, 1993).

In trying to support this idea many laboratories have studied the methylation status of transgenes and endogenous genes in the gametes and developing embryos. If methylase differences exist in the gametes then one would expect to find different DNA methylation patterns in sperm and

eggs. These differences should be perpetuated at least through early development where the final somatic tissue methylation pattern is established. Monk *et al.* (1987) and Sanford *et al.* (1987) demonstrated in mice that sperm DNA is more methylated than oocyte DNA. Following fertilization in the mouse until approximately the blastocyst stage a period of overall demethylation has been observed followed by extensive *de novo* methylation beginning at gastrulation ultimately producing levels higher than those observed in either of the gametes (reviewed in Monk, 1990). Later, similar methylation changes were observed during development when the methylation phenotype of a variety of transgenes in mice were examined (Chaillet *et al.*, 1991; Ueda *et al.*, 1992). Although there is a general pattern of methylation events occurring during development, Reik *et al.* (1987), Allen *et al.* (1990) and McGowan *et al.* (1990) showed variability in the methylation of a variety of murine transgenes in mature sperm. The transgene methylation pattern in sperm upon entering the oocyte is not equivalent in all cases.

It was shown in the previous chapter that a system of dominance modification exists in the zebrafish *Danio rerio* that can affect the methylation and expression of a transgene locus. The purpose of this investigation was to investigate the possibility that this process of dominance modification was associated with genome imprinting as it is in mammals.

Genomic imprinting can be loosely defined as the gamete-of-origin dependent modification of phenotype. That is, the phenotype elicited from a locus is differentially modified by the sex of the parent contributing that particular allele. This process results in a reversible gamete-of-origin specific marking of the genome, that ultimately produces a functional

difference between the genetic information contributed by each parent.

The significance and relevance of imprinting to mammalian gametogenesis and developmental regulation has been clearly demonstrated (McGrath and Solter, 1984; Surani *et al.*, 1984; 1986; Cattanach, 1986). Furthermore, considerable interest has been generated by the realization that the penetrance and severity of many complex human diseases can be affected by the sex of the contributing parent including Huntington's disease, cystic fibrosis, Prader-Willi and Angelman syndromes; cancers including Wilm's tumour, rhabdomyosarcomas, and osteosarcomas (reviewed in Hall, 1990); and birth defects such as spina bifida (Chatkupt *et al.*, 1992).

The pronuclear transplantation experiments of McGrath and Solter (1984) and Surani *et al.* (1984; 1986) using mice have unequivocally demonstrated that there is an absolute requirement for a genetic contribution from both sexes in order for development to proceed normally, *ie.*, maternal and paternal contributions are not equivalent. Embryos containing only maternal contributions develop minimal extraembryonic tissues (trophectoderm), while a poorly developed embryo proper is characteristic of embryos containing only paternal genomes. The importance to the mammalian embryo of both maternal and paternal genomes is most apparent in the very low viability of mammalian parthenogenotes.

A variety of different organisms have been shown to have some sort of gamete-of-origin effect (Crouse, 1960; Brown and Nelson-Rees, 1961; Spofford, 1961; Klar, 1987). It is not at all clear at this point whether the phenomena observed in this wide range of phyla are acting under the same mechanism. The genome imprinting phenomenon, however, has been most

intensively studied in mammals, particularly mice. Several endogenous genes, as well as regions of many mouse chromosomes (Cattanach and Beechey, 1990) have been shown to be imprinted. Work by a number of laboratories has demonstrated a gamete-of-origin influence on the level of methylation of a variety of transgenes in mice (Hadchouel *et al.*, 1987; Reik *et al.*, 1987; Sapienza *et al.*, 1987; Swain *et al.*, 1987). With one exception, methylation tends to be increased after passage of a transgene locus through the female germline and decreased after passage through male gametogenesis. In a number of cases an accompanying effect has been found with respect to expression of the transgene; expression is decreased after passage through a female and increased by passage through a male (Swain *et al.*, 1987; McGowan *et al.*, 1989; Allen *et al.*, 1990). A mosaic pattern of transgene expression has also been found in several transgenic lines of mice suggesting that the imprinted state may not be retained in all cells (McGowan *et al.*, 1989) or that only a sub-set of early stem cells is affected by the imprint.

As discussed here, and previously, the phenomenon of genomic imprinting appears to possess the following characteristics: 1) genotype- or strain-specific modification of a locus, 2) a relationship between the methylation and expression of a locus, and 3) gamete-of-origin specific modification of a locus. I have previously demonstrated that the zebrafish transgene methylation phenotype can be modified by genotype-specific modifiers that act in *trans* and that a correlation exists between the proportion of expressing cells and the transgene methylation pattern. The similarity between the behaviour of the zebrafish transgene and the imprinted mouse transgenes suggested that genomic imprinting may be operating on the zebrafish transgene. Genomic imprinting was studied in

mice by maintaining the transgene in a hemizygous state so that the parental origin of the transgene could be controlled. I employed a similar strategy to demonstrate that the zebrafish transgene is also affected by gamete-of-origin specific modification. I also determined the methylation patterns of a variety of tissues including the gametes which may provide information on the time and place of the epigenetic modification of the transgene. It will further be determined whether variation exists in the methylation phenotype of the transgene prior to fertilization.

The identification of an imprinting process in the zebrafish has important ramifications in deciphering the evolutionary origin of imprinting since most of the present theories are based on imprinting effects as seen in mammals.

Materials and Methods

Tissue Dissection

Adult transgenic zebrafish (a homozygous transgenic female, a homozygous transgenic male, and a hemizygous F1 carrying a male derived transgene) were anaesthetized with 2-phenoxyethanol. The following tissues were removed for processing under a dissecting microscope: skin and scales, eyes, muscle, brain and braincase. The tissues were used for DNA extraction and Southern blot analysis.

Mature sperm samples were obtained by anesthetizing male transgenic fish and placing them under a dissecting microscope. A slight pressure was applied to the abdominal region with the finger and the zebrafish milt was collect from the genital pore region using a 20 ul pipetteman. The sperm were immediately transferred into a 1.5 ml eppendorf tube and kept on ice until processing. Sperm samples were pooled from 6-8 sibling fish from

the same breeding. Samples were collected from 1) hemizygous transgenic males from a cross between a wild type longtail phenotype female and a homozygous transgenic male, and F1 males hemizygous for the transgene locus containing a 2) male derived transgene, and a 3) female derived transgene.

Unfertilized eggs were collected from gravid females by first anesthetizing them with 2-phenoxyethanol. The fish were placed in a small petri dish and were viewed under a dissecting microscope. The eggs were stripped from gravid females by applying a gentle pressure to the abdominal region. The eggs were placed in 1.5 ml. eppendorf tubes and stored at -70°C until processing. Because of the large number of eggs required to obtain adequate amounts of DNA for southern analysis, samples were pooled from females belonging to a number of different crosses (representing different somatic transgene methylation phenotypes).

Results

An examination of the transgene methylation pattern of a variety of tissues within individual fish (homozygous and hemizygous) gave the result (Figure 2.1) predicted from studies on transgenic mice. The methylation phenotype of the transgene locus was similar in all somatic tissues with only minor variation. The exception was the germ cells. The male gametes displayed a markedly different and very hypomethylated pattern than all other tissues and/or individuals. The *Hpa II* digestion of pooled, unfertilized egg DNA from transgenic zebrafish on the other hand revealed a very hypermethylated transgene locus. The egg transgene was more methylated than all sperm, and adult DNAs sampled (with the possible

exception of some individuals with very highly methylated transgenes). This suggests that at least some of the methylation that appears to be gamete-of-origin specific actually occurs post-fertilization. A similar result has been found with transgenes (McGowan et al., 1990: Chaillet et al., 1991) and endogenous genes (Stoger et al., 1993) in mice, at least for the male derived allele.

In sperm taken from groups of individuals also possessing different somatic transgene methylation phenotypes, the locus displayed a methylation pattern that was different from somatic patterns, and were relatively hypomethylated in all cases. However, it is apparent that there is some variation in the transgene methylation between these different groups. Similar variations in the degree of transgene methylation are observed in both BstU I (LTR probe) and Hpa II (CAT probe) cleaved DNAs (Figure 2.2a,b). From these autoradiographs it is difficult to determine which sperm DNAs are more or less methylated. It is apparent, however, that variation exists in the base-line methylation phenotype of mature sperm that was established during gametogenesis (cf. Ueda *et al.*, 1992). That is, there is genetic variation in the methylation events that take place during gametogenesis. A similar effect was observed with murine transgene loci (Chaillet *et al.*, 1991; Reik *et al.*, 1990). If methylation is in fact the epigenetic marking used to establish the male-specific imprint, these data suggest that all males do not imprint their DNA to the same degree or in the same way.

Analysis of the zebrafish transgene methylation phenotype after passage of the transgene locus through male and female gametogenesis revealed a distinct gamete-of-origin specific effect on the transgenes methylation. Consistently, passage of the transgene through male

gametogenesis resulted in a higher methylation pattern compared to a decrease in transgene methylation after passage through a female (Figures 2.3,2.4,2.5). Figure 2.3 illustrates the gamete-of-origin effect on transgene methylation resulting from breedings between randomly chosen male and female homozygous transgenics and non-transgenic fish. Offspring that have inherited their transgene from a male have a more hypermethylated transgene than offspring whose transgene locus was received from a female. These crosses were repeated with a variety of randomly selected parents with consistent results of varying magnitude. This effect on transgene methylation is analogous to the gamete-of-origin specific modification phenomenon observed in some murine transgenes and is indicative of genomic imprinting. Since these initial results represent only a single generation it is formally possible that this effect is just the result of differences in the original transgenic and non-transgenic mates (*ie.* randomly selected males always had a higher transgene methylation phenotype than the randomly selected female zebrafish). Therefore, some of the F1 hemizygous offspring were again crossed to the same non-transgenic mates, the offspring were examined, and the pedigree then extended to a third generation. A number of individuals as well as pooled samples were examined in each generation except in the last generation in which only pooled samples were analyzed. In all cases nearly identical methylation phenotypes were observed within a single group. Figure 2.4 represents one example of the passage of the transgene through three generations were the transgene was initially maternally derived and was then passed on through the second generation backcross male and female siblings. Because both the second and third generation backcross offspring shared the same hemizygous transgenic grandparent and non-transgenic

mates, the transgene was, therefore, identical by descent. Any differences in the transgenes methylation must result from the sex of the parent for whom the offspring received the transgene locus. Again, offspring from the female siblings had a consistently less methylated transgene locus than offspring of the male siblings. The reversibility of this epigenetic modification, the imprint, is an important feature of the genomic imprinting system. That is, males receiving their transgene locus from a female must be able to erase the female generated imprint and pass the transgene locus with the male imprint (and *visa versa*). The ability of one sex to reverse the epigenetic modification imposed on the locus by the opposite sex was evident in the offspring (F3) of a male (F2) that inherited the transgene locus from a female (Figure 2.5).

The results of passing the transgene in a hemizygous state through male and female zebrafish over several generations clearly suggests the presence of a genomic imprinting system similar to that observed in mouse for both endogenous genes and transgenic loci. Interestingly, the imprint is in the opposite direction from that observed in mice where passage through a female usually resulted in a higher methylation pattern among the offspring as compared to a lower transgene methylation pattern seen after passage through male gametogenesis.

Figure 2.1: Tissue differences in transgene methylation. DNAs were extracted from a variety of tissues obtained from a homozygous female zebrafish including skin and scales (lane 1), eyes (lane 2), muscle (lane 3) and head (lane 4). The DNAs were digested with *HpaII* restriction endonuclease, electrophoresed in a 0.8% agarose gel, transferred to nylon membrane and hybridized with the same DNA probe used in the previous figures. Also shown in the figure is sperm DNA (pooled from several individual homozygous males), lane 5 and unfertilized egg DNA (pooled from a large number of homo- and hemizygous females), lane 6. All somatic tissues display a similar methylation pattern whereas egg DNA is hypermethylated and sperm DNA is hypomethylated.

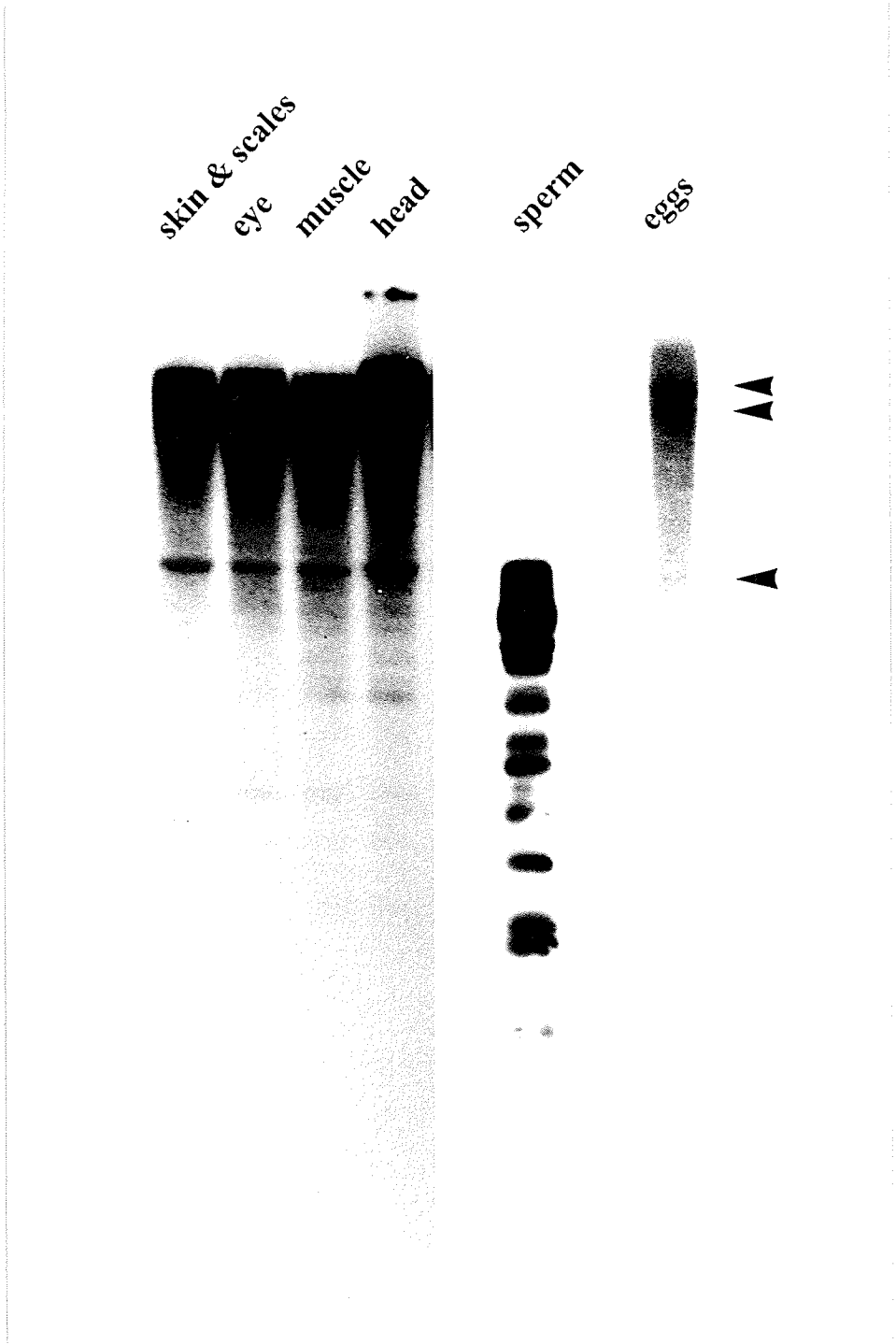

skin & scales
eye
muscle
head
sperm
eggs

Figure 2.2: Sperm DNA methylation. DNA was extracted from pooled sperm samples collected from sibling males, digested with *BstUI* (left panel) or *HpaII* (right panel) restriction endonucleases, electrophoresed through a 0.8% agarose gel, transferred onto nylon membrane. The *BstUI* digests were hybridized with a RSV-LTR promoter-specific probe and the *HpaII* digests were hybridized with the same probe as in the previous figures. Each lane represents the sperm from F1 males resulting from a cross of a homozygous transgenic parent to; lane 1, a longtail female; lane 2, a wildtype female; and lane 3, a wildtype male. The figure illustrates that there are differences in the methylation pattern between sperm samples in both the promoter region and the CAT gene itself.

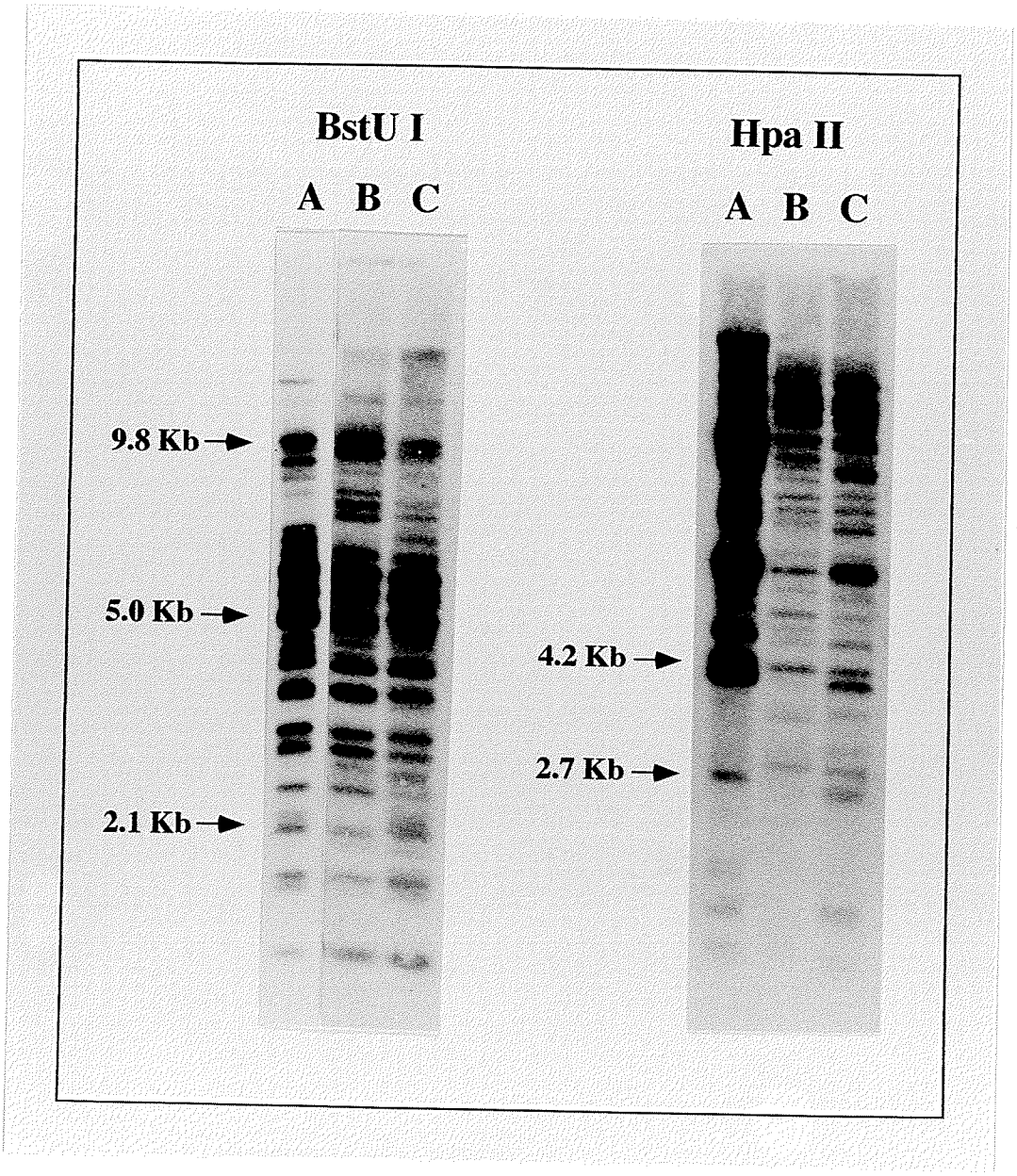

BstU I
A B C
Hpa II
A B C
9.8 Kb
5.0 Kb
2.1 Kb
4.2 Kb
2.7 Kb

Figure 2.3: Reciprocal crosses of transgenic zebrafish to non-transgenic mates. *(Top)* The filled triangle represents the homozygous transgenic parent of both sexes crossed to either a non-transgenic male (open square) or female (open circle). *(Bottom)* DNAs from individual offspring produced from the above crosses digested with *BglI/HpaII* and hybridized with transgene-specific probes. The methylation of the transgene locus is lower in the offspring of transgenic females (lanes 1 and 2) as compared to the offspring of transgenic males (lanes 3 and 4).

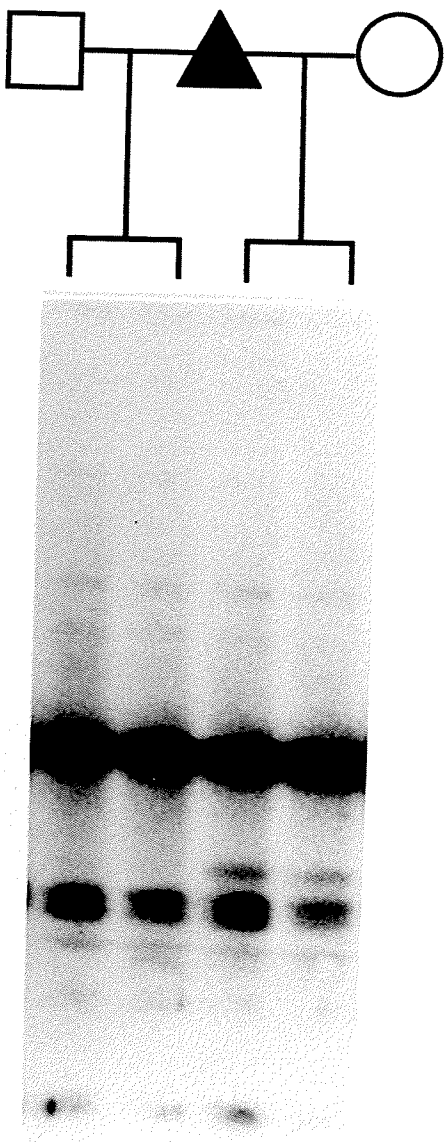


Figure 2.4: Inheritance of the transgene methylation. A three generation pedigree is shown in which the grandfather shown at the top of the pedigree is an F1 hemizygote (see figure 1) carrying a single maternally-derived transgene allele. Shown beneath the pedigree is DNA extracted from two generations of offspring, digested with *HpaII*, electrophoresed through a 0.8% agarose gel, transferred to nylon membrane and hybridized with a CAT-specific probe. The middle lane illustrates the methylation of the transgene locus in both male and female siblings of the 2nd generation backcross. The lane on the left (pooled sample) illustrates the decrease in methylation after passage of the locus through a female and the lane on the right (pooled sample) the increase in methylation after passage through her male sibling. The arrows indicate specific fragments which are diagnostic in determining the direction of the change in methylation level.

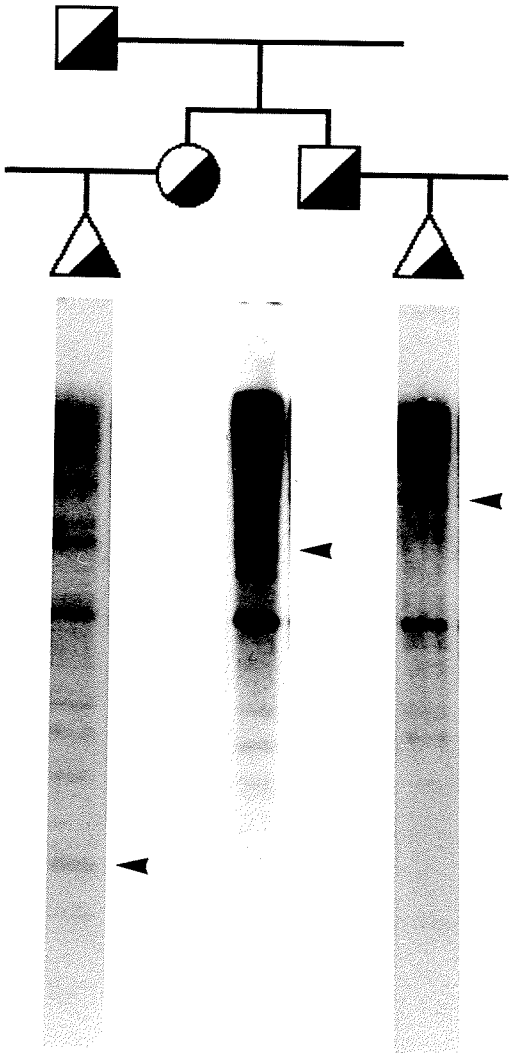


Figure 2.5a and 2.5b: Inheritance of the transgene methylation. A three generation pedigree is shown in which the grandfather shown at the top of the pedigree is an F1 hemizygote (see figure 1) carrying a single paternally-derived transgene allele. Shown beneath the pedigrees is DNA extracted from three generations of offspring, digested with *HpaII* , and hybridized with a CAT-specific probe (Panel A). The middle lane illustrates the methylation of the transgene locus in the female siblings of the 2nd generation backcross. Passage of the transgene from a male resulted in an increase in transgene methylation in the second generation. The lane on the right (pooled sample) illustrates the decrease in methylation after passage of the locus through this 2nd generation female to her 3rd generation offspring. The promoter region of the transgene shows the same decrease in transgene methylation after through a female. DNA was digested with BstU I, and hybridized to the SV-40 -RSV LTR probe. The arrows indicate specific fragments which are diagnostic in determining the direction of the change in methylation level.

B
A

Discussion

Genomic imprinting is a phenomenon that differentially marks the genome in contributing parents during gametogenesis, ultimately producing functional differences in the genetic information contributed by each parent. Work by a number of laboratories has suggested that the methylation of gametic DNA is either the marking mechanism of genomic imprinting or reflects that marking system.

The methylation phenotype of the transgene was similar in all tissues examined with the exception of the gametes, sperm and egg. Slight variations were observed between the methylation phenotype of adult somatic tissues assayed, but a different methylation pattern was observed in egg, and even more dramatically in sperm. The methylation patterns identified in sperm and eggs were unique to those cells and were not observed in any other tissues of any individual examined. These results, too, are as predicted from observations in transgenic mice. In those lines having an imprinting effect, the male germ cells tend to have the transgene locus in a unique and hypomethylated state. This suggests that at least some epigenetic modification occurs post-fertilization (Sapienza *et al.*, 1987; Chaillet *et al.*, 1991). Further, the markedly different methylation patterns of the transgene between sperm and egg may reflect the different chromatin states of the two gametes. In any case, because of the difficulty in obtaining large numbers of eggs from mouse, this is the first information on the methylation status of an imprinted transgene in eggs. Clearly, the differences in the transgene between the two gametes may allow for the differentiation of the maternal and paternal genomes post-fertilization and facilitate the imprinting phenomenon. Although in

mammals at least DNA is globally demethylated pre-gastrulation, it is interesting that the direction of the imprint is opposite to the methylation patterns found in the gametes.

McGowan *et al.* (1989), studied a line of transgenic mice containing a *lacZ* construct (Tg4), and demonstrated an imprinting effect associated with that transgene. More importantly, they showed that the transgene was expressed in a mosaic pattern in the ventral floor plate of the neural tube, a tissue in which the gene was spontaneously expressed. They concluded that the observed variation in transgene expression in the neural tube and the variation in the methylation of all tissues was the result of changes in the proportion of two populations of cells that have different methylation phenotypes, *ie.* high and low. Due to the similarity in methylation between all somatic tissues examined they suggested that the methylation mosaicism may already be established in the last common precursor of those tissues (McGowan *et al.*, 1990). It is not clear from that study whether the methylation pattern is established once in all precursor cells and faithfully propagated, or if the methylation phenotype is unstable and continuously modified.

The results shown in Figures 2.3, 2.4 and 2.5 suggest that the transgene methylation differences observed in the offspring of the various zebrafish pedigrees were caused by the sex of the parent contributing the transgene locus rather than the sex or genotype of the offspring. Breeding the offspring of each generation to the same non-transgenic mates also minimizes the effects of normal population variation. The presented data strongly suggests that genomic imprinting is operating in a non-mammalian vertebrate, the zebrafish, and can be monitored at the level of transgene methylation. A significant difference between the mechanism of genomic

imprinting in mice and zebrafish is in the direction of the imprint. In the zebrafish passage of the transgene through male gametogenesis causes an increase in transgene methylation, while passage through female gametogenesis decreased the transgene methylation. In all imprinted murine transgenes except one (Sapienza *et al.*, 1987), passage of the transgene through the male germ line caused a decrease in transgene methylation while passage through female gametogenesis causes an increase in transgene methylation. Mice have a heterogametic sex system of the XY/XX type with the male being the heterogametic sex. The type of sex chromosome system used by the zebrafish is unknown but a variety of different sex chromosome systems can be found among the fishes. For example, the mollies *Poecilia sphenops* and *P. velifera* have a ZZ/ZW chromosome configuration whereas the XY/XX system is used by the guppies *Poecilia reticulata* and the platyfish *Xiphophorus maculatus* (Nanda *et al.*, 1993). If imprinting results from sex-linked, dosage sensitive modifiers as proposed by Sapienza (1990) then our results, when compared to the mouse, suggests that the zebrafish may be using an opposite heterogametic arrangement to the mouse, *ie.*, the ZZ/ZW, in which the female is the heterogametic sex. Locke *et al.* (1990) defined two classes of these dosage sensitive modifiers: a) Class I - suppressors of variegation and b) Class II -enhancers of variegation. By these definitions it is also formally possibly that the zebrafish does have a XY/XX sex chromosome system but the modifiers involved in the imprinting process are X-linked class II modifiers as opposed to the suggested X-linked class I modifiers thought to be operating in the mouse. Conversely, the imprinting effects observed in the zebrafish could also be accomplished by the presence of Y-linked class I modifiers.

The recognition that an imprinting process is acting in the zebrafish provides a possible alternative to mice as a model system for analyses into the mechanism of imprinting as well as what role it may play in human disease. Although the zebrafish is more evolutionarily distinct from humans than the mouse it has characteristics that make it ideal for asking certain types of questions. It is possible to obtain large numbers of offspring that develop externally and are optically clear so that they can be monitored throughout development. Therefore, the zebrafish may provide a unique opportunity to examine the behaviour of imprinted genes or even imprinted cells during the development of an organism.

Chapter 3: *Snrpn,* A Possible Candidate for the Study of Genomic Imprinting of an Endogenous Gene in the Zebrafish, *Danio rerio.*

Introduction

Using Robertsonian and reciprocal translocation genetic studies, Cattanach and Beechley (1990) identified, in a number of the 19 mouse chromosomes, many autosomal regions that are subject genomic imprinting. Although many transgenes in mice have been shown to be imprinted, very few endogenous genes so far have been proven to show the gamete-of-origin phenomenon. To date, although many genes are suspect, only four endogenous genes (mouse and human) have unequivocally been shown to be imprinted: *IGF2*, *IGF2r*, *H19* and *Snrpn.*

The *IGF2* gene encodes a single polypeptide that has a direct role in embryonic growth (DeChiara *et al.*, 1990; DeChiara *et al.*, 1991) and, therefore, may mediate developmental pathways by providing a product necessary for growth and cell differentiation (Whitman and Melton, 1989). The recognition that genes of this type can be imprinted goes a long way in explaining the failure of uniparental embryos. Interestingly, studies by DeChiara *et al.* (1990) and Barlow *et al.* (1991) concluded that the gene for insulin-like growth factor II (*IGF2*) and the insulin-like growth factor receptor(*IGF2r*) were imprinted in opposite directions. That is, in mouse embryos, the maternal gene for *IGF2* is poorly transcribed while the paternal copy of the *IGF2r* is also poorly transcribed. The maternal copy of *IGF2r* and the paternal copy of *IGF2* are actively expressed. By studying the methylation patterns of various regions of both *IGF2* and *IGF2r* loci, Ströger *et al.* (1993) attributed the cause of the imprint to the

regional methylation differences in both loci of the maternal and paternal contributions. Disruption or relaxation of the normal imprinting of *IGF2* and *IGF2r* has been implicated in the cause of a number of cancers such as Wilms' tumors (Schroeder *et al.*, 1987; Pal *et al.*, 1990; Ogawa *et al.*, 1993) and rhabdomyosarcoma (Scable *et al.*, 1989) (reviewed in Rainer *et al.*, 1993).

The *H19* gene maps to the distal segment of the mouse chromosome 7 and is genomically imprinted such that the maternally inherited gene is expressed and the paternally derived gene is repressed (Bartolomei *et al.*, 1991). The transcripts of the *H19* gene are one of the most abundant RNAs (although of presently unknown function) in the developing mouse embryo from blastocyst to birth, where expression is then limited to skeletal muscle (Poirier *et al.*, 1991). Ferguson-Smith *et al.* (1993) showed that the 5' portion of the *H19* gene is only methylated on the paternal allele, suggesting that methylation may be the epigenetic modification responsible for the imprinting phenomena. The *H19* gene has also been implicated in numerous cancers (reviewed in Rainer *et al.*, 1993). Interestingly, both *IGF2* and *H19* are closely linked and map to the same human chromosome 11 region, 11p15.5 (Zemel *et al.*, 1992).

The *Snrpn* gene encodes one of a group of polypeptides, called small nuclear ribonucleoprotein particles, which play a functional role in mRNA splicing (Luhrmann *et al.*, 1990). The transcripts produced from the *Snrpn* gene are most abundant in the brain, to a lesser degree the heart, and virtually undetectable in all other tissues (McAllister *et al.*, 1988; Li *et al.*, 1989). These data have led to the suggestion that *Snrpn* may be involved in regulating tissue-specific mRNA splicing. The *Snrpn* gene maps to an area of mouse chromosome 7 which is homologous to the

region of human chromosome 15q11-13. This region of the human chromosome has been implicated in the human disorders: Prader-Willi syndrome and Angelman syndrome (Nicholls *et al.*, 1991; Chaillet *et al.*, 1991b; Wagstaff *et al.*, 1991). Furthermore, the human *Snrpn* gene is expressed in the affected tissues and maps specifically to the Prader-Willi syndrome critical region, but not that of Angelman syndrome (Ozcelik *et al.*, 1992). Paternal deletions or maternal disomy in this region, which causes a paternal gene deficiency of this region, have been associated with Prader-Willi syndrome (Ledbetter *et al.*, 1981; Nicholls *et al.*, 1989). Though this information suggested that imprinting of the *Snrpn* gene might be involved in the disorder of Prader-Willi syndrome (Hall, 1990), Leff *et al.* (1992) only recently showed by RNase protection assay that *Snrpn* is only expressed from the paternal allele (maternally imprinted), in agreement with the deletion studies.

Zebrafish RNA and genomic DNA have been screened by northern and Southern blot analysis for homologous sequences with mouse *IGF2* and *Snrpn* cDNA clones in an attempt to identify possible zebrafish homologs of these genes that would, therefore, be candidates for endogenous, imprinted genes. The imprinted gene *H19* was not investigated due to the poor likelihood of cross-species homology (S. Tilgman, personal communication). The previous evidence of an imprinted transgene in zebrafish suggests that there are regions of the zebrafish genome subjected to genomic imprinting. Demonstration that similar genes are imprinted within vertebrates would be evolutionarily and developmentally significant.

Materials and Methods

Probes for Cross Species Hybridization

Plasmid containing the gene for mouse insulin-like growth factor (*IGF2*) was generously supplied by Dr. L. Murphy (Department of Physiology, University of Manitoba). Southern blots were probed with a 1.0 Kb EcoRI fragment containing the mouse *IGF2* gene.

The plasmid pMN.13 which contains a full length mouse small nuclear ribonucleoprotein particle polypeptide-N (*Snrpn*) cDNA was a generous gift from Dr. Nancy A. Jenkins (Mammalian Genetics Laboratory, NCI, Frederick, Maryland). The cDNA clone was isolated from a mouse forebrain cDNA library. Southern and northern blots were probed with a 0.85 Kb Hind III fragment containing the complete coding sequence.

Southern blots containing DNAs from wildtype zebrafish, longtail zebrafish, the leopard danio *Brachydanio rerio frankei* were prepared as previously described. DNAs from these various strains were used to determine the presence of RFLPs. Lanes containing DNA from BALB/c mice were included as positive controls. DNAs were digested according to the manufacturers instruction with the restriction enzymes: Hind III, EcoR I, Pst I, and Hpa II. *IGF2* and *Snrpn* probes were labeled by random priming, as previously described. Blots were prehybridized overnight at 43°C in a solution containing: 6X SCC, 10X Denhardts solution, 100 ug/ml denatured *E.coli* DNA, 0.5% SDS. Blots were hybridized overnight at 50°C in a solution containing: 4X SET, 1X Denhardt's, 100 ug/ml denatured *E.coli* DNA, 0.5% SDS, and ^{32}P labeled probe (4,000,000 cpm/ml). The blots were washed at 60°C in several changes of 1X SCC and 0.1% SDS. Autoradiographs were obtained from labeled blots as previously described.

Mouse (brain and kidney) and zebrafish (whole fish) total RNA were prepared by guanidine HCl method (Cox, 1968). The concentration of the

RNA was determined spectrophotometrically. The RNA was denatured with glyoxal and dimethyl sulfoxide buffer (1.0M glyoxal, 50% DMSO (vol/vol), 10mM $NaPO_4$ and 0.1% SDS) at 50°C for 30-60 minutes (Thomas, 1980) and placed on ice. The RNA was electrophoresed in a 1% agarose gel made with 10mM $NaPO_4$ running buffer pH 6.8 with circulation of the running buffer. The RNA was transferred onto nylon membrane. The prehybridization and hybridization conditions were the same as described above for Southern blots.

Screening Zebrafish cDNA Libraries and Plasmid Excision

A Lambda ZapII library containing cDNAs from 20-28 hour zebrafish embryos was obtained from D. Grunwald of the University of Utah. This cDNA library was screened with the mouse *Snrpn* cDNA by methods from Sambrook *et al.* (1989). This library was used since during this stage the brain comprises a large portion of the whole embryo. Approximately 250 ml. of NZCYM agar medium (Sigma) containing 50 mg/ml tetracycline was poured into three large square culture plates to form the bottom agar layer. The bottom agar was allowed to dry for 1-2 days.

A single colony of XL-1 Blue cells was inoculated into sterile LB-media supplemented with 0.2% maltose and grown overnight at 37°C. The cells were pelleted by centrifugation at 4000g for 10 minutes. The supernatant was discarded and the cells resuspended in sterile 0.01 M $MgSO_4$ to a dilution of 1.6 X 10^9 cells/ml. (OD_{600}=0.5). The resuspended cells were then stored at 4 °C until later used.

The XL-1 Blue cells were infected with the Lambda phage by adding 100 ul. of diluted bacteriophage stocks of a predetermined concentration (diluted in SM solution: 0.1 M NaCl, 0.01 M $MgSO_4$, 0.05 M Tris-Cl (pH

7.5), 0.01% gelatin) into 100 ul of diluted XL-1 Blue cells. This mixture was incubated for 20 minutes at 37°C to allow bacteriophage particles to infect the bacterial cells. After infection, 30 mls. of molten 0.7% agarose (47°C) supplemented with 50 μg/ml tetracycline was mixed with the infected cells. This solution was immediately poured and spread over the bottom agar previously prepared. These plates were incubated at 37 °C overnight during which time plaques appeared.

Plaque lifts were made using nylon hybridization membrane and approximately 75 000 plaques were screened using the mouse *Snrpn* clone as a probe (methods the same as for previously described southern blots). Five positive clones were observed on the autoradiographs. Plugs of the positive plaques were picked and placed in 1 ml. SM buffer with 50 ul of chloroform. These were allowed to sit at room temperature for 1-2 hours, centrifuged, and the supernatant containing the *Snrpn* positive phage was removed. This supernatant was stored at 4 °C and used for a second screen.

The phagemid Bluescript containing the *Snrpn* cDNA insert was excised from the Lambda ZapII vector by co-infection of XL1-Blue MRF' cells with ExAssist interference-resistant helper phage (Stratagene). The excision process produces an infectious single-stranded filamentous phage which when infected into SOLR bacteria strain replicates as the double stranded Bluescript phagemid.

Dideoxy-Sequencing of cDNA Clone and Sequence Analysis

The cDNA inserts within the excised Bluescript were sequenced by the dideoxy-nucleotide termination method using the dsDNA Cycle Sequencing System (BRL). Both the forward primer (5'-CCCAG TCACG ACGTT GTAAA ACG-3') and the reverse primer (5'-AGCGG ATAAC AATTT

CACAC AGG-3') were used to sequence the complete 350 bp. cDNA insert. The sequence was compared to all registered sequences stored at a number of data banks using the BLAST Network Service (Blaster) (Altschul *et al.*, 1990) available via e-mail at National Center for Biotechnology Information (NCBI) (e-mail address: blast@ncbi.nlm.nih.gov). This program aligns DNA or protein sequences and suggests registered sequences with the greatest degree of homology.

Results:

Hybridization with the mouse *IGF2* clone resulted in no distinct bands on Southern blots of zebrafish DNAs when screened under low stringency conditions. This suggested that the zebrafish does not possess any sequences with significant homology to the mouse *IGF2* gene. Further investigation of a zebrafish *IGF2* homologue was discontinued because of this fact.

Hybridization of the mouse *Snrpn* to the same Southern blot used for *IGF2* showed distinct bands of hybridization (Figure 3.1) even at relatively high stringency conditions (1X SSC at 60°C). The zebrafish genome, therefore, must possess a sequence with a relatively high degree of homology to the mouse *Snrpn* gene. Almost as important, distinct restriction length polymorphism's (RFLPs) were observed between the different zebrafish strains. The presence of possible exon polymorphism's between zebrafish strains is essential for the use of RNase protection assays in determining whether the *Snrpn* gene is imprinted. A large number of bands are present in these Southern blots that are unlikely to be accounted for by the hybridization of the probe to a single homologous gene or DNA sequence. The implication is that, at this specific stringency, the probe

must be hybridizing to a number of highly homologous DNA sequences.

The mouse and human *Snrpn* genes are almost exclusively expressed in the brain. Northern blot analysis of mouse brain total RNA and whole zebrafish total RNA hybridized to the murine *Snrpn* showed, as expected, very strong expression in the mouse brain and a weak signal in the whole zebrafish RNA (Figure 3.2). The weak signal in the total zebrafish RNA is consistant with the idea that the expression of Snrpn in the zebrafish may be restricted to a discrete tissue and not the whole body.

Indications that *Snrpn* may be expressed in the zebrafish suggested that a zebrafish clone could be obtained by screening cDNA libraries. By screening approximately 75,000 plaques with the mouse *Snrpn* cDNA, a cDNA clone of approximately 350-400 bp was obtained from a 20-28 hour zebrafish (post-somatogenesis) cDNA library. The complete sequence of the cloned zebrafish cDNA (350 bp.) is shown in the sequence alignment (Figure 3.6).

Southern blots using DNA from the three different previously mentioned zebrafish strains hybridized with the zebrafish cDNA clone revealed distinct bands and RFLP's between the different strains (Figure 3.3). Some of these bands can be aligned with the Southern blot using the mouse SmN probe. It is also apparent that the probe is hybridizing to other fragments which must also have a relatively high degree of homology. Northern blots with equal amounts of zebrafish total brain enriched RNA and total tail RNA hybridized to the fish cDNA shows a single band of near equal intensity in both samples (Figure 3.4). No hybridization occurred in the lane containing mouse brain RNA. The expression in zebrafish of the cloned cDNA is therefore not limited to the brain, and is probably ubiquitous. Hybridization of an identical northern

blot with the mouse Sm N gene revealed a band in the mouse brain lane which appears to be of identical size to the band found in the fish lanes hybridized to the zebrafish cDNA clone.

The BLAST Network Service (Blaster) (Altschul *et al.*, 1990) determined that the zebrafish clone was most homologous to the murine SmB protein gene (GenBank #M58761) with 81% sequence homology (Figure 3.5) and the rat *Snrpn* (GenBank #M29295) with 80% sequence homology (Figure 3.6). The region of homology between the zebrafish clone and these two mammalian genes is in the 3' region. The zebrafish *Snrpn* was most homologous with the murine SmB protein gene (1300 bp) between the1100 - 1300 bases, and most homologous with the rat Sm N gene (913 bp) between the 460 - 810 bp region.

Figure 3.1: Southern blot of DNA from wildtype, leopard danio, and longtail zebrafish digested with Hind III, EcoRI, and Pst I and hybridized to the mouse *Snrpn* cDNA and washed at low stringency. A single lane of Pst I digested BALB/c mouse DNA is included as a positive control. The number of bands suggests that the probe is hybridizing to more than a single gene or homologous sequence.

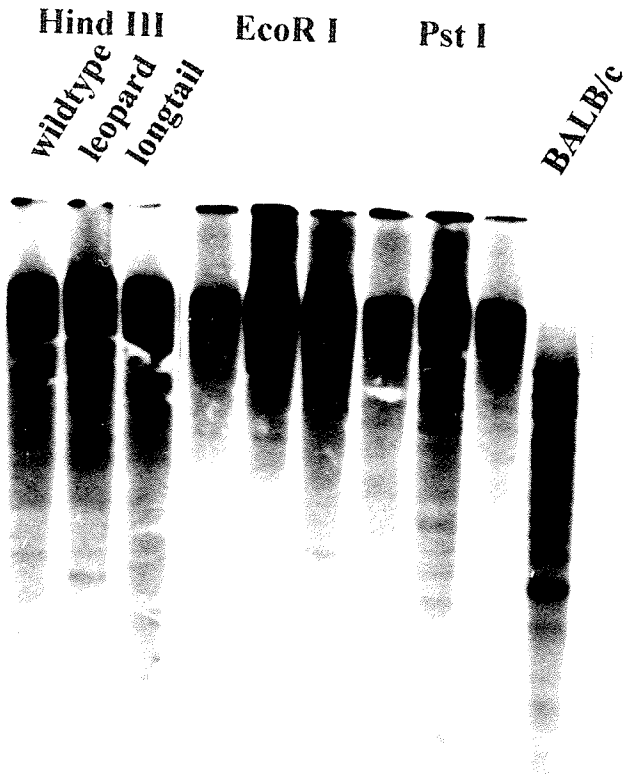

Hind III
EcoR I
Pst I
wildtype
leopard
longtail
BALB/c

Figure 3.2: Northern blot of BALB/c mouse brain and kidney total RNA (20μg) and male and female zebrafish total RNA (40μg) hybridized to the mouse *Snrpn* cDNA. The tissue specificity of *Snrpn* is apparent by the strong message in the mouse brain and weak message in the mouse kidney. There is also evidence of a very weak signal in the zebrafish total RNA.

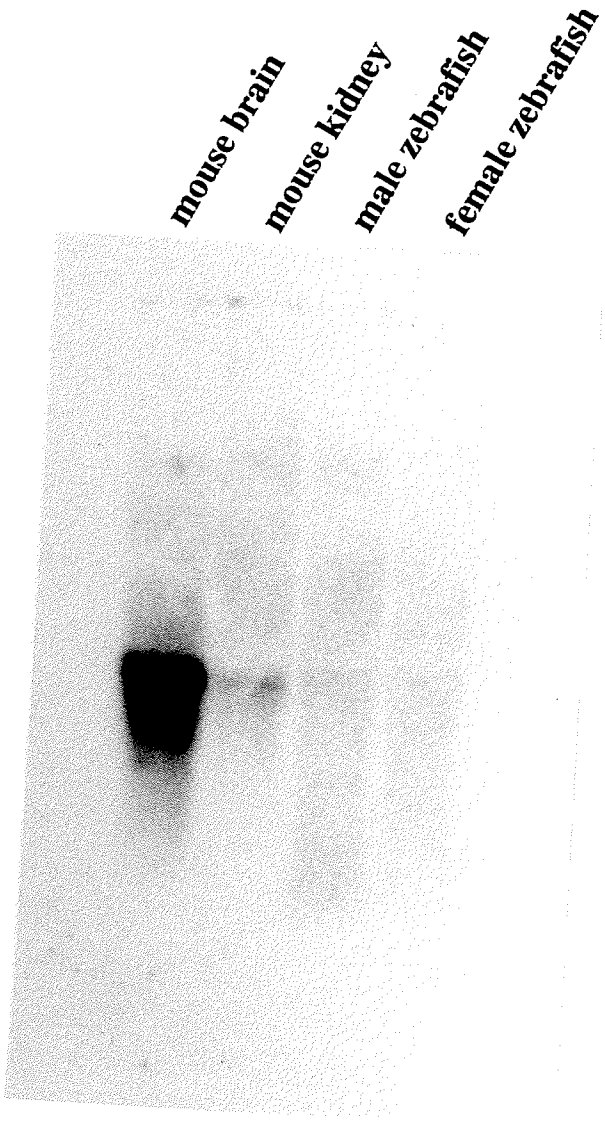

mouse brain
mouse kidney
male zebrafish
female zebrafish

Figure 3.3: Southern blot of DNA from wildtype, leopard danio, and longtail zebrafish digested with Hind III, EcoRI, Pst I and Hpa II and hybridized to the cloned zebrafish *Snrpn* partial cDNA and washed at high stringency. The zebrafish *Snrpn* probe hybridizes to specific restriction fragments, while it is obvious that there exists restriction fragment polymorphism's between the different strains of zebrafish. Some faint bands are also apparent possibly indicating some probe homologous to other sequences within the zebrafish genome.

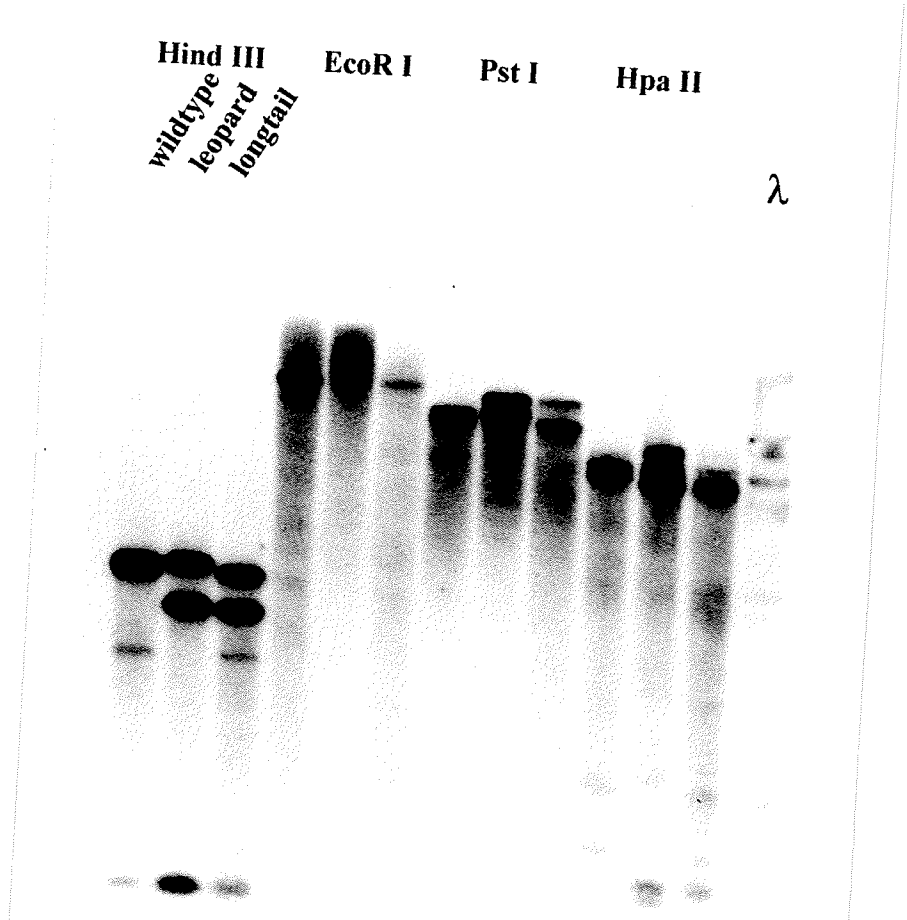

Hind III
wildtype
leopard
longtail
EcoR I
Pst I
Hpa II
λ

Figure 3.4: Northern blot of total tail RNA (30μg), total brain-enriched head RNA (30ug) and BALB/c mouse brain total RNA (20 ug) hybridized to the cloned zebrafish *Snrpn* cDNA. The cloned zebrafish *Snrpn* appears to be expressed in equal proportions in these different tissues. The cloned zebrafish *Snrnp* is ubiquitously expressed in the zebrafish.

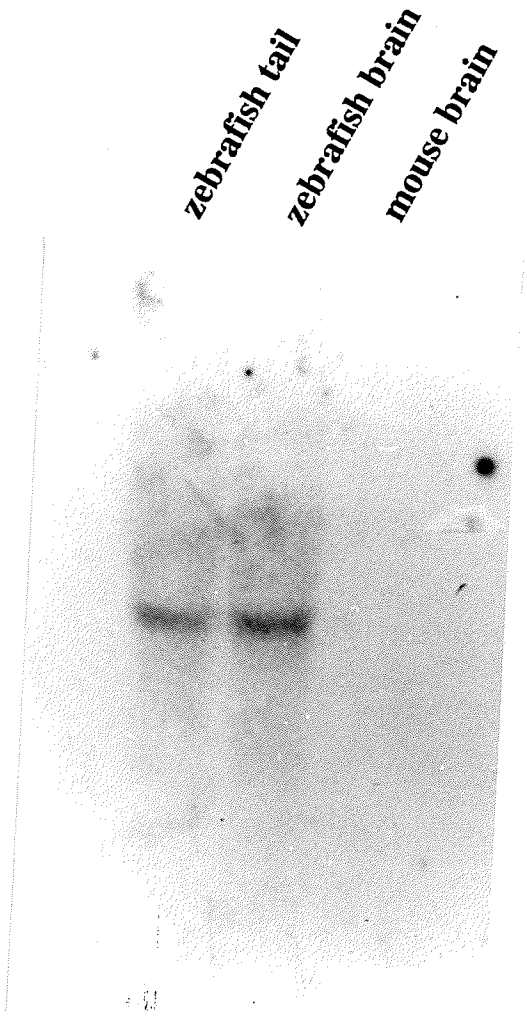

zebrafish tail
zebrafish brain
mouse brain

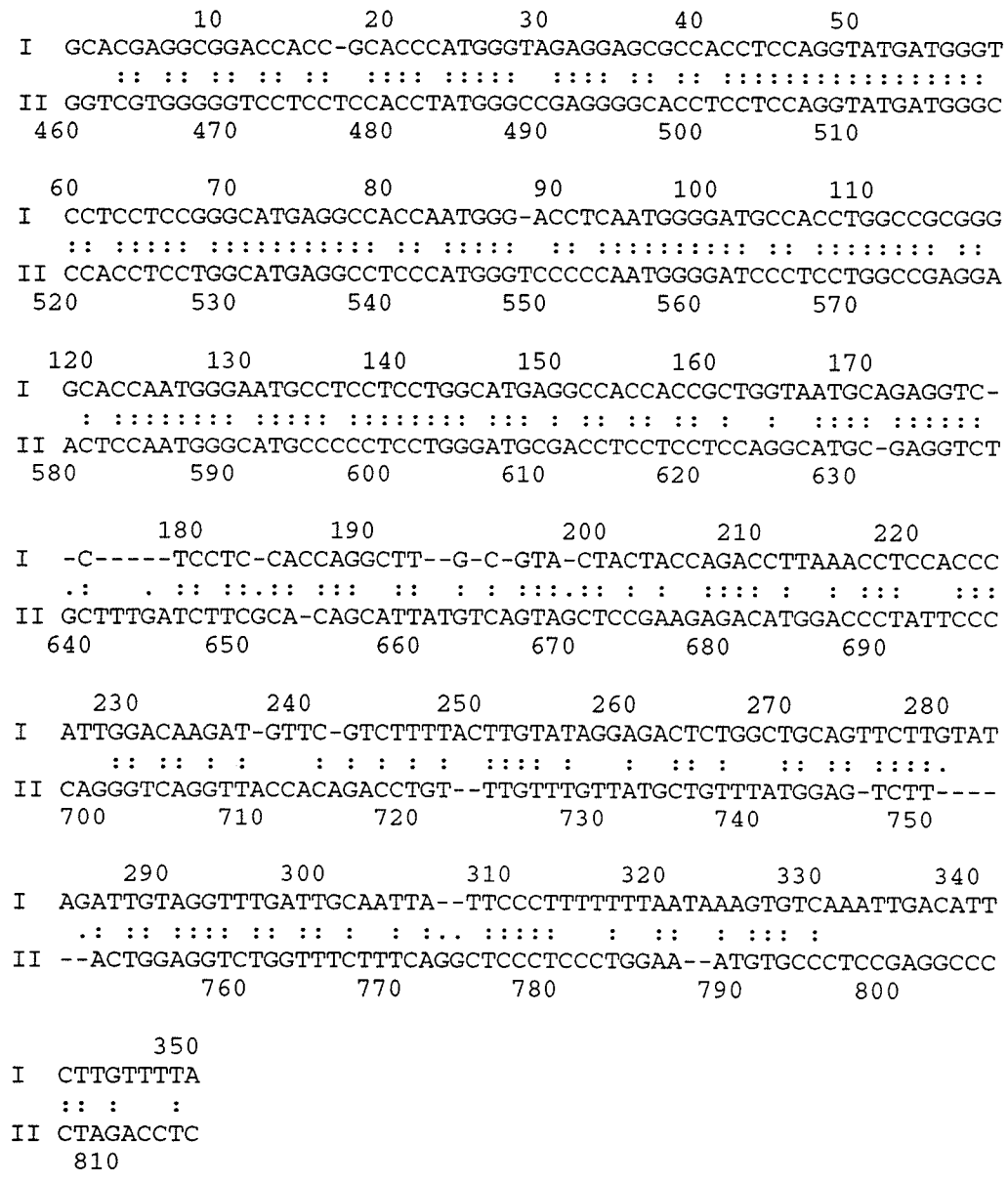
Figure 3.5: DNA sequence alignment between the presumed partial 350 bp cDNA of the cloned zebrafish*zfsnrpn* and the murine Sm B protein gene (GenBank #M58761). The global sequence aligner Blastn rated these two sequence as having 81% homology.

Alignment of zebrafish snrpn (350 NT) (I) and mouse SmB (1300 NT) (II)

```
           10         20         30         40         50
I  GCACGAGGCGGACCACC-GCACCCATGGGTAGAGGAGCGCCACCTCCAGGTATGATGGGT
      :: :: :: :: ::  :::: :::::  ::::::: :: ::::::::::::::::::
II GGTCGTGGGGGTCCTCCTCCACCTATGGGCCGAGGAGCCCCTCCTCCAGGTATGATGGGC
      1100       1110       1120       1130       1140       1150

  60         70         80          90        100        110
I  CCTCCTCCGGGCATGAGGCCACCAATGGG-ACCTCAATGGGGATGCCACCTGGCCGCGGG
   :: ::::: :::::: :::: :: :::::  :: :::::::: : :: :::::::: ::
II CCACCTCCTGGCATGCGGCCTCCCATGGGTCCCCCAATGGGGCTCCCTCCTGGCCGAGGA
      1160       1170       1180       1190       1200       1210

  120        130        140        150        160        170
I  GCACCAATGGGAATGCCTCCTCCTGGCATGAGGCCACCACCGCTGGTAATGCAGAGGTC-
    : :::::::: ::::: :::::::::::: :::: :: :: :  :  :::: ::::::
II ACTCCAATGGGCATGCCCCCTCCTGGCATGCGGCCTCCTCCTCCAGGCATGC-GAGGTCT
      1220       1230       1240       1250       1260       1270

        180        190
I  -C--T--CCTCCAC-CAGGCTTGCGTAC
   .:  :.  :: ::: :::  ::  ::
II GCTTTGATCTTCACACAGCATTATGTCA
       1280       1290       1300
```

Figure 3.6: DNA sequence alignment between the presumed partial 350 bp cDNA of the cloned zebrafish *zfsnrpn* and the rat *Snrpn*. (GenBank #M29295). The global sequence aligner Blastn rated these two sequence as having 80% homology.

Alignment of zebrafish snrpn (350 NT) (I) and rat snrpn (913 NT) (II)



Discussion:

Spliceosomes, the mRNA processing machinery, are composed of a group of small nuclear RNAs (snRNAs) termed U1, U2, U4, U5, and U6 complexed to a family of proteins known as the snRNP-associated proteins (Steitz *et al.*, 1988). The mammalian snRNP-associated proteins B, B' and N have been shown to be highly homologous. Because these proteins were found to be so similar it was once thought that they were derived from a single precursor RNA by differential processing (Schmauss and Lerner, 1990). The snRNP-associated polypeptide N gene shows tissue specificity, being expressed in the brain and heart, while the B/B' protein gene is expressed in all tissues. The tissue specific snRNP gene has been shown to be maternally imprinted in humans and mice. A deletion in the region of this gene is thought to be involved in the pathogenesis of Prader-Willi syndrome (Glenn *et al.*, 1993; Reed and Leff, 1994).

Sequence analysis of the cloned zebrafish *Snrpn* cDNA revealed very high sequence homology to the mammalian genes encoding snRNP-associated polypeptides B, B' and N. This is not surprising since the genes encoding the B and N type polypeptides are themselves highly homologous in mammals, and furthermore, the Sm B protein of *Drosophila melanogaster* is 80% homologous with murine and human forms when conservative amino acid substitutions are taken into account (Brunet *et al.*, 1993). The high sequence homology between the members of this gene family probably account for the observation of multiple bands in the cross-species hybridization of Southern blots. The presence of such a high sequence homology between the tissue-specific and ubiquitously expressed genes, along with only a partial zebrafish cDNA, makes it difficult to

determine the actual gene type that has been cloned from the zebrafish.

Northern analysis of total zebrafish RNAs enriched for brain RNAs and muscle RNAs revealed signals of equal intensity from both tissues. Apparently, the cloned zebrafish cDNA is expressed ubiquitously in all tissues. Very little or no cross-hybridization was noted between the zebrafish cDNA probe and mouse RNA even though cDNA homology is high, as was similarly noted by Schmauss *et al.* (1989). It must be concluded from this ubiquitous expression that the cloned zebrafish cDNA is most characteristic of the snRNP-associated polypeptide B form. This is unfortunate since the Sm B protein gene has not been shown to be imprinted, even though the N and B forms of this gene probably evolved from a common ancestor by gene duplication (Schmauss and Lerner, 1990). The SmB gene is of interest due to its relevance to human diseases. Patients with systemic lupus erythematosus develop autoimmune responses against epitopes of the Sm B proteins (Huntriss et al., 1993) and other connective tissue diseases (Lerner and Steitz, 1979).

It may still be possible to clone a zebrafish snRNP polypeptide N homolog by using the presently cloned polypeptide B gene. Since the snRNP-N and -B genes show very high homology between themselves (at least among mammalian forms), screening zebrafish cDNA libraries using the zebrafish Sm B gene may provide the best way of obtaining the desired clone. It may be possible, however, that the tissue specific snRNPs are limited to mammals, although the presence of the snRNP-N proteins in organisms other than mammals has not been addressed in the literature.

I have previously demonstrated the existence of a genomically imprinted transgene in the zebrafish *Danio rerio*. It seems unlikely that the observation of an imprinted transgene locus in the zebrafish is strictly an

artifact of the foreign transgene DNA. It is likely that there are endogenous genetic loci that are also subject to the epigenetic modification of genomic imprinting. As the zebrafish becomes an increasingly popular model for genetic and developmental research it may become easier to identify endogenous loci, which by their genetic behavior, suggest that they are imprinted. The snRNP protein genes maintain essential function in RNA processing and have been conserved throughout evolution. For this reason, the snRNP polypeptide N gene remains an excellent candidate for an endogenous imprinted locus in the zebrafish *Danio rerio*.

Chapter 4: The Sins of the Father - Theory on the Evolution of Genomic Imprinting

Genomic imprinting can be defined as an epigenetic, gamete-of-origin dependent modification of the genome that produces a functional difference in the expressivity of an allele. That is, the ability of an imprinted allele to be expressed is dependent on the sex of the parent contributing that particular allele, the final result being an individual that is functionally hemizygous for the imprinted locus.

Imprinting-type phenomena have been recognized in a wide range of phyla from both plant and animal kingdoms. Some examples of this gamete-of-origin dependent modification include: parental dominance in hybrid plants (Heslop-Harrison, 1990), paternal chromosome elimination in *Sciaria* (Metz,1938; Crouse et al., 1971), inactivation of the paternal genome by heterochromatization in the scale insects (coccids) (Nur, 1990), parent-of-origin specific modification of position effect variegation in Drosophila (Spofford, 1976), preferential inactivation of the paternally derived X chromosome in marsupials and rodent extraembryonic tissues (VandeBerg et al., 1987), parent-of-origin dependent switching of yeast mating types (Klar, 1987), and allelic exclusion of immunoglobulin genes (Pernis et al., 1965; Holliday, 1990). Probably the most extensively studied and well understood example of genomic imprinting is the gamete-of-origin dependent modification of transgene methylation and expression in mice (Swain et al., 1987; Reik et al., 1987; Sapienza et al.,1987) and most recently in a line of transgenic zebrafish (this report). It is not at all clear at this time whether all these phenomena are occurring via a similar molecular mechanism. That it occurs, however, in such a wide range of

organisms suggests that there is strong evolutionary conservation of this phenomenon.

A number of theories have been put forth to explain the evolution of the process known as genomic imprinting. Holliday (1990) suggests that the function of genomic imprinting is to produce functionally haploid gene sets in order for the cell's regulatory machinery to be able to 'fine tune' itself on just a single gene copy. It was further implied that imprinting would also serve to prevent possible detrimental cross-talk between maternal and paternal transcripts. Chandra and Nanjundiah (1990) put forward several postulates describing in a general way the kinds of influences that could produce imprinting although they do not address what specific influences might be involved: 1) imprinted and non-imprinted alleles of a locus may confer different phenotypes; 2) modifiers of imprinting - "imprintor genes" may have pleiotropic effects such that they are selected for reasons other than their action on the imprinted locus; and 3) imprinting could have co-evolved with other traits. There are two theories of the evolution of imprinting with some experimental support. These theories suggest that imprinting evolved to either: 1) maintain sex by preventing parthenogenesis or, 2) to maximize the paternal genetic contribution to a population by forcing the female to increase her maternal investment to the males offspring. Both of these ideas are consistent with the characteristics of genome imprinting as seen in mammals; however, they are less consistent with imprinting in fish.

It is apparent from the present study that genomic imprinting is operating in the zebrafish *Danio rerio*, and that the effects of imprinting on a locus can be monitored at the level of DNA methylation. The epigenetic modification system identified in the zebrafish is nearly identical, at all

levels studied, to the modification system found operating in mice and humans. The similarities in the imprinting process suggests evolutionary conservation of this modification system, at least among vertebrates. The presence of imprinting suggests that it is playing an important role in other organisms besides mammals. At this point it is difficult to suggest what this role might be. However, the fact that imprinting is occurring in a non-mammalian vertebrate provides us with an opportunity to re-evaluate the above mentioned theories put forth to explain the evolution of imprinting, both of which are based primarily on mammalian data.

One of the most obvious effects of genomic imprinting, as observed in mammals, is the inviability of parthenogenetic embryos. Mouse embryos that contain only two copies of either the maternal or paternal genome almost invariably do not develop to adulthood. This poses a problem since a proportion of gynogenetic fish, containing two maternal copies of the genome, do develop to reproduce. It is apparent however, from the production of parthenogenetic and gynogenetic mouse embryos that there is a range in the associated severity of developmental defects and the time of termination. Developmental abnormalities and termination have been observed as early as first cleavage, or at the blastula stage, and even as late as the 25 somite stage. The DDK mouse strain is particularly interesting because it is thought to be an imprinting mutant (Peterson and Sapienza, 1993). When male DDK mice are mated with females of another mouse inbred line their offspring develop normally. However, when female DDK are bred to BALB/c males for example, only 5% of the offspring develop normally. The fact that 5% of these ovum mutants do survive coupled with the observations of a wide range of strain-specific detrimental phenotypes associated with parthenogenetic embryos suggest that the imprint imposed

on a locus may be leaky. This is further supported by the observation of low levels of mRNA from imprinted endogenous loci in parthenogenetic and androgenetic embryonic stem cell lines (Szabo and Mann, 1994). In the zebrafish, only about 10-20% of gynogenetic embryos actually survive to maturity (Westerfield, 1990). This suggests that imprinting may still be having a detrimental effect but it is not as drastic as that observed in mammals. That may not be surprising given the difference in breeding mechanisms between the two. The zebrafish develop externally in a potentially variable environment, it might be expected that they would have to be more developmentally plastic than mammals which develop in the relatively invariant environment of the mother's womb. Given the effect of genetic and environmental factors on the methylation and, by analogy, the expression of the imprinted pUSVcat locus that I have previously demonstrated, it is clear that a variety of factors may affect the imprint and play a role in the survival of parthenogenetic embryos of the zebrafish.

McGrath and Solter (1984) noted that there was a phenotypic difference in the developing embryos of mouse andro- and gynogenotes. More specifically, androgenetic embryos showed poorly developed embryos and excessive trophectoderm (extraembryonic supportive tissue), while gynogenetic embryos had well developed embryos and poorly developed trophectoderm. These "imprinting" phenotypes of andro- and gynogenetic embryos seem to suggest the possibility of the male using imprinting to maximize maternal input to the embryo (Haig and Westoby,1989). Since fertilization and development is external in zebrafish and, therefore, the maternal contribution stops at the point of egg release this theory seems less tenable in this case. Unless fish evolved from mammals, it is unlikely that imprinting evolved to maximize maternal

investment to a male's offspring. Male/female competition at the level of maternal input into offspring, therefore, is unlikely to be the force behind the evolution of imprinting. It may be, however, that males (mammals) have later capitalized on imprinting to further their own end by selectively imprinting genes which increase maternal investment.

I wish to suggest a possible alternative theory that is based on the idea that imprinting may be a means of increasing an organism's ability to change and evolve its genetic loci relatively quickly and efficiently while at the same time maintaining a portion of the population safe from selection against the altered and possibly deleterious new alleles. Figure 4.1 illustrates the evolving scenario of a single paternally imprinted gene with only one allele. Within a sexually reproducing group of organisms that has a single paternally imprinted locus, then every individual within that population would carry one active allele (the maternally derived one) and one inactive allele (the paternally derived one). In general terms imprinting causes a locus to essentially exist in a functionally hemizygous state. Since in this basic model there is only one type of allele in the population, all individuals within the population would have the same phenotype elicited from that allele. They would all, however, have a second silent allele (not subject to natural selection) that is free to mutate differently in each individual. In fact, the DNA methylation associated with the inactivation of imprinted alleles would significantly increase the mutation rate of an imprinted allele by 5-methylcytosine deamination. Therefore, an imprinted allele can be genuinely called a "mutational hotspot", particularly vulnerable to cytosine to thymine transitional mutations (Wiebauer et al., 1993).

In the next generation there will actually be four different alleles

present (Figure 4.1). Two of these would be of maternal origin and would now be active because the female does not imprint this locus. One of those alleles was active in the mother and, therefore, was subject to natural selection and is presumably unchanged. The other allele was inactive in the mother, was not subject to selection and could potentially have been altered. The other two alleles would come from the father and both of these would now be inactive and imprinted. Moreover, only one of them was active in the father and subject to natural selection. The other allele was already inactive for one generation, remains inactive for another generation and is free to mutate even further. The allele that was inactive in the mother and had the potential to mutate is now active in 50% of her offspring. The allele is now subject to selection because the homologous allele received from the father is inactive due to imprinting. The only allele that is being expressed is the newly activated allele from the mother. This potentially altered allele can now be checked by nature and either accepted or rejected. If the alteration is bad, the organism is selected against and that particular allele disappears from the population. The population itself is not critically affected though, because only one individual had that mutation to begin with and 50% of that individual's offspring have a "good" original allele and thus retain their fitness. As shown in Figure 4.1, after two generations 25% of alleles have never been inactivated or 50% have never been inactivated or only inactivated once. This pattern will continue to the next generation with 38% of alleles which have either never been inactivated or only been inactivated for one generation. In Figure 4.1 a single bar was removed from an allele if it was returned to the selectable pool of genes. Since an allele reintroduced into the selectable pool must be accepted (due to no genetic change from

original, a silent mutation, or beneficial mutation) or eliminated from the population, it would be more appropriate to remove all bars for a newly activated allele since its past history of inactivation would be irrelevant. A newly activated allele carrying a beneficial mutation would eventually increase in frequency within a population.

Every breeding female in this example could, potentially, have mutated its silent allele in a different manner and, therefore, at each generation a large number of different changes could be scrutinized by natural selection. The imprinting system, therefore, allows some alleles never to be silenced (those that are continually passed though a female), some alleles to be silenced for a generation or two, and some alleles to be silenced for a large number of generations depending on how long they are passed through a male. In every generation some previously silent alleles are activated and become selectable, while some previously active alleles become silenced and become free to change.

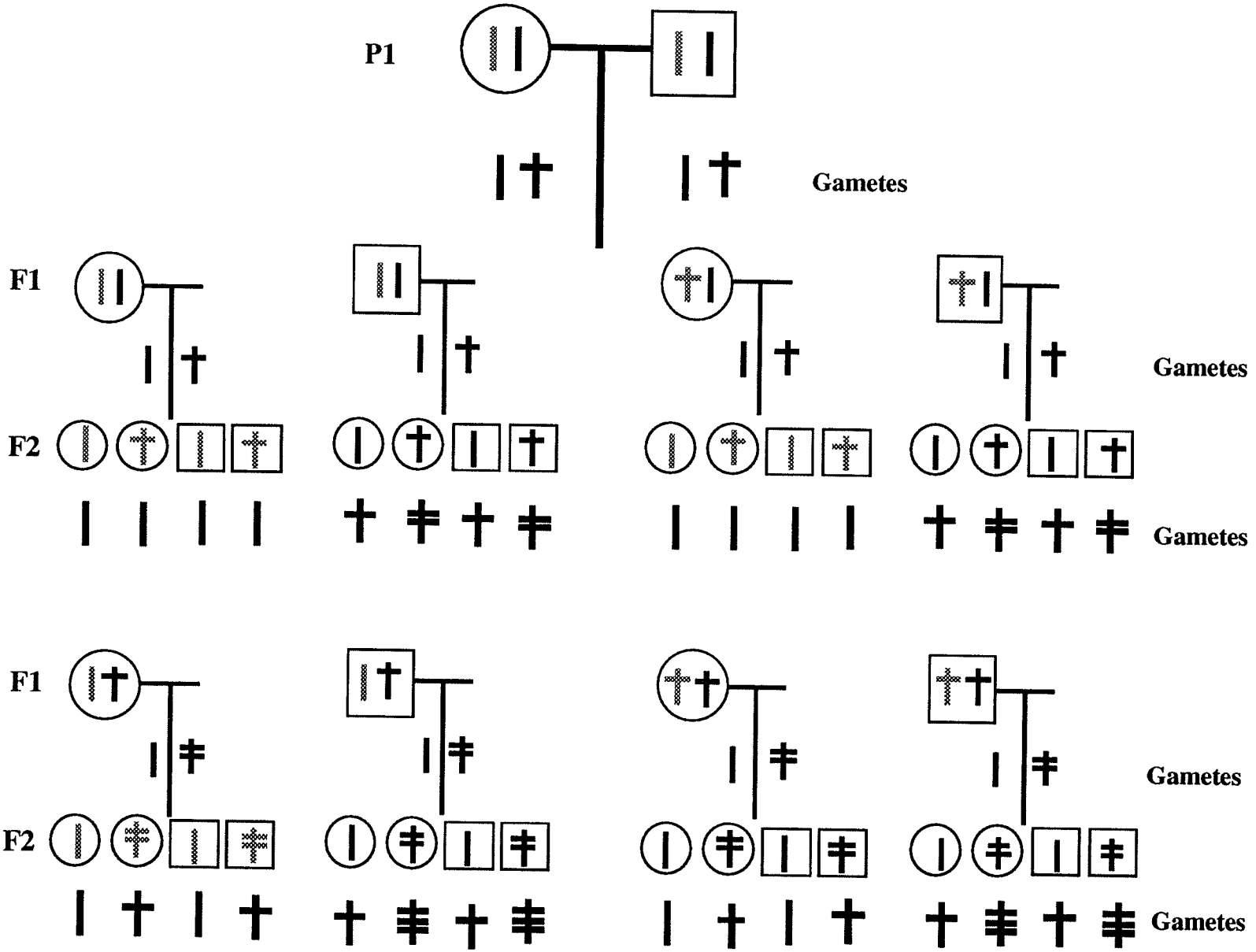
The genetic advantages noted above are similar to the advantages achieved by gene duplication. An imprinted locus has an added advantage, however, in that there is no change in the physical length of the DNA molecule eliminating possible difficulties encountered during crossing-over. Another advantage of the gamete-of-origin dependent silencing or imprinting of the alleles is that there is no selection against one sex or the other. All individuals both male and female have one paternally derived silent copy and one maternally derived active copy which may or may not have been previously silenced. Equal numbers of both sexes carry the newly activated allele that may be selected for or against. The imprinting system continuously allows alleles at a locus to move back and forth between an unselectable and selectable pool of alleles. Some alleles will

stay in the unselectable pool for a long time and potentially change a great deal, while

some will stay only a short time with little or no functional change to the allele. Another characteristic imposed by this system is that altered alleles that have been reintroduced into an active, selectable pool and accepted by natural selection can move back into the non-selectable pool at a later date to attempt even more changes. The obvious effect of imprinting on a locus would be to greatly increase the possible genetic variation in a population compared to loci operating under normal Mendelian inheritance. Genomic imprinting can provide a significant evolutionary force for genetic change. It is thought that genes can switch their imprint (from maternal to paternal, and visa versa) or can become imprinted spontaneously (or de-imprinted) (Varmuza, 1993). If a large number of developmentally significant genes become imprinted in a particular population it would provide a marvelous genetic environment for punctuated equilibrium by providing a mechanism for accelerated genetic change. In fact, punctuated equilibrium has been thought to have occurred during the geologically rapid radiation of vertebrate taxa. Further, imprinting would also facilitate a much more efficient coevolution of genes that are functionally linked. In this case, the two genes in question could mutate and evolve independently but changes in one gene would periodically be tested for compatibility with the other gene. This aspect of imprinting would greatly increase the functional coevolution of gene groups such as gene products and their receptors. It seems clear that genomic imprinting can act to increase the variability of a locus and to effectively increase that rate of evolution of an imprinted locus. The increasing list of endogenous imprinted loci in mice and humans and their characterization, and the future identification of

imprinted genes in other organisms may aid in determining the result of this proposed effect. I have presently suggested some possible results of silencing an imprinted locus. Testing these hypotheses would be extremely difficult since the iterations of the generation's testing and acceptance of a functionally mutated locus would have to occur over a considerable length of evolutionary time.

Figure 4.1: The evolution of a paternally imprinted locus depicting in the gametes the number of generations a particular imprinted allele has been inactive (denoted by the addition of a cross-bar). Each pass of an allele through a male causes that allele to be inactivated (black bars), not subject to natural selection and free to mutate. The allele will only be later selected for when it is passed through a female (grey bars). After several generations it is apparent that a large proportion of the alleles passed in the gametes will have never been inactivated or inactivated for only one generation (and probably not altered). *Note: a single bar indicating a generation of silencing was removed from an allele reintroduced into the selectable gene pool.*


P1
Gametes
F1
F2
Gametes
Gametes
F1
F2
Gametes
Gametes

Literature Cited

Allen, N.D., Norris, M.L., and Surani, M.A.H. (1990). Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61**: 853-861.

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

Anton, N.K. and D. Vapnek (1979). Nucleotide sequence analysis of the chloramphenical resistance transposon Tn9. *Nature* **282**: 864-869.

Barlow, D.P., Ströger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 recepter is imprinted and closely linked to the *Tme* locus. *Nature* **349**: 84-87.

Bartolomei, M.S., Zemel, S., and S.M. Tilghman (1991). Parental imprinting of the mouse H19 gene. *Nature* **351**: 153-155.

Braunstein, M., Rose, A.B., Holmes, S.G., Allis, C.D., and J.R. Broach (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes and Development* **7**: 592-604.

Brown, S.W., and W.A. Nelson-Rees (1961). Radiation analysis of a lecanoid genetic system. *Genetics* **46**: 983-1007.

Brunet, C., Quan, T. and J. Craft (1993). Comparison of the *Drosophila melanogaster*, human and murine Sm B cDNAs: evolutionary conservation. *Gene* **124**: 269-273.

Cattanach, B.M., and Beechey, C.V. (1990). Autosomal and X-chromosome imprinting. *Development Supplement*, 63-72.

Cattanach, B.M. (1986). Parental origin effect in mice. *J. Embryol. exp. Morph.* **97 supplement**, 137-150.

Ceder, H. (1988). DNA methylation and gene activity. *Cell* **53**: 3-4.

Chaillet, J.R., Vogt, T.F., Beier, D.R., and Leder, P. (1991a). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell* **66**, 77-83.

Chaillet, J.R., Knoll, J.H.M., Horsthemke, B., and M. Lalande (1991). The syntenic relationship between the critical deletion region for the Prader-Willi/Angelman syndromes and proximal mouse chromosome 7. *Genomics* **11**: 773-776.

Chatkupt, S., Lucek, P.R., Koenigsberger, M.R., and Johnson, W.G. (1992). Parental sex effect in spina bifida: A role for genomic imprinting? *Am. J. Med. Genet.* **44**: 508-512.

Cox, R.A. (1968). The use of guanidinium chloride in the isolation of nucleic acids. In Grossman, L., Moldave, K. eds. *Methods in Enzymology - Vol. 12b.* Academic Press, New York, New York, pp. 120-129.

Crouse, H.V. (1960). The controlling element in sex chromosomes behaviour in *Sciara. Genetics* **45**: 1429-1443.

DeChiara, T.M., Efstratiadis, A., and Robertson, E.J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**: 78-80.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**: 849-859.

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., and M. Surani (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* **362**: 751-755.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and M. Ysebaert (1978). Complete nucleotide sequence of SV40 DNA. *Nature* **273**: 113-120.

Fisher, R.A. (1928). The possible modification of the response of the recurrent mutations. *American Naturalist* **62**: 115-126.

Forejt, J., and S. Gregorova (1992). Genetic analysis of genomic imprinting: An *Imprintor-1* gene control inactivation of the paternal copy of the mouse *Tme* locus. *Cell* **70**: 443-450.

Glenn, C.C., Porter, K.A., Jong, M.T., Nicholls, R.D. and D.J. Driscoll (1993). Functional imprinting and epigenetic modification of the human SNRPN gene. *Human Molecular Genetics* **2**: 2001-2005.

Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I., and B.H. Howard (1982). The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc. Natl. Acad. Sci. USA* **79**: 6777-6781.

Gowen, J.W., and E.H. Gay (1934). Chromosome constitution and behavior in ever-sporting and mottling in *Drosophila melanogaster. Genetics* **19**: 189-208.

Hadchouel, M., Farza, H., Simon, D., Tiollais, P., and C. Pourcel (1987). Maternal inhibition of hepatitis B surface antigen gene expression in trangenic mice correlates with *de novo* methylation. *Nature* **329**: 454-456.

Hall, J.G. (1990). Genomic imprinting: review and relevance to human diseases. *Am. J. Hum. Genet.* **46**: 857-873.

Henikoff, S. (1990). Position-effect variegation after 60 years. *Trends in Genetics* **6**: 422-426.

Huntriss, J.D., Latchman, D.S. and D.G. Williams (1993). Lupus autoantibodies discriminate between the highly homologous Sm polypeptides B/B' and SmN by binding an epitope restricted to B/B'. *Clinical & Experimental Immunology* **92**: 263-267.

James, T.C., and S.C.R. Elgin (1986). Identification of a non-histone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell Biol.* **6**: 3862-3872.

Karlsson, S., Humphries, R.K., Gluzman, Y., and A.W. Nienhuis (1985). Transfer of genes into hematopoietic cells using recombinant DNA viruses. *Proc. Natl. Acad. Sci. USA* **82:** 158-162.

Klar, A.J.S. (1987). Differentiated parental DNA strands confer developmental asymmetry on daughter cells in fission yeast. *Nature* **326**: 466-470.

Laird, C. (1990). Proposed genetics basis of Huntington's disease. *Trends in Genetics* **6**: 242-247.

Ledbetter, D.H. *et al.* (1981). Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. *New Engl. J. Med.* **304**: 325-329.

Lee, S., and D.S. Gross (1993). Conditional silencing: the *HMRE* mating-type silencer exerts a rapidly reversible position effect on the yeast *HSP82* heat shock protein. *Molecular and Cellular Biology* **13**: 727-738.

Leff, S.E., Brannan, C.I., Reed, M.L., Ozcelik, T., Francke, U., Copeland, N., and N.A. Jenkins (1992). Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nature Genetics* **2**: 259-264.

Lerner, M.R. and J.A. Steitz (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U.S.A.* **76**: 5495-5499.

LeRoith, D. (1991). Insulin-like Growth Factors: Molecular and Cellular Aspects. CRC Press, Boca Raton, Florida.

Levy-Wilson, B., Watson, D.C., and G.H. Dixon (1979). Multi-acetylated forms of H4 are found in a putative transcriptionally competent chromatin fraction from trout testes. *Nucleic Acid Res.* **6**: 259-274.

Li, S., Klein, E.S., Russo, A.F., Simmons, D.M., and M.G. Rosenfeld (1989). Isolation of cDNA clones encoding encoding small nuclear ribonucleoparticle-associated proteins with different tissue specificities. *Proc. Natl. Acad. Sci. USA* **86**: 9778-9782.

Locke, J., Kotarski, M.A., and K.D. Tartoff (1988). Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics* **120**: 181-198.

Luhrmann, R., Kastner, B., and M. Bach (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta* **1087**: 265-292.

McAllister, G., Amara, S.G., and M.R. Lerner (1988). Tissue-specific expression and cDNA cloning of small nuclear ribonucleoprotein-associated polypeptide N. *Proc. Natl. Acad. Sci. USA* **85**: 5296-5300.

McGowan, R., Campbell, R., Peterson, A., and C. Sapienza (1989). Cellular mosaicism in the methylation and expression of hemizygous loci in the mouse. *Genes and Development* **3**: 1669-1676.

McGowan, R., Tran, T.H., Paquette, J. and Sapienza, C. (1990). Transgene methylation imprints are established post-fertilization. In: *Progress in Nuclei Acid Research and Molecular Biology*, Volume 36, Academic Press, Inc., pp. 367-379, San Diego, California.

McGrath, J., and D. Solter (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**: 179-183.

Miller, O.J., Schedl, W., Allen, J., and B.F. Erlanger (1974). 5'-Methyl-cytosine localized in mammalian constitutive heterochromatin. *Nature* **251**: 636-637.

Monk, M. (1990). Changes in DNA methylation during mouse embryonic development in relation to X-chromosome activity and imprinting. *Phil. Trans. R. Soc. Lond. B* **326**: 299-312.

Monk, M., Boubelik, M. and S. Lehnert (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**: 371-382.

Mottus, R., Reeves, R., and T.A. Grigliatti (1980). Butyrate suppression of position-effect variegation in *Drosophila melanogaster*. *Molec. gen. Genet.* **178**: 465-469.

Nanda, I., Schartl, M., Epplen, J.T., Feichtinger, W. and M. Schmid (1993). Primitive sex chromosomes in poeciliid fishes harbor simple repetitive DNA sequences. *J. Exp. Zool.* **265**: 301-308.

Nicholls, R.D. *et al.* (1989). Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *Nature* **342**: 281-285.

Nicholls, R.D., Gottleib, W., Avidano, K., Williams, C.A., and D. Driscoll (1991). Mouse chromosome mapping of clones from the PWS/AS genetic region. *Mouse Genome* **89**: 254.

Ogawa, O., Eccles, M.R., Szeto, J., McNoe, Y.K., Maw, M.A., Smith, P.J., and A.E. Reeve (1993). Relaxation of insulin-like growth factor II gene imprinting Wilms' tumour. *Nature* **362**: 749-751.

Ozcelik, T., Leff, S., Robinson, W., Donlon, T., Lalande, M., Sanjines, E., Schinzel, A., and U. Francke (1992). Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. *Nature Genetics* **2**: 265-269.

Pal, N., Wadey, R.B., Buckle, B., Yeomans, E., Pritchard, J., and J.K. Cowell (1990). Preferential loss of maternal alleles in sporadic Wilms' tumour. *Oncogene* **5**: 1665-1668.

Paro, R., and D.S. Hogness (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila. Proc. Natl. Acad. Sci. USA* **88**: 263-267.

Peter, E., Candido, M., Reeves, R., and J.R. Davie (1978). Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* **14**: 105-113.

Poirier, F., Chan, C.T., Timmons, P.M., Robertson, E.J., Evans, M.J., and P.W. Rigby (1991). The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development* **113**: 1105-1114.

Rainer, S., Johnson, L.A., Dobry, C.J., Ping, A.J., Grundy, P.E., and A.P. Feinberg (1993). Relaxation of imprinted genes in human cancer. *Nature* **362**: 747-749.

Razin, A. and A.D. Riggs (1980). DNA methylation and gene function. *Science* **210**: 604-610.

Reed, M.L. and S.E. Leff (1994). Maternal imprinting of human SNRPN, a gene deleted in Prader-Willi syndrome. *Nature Genetics* **6**: 163-167.

Reik, W., Howlett, S.K., and M.A. Surani (1990). Imprinting by DNA methylation: from transgenes to endogenous gene sequences. *Development Suppl.*, 99-106.

Reik, W., Collick, A., Norris, M.L., Barton, S.C., and Surani, M.A.H. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328**: 248-251.

Reuter, G. and P. Spierer (1992). Position effect variegation and chromatin proteins. *BioEssays* **14(9)**: 605-612.

Sambrook, J., Fritsch, E.F., and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual (2nd Ed.)*, Cold Spring Harbor Laboratory Press, New York, New York.

Sapienza, C., Peterson, A., Rossant, J. and Balling, R. (1987). Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**, 251-254.

Sapienza, C., Peterson, A., Rossant, J. and R. Balling (1987). Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**: 251-254.

Sapienza, C., Paquette, J., Tran, T.H., and A. Peterson (1989). Epigenetic and genetic factors affect transgene methylation imprinting. *Development* **107**: 165-168.

Sapienza, C. (1990a). Sex-linked dosage-sensitive modifiers as imprinting genes. *Development Supplement*, 107-113.

Sapienza, C. (1990b). Genome imprinting, cellular mosaicism and carcinogenesis. *Molecular Carcinogenesis* **3**: 118-121.

Sanford, J.P., Clark, H.J., Chapman, V.M. and J. Rossant (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development* **1**: 1039-1046.

Sasaki, H., Allen, N.A., and M.A. Surani (1993). DNA methylation and genomic imprinting in mammals. *In:* DNA Methylation: Molecular Biology and Biological Significance (J.P. Jost and H.P. Saluz, eds.). Birkhauser Verlag, Basel, Switzerland. pp. 469-486.

Schmauss, C., McAllister, G., Ohosone, Y., Hardin, J.A. and M.R. Lerner (1989). A comparison of snRNP-associated Sm-autoantigens: human N, rat N and human B/B'. *Nucleic Acids Research* **17**: 1733-1743.

Schmauss, C. and M.R. Lerner (1990). The closely related small nuclear ribonucleoprotein polypeptide N and B/B' are distinguished by antibodies as well as by differences in their mRNAs and gene structures. *The Journal of Biological Chemistry* **265**: 10733-10739.

Schmid, M., Hoff, T., and D. Grunert (1984). 5'-Azacytidine undercondensation in human chromosomes. *Hum. Genet.* **67**: 257-263.

Schroeder, W.T., Chao, L-Y., Dao, D.D., Strong, L.C., Pathak, S., Riccardi, V., Lewis, W.H., and G.F. Saunders (1987). Nonrandom loss of maternal chromosome 11 alleles in Wilms' tumors. *Am. J. Hum. Genet.* **40**: 413-420.

Scrable, H., Cavenee, W., Ghavimi, F., Lovell, M., Morgan, K., and C. Sapienza (1989). A model for embryonal rhabdomyosarcoma tumorigenesis that involves genomic imprinting. *Proc. Natl. Acad. Sci. USA* **86**: 7480-7484.

Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., Paro, R., James, T.C., and S.J. Gaunt (1991). A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants. *Nucleic Acid Research* **19**: 789-794.

Spofford, J.B. (1976). Position effect variagation in *Drosophila*. *In:* The Genetics and Biology of *Drosophila* (Ashburner, M., Novitski, E., eds). Academic Press, New York, New York. pp. 955-1018.

Spofford, J.B. (1961). Parental control of position-effect variegation. II. Effect of parent contributing white mottled rearrangement in *Drosophila melanogaster. Genetics* **46**: 1151-1167.

Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Kramer, A., Frendewey, D. and W. Keller (1988). *Structure and Function of Major and Minor Small Nuclear Ribonuclear Ribonuclear Ribonucleoprotein Particles* (Birnstiel, M.L., ed.). Springer Verlag, New York, New York.

Stöger, R., Kubicka, P., Liu, C.G., Kafri, T., Razin, A., Cedar, A. and Barlow, D.P. (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61-71.

Stuart, G.W., Vielkind, J.R., McMurry, J.V., and M. Westerfield (1990). Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression. *Development* **109**: 577-584.

Surani, M.A.H., Barton, S.C., and M.L. Norris (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**: 548-550.

Surani, M.A.H., Barton, S.C., and M.L. Norris (1986). Nuclear transplantation in the mouse: Heritable differences between parental genomes after activation of the embryonic genome. *Cell* **45**: 127-136.

Swain, J.L., Stewart, T.A., and P. Leder (1987). Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. *Cell* **50**: 719-727.

Szabo, P., and J.R. Mann (1994). Expression and methylation of imprinted genes during in vitro differentiation of mouse parthenogenetic and androgenetic embryonic stem cell lines. *Development* **120**: 1651-1660.

Tartoff, K.D., and M. Bremer (1990). Mechanisms for the construction and developmental control of heterochromatin formation and imprinted chromosome domains. *Development Suppl.* 35-45.

Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**: 5201-5205.

Ueda, T., Yamazaki, K., Suzuki, R., Fujimoto, H., Sasaki, H., Sakaki, Y., and T. Higashinakagawa (1992). Parental methylation patterns of a transgenic locus in adult somatic tissues are imprinted during gametogenesis. *Development* **116**: 831-839.

VandeBerg, T.L., Robinson, E.S., Samallow, P.B., and P.G. Johnston (1987). X-linked gene expression and X-chromosome inactivation: Marsupials, mouse and man compared. In: *Isozymes: Current Topics in Biological and Medical Redsearch* (ed. C.L. Markert), Alan R. Liss, New York, New York.

Varmuza, S. (1993). Gametic imprinting as a speciation mechanism in mammals. *Journal of Theoretical Biology* **164**: 1-13.

Wagstaff, J., Chaillet, J.R., and M. Lalande (1991). The $GABA_A$ receptor B3 subunit gene: Characterization of a human cDNA from chromosome 15q11q13 and mapping to a region of conserved synteny on the mouse chromosome 7. *Genomics* **11**: 1071-1078.

Westerfield, M. (1989). *The Zebrafish Book.* University of Oregon Press, Eugene, Oregon.

Whitman, M., and Melton, D.A. (1989). Growth factors in early embryogenesis. *Annu. Rev. Cell Biol.* **5**: 93-117.

Wiebauer, K., Neddermann, P., Hughes, M. and J. Jiricny (1993). The repair of 5-methylcytosine deamination damage. *In:* DNA Methylation: Molecular Biology and Biological Significance (J.P. Jost and H.P. Saluz, eds.). Birkhauser Verlag, Basal, Switzerland. pp. 510-522.

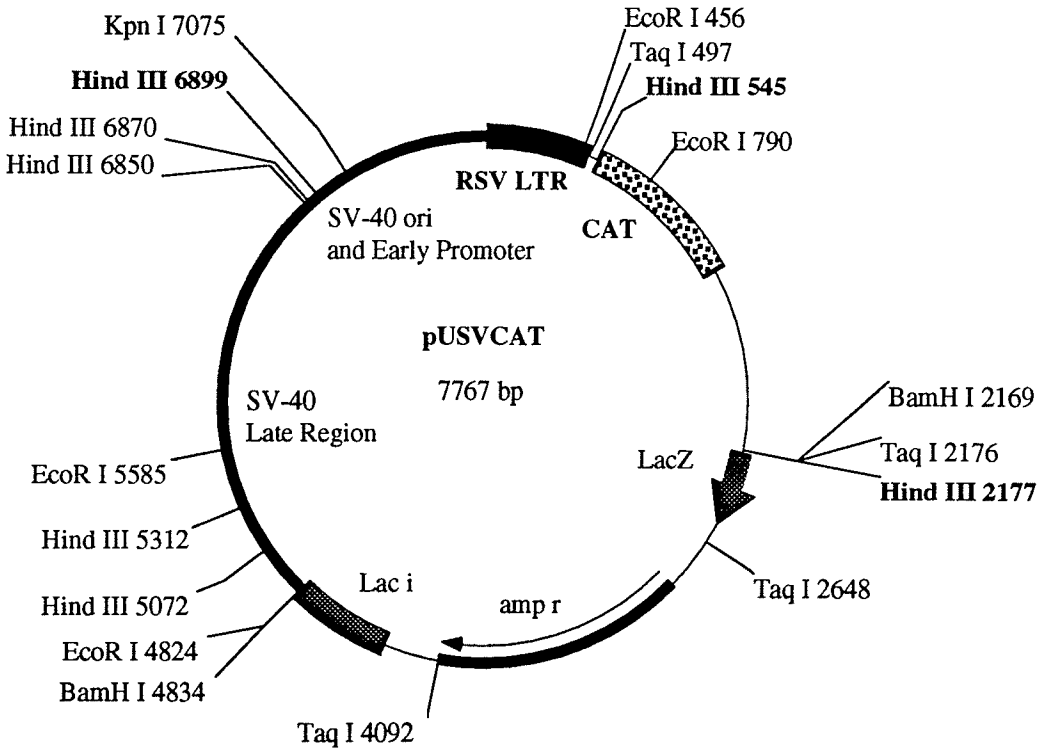
Wreggett, K.A., Hill, F., James, P.S., Hutchings, A., Butcher, G.W., and P.B. Singh (1994). A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenet. Cell. Genet.* **66**: 99-103.

Yamamoto, T., Jay, G., and I. Pastan (1980). Unusual features in the nucleotide sequence of a cDNA clone derived from the common region of avian sarcoma virus messenger RNA. *Proc. Natl. Acad. Sci. USA* **77**: 176-180.

Zemel, S., Bartolomei, M., and S. Tilghman (1992). Physical linkage of two mammalian imprinted genes, H19 and insulin-like growth factor 2. *Nature Genet.* **2**: 61-65.

Appendix 1: Plasmid Map of pUSVCAT

The plasmid map of pUSVCAT depicted below was created using MacPlasmap Version 1.82 software (Jingdong Liu, Biology Department, University of Utah). The transgenic zebrafish used in this study contained this 7831 bp plasmid. The probes used in the Southern hybridizations experiments of transgenic zebrafish were: 1) a 1.4 Kb Hind III fragment containing the Rous Sarcoma Virus LTR (long terminal repeat) and a portion of the SV-40 early promoter (filled region) and 2) a 1.6 Kb Hind III fragment containing the reporter gene for chloroamphenicol acetyltransferase (speckled region). The information required to construct the following plasmid map, restriction endonuclease site list, and the complete DNA base pair sequence was taken from Anton and Vapnek (1979), Fiers*et al.* (1978), Gorman*et al.* (1982), Karlsson*et al.* (1985), Yamamoto *et al.* (1980).



Appendix 2: Restriction Endonuclease Sites in the Plasmid pUSVCAT

This list of popular restriction endonuclease cutting sites in the plasmid pUSVCAT was computer generated using the computer software DNAid+ (F. Dardel and P. Bensoussan, Laboratoire de Biochimie, Ecole Polytechnique, 91128 Palaiseau cedex, France) by supplying the complete DNA code (Appendix 3). The numbers indicated as restriction cutting sites correspond to the same base pair numbering system shown in the circular plasmid map (Appendix 1) and the DNA code (Appendix 3). A list of restriction endonucleases that have no cutting sites in pUSVCAT is shown at the end of Appendix 2.

Aat II GACGTC
1 site
4363

Acc I GTMKAC
2 sites
2176 5740

Acc II TCCGGA
1 site
786

Afl II CTTAAG
1 site
5668

Afl III ACRYGT
2 sites
169 2548

Alu I **AGCT**
52 sites

4 20 75 495 547 562 571 690 819 1391
1720 1874 2054 2191 2213 2308 2372 2490 2716 2806
2852 3109 3630 3730 3793 4472 4491 4730 4847 5091
5098 5384 5560 5653 5661 5736 5789 5827 5876 6020
6273 6323 6566 6578 6731 6758 6935 6964 7097 7369
7423 7469

Alu N I **CAGNNNCTG**
2 sites

2964 6734

Aos II **GRCGYC**
3 sites

3978 4360 4664

Apa I **GGGCCC**
1 site

5113

Apa L I **GTGCAC**
4 sites

537 2862 4108 4605

Asp 718 **GGTACC**
1 site

7071

Ava II **GGWCC**
7 sites

3578 3800 5354 6349 6779 6810 7474

Ava I **CYCGRG**
1 site
4829

Avr II **CTTAGG**
2 sites
6289 7405

Bal I **TGGCCA**
1 site
1057

Bam H I **GGATCC**
2 sites
2169 4834

Ban I **GGYRCC**
5 sites
1256 2292 3389 4668 7071

Ban II **GRGCYC**
2 sites
5113 6595

Bbe I **GGCGCC**
1 site
4667

Bcn I **CCSGG**
6 sites
2629 3976 4477 4512 4831 4832

Bgl I **GCCNNNNNGGC**
2 sites
3561 4679

Bsp 1286 **GDGCHC**
9 sites
541 1261 2866 4027 4112 4609 5113 6083 6595

Bsp H I **TCATGA**
4 sites
3268 4276 4381 5455

Bsp M II **TCCGGA**
1 site
786

Bsp N I **CCWGG**
18 sites
967 1023 1271 2288 2576 2697 2710 4783 4841 5834 5961
6405 6459 7011 7135 7190 7502 7702

Bst U I **GCGC**
21 sites
58 139 2330 2395 2423 2456 2726 2793 2893 3067 3176
3569 3662 3999 4331 4431 4534 4664 6536 7026

Cfr 101 **RCCGGY**
2 sites
3521 7022

Cfr I **YGGCCR**
4 sites
1055 2387 3829 4816

Dde I **CTNAG**
27 sites
49 121 177 241 543 563 1011 1394 1533 1645 2823
3232 3398 3938 4364 4599 4993 5511 5657 5727 5932 6031
6646 7079 7365 7675 7735

Dpn I **GATC**
21 sites
400 1324 1934 2171 3116 3191 3202 3210 3288 3300 3405
3746 3764 3810 4068 4085 4121 4706 4836 5232 6496

Dra I **TTTAAA**
12 sites
710 1049 1387 1977 3307 3326 4018 5006 5571 5712 7447
7537

Eco 47 I **GGWCC**
7 sites
3578 3800 5354 6349 6779 6810 7474

Eco 47 III **AGCGCT**
1 site
6537

Eco 0109 **RGGNCCY**
3 sites
4417 5109 6780

Eco R I **GAATTC**
4 sites
456 790 4824 5585

Eco R V **GATATC**
1 site
6601

Esp I **GCTNAGC**
1 site
5657

Fnu 4 H I **GCNGC**
43 sites
21 142 1132 1721 2052 2184 2373 2454 2472 2475 2593
2748 2891 2956 2959 3165 3493 3832 3859 3954 4183 4470
4579 4683 4756 4908 5382 5654 5828 6567 6585 6636 6762
6705 6708 6720 6750 6753 6756 6801 6927 7029 7467

Fnu D II **CGCG**
14 sites
60 149 171 226 2395 2397 2595 3176 3506 3999 4331
4431 4433 4536

Fsp I **TGCGCA**
2 sites
3663 4686

Hae II **RGCGCY**
4 sites
2426 2796 4667 6539

Hae III **GGCC**
33 sites
160 703 748 970 1057 2389 2563 2574 2592 3026 3484
3564 3625 3831 4418 4716 4818 5111 5863 5908 6135 6435
6614 6628 6661 6710 7009 7037 7345 7351 7361 7404 7733

Hgi A I **GWGCWC**
5 sites
541 2866 4027 4112 4609

Hha I **GCGC**
21 sites
60 141 2332 2397 2428 2458 2728 2795 2895 3069 3178
3571 3664 4001 4333 4433 4536 4666 4687 6538 7028

Hinc II **GTYRAC**
7 sites
2038 2177 5072 5312 6850 6870 6899

Hind III **AAGCTT**
6 sites
545 2177 5072 5312 6850 6870 6899

Hin f I **GANTC**
14 sites
113 234 1140 1440 1464 1573 1656 2383 2448 2523 2919
3436 5629 7458

Hin P I **GCGC**
21 sites
58 139 2330 2395 2423 2456 2726 2793 2893 3067 3176
3569 3662 3999 4331 4431 4534 4664 6536 7026

Hpa I **GTTAAC**
3 sites
2038 6850 6870

Hpa II **CCGG**
16 sites
745 787 915 2266 2755 2902 2928 3118 3522 3556 3733
3975 4476 4510 4830 7023

Kpn I **GGTACC**
1 site
7075

Mae I **CTAG**
15 sites
186 492 1550 1928 2124 3043 3296 3631 5209 5590 5759
5910 6290 7406 7482

Mae II **ACGT**
7 sites
877 1052 3251 3667 4040 4360 4802

Mae III **GTNAC**
26 sites
320 847 952 1353 1758 2063 2904 2967 3083 3366 3697
3755 3908 4096 4484 4775 4795 4943 5331 5496 5523 5890
6034 6501 6547 7015

Mbo I **GATC**
21 sites
398 1322 1932 2169 3114 3189 3200 3208 3286 3298 3403
3744 3762 3808 4066 4083 4119 4704 4834 5230 6494

Mlu I **ACGCGT**
1 site
169

Msp 1 **CCGG**
16 sites
745 787 915 2266 2755 2902 2928 3118 3522 3556 3733
3975 4476 4510 4830 7023

Mst I **TGCGCA**
2 sites
3663 4686

Mva I **CCNGG**
24 sites
967 1023 1271 2288 2576 2697 2710 2928 3975 4476 4511
4783 4830 4831 4841 5834 5961 6465 7011 7135 7190 7502
7702

Nac I **GCCGGC**
1 site
7024

Nar I GGCGCC
1site
4664

Nci I **CCSGG**
6 sites
2928 3975 4476 4511 4830 4831

Nco I **CCATGG**
4 sites
1091 6807 7032 7311

Nde I **CATATG**
1 site
4612

Nla III **CATG**
29 sites
110 320 342 373 439 865 1095 1156 1177 1663 2165
2206 2552 3272 3763 3773 3851 3887 3892 4280 4385 4469
5459 6811 6988 7036 7169 7224 7315

Nla IV **GGNNCC**
27 sites
1258 1337 2171 2294 2580 2619 3391 3485 3526 3624 3737
4327 4665 4836 5111 5198 5238 5394 5808 5862 5965 6352
6592 6782 7073 7139 7194

Nru I **TCGCGA**
1 site
149

Nsi I **ATGCAT**
2 sites
7171 7226

Nsp 75241 **RCATGY**
6 sites
342 373 2552 4469 7169 7224

Nsp B II **CMGCKG**
13 sites
20 690 819 1134 2372 2890 3135 4076 4542 4736 5653
6707 7097

Oxa N I **CCTNAGG**
1 site
241

Pfl M I **CCANNNNNTGG**
3 sites
1022 1474 6361

Ppu M I **RGGWCCY**
1 site
6780

Pss I **RGGWCCY**
3 sites
4420 5112 6783

Pst I **CTGCAG**
3 sites
13 2185 5383

Pvu I **CGATCG**
3 sites
401 3811 4707

Pvu II **CAGCTG**
7 sites
20 690 819 2372 4736 5653 7097

Rsa I **GTAC**
13 sites
155 223 396 669 1207 1866 3921 4597 5903 6400 7073
7608 7719

Sal I GTCGAC
1 site
2175

Sau 96A GGNCC
17 sites
158 3483 3562 3579 3623 3801 4417 4714 5109 5110 5355
5861 6350 6780 6811 7350

Sca I AGTAGT
2 sites
1207 3921

Scr F I CCNGG
24 sites
967 1023 1271 2288 2576 2697 2710 2928 3975 4476 4511
4783 4830 4831 4841 5834 5961 6405 6459 7011 7135 7190
7502 7702

Sdn I GDGCHC
9 sites
541 1261 2866 4027 4112 4609 5113 6083 6595

Sln I GGWCC
7 sites
3579 3801 5355 6350 6780 6811 7475

Sma I CCCGGG
1 site
4831

Sph I GCATGC
3 sites
373 7169 7224

Ssp I **AATATT**
7 sites
1102 1753 1906 4245 5546 6066 7511

Stu I **AGGCCT**
7 sites
5908 6135 6614 6628 6661 7009 7404

Sty I **CCWGG**
7 sites
1091 1622 6289 6807 7032 7311 7405

Taq I **TCGA**
4 sites
497 2176 2648 4092

Xho II **RGATCY**
9 sites
1322 2169 3189 3200 3286 3298 4066 4083 4834

Xma I **CCCGGG**
1 site
4829

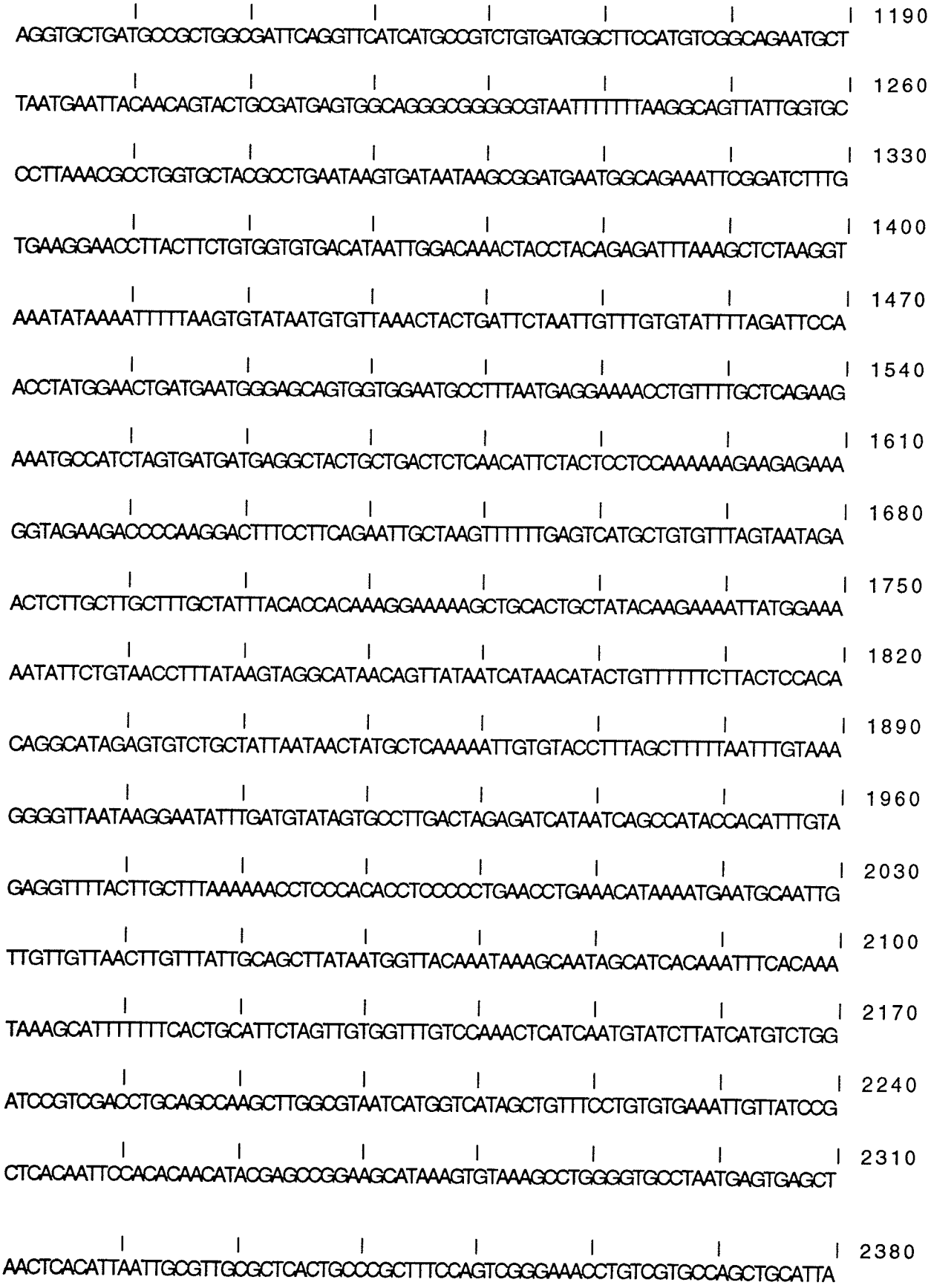
Xmn I **GAANNNNTTC**
1 site
4040

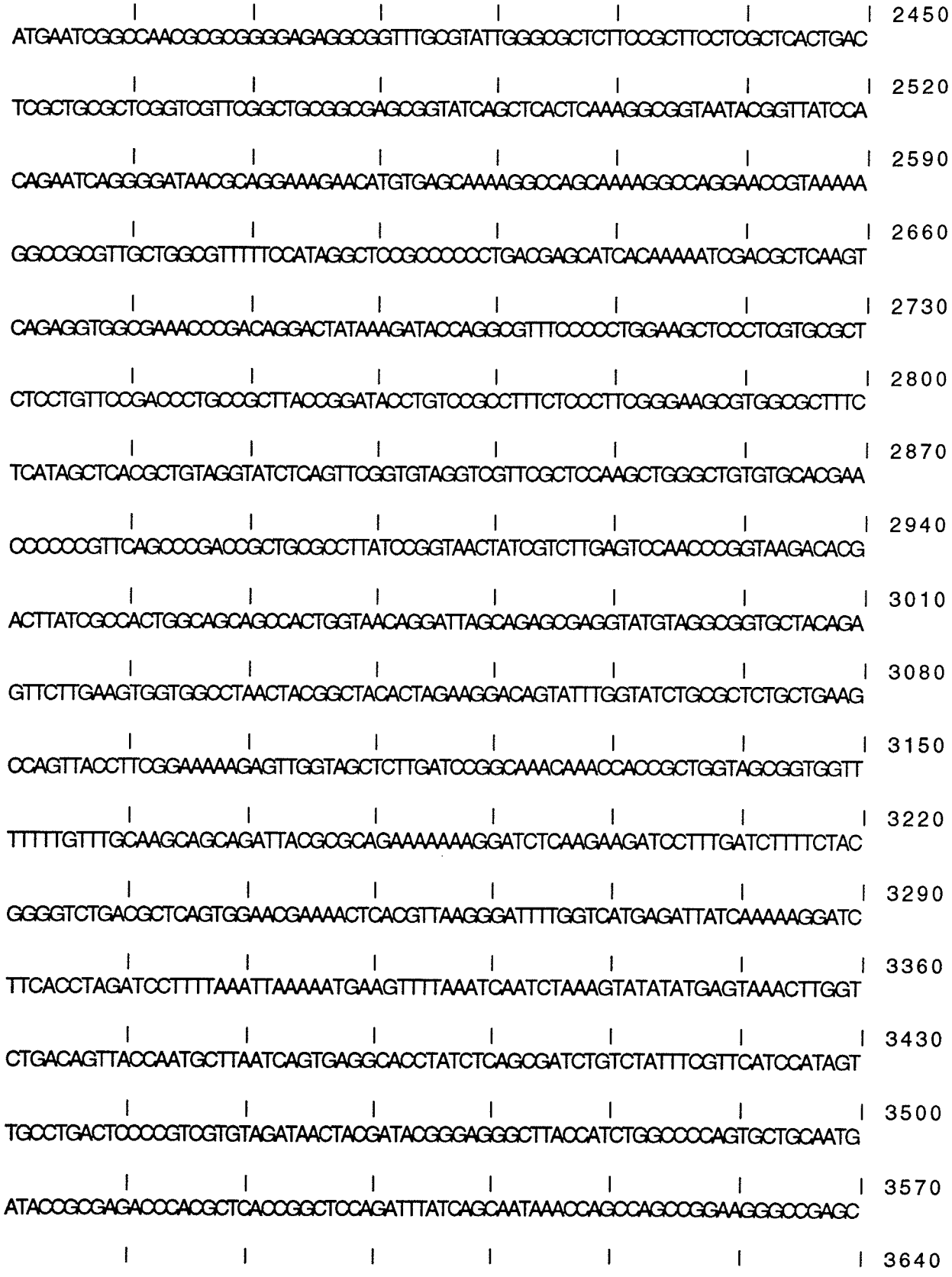
The following enzymes have no sites:

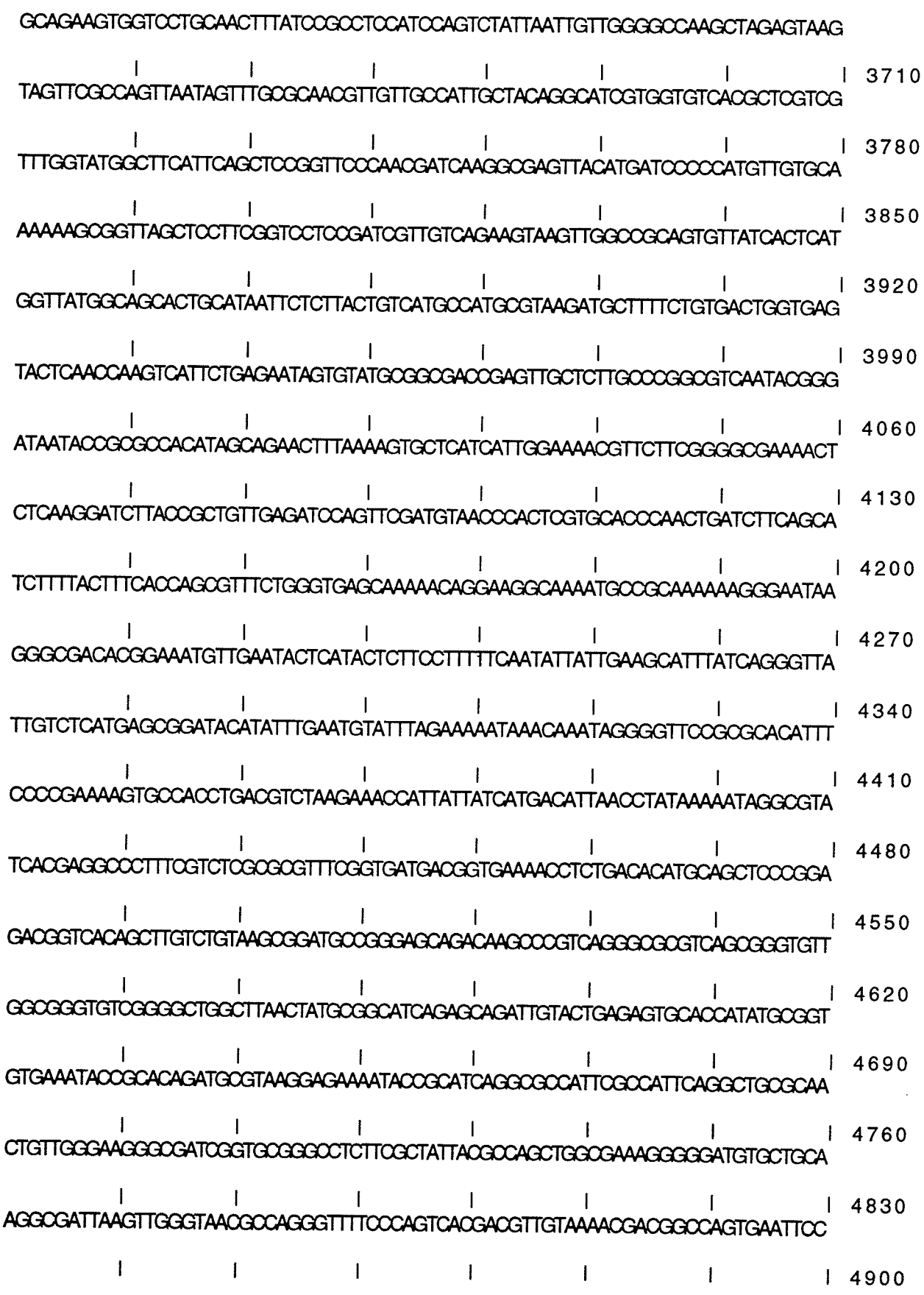
Asu II	Bcl I	Bgl II	Bss H II	Bst E II	Bst XI
Cla I	Dra III	Nhe I	Not I	Rsr II	Sac I
Sac II	Sfi I	Sna B I	Spe I	Spl I	Tth 111 I
Xba I	Xca I	Xho I	Xma I		

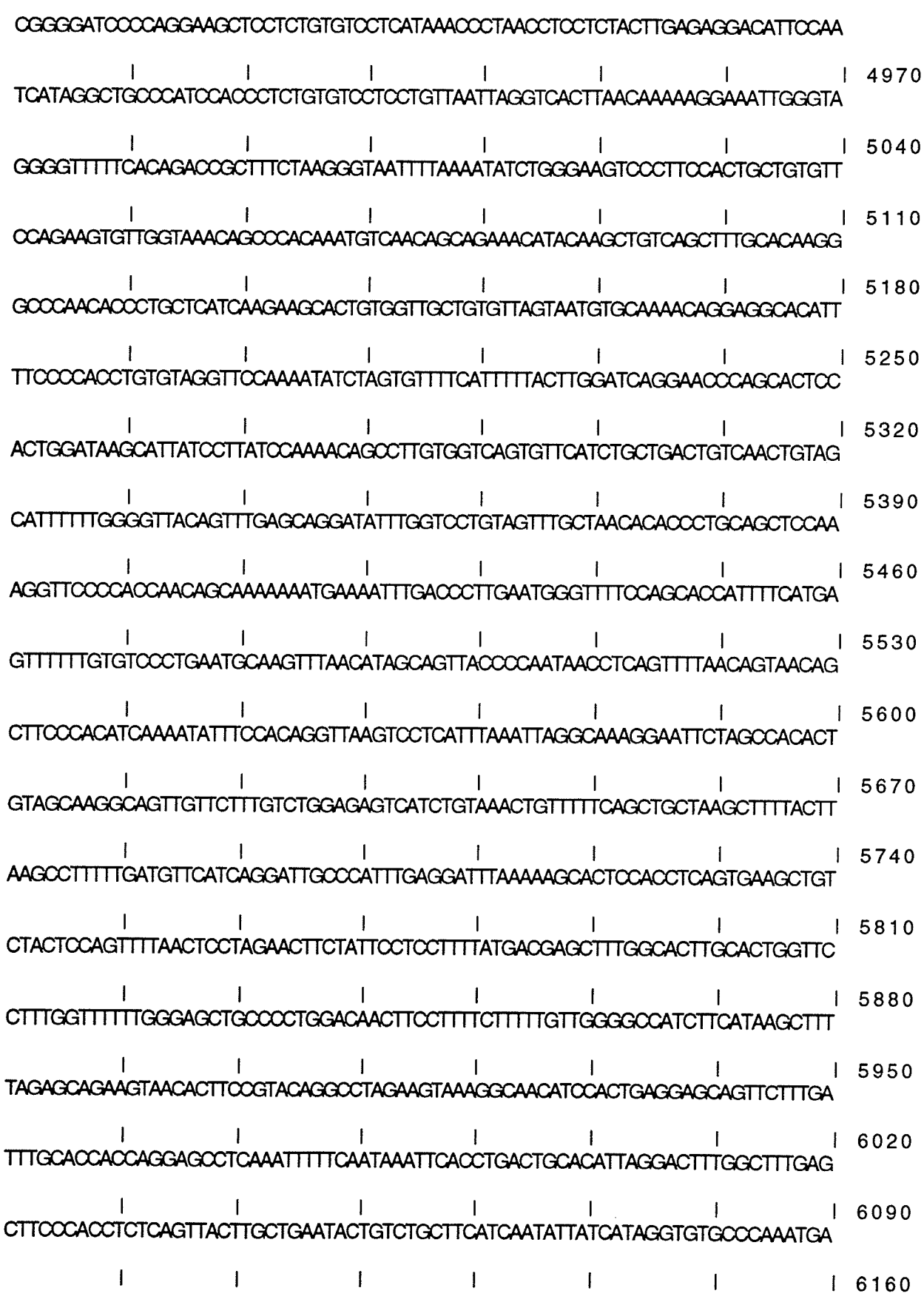
Appendix 3: Complete DNA Code for the Plasmid pUSVCAT 5'→3'

```
                                                                    70
CGAGCTACCTGCAGGTTCAGCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAAT

                                                                   140
TTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCCTGCTTAGGGTTAGGCGTTTTGC

                                                                   210
GCTGCTTCGCGATGTACGGGCCAGATATACGCGTATCTGAGGGGACTAGGGTGTGTTTAGGCGAAAAGCG

                                                                   280
GGGCTTCGGTTGTACGCGGTTAGGAGTCCCCTCAGGATATAGTAGTTTCGCTTTTGCATAGGGAGGGGGA

                                                                   350
AATGTAGTCTTATGCAATACACTTGTAGTCTTGCAACATGGTAACGATGAGTTAGCAACATGCCTTACAA

                                                                   420
GGAGAGAAAAAGCACCGTGCATGCCGATTGGTGGAAGTAAGGTGGTACGATCGTGCCTTATTAGGAAGGC

                                                                   490
AACAGACAGGTCTGACATGGATTGGACGAACCACTGAATTCCGCATTGCAGAGATAATTGTATTTAAGTG

                                                                   560
CCTAGCTCGATACAATAAACGCCATTTGACCATTCACCACATTGGTGTGCACCTAAGCTTGATTTTCAGG

                                                                   630
AGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGT

                                                                   700
AAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATACCCAGACCGTTCAGCTGGATATTA

                                                                   770
CGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCG

                                                                   840
CCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTCAGCTGGTGATATGGGATAGTGTT

                                                                   910
CACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACG

                                                                   980
ATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCC

                                                                  1050
TAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTA

                                                                  1120
AACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACA
```


1190
AGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCT
1260
TAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGC
1330
CCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGGATCTTTG
1400
TGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGT
1470
AAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCA
1540
ACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAG
1610
AAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAA
1680
GGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGA
1750
ACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAA
1820
AATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACA
1890
CAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAA
1960
GGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTTGTA
2030
GAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTG
2100
TTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAA
2170
TAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGG
2240
ATCCGTCGACCTGCAGCCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
2310
CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
2380
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA


2450
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGAC
2520
TCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCA
2590
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
2660
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
2730
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
2800
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
2870
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
2940
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACG
3010
ACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
3080
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAG
3150
CCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
3220
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
3290
GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATC
3360
TTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
3430
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT
3500
TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATG
3570
ATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
3640


GCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGGGGCCAAGCTAGAGTAAG
3710
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCG
3780
TTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
3850
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCAT
3920
GGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATGCGTAAGATGCTTTTCTGTGACTGGTGAG
3990
TACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGG
4060
ATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACT
4130
CTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA
4200
TCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
4270
GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTA
4340
TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTT
4410
CCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTA
4480
TCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGA
4550
GACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTT
4620
GGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGT
4690
GTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAA
4760
CTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCA
4830
AGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCC
4900


CGGGGATCCCCAGGAAGCTCCTCTGTGTCCTCATAAACCCTAACCTCCTCTACTTGAGAGGACATTCCAA
4970
TCATAGGCTGCCCATCCACCCTCTGTGTCCTCCTGTTAATTAGGTCACTTAACAAAAAGGAAATTGGGTA
5040
GGGGTTTTTCACAGACCGCTTTCTAAGGGTAATTTTAAAATATCTGGGAAGTCCCTTCCACTGCTGTGTT
5110
CCAGAAGTGTTGGTAAACAGCCCACAAATGTCAACAGCAGAAACATACAAGCTGTCAGCTTTGCACAAGG
5180
GCCCAACACCCTGCTCATCAAGAAGCACTGTGGTTGCTGTGTTAGTAATGTGCAAAACAGGAGGCACATT
5250
TTCCCCACCTGTGTAGGTTCCAAAATATCTAGTGTTTTCATTTTTACTTGGATCAGGAACCCAGCACTCC
5320
ACTGGATAAGCATTATCCTTATCCAAAACAGCCTTGTGGTCAGTGTTCATCTGCTGACTGTCAACTGTAG
5390
CATTTTTTGGGGTTACAGTTTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGCAGCTCCAA
5460
AGGTTCCCCACCAACAGCAAAAAAATGAAAATTTGACCCTTGAATGGGTTTTCCAGCACCATTTTCATGA
5530
GTTTTTTGTGTCCCTGAATGCAAGTTTAACATAGCAGTTACCCCAATAACCTCAGTTTTAACAGTAACAG
5600
CTTCCCACATCAAAATATTTCCACAGGTTAAGTCCTCATTTAAATTAGGCAAAGGAATTCTAGCCACACT
5670
GTAGCAAGGCAGTTGTTCTTTGTCTGGAGAGTCATCTGTAAACTGTTTTTCAGCTGCTAAGCTTTTACTT
5740
AAGCCTTTTTGATGTTCATCAGGATTGCCCATTTGAGGATTTAAAAAGCACTCCACCTCAGTGAAGCTGT
5810
CTACTCCAGTTTTAACTCCTAGAACTTCTATTCCTCCTTTTATGACGAGCTTTGGCACTTGCACTGGTTC
5880
CTTTGGTTTTTTGGGAGCTGCCCCTGGACAACTTCCTTTTCTTTTTGTTGGGGCCATCTTCATAAGCTTT
5950
TAGAGCAGAAGTAACACTTCCGTACAGGCCTAGAAGTAAAGGCAACATCCACTGAGGAGCAGTTCTTTGA
6020
TTTGCACCACCAGGAGCCTCAAATTTTTCAATAAATTCACCTGACTGCACATTAGGACTTTGGCTTTGAG
6090
CTTCCCACCTCTCAGTTACTTGCTGAATACTGTCTGCTTCATCAATATTATCATAGGTGTGCCCAAATGA
6160

TATTTGCAACCCTTCCCTGTTGGCTACTTGTCTCACCATTGTAGGCCTAATGGGAGACAAAGTAGAGTAG

TAATCTTGTAAAGAGTTATACCAATTAACAGGAGCATTAATTACTGTCCAAGTAGTTTCCTCTAAAAACC 6230

TTGCCAAACTGTCCCTTAAATATCTTTGGGTTCTTCTTTCAAGCTCCTGTGAGGTGAGCCTAGGAATGTC 6300

ATTTTGTATTACACGCCAAAAAGCTTGAGAAATGGCATTAAAAAGTGTTGGACCCCAATGTCTGGGGTCA 6370

AGATACTGAACACTGTGAACAAAGGTTTGTACTCCAGGAAATAAAATATCATAGTAATCATCTGGCCTAT 6440

ACAAATCTACAGCCATTCCTGGTTGTTGATATAAACCAACAGTAGAAACTTTGTGATCCCAGTCACTAAA 6510

AAATCTATACCCCACTTGAGCAACAGCGCTCACACCAGTCACAGTTTGCAGTAAAGCTGCAAATCCAGCT 6580

ATAGCAGCAGGAGCCCCAGATATCACAGCATAGGCCTGTGGAGTGAGGCCTATAGCAGCAATTGCCTCAG 6650

AGGTTGTTAGGCCTTCAACAGTAGCAACAGATGCAAGTTGCACTTCAATTGCAGCAGCGGCCTCTCCAGC 6720

AGCAATTTCAGCTACTGAAAATCCAGTAGCAGCAGCAGCTTCAGACACAGTAGCAATTAGGTCCCCCAAC 6790

AGTGTTAAAGCAGCACCCATGGACCTGAAATAAAAGACAAAAAGACTAAACTTACCAGTTAACTTTCTGG 6860

TTTTTCAGTTAACCTTTCTGGTTTTTTGCGTTTCCCGTCAACAGTATCTTCCCCTTCACAAAATTGCAGC 6930

AAAAGCTCTAAAACAAACACAAAAAGGCGTTGAGCTGTTTTTTTACTTTCAGTCCATGACCTACGAACCT 7000

TAACGGAGGCCTGGCGTGACAGCCGGCGCAGCCATGGCCTGAAATAACCTCTGAAAGAGGAACTTGGTTA 7070

GGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCT 7140

CCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGCTCCCCAGCAGGCAGAA 7210

GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCT 7280

AACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAG 7350

7420

GCCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAA

| | | | | | | 7490
AAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGACCTTCTAGGTCTT

| | | | | | | 7560
GAAAGGAGTGCCTGGGGGAATATTCCTCTGATGAGAAAGGCATATTTAAAAAAATGCAAGGAGTTTCATC

| | | | | | | 7630
CTGATAAAGGAGGAGATGAAGAAAAAATGAAGAAAATGAATACTCTGTACAAGAAAATGGAAGATGGAGT

| | | | | | | 7700
AAAATATGCTCATCAACCTGACTTTGGAGGCTTCTGGGATGCAACTGAGGTATTTGCTTCTTCCTTAAAT

| | | | | | 7770
CCTGGTGTTGATGCAATGTACTGCAAACAATGGCCTGAGTGTGCAAAGAAAATGTCTGCTAACTGCA