THE UNIVERSITY OF CALGARY

Regulation of the ges-1 gene from the nematode Caenorhabditis briggsae

by

Sean David Goldie Marshall

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

CALGARY, ALBERTA

APRIL, 1998

Abstract

The *ges-1* gene has been used to investigate how a particular gene can be expressed in only a subset of cells within a complex multicellular organism. The *C. elegans ges-1* (*Ce-ges-1*) gene is only expressed in the E lineage (i.e. gut) of embryos. Previous analysis of the *Ce-ges-1* regulatory regions uncovered a gut activator-pharynx/rectum repressor regulatory switch. This expression switch centres on a tandem pair of GATA sites. A *ges-1* homologue from a closely related nematode, *C. briggsae*, is also expressed in the gut. Study of the *C. briggsae ges-1* (*Cb-ges-1*) regulatory regions revealed a similar gut to pharynx/rectum regulatory switch. A putative pharynx/rectum activator element was located 57 bp downstream from the poly(A) signal site of *Cb-ges-1*, while the putative *Ce-ges-1* pharynx/rectum activator element was found 70 bp upstream from the *Ce-ges-1* initiation codon.

Acknowledgements

I thank the following people who helped me during the pursuit of my degree:

To my supervisor, Dr. Jim McGhee for his support and guidance throughout my degree training. His enthusiasm for and dedication to science provided much inspiration. I appreciated the opportunity to work with you.

To my supervisory committee, Drs. Paul Mains and Kostas Iatrou. Thank you for all your helpful suggestions and constructive criticisms. I also extend my thanks to Dr. Mike Bentley for being a part of my examination committee.

To the members of the McGhee lab, both past and present, for sharing your ideas, protocols, and advice. I especially thank: Barb Goszczynski for helping me get started in the lab, and sharing her expertise throughout my degree (thanks for all of your help); May Chung for teaching me the how to use a worm pick and transform worms, and for sharing your wisdom with me; Shelley Culp, John Kalb, and Karen Lau for your help and friendship over the years.

To the Medical Biochemistry secretarial staff for helping with the administration portion of this degree.

To my family and friends for all of your support and encouragement while I pursued my Master's degree.

Finally I acknowledge the financial support I received from the Faculty of Medicine (1997) and the Department of Medical Biochemistry (1995-1996).

To my parents,

David and Betty Marshall

Table of Contents

Table of Contents (continued)

List of Tables

List of Figures

List of Figures (continued)

List of Figures (continued)

List of Abbreviations

bp - base pair

Cb-ges-1 - *C. briggsae* gut esterase #1 gene

cDNA - complimentary deoxyribonucleic acid

Ce-ges-1 - *C. elegans* gut esterase #1 gene

cpm - counts per minute

dH_20 - distilled water

DMF - dimethylformamide

DNA - deoxyribonucleic acid

dNTP - deoxynucleotide triphosphate

DTT - dithiothreitol

γ-P^{32}-ATP - gamma phosphorous-32 adenosine triphosphate

$GATA_1$ - *Ce-ges-1* upstream GATA site

$GATA_2$ - *Ce-ges-1* downstream GATA site

$GATA_{Dn}$ - *Cb-ges-1* downstream GATA site

$GATA_{Up}$ - *Cb-ges-1* upstream GATA site

GCG - Genetics Computer Group

GFP - green fluorescent protein

HEPES - N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)

HSE - heat shock element

Na_2EDTA - disodium ethylenediaminetetraacetic acid

NGM - nematode growth media

PCR - polymerase chain reaction

PEG - polyethylene glycol

pmol - picamole

PRA -pharynx/rectum activator

SDS - sodium dodecyl sulfate

Tris-HCl - tris(hydorxymethyl)aminomethane-hydrochloric acid

2xYT - two times yeast tryptone media

V/cm - volts per centimetre

X-Gal - 5-bromo-4-choloro-3-indolyl-β-D-galactopyranoside

Introduction

The aim of developmental biology is to understand the processes and mechanisms that are involved in transforming a fertilized egg into a complex multicellular organism. Ultimately, the underlying mechanisms involved in this transformation require the expression of specific genes in the right place at the right time. How does a developing embryo "know" which genes to express, where and when these genes should be expressed, and how much of the gene product is required? Furthermore, how are the regulatory networks organized, and in what manner do these networks evolve?

Transcriptional Regulation of Gene Expression

Transcription seems to be the primary level of gene regulation, and is a highly complex and tightly regulated process. In eukaryotes, genes are expressed by the coordination of events at the core promoter (comprised of a TATA box and a transcription start site) and the surrounding regulatory regions to allow transcription (reviewed in Ptashne and Gann, 1997; Greenblatt, 1997). The core promoter region is the site where the basal transcription machinery binds prior to receiving the signal for initiating mRNA transcription. The basal transcription machinery is a protein complex composed of a TATA-binding-protein (TBP) and several general TBP associated factors (TAFs), which are required for transcription (reviewed in Roeder, 1996); TAFs composition in the basal transcription machinery varies with each cell type and developmental

stage. Assembly of the basal transcription machinery at the core promoter region does not provide a large degree of specificity for regulating the pattern of gene expression, but does play a role in recruiting other transcription factors to the promoter (reviewed in Verrijzer and Tjian, 1996).

Specificity of gene expression is largely controlled by the regulatory region(s) of the promoter with control being cell type, developmental stage, and gene specific (Ptashne and Gann, 1997). Located upstream and/or downstream from the coding sequence of a gene, regulatory regions consist of sequences that are referred to as enhancers (activating transcription) and silencers (repressing transcription) (Kamakaka, 1997). Some DNA binding proteins bind to specific sequences within the regulatory regions and allow either transcriptional activation or repression. Specific combinations of transcription activators and repressors allow subsets of genes within the genome to be expressed in specific spatial and temporal patterns. The ability of each cell to express a specific set of genes is poorly understood.

Regulatory regions located relatively close to the core promoter (within ~150 bp) can aid in transcriptional activation by altering the local DNA topology. Alteration of the DNA structure can cause a change in energy levels that favour transcription initiation (van der Vliet and Verrijzer, 1993). Binding of DNA by transcription factors can be aided through cooperative binding of other proteins to the DNA, i.e. a single transcription factor may require the presence of other protein-DNA interactions to form a stable protein-DNA complex. The Drosophila

Ubx protein provides an example of cooperative binding that facilitates transcriptional activation. A yeast expression system was used to demonstrate that Ubx binding site number and sequence orientation were involved in a synergistic increase in transcriptional activation of a reporter gene (Beachy *et al.*, 1993).

Many transcription factors bind to specific sequences in a regulatory region(s), which is often located far from the core promoter. If a large distance separates a regulatory region from the core promoter (~1000 bp), the DNA may be flexible enough to allow the DNA bound transcription factor to interact directly with the pre-initiation complex (Rippe *et al.*, 1995). The degree of superhelicity, natural curvature of the DNA sequence, and protein binding site placement have all been suggested to influence DNA flexibility. Alternatively, the regulatory and core promoter regions can be brought together through DNA bending (van der Vliet and Verrijzer, 1993). Several DNA binding proteins are known to bend DNA (to varying degrees). The high mobility group (HMG) proteins provide one of the more dramatic examples of DNA bending (reviewed in Grosschedl *et al.*, 1994). The HMG domain in the mammalian protein LEF-1 can induce a 130° bend in the DNA upon binding (Giese *et al.*, 1992). It has been postulated that protein directed DNA bending facilitates the interaction of distantly (and occasionally closely) positioned regulatory factors with the core promoter to stabilize weak protein-protein and/or protein-DNA interactions (Grosschedl *et al.*, 1994).

Formation of heterochromatin along chromosomes is also a potential method of controlling gene expression by making genes within the area of heterochromatin transcriptionally silent (Beato and Eisfeld, 1997). The genome of eukaryotes is highly organized in order to achieve the level of compaction necessary for the nuclear DNA to fit into the nucleus. The lowest order of genome organization is the nucleosome, where DNA is wrapped around nucleosome core particles (Beato and Eisfeld, 1997). Assembly of DNA into nucleosomes (and higher order structures) limits the ability of transcription factors to bind enhancer sequences, which often results in transcriptional repression. The yeast PHO phosphate metabolism pathway is one example of how nucleosome repression can be overcome to allow transcription . Evidence from *in vivo* footprinting and transcriptional activation assays (using hybrid protein domain fusions) suggests that the PHO4 transcription factor binds cooperatively to the *pho5* promoter to disrupt the local nucleosome organization, thereby allowing *pho5* transcription (reviewed in Svaren and Horz, 1997). Alternatively, insertion and deletion of DNA sequences was used to demonstrate the need for $(CT)_n$ repeats in the promoter to keep the Drosophila *hsp26* free of nucleosomes (Lu *et al.*, 1993).

Another level of transcriptional control can be achieved through DNA methylation. Siegfried and Cedar (1997) have suggested that gene expression via DNA methylation can occur in three ways: 1) by interfering directly with the ability of specific transcription factors to bind DNA, 2) by facilitating interaction of

transcription factors with DNA, and 3) by altering nucleosome positioning. Transcriptional inhibition via DNA methylation was previously demonstrated by measuring the levels of luciferase transcription from a minichromosome with known levels of DNA methylation (Hsieh, 1994). Transcriptional inhibition was correlated with increased levels of DNA methylation. It is interesting to note that to date, no evidence for DNA methylation in the nematode, *C. elegans*, has been found (Simpson *et al.*, 1986; reviewed in Hodgkin, 1994).

Gene Regulation Across Species

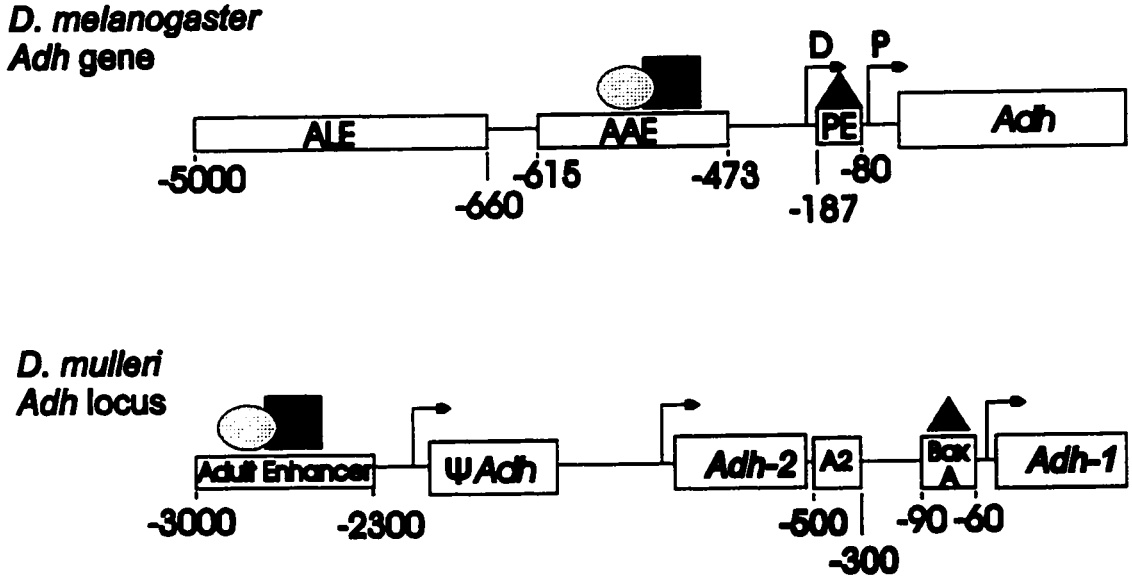
While much work has been done to elucidate regulatory mechanisms involved for regulating identified genes, this work has been largely restricted to studying particular genes within a single species. Only a limited number of studies have directly compared gene regulation between species (either related or divergent). Comparing the regulatory elements of a specific gene between two or more species can lead to a better overall understanding of how a particular gene is regulated. Furthermore, interspecies gene comparison may provide insights into the changes that regulatory networks undergo over "defined" periods of time.

Regulation of the *alcohol dehydrogenase* (*Adh*) gene from Drosophila has been investigated in some detail across several Drosophila species. Of particular interest are the *Adh* loci from *D. melanogaster* and *D. mulleri*, which are two species separated by approximately 40 million years (Throckmorton,

1982). The *D. mulleri Adh* locus is comprised of two closely linked *Adh* genes; *Adh-1* is expressed in larvae, and *Adh-2* expressed in late larvae and adults (figure 1). A pseudogene (Ψ-*Adh*; transcripts are detected in adults) is present upstream from *Adh-2* and *Adh-1* (figure 1). In larval *D. mulleri*, *Adh-1* and *Adh-2* are expressed in the fat body and Malpighian tubules, with *Adh-1* also being expressed in the midgut (Fischer and Maniatis, 1986). In adult *D. mulleri*, *Adh-2* is expressed in the fat body, hindgut and rectum (Fischer and Maniatis, 1986). The *D. melanogaster Adh* locus has a single *Adh* gene that is regulated by a proximal promoter (larval expression) and a distal promoter (adult expression), as illustrated in figure 1. During early *D. melanogaster* larval development, the proximal promoter controls *Adh* expression in the fat body, midgut, and Malpighian tubules (Corbin and Maniatis, 1990; Shen *et al.*, 1991). During late larval (and adult) development, the distal promoter takes over from the proximal promoter to allow *Adh* expression in the fat body, hindgut, rectum, Malpighian tubules and some sex specific tissues (Corbin and Maniatis, 1990; Shen *et al.*, 1991).

It is interesting that P-element mediated transformation of *D. melanogaster* using the *D. mulleri Adh-1* and *Adh-2* genes resulted in a near perfect "*D. mulleri*" expression pattern in *D. melanogaster* (Fischer and Maniatis, 1986). This observation suggested that the regulatory sequences and proteins were conserved between the two species, even though the arrangement of the *Adh* loci had changed dramatically. Further investigation of the *D. mulleri Adh*

Figure 1: Diagram of regulatory binding sites in the *D. melanogaster* and *D. mulleri Adh* genes. The binding sites shown are for C/EBP (box), AEF-1 (oval), ABF-2 (triangle). Transcriptional start sites are indicated by arrows, with D and P being the distal and proximal transcriptional start sites for *D. melanogaster Adh*. AAE = *Adh* Adult Enhancer, ALE = *Adh* Laval Enhancer, PE = Proximal Enhancer, and A2 = *Adh-2* Malpighian tubules 3' enhancer. Numbers show distances from the transcriptional start site. Figure was modified from Abel *et al.* (1992).


D. melanogaster
Adh gene
D
P
ALE
AAE
PE
Adh
-5000
-660
-615
-473
-187
-80
D. mulleri
Adh locus
Adult Enhancer
ΨAdh
Adh-2
A2
Box A
Adh-1
-3000
-2300
-500
-300
-90
-60

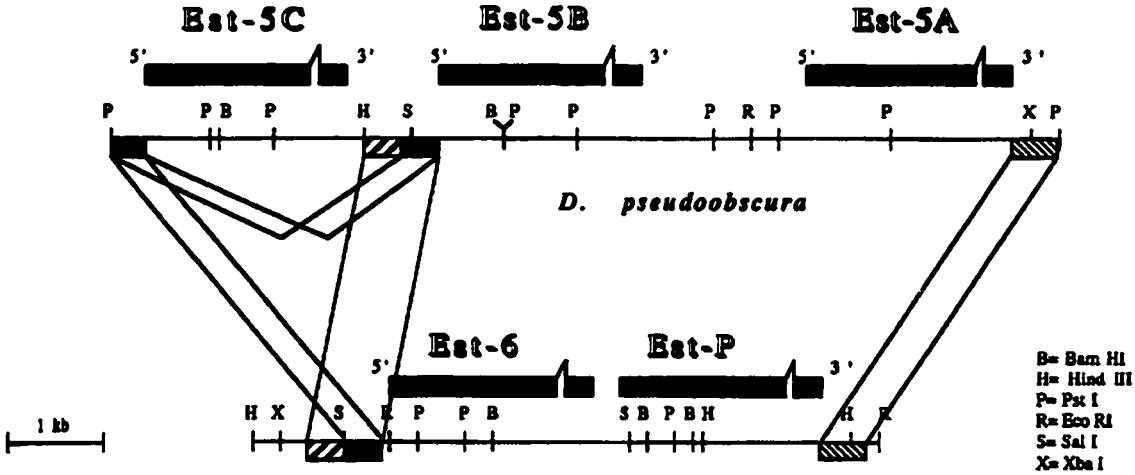
locus demonstrated that separate tissue specific control elements were involved. Expression of variously deleted *Adh-1* and *Adh-2* constructs within a *D. melanogaster* background allowed for mapping and characterization of the *cis*-regulatory elements within the *D. mulleri Adh* locus. Sequences responsible for fat body expression of both *Adh-1* and *Adh-2* were found upstream from the respective genes (Fischer and Maniatis, 1986). A Malpighian tubule specific enhancer for *Adh-2* expression was found at the 3' end of the *Adh-2* gene (Fischer and Maniatis, 1986). Characterization of the larval fat body enhancer from *D. melanogaster* and *D. mulleri* revealed the presence of a WGATAR site that was recognized by a single zinc finger DNA-binding protein (ABF-1/Serpent), which shares similarity with the GATA factor family of transcription factors (Abel *et al.*, 1993; Reuter, 1994). Further investigation of *Adh* regulation has demonstrated that a C/EBP-like *trans*-acting factor and a *trans*-repressing factor (AEF-1) were also involved in *Adh* expression (Abel *et al.*, 1992). Figure 1 diagrams the arrangement of AEF-1 and C/EBP binding sites within the *Adh* loci from *D. melanogaster* and *D. mulleri*. Together the *Adh* studies demonstrate that although the *Adh* regulatory arrangement is different in each Drosophila species, the *D. mulleri Adh* regulatory elmenents are still recognized by *Drosophila melanogaster* transcription factors to produce a "*D. mulleri*" expression pattern.

Another gene locus where the regulatory elements have been studied across several species are the Drosophila esterase loci, *Est-6/Est-5*. Of

particular interest are the studies that have characterized the *D. melanogaster Est-6* and *D. pseudoobscura Est-5* loci. The *D. pseudoobscura* locus contains three *Est-5* genes (C, B, and A, as arranged on the chromosome; see figure 2), with *Est-5B* encoding the active EST-5 protein (Brady and Richmond, 1992). The *D. melanogaster* locus encodes *Est-6* and *Est-P*, with *Est-6* producing the active EST-6 protein (see figure 2; Brady and Richmond, 1992). Duplication events are believed to be responsible for the presence of the multiple esterase genes in each homologous locus (Brady and Richmond, 1992). Both esterase loci encode β-carboxylesterase proteins that are mainly expressed in the hemolymph and some portions of the male reproductive tract, with other tissues expressing these genes at significantly lower levels (Healy *et al.*, 1996). While the spatial expression patterns of the esterases were similar between the two Drosophila species, the expression levels were quite different. In the male reproductive tract, high levels of EST-6 are present in the anterior ejaculatory ducts of *D. melanogaster*, while extremely low levels of EST-5 activity are detected in *D. pseudoobscura* (Brady and Richmond, 1992; Healy *et al.*, 1996).

A detailed study of the *D. melanogaster Est-6* regulatory regions was carried out to identify regulatory sequences responsible for the ancestral expression pattern in the hemolymph and the ejaculatory bulb, plus the recently acquired anterior ejaculatory duct esterase expression pattern. Most of the qualitative regulatory regions (i.e. tissue specific control sequences) were found in the first 450 bp upstream from the translation start site (Healy *et al.*, 1996).

Figure 2: Schematic diagram of the *D. melanogaster* and *D. pseudoobscurra Est-6/Est-5* sequence alignment. Flanking sequences with significant identify are connected by lines and have the same shading. Figure taken from Brady and Richmond (1992).



D. melanogaster

The quantitative regulatory regions (i.e. controlling expression levels in the expressing tissues) were found far from the translational start site (Healy *et al.*, 1996). Furthermore, sequences were identified that contained both positive and negative information for regulating *Est-6* transcription (Healy *et al.*, 1996). Of particular interest, in view of the later results, was the observation that a positive acting quantitative control element for expression in the ejaculatory duct was mapped to the 3' end of the *Est-6* gene. Located approximately 200 bp from the 3' end of the *Est-6* gene, the 3' control element for ejaculatory duct expression was suggested to overlap with the *Est-P* pseudogene, either in the presumed *Est-P* promoter region or the *Est-P* transcriptional unit itself (Healy *et al.*, 1996).

The compative studies of the *Adh* and *Est-6/Est-5* genes were used as examples to illustrate several themes that are found in the present analysis of the *Caenorhabditis elegans* and *Caenorhabditis briggsae ges-1* (*gut esterase #1*). The *Adh* genes contain several regulatory elements that are clustered within a relatively short distance from each other; two of these binding sites, AEF-1 and C/EBP, are overlaping. Two apparently linked regulatory elements that centre around a tandem pair of GATA sites are found in both *ges-1* genes. The *Est-6* gene contains a regulatory region that is located at the 3' end of the gene; a 3' located regulatory element was identified for the *C. briggsae ges-1* gene.

Why Study Nematodes?

C. elegans was chosen as a model organism by Sydney Brenner in 1965 to investigate various biological questions about development (reviewed in Wood, 1988). Several features of this nematode make it an amenable model organism for studying developmental processes. *C. elegans* is easy to grow and maintain, has a short generation time (about 3 days) and high fecundity (produces ~300 eggs per hermaphrodite); the existence of both males and self-fertilizing hermaphrodites make *C. elegans* ideal for use in classical genetic studies. *C. elegans* is transparent through its entire life cycle (egg to adult), allowing structural morphology and cell movements to be described in live animals. Additionally, the entire somatic cell lineage, which is essentially invariant, has been mapped from the time of fertilization to a mature adult, and the entire nuclear genome will soon be sequenced (Sulston and Horvitz, 1977; Sulston *et al.*, 1983; Sulston *et al.*, 1992). These features provide a wealth of information for studying any number of biological processes.

Why Study Nematode Intestines?

Figure 3 is a schematic diagram illustrating the early cell divisions of the *C. elegans* embryo. By the time an embryo has reached the 28 cell stage, five somatic founder cells (AB, MS, E, C, and D) plus one germline founder cell (P4) have been established (Sulston *et al.*, 1983). The digestive tract of *C. elegans* is derived from four different cell lineages (figure 4): the ABa lineage contributes

Figure 3: Schematic diagram of the early *C. elegans* embryo cell lineage showing birth of the six founder cells. The time scale refers to developmental stage (at 20°C) after fertilization. The cell # column indicates the number of cells present after each cell division (prior to gastrulation). Blackened circles indicate the position of nuclei for a particular cell lineage. Figure taken from McGhee (1995), which was redrawn from Sulston *et al.* (1983).

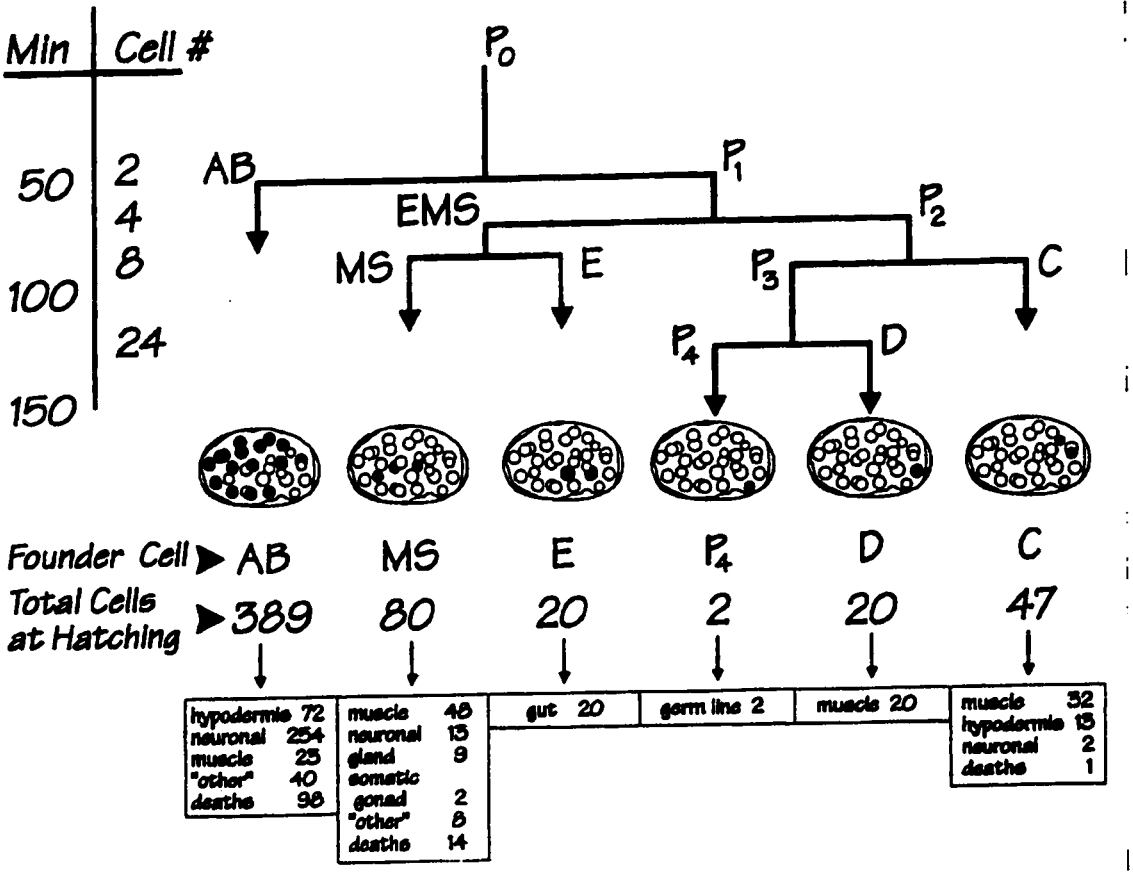
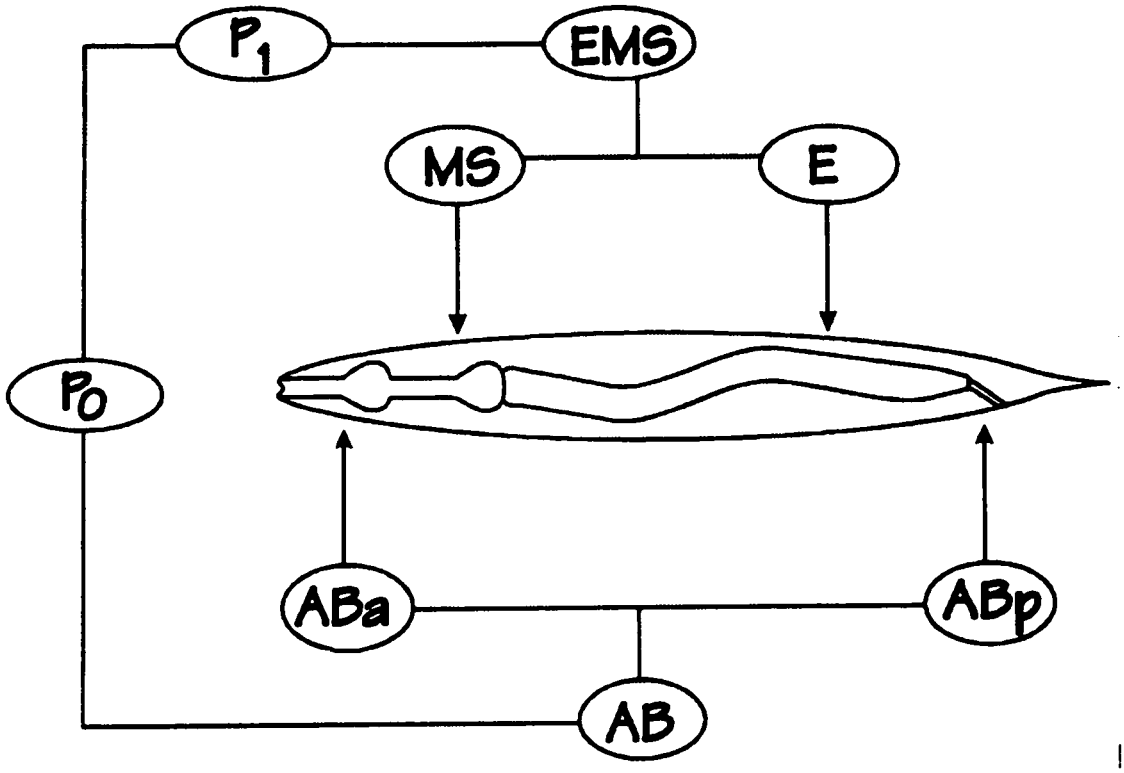

Min
Cell #
50
100
150
2
4
8
24
P0
AB
P1
EMS
P2
MS
E
P3
C
P4
D
Founder Cell ▶ AB MS E P4 D C
Total Cells at Hatching ▶ 389 80 20 2 20 47
hypodermis 72
neuronal 254
muscle 25
"other" 40
deaths 98
muscle 48
neuronal 13
gland 9
somatic gonad 2
"other" 8
deaths 14
gut 20
germ line 2
muscle 20
muscle 32
hypodermis 15
neuronal 2
deaths 1

Figure 4: Lineages contributing to the *C. elegans* digestive tract. The anterior pharynx is made from Aba derived cells, while the posterior pharynx has cells contributed from the MS lineage. The E lineage is the sole source of gut (i.e. intestine) cells. The rectum is derived from cells contributed by the Abp lineage. Figure taken from Fukushige *et al.* (1996).


P1
EMS
MS
E
P0
ABa
ABp
AB

some cells to form the anterior pharynx, some cells from the MS lineage make up the posterior pharynx, a few cells from ABp lineage contribute to the formation of the rectum, and the gut (also referred to as the intestine) is formed entirely from the E cell lineage (Sulston *et al.*, 1983).

Unlike other somatic structures and tissues in *C. elegans*, the entire gut (20 polyploid cells in the adult) is derived solely from the E lineage and the E lineage produces only gut (Sulston and Horvitz, 1977; Sulston *et al.*, 1983). The connection of the gut and the E lineage provides a unique opportunity for understanding what defines one tissue type from other tissues in terms of gene regulatory networks. What is required for a specific gene to be expressed in a single cell lineage (e.g. gut) within a multi-lineage structure (e.g. digestive tract)?

Summary of Previous Work Done On ges-1

To investigate the mechanisms of gut differentiation, a gut specific marker needed to be identified. The general idea was to identify a gut specific gene and analyse how it was regulated, then study the regulators of the regulators, etc. Eventually a pathway could be constructed that would illustrate how a specific gene at the end of a regulatory pathway was specified to be turned on in a particular cell or tissue type (discussed in McGhee, 1992).

Based on the histochemical staining pattern, *ges-1* (gut esterase #1, a non-specific serine carboxylesterase) was chosen as a gut specific marker that could be used to study gene regulation during gut differentiation (see McGhee,

1992). The *ges-1* gene is expressed in the gut of *C. elegans* and is easily detected through histochemical staining for esterase activity (Edgar and McGhee, 1986). The *ges-1* esterase activity is the major esterase activity found during embryogenesis. This was demonstrated by: 1) separating embryo protein extracts on isoelectric focusing gels and staining for esterase activity, and 2) lack of any detectable esterase staining in embryos from the *C. elegans ges-1* null mutant strain, JM1041 (McGhee *et al.*, 1990).

After cloning the *ges-1* gene, promoter analysis was done to identify the important regulatory elements through a series of unidirectional deletions (Kennedy *et al.*, 1993; Aamodt *et al.*, 1991). Results from the promoter analyses indicated that regulation of *ges-1* was not simple. A regulatory region was revealed that turned *ges-1* expression off in the gut but simultaneously allowed expression in the pharynx/rectum portions of the digestive tract (Aamodt *et al.*, 1991). Fine mapping of the *ges-1* regulatory elements revealed that gut activation-pharynx/rectum repression centred around a tandem pair of GATA sites, located approximately 1.1 kb upstream from the translation start site (Egan *et al.*, 1995). *ges-1* promoter analysis showed that sequences upstream and downstream from the tandem GATA sites were also involved in regulating *ges-1* expression (Egan *et al.*, 1995). The nature of these flanking sequences remains unknown.

Study of a homologous gene from two or more species (e.g. *C. elegans* and *C. briggsae*) allows for the identification of conserved (and therefore

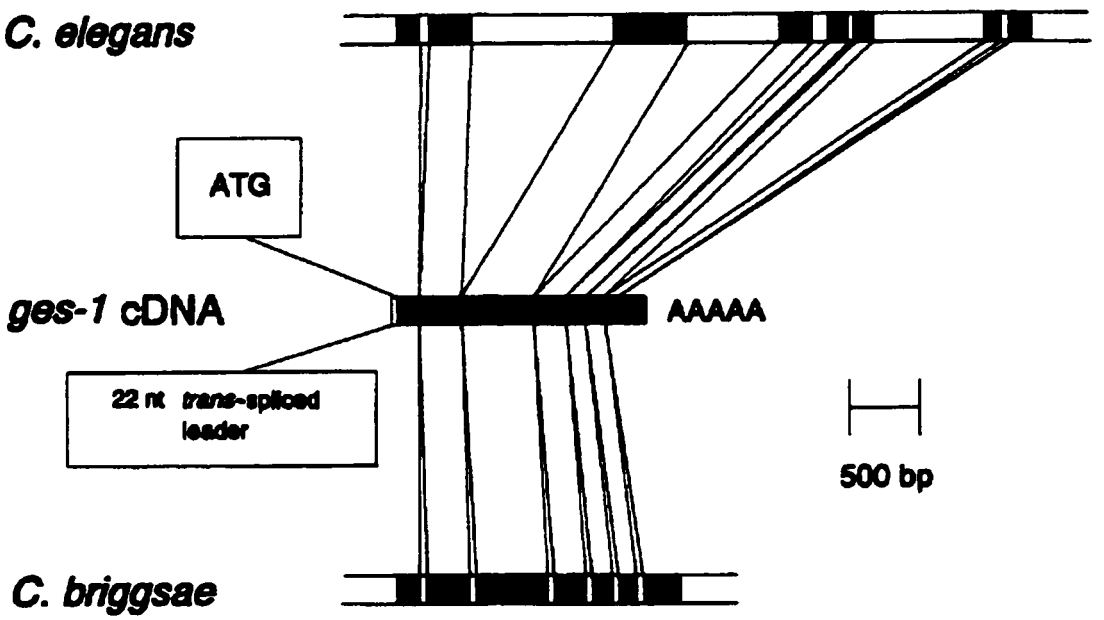
presumed important) *cis*-regulatory elements. This sequence comparison has been done for a few *C. elegans* genes in the past, including: *hsp-3*, *vit-1* to *-6*, *gpd-2* and *-3* (Heschl and Baillie, 1990; Zucker-Aprison and Blumenthal, 1989; Lee *et al.*, 1992). The *C. elegans ges-1* (*Ce-ges-1*) homologue from *C. briggsae* has also been cloned (Kennedy *et al.*, 1993).

Cloning of the *C. briggsae ges-1* gene (*Cb-ges-1*) was done to study *ges-1* regulation in both *C. elegans* and *C. briggsae*, and to determine how the regulatory mechanisms controlling expression have evolved. The *Cb-ges-1* gene is expressed in the gut, similar to the *Ce-ges-1* expression pattern (Kennedy *et al.*, 1993).

Comparison of the *ges-1* coding region between the two nematode species showed a high degree of sequence conservation, which is in agreement with other *C. elegans*/*C. briggsae* gene comparisons; examples include: *unc-119* (90% nucleotide identity), *hsp-3* (92.6%nucleotide identity), the *vit* genes (85-90% nucleotide identity), *dpy-7* (92.5% nucleotide identity) and *col-12* (89.7% nucleotide identity) (Maduro and Pilgrim, 1996; Heschl and Baillie, 1990; Zucker-Aprison and Blumenthal, 1989; Gilleard *et al.*, 1997a; Gilleard *et al.*, 1997b). The *ges-1* gene is 75% identical at the nucleotide level and 83% identical (93% similar) at the amino acid level (Kennedy *et al.*, 1993). Splice donor and acceptor sites were not obviously different, and the first six introns between the two nematode species were located at the same positions; *Cb-ges-1* is missing a seventh intron that is found in *Ce-ges-1* (see figure 5;

Figure 5: Genomic and cDNA organization of the *C. elegans* and *C. briggsae* *ges-1* genes.

"The cDNA clone (middle) is from *C. elegans* and shows the 22 nucleotide *trans*-spliced leader (SL1) added immediately upstream from the ATG initiation codon. Filled regions represent exons; open regions represent introns or flanking sequences." Figure and text taken from Kennedy *et al.* (1993). Unpublished experiments show that *Cb-ges-1* is also trans-spliced (Dr. J. D. McGhee, personal communication).

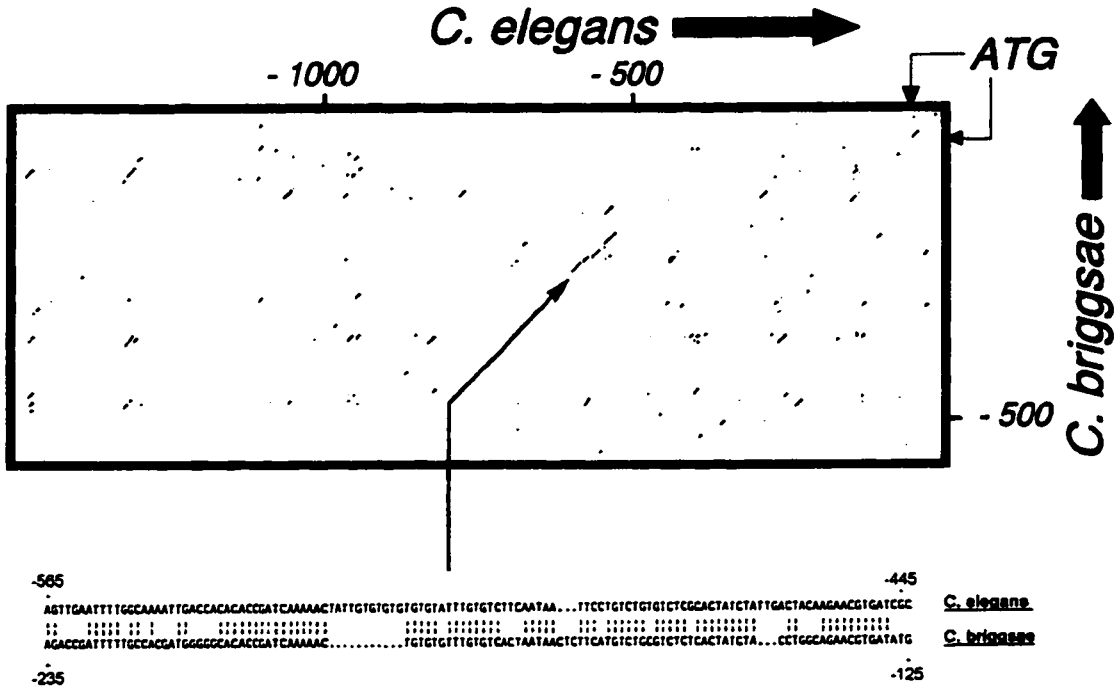

C. elegans
ATG
ges-1 cDNA
AAAAA
22 nt trans-spliced leader
500 bp
C. briggsae

Kennedy *et al.*, 1993). Comparison of the 5' flanking DNA sequence reveals a limited degree of sequence identity; only a small area encompassing approximately 100 bp at the 5' end of both *Ce-ges-1* and *Cb-ges-1* has stretches of 70% or more sequence identity (figure 6). Kennedy *et al.* (1993) reported that *C. briggsae* could be transformed in the same manner as *C. elegans*. Preliminary cross-species transformation experiments suggested conservation of some *ges-1* regulatory mechanisms existed between *C. briggsae* and *C. elegans* (Kennedy *et al.*, 1993).

What's In Store?

Comparative study of *ges-1* regulation in *C. elegans* and *C. briggsae* provides a unique opportunity for understanding how a gene can be specifically expressed in a tissue whose cell lineage is completely defined. Characterization of the important *ges-1* regulatory elements in each species will provide a better understanding of gene regulation. The work presented in this thesis provides details on the identification and characterization of the regulatory elements from *Cb-ges-1*. Comparison with the *C. elegans* regulatory elements is also presented and will be discussed.

Figure 6: Sequence alignments of the *C. elegans* and *C. briggsae ges-1* genes. Top, GCG dot matrix analysis (window = 21, stringency = 14) of the *Ce-ges-1* 5' flanking region (-1500 - 0 bp upstream from the ATG initiation codon) and the *Cb-ges-1* 5' flanking sequence (-500 - 0 bp upstream from the ATG initiation codon). Bottom, GCG GAP alignment of the region of highest similarity from the dot plot. Figure taken from Kennedy *et al.* (1993).


C. elegans
ATG
- 1000
- 500
- 500
C. briggsae
-565
-445
C. elegans
C. briggsae
-235
-125

Materials and Methods

Laboratory disposables (e.g. Pasteur pipettes, petri plates, Pipetteman tips, etc.), enzymes, and reagent grade chemicals and solvents were obtained from the following companies and suppliers: Amersham, BioRad, Boehringer-Mannheim, Fisher Scientific, Gibco-BRL, New England Biolabs, Pharmacia, Promega, Qiagen, Sigma, and VWR. All oligonucleotides used in this thesis were purchased from either Gibco-BRL or The University of Calgary *UCDNA* Services. Oligonucleotide names and sequences are listed in table 1.

Nematode strains used are as follows: *C. elegans* wildtype (var. Bristol N2), *C. elegans ges-1* null allele strain (JM1041), *C. briggsae* wildtype (var. Gurarat AF16). Nematode strains were obtained from Dr. J. D. McGhee (The University of Calgary, Calgary), and the *Caenorhabditis* Genetics Center (University of Minnesota, St. Paul).

Construction of the enhancer assay vector used in this thesis was made possible through the generosity of the following people: pPD96.04 was a gift from Dr. A. Fire (Carnegie Institution of Washington, Baltimore); the *hsp-16* promoter was removed from pPC16.48-1 (see Stringham *et al.*, 1992), which was donated by Dr. E. P. Candido (University of British Columbia, Vancouver).

Table 1: List of oligonucleotide names and sequences used in this thesis. Sequences are written 5' to 3'. Numbers after the oligonucleotide name refer to the nucleotide coordinates based on the *Cb-ges-1* clone, pJM102. Orientation of each primer is indicted by the order of coordinates. Lower case letters refer to sequences outside of the pJM102 clone; underlined sequences correspond to the GATA sites.

MC1 [a.k.a. 1474-1495A (Cbriggs)]:
1495-1474
CGGTCTATGGGGATAAATTGAG
SM3: 772-788
TCGAGCCTCATTGTGGG
SM4: 1022-1040
GAGTAACAAAGTGGTATGG
SM5: 988-971 (5' tail is antisense SM4)
CCATACCACTTTGTTACTCAGCAG TTTTTCTTCACG
SM6: 5004-4986
CAGAAGTCTGGACCTTGGC
SM7: 3572-3588
CTGGAATAATCTTCGTGAGAG
SM8: 4031-4047
GAATGTTTCCAGGTTGC
SM9: 3746-3730 (5' tail is antisense SM8)
GCAACCTGGAAACATTCGCAAAA AGGGCGATAGG
SM10: 1331-1349
CACATATGAGATGGGCTTC
SM11: 1025-1008 (5' tail is antisense SM4)
GAAGCCCATCTCATATGTGACTCT ACGTTCTTATCAG
SM12: 1006-985 (5' tail is antisense SM4)
CCAATACCACTTTGTTACTCATAA CACTTTTCTTATCACAGC
SM13: 988-971 (5' tail is antisense CbGATA1A)
CTTATCAGTATAACACTTTTCGCA GTTTTTCTTCACG
SM14: 2091-2072
GTGAACCCAGAAAAGAACTG
SM15: 1687-1703
GAAATTCCAGATGCGGG
SM16B: 1321-1340 (5' tail is antisense SM15)
CCCGCATCTGGAATTTCCTCATAT GTGCACTTCGAGC
SM17: 3837-3855 (5' tail is antisense SM19)
CCTATCGCCCTTTTTGCTTTGTGC CGTTGGACGACG
SM18: 3950-3931 (5' tail is antisense SM8)
GCAACCTGGAAACATTCCCTAATT CTCCTGCTTATCC
SM19: 3746-3730
GCAAAAAGGGCGATAGG
CbGATA1A: 995-1015 ($GATA_{Down}$)
GAAAAGTGTTATACTGATAAG
CbGATA1B: 1008-995 (mutant $GATA_{Up}$)
GTATAACACTTTTCCCGCTGCAG CAGTTTTTC
CbGATA2B: 977-1024 (mutant $GATA_{Down}$)
GAAAAGTGTTATACGACGCCGAA CGTAGAG
CbGATA2C: 1008-986 ($GATA_{Up}$)
GTATAACACTTTTCTTATCACAG
CbGATA-F: 984-1019
atgcTGCTGTGATAAGAAAAGTGTT ATACTGATAAGAACG
CbGATA-R: 1019-984
gcatCGTTCTTATCAGTATAACACT TTTCTTATCACAGCA
CbUpGATA-F: 985-999
atgcGCTGTGATAAGAAAAG
CbUpGATA-R: 999-985
gcatCTTTTCTTATCACAGC
CbDnGATA-F: 1004-1019
atgcTATACTGATAAGAACG
CbDnGATA-R: 1019-1004
gcatCGTTCTTATCAGTATA
CeGATA-F
atgcCAACTGATAGCAAAACTGATA AGGGTCAAA
CeGATA-R
gcatTTTGACCCTTATCAGTTTTGC TATCAGTTG

Microscopy and Photography

Worms were observed using a Wild Heerbrugg M5A dissecting microscope. Microinjections were performed on a Zeiss IM 35 inverted compound microscope equipped with differential interference contrast (DIC) optics. Histochemically stained embryos were observed using a Zeiss Universal microscope equipped with Nomarski optics at 400x. GFP fluorescence was detected on the Zeiss Universal microscope using UV filters (Blue 450-490 filter set: excitation - BP450-490; beam splitter - FT510; barrier - LP520) and exposing the embryos to a UV light from a Zeiss UV light source.

Photographic images were recorded on Kodak Ektachrome 400 film at ASA 100 for bright field and ASA 400 for GFP fluorescence. Some of the developed slide pictures were then digitized using a slide scanner (Polaroid Sprint Scan 35 Plus). Scanned images were further manipulated using Adobe Photoshop (version 3.0).

Nematode Care

Worm strains were maintained essentially as described by Brenner (1974). Worms were grown on nematode growth media (NGM) agar plates seeded with a lawn of *E. coli* OP50. Stocks were generally maintained at either 10°C or 20°C.

Microinjections and Selection of Transformants

Plasmids used for microinjection were usually prepared by alkaline lysis of the bacterial cells (containing the plasmid of interested) followed by purification on Qiagen columns (Qiagen), WizardPrep columns (Promega), or PEG precipitation (as described in protocols provided by Qiagen, Promega, and ABI Systems, respectively). A plasmid containing a particular test construct was mixed in 1x TE buffer (10 mM Tris-HCl, 1 mM Na_2EDTA, pH 8.0) with the *rol-6* (su1006) marker plasmid (pRF4 from Dr. A. Fire, Carnegie Institution of Washington, Baltimore) at a combined concentration of 50 to 100 μg/mL and injected into the syncytial gonad arms of young adult hermaphrodite worms (described in Mello *et al.*, 1991). *Cb-ges-1* constructs were injected into *C. elegans* JM1041 (the *Ce-ges-1* null strain) or *C. briggsae* AF16 (wildtype for *Cb-ges-1*; a *ges-1* null mutant strain was unavailable). Wildtype *C. elegans* (N2) was used when esterase activity detection was not required (e.g. GFP or β-galactosidase reporter constructs). F1 progeny were screened for the rolling phenotype caused by the pRF4 marker plasmid, with transformed progeny transferred to a separate plate. Lines in which F2 (and subsequent generations) displayed the rolling phenotype were considered to be heritable.

Histochemistry

Embryos were removed from NGM agar plates, placed on a two-well microscope slide, and soaked in 0.5% NaOCl for 3 minutes to permeablize the

egg shell, then rinsed in an M9/0.5% bovine serum albumin solution (Edgar and McGhee, 1986; Wood, 1988). Embryos were transferred to subbed microscope slides (10 μL of 2% gelatin spread on the slide, and air dried) with a glass coverslip placed over top of the embryos (3M double-sided tape was applied to the dried subbed slides to accommodate an 22 mm glass microscope coverslip). The egg shells surrounding the embryos were cracked by applying gentle pressure to the coverslip. Embryo fixative (2.25% paraformaldehyde, 0.1% glutaraldehyde) was added for 3 minutes at 4°C. Fixed embryos were rinsed with 0.25 M phosphate buffer, pH 7.2. A full description of embryo fixation can be found in Egan and McGhee (1986).

For detection of GES-1 activity, fresh esterase staining solution (10 μL 4% $NaNO_2$, 10 μL pararosaniline stock [400 mg pararosaniline-HCl+8 mL dH_2O+2 mL concentrated HCl], 400 μL 2.8% Na_2HPO_4, 10 μL 0.2 M NaOH, 10 μL α-napthyl acetate [20 mg/mL in acetone]) was prepared and added to the fixed embryos and incubated for 3 minutes (exceptions are noted in figures) at room temperature (Edgar and McGhee, 1986).

For detection of β-galactosidase activity, fresh β-galactosidase staining solution (40 mM NaH_2PO_4, 210 mM Na_2HPO_4, 1 mM $MgCl_2$, 5 mM $K_4Fe(CN)_6$, 5 mM $K_3Fe(CN)_6$, 40 μg/mL SDS, 240 μg/mL X-Gal in DMF) was prepared and added to the fixed embryos and incubated for 12 hours at 37°C (Fire *et al.*, 1990).

For GFP detection, embryos were fixed and then exposed to a UV light source (Zeiss) on a Zeiss Universal microscope equipped with Nomarski optics (also see Chalfie *et al.*, 1994, and). See Microscopy and Photography for UV filter set used to detect GFP fluorescence.

Molecular Biology Techniques

General DNA Manipulations

Standard molecular biology techniques, such as isolation of plasmid DNA, restriction enzyme digests, enzyme modifying reactions, separation of DNA samples on agarose and polyacrylamide gels, etc. were performed as described in Sambrook *et al.* (1989) and *Current Protocols in Molecular Biology* (Ausubel *et al.*, 1987-present). Subcloning manipulations were performed in pBluescript SK+ vectors (Stratagene). Plasmids were propagated in *E. coli* strains (JM109 or DM1) grown on 2xYT media with ampicillin (100 μg/mL) antibiotic (Sambrook *et al.*, 1989).

Restriction Enzyme Based Deletions

The parent plasmid (pJM102) for the promoter analysis contained: 1.69 kb of *Cb-ges-1* 5' flanking region (upstream of the ATG translation initiation codon); 2.00 kb containing the entire *Cb-ges-1* coding region; and 2.22 kb of 3' flanking region (downstream from the poly-adenylation addition site). pJM102

was cloned into the SalI/EcoRI sites of pBluescript SK+ (Kennedy *et al.*, 1993). Figure 7 is a schematic diagram of the pJM102.

The following deletion series was created by deleting between unique or rare-cutting restriction enzyme sites: pJM102ΔSA, pJM102ΔRI*, pJM102ΔAN, pJM102ΔBS, JM102ΔE_{II}6, pJM102ΔE_{II}4, pJM102Δ3', pJM102ΔNS. Table 2 shows the restriction sites used to create these deletions. All deletions were confirmed by sequencing.

Splicing By Overlap Extension (SOEing)

Figure 8 outlines the principle behind the SOEing technique used to create deletions within a desired gene region (White, 1993). Two rounds of PCR are required. In the first round, primer pairs a/b and c/d are used to create products AB and CD, respectively. The design of primer b is such that in addition to containing gene sequence I, primer b also contains the antisense sequence to primer c (of gene sequence II). This allows products AB and CD to anneal with each other during the second round of PCR. Upon completion of the second round, products AB and CD will be joined together with the desired fragment having been deleted. Unique or rare restriction sites are used to aid in subcloning the PCR deleted product ABCD into the parental construct.

PCR reactions were performed in 20 μL volumes using the following conditions: 1x PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 20 μg/mL gelatin), 3 nmol dNTP, 15 pmol each primer, 0.5 units Taq, 0.5-1 μL

Figure 7: The longest genomic *C. briggsae ges-1* clone, pJM102. The pJM102 clone begins at the SalI restriction site and ends at the EcoRI restriction site. The thickest lines represent non-coding DNA; exons are indicated by lines of medium thickness; introns are shown by the thinnest lines. The numbers listed correspond to the nucleotide position of restriction sites, ATG codon, stop codon, poly A signal site, and exon/intron splice sites. EcoRI* refers to an EcoRI star activity induced cleavage site. Please note that SmaI is actually part of the multiple cloning site in pBluescript and the number in quotes is a continuation from the EcoI site distance. $GATA_{Up}GATA_{Dn}$ and PRA indicate the positions of the gut activator-pharynx/rectum repressor and pharynx/rectum activator regulatory elements, respectively.

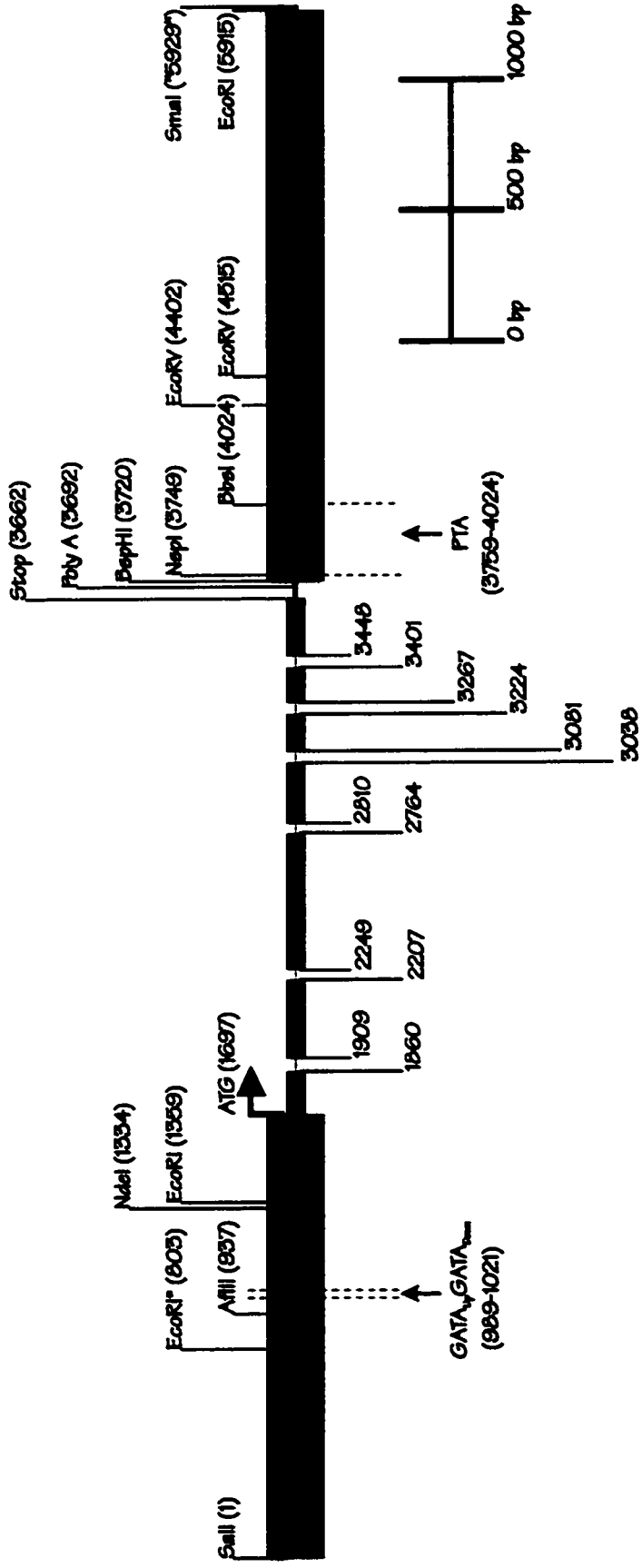

SmaI (5923*)
EcoRI (5915)
EcoRV (4402)
EcoRV (4515)
BbsI (4024)
NspI (3749)
BspHI (3720)
Poly A (3692)
Stop (3662)
PTA (3759-4024)
3448
3401
3267
3224
3061
3038
2810
2764
2249
2207
1909
1860
ATG (1697)
NdeI (1334)
EcoRI (1359)
EcoRI* (805)
MfiI (857)
GATA1,GATA2 (989-1021)
SalI (1)
1000 bp
500 bp
0 bp

Table 2: Summary of the *Cb-ges-1* restriction enzyme deletion and multimerization constructs.

A) Restriction enzyme based deletion constructs. 5' and 3' cut sites with enzyme used are shown.

B) Multimerization of *C. briggsae* GATA sites for enhancer assay experiments. Primer pairs used in multimerization step are listed.

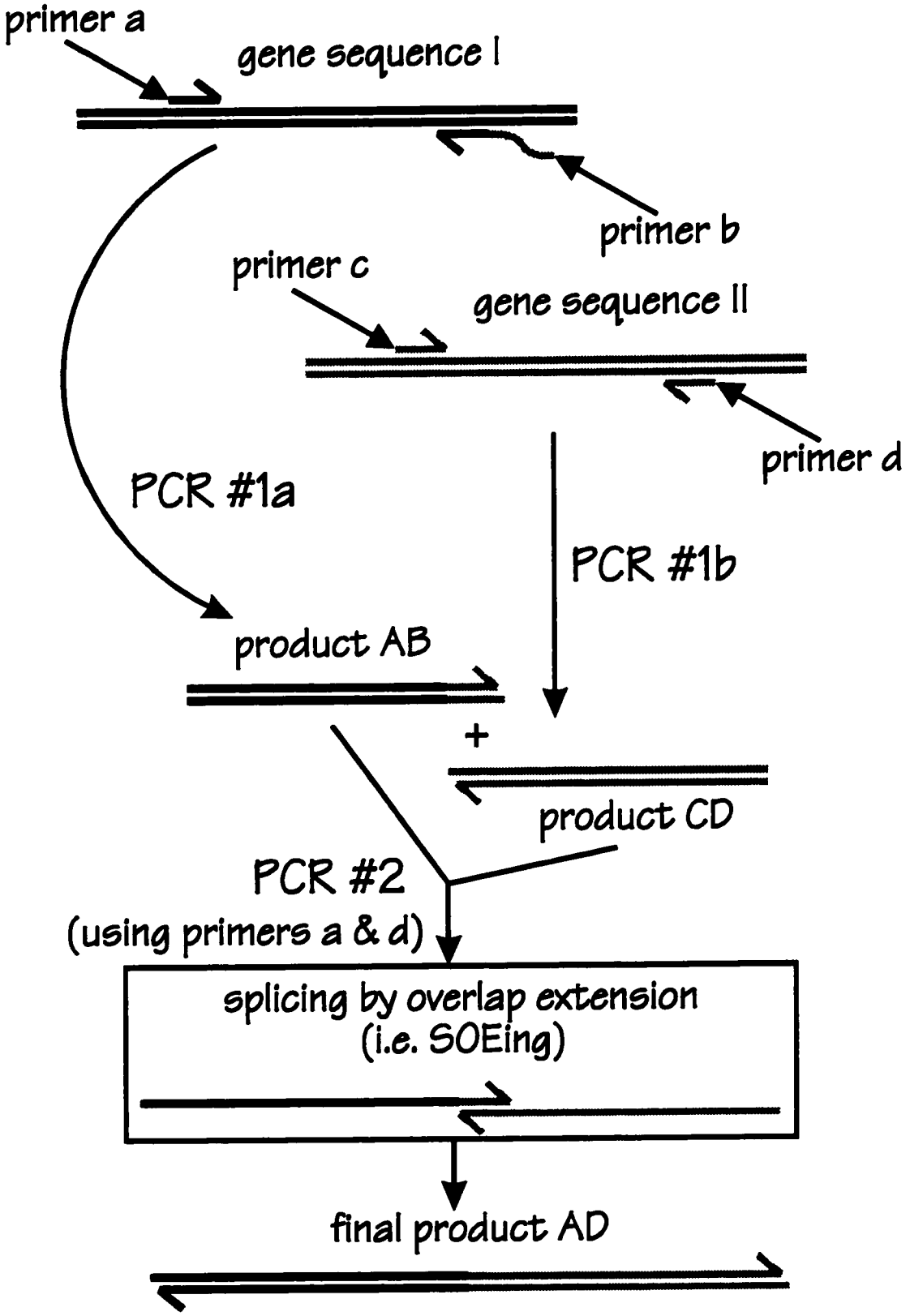
A)

Construct	5' Enzyme (Cut Site)	3' Enzyme (Cut Site)
pJM102ΔSA	SalI (1)	AflII (937)
pJM102ΔRI*	EcoRI* (803, star activity)	EcoRI (1359)
pJM102ΔAN	AflII (937)	NdeI (1334)
pJM102ΔBS	BbsI (4024)	SmaI (5915)
pJM102ΔE$_{II}$6	EcoRV (4402)	SmaI (5915)
pJM102ΔE$_{II}$4	EcoRV (4515)	SmaI (5915)
pJM102ΔNS	NspI (3749)	SmaI (5915)
pJM102Δ3'	BspHI (5092)	SmaI (5915)

B)

Multimerized GATA Region	Sense Primer	Anti-Sense Primer
p6xFGATAs p6xRGATAs	CbGATA-F	CbGATA-R
p7xFUpGATA	CbUpGATA-F	CbUpGATA-R
p7xRDnGATA	CbDnGATA-F	CbDnGATA-R

Figure 8: Diagram of the Splicing by Overlap Extension (SOEing) technique. Primer a and b have sequences unique to gene sequence I. Primer b also contains sequences from gene sequence II. Primer c and d have sequences unique to gene sequence II. After the first round of PCR, products AB and CD and primers a and d are added to a second PCR reaction (the SOEing step) to create the hybrid product, AD. Figure adapted from White (1993, p.253).


primer a
gene sequence I
primer b
primer c
gene sequence II
primer d
PCR #1a
PCR #1b
product AB
+
product CD
PCR #2
(using primers a & d)
splicing by overlap extension
(i.e. SOEing)
final product AD

(~0.5-1 ng) DNA template. Reactions were incubated at 94°C for 5 minutes; 25 cycles of 1 minute at 94°C, 1 minute at 34-68°C (primer dependent), 2 minutes at 72°C; and finally 5 minutes at 72°C. PCR products from the first amplification were run on a 1% agarose gel; the band of interest was cut out and purified using the Prep-A-Gene protocol (BioRad). Aliquots of the purified PCR products (typically 1-2 μL) were then diluted to 1:10 and 1:100 in sterile distilled water. The second PCR amplification was done using the same conditions described above, except the PCR products from the first PCR amplification (undiluted, 1:10, and 1:100 dilutions) were used as the template. PCR products from the second PCR amplification were run on a 1% agarose gel with the band of interest cut out and purified using the Prep-a-gene protocol (BioRad). The final PCR product was digested with the appropriate restriction enzymes (e.g. AflII and NdeI for GATA site deletions), and ligated into the appropriate sites of pJM102 (e.g. AflII and NdeI sites for GATA site deletions).

PCR Based Deletions

PCR based deletion (also known as SOEing) was used to create the following deletions: pJM102ΔGATAs (989 to 1021); pJM102ΔUpGATA (989 to 996); pJM102ΔDnGATA (1007 to 1021); pJM102Δ1 (1025 to 1330); pJM102ΔPRA1 (3745 to 4030). See table 3 for list of oligonucleotides used to create deletions (specific primer sequences are listed in table 1). All constructs were confirmed by sequencing.

Table 3: Summary of the *Cb-ges-1* PCR based deletions and mutation constructs. Primers used for each PCR reaction are indicated.

Construct	PCR#1a	PCR#1b	PCR#2	Deletion End Points
pJM102ΔGATAs	SM3-SM5	SM4-MC1	SM3-MC1	989 - 1021
pJM102ΔUpGATA	SM3-SM13	CbGATA1A-MC1	SM3-MC1	989 -996
pJM102ΔDnGATA	SM3-SM12	SM4-MC1	SM3-MC1	1007 - 1021
pJM102Δ1	SM3-SM11	SM10-MC1	SM3-MC1	1025 - 1330
pJM102ΔPRA1	SM7-SM9	SM8-SM6	SM6-SM7	3745 - 4030
pJM102mGATAs	SM3-CbGATA1B	CbGATA2B-MC1	SM3-MC1	N/A
pJM102mUpGATA	SM3-CbGATA1B	CbGATA1A-MC1	SM3-MC1	N/A
pJM102mDnGATA	SM3-CbGATA2C	CbGATA2B-MC1	SM3-MC1	N/A

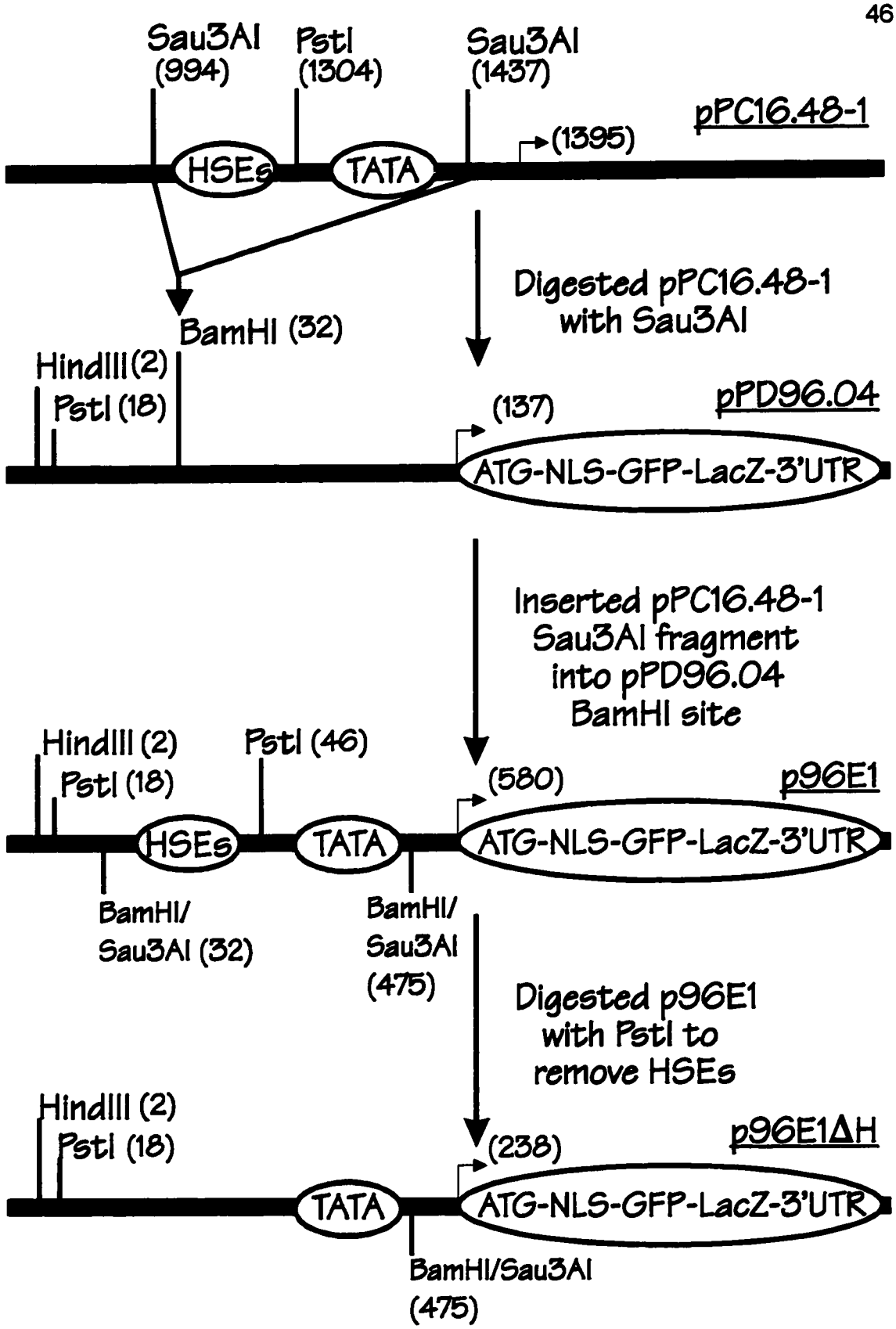
PCR Based Mutation

A series of site specific mutations were created using the same SOEing methodology described for the PCR directed deletions; however, primers were designed that contained mutated GATA sites in the middle of the primer sequence. Only the six bases of the canonical WGATAR sites were altered; the TGATAA sequence was converted to CCGCTG. The following are the contructs containing the mutated GATA site(s): pJM102mGATAs in which both GATA sites between 989 and 1021 were mutated; pJM102mUpGATA in which the upstream GATA site (989 to 994) was mutated; pJM102mDnGATA in which the downstream GATA site (1009 to 1014) was mutated. See table 3 for primer combinations used to create the mutation constructs (specific primer sequences are listed in table1). All constructs were confirmed by sequencing.

Construction of Enhancer Assay Vectors for C. briggsae ges-1 Promoter Elements

To create the basal promoter-reporter gene construct used to test for enhancer activity (p96E1ΔH), several vectors were spliced together (summarized in figure 9). A Sau3AI fragment from the pPC16.48-1 vector, which contains the *hsp-16* basal promoter plus two heat shock response elements (HSEs), was inserted into the BamHI site of pPD96.04. This new vector, p96E1, contains the *hsp-16* promoter from pPC16.48-1, and the GFP::β-galactosidase reporter gene from pPD96.04. p96E1 was then digested with PstI to remove the

Figure 9: Construction of the p96E1ΔH vector used for enhancer assay experiments. The Sau3AI fragment, which contains the *hsp-16* basal promoter and two heat shock response enhancer elements (HSEs), was inserted into the BamHI site of pPD96.04 GFP-LacZ reporter gene vector to create p96EI. The two HSEs were removed by digestion of p96EI with the restriction enzyme PstI to create p96EIΔH.


Sau3AI (994)
PstI (1304)
Sau3AI (1437)
pPC16.48-1
(1395)
HSEs
TATA
Digested pPC16.48-1 with Sau3AI
BamHI (32)
HindIII (2)
PstI (18)
pPD96.04
(137)
ATG-NLS-GFP-LacZ-3'UTR
Inserted pPC16.48-1 Sau3AI fragment into pPD96.04 BamHI site
HindIII (2)
PstI (18)
PstI (46)
(580)
p96E1
HSEs
TATA
ATG-NLS-GFP-LacZ-3'UTR
BamHI/ Sau3AI (32)
BamHI/ Sau3AI (475)
Digested p96E1 with PstI to remove HSEs
HindIII (2)
PstI (18)
(238)
p96E1ΔH
TATA
ATG-NLS-GFP-LacZ-3'UTR
BamHI/Sau3AI (475)

two HSEs, thereby leaving only the basal promoter from pPC16.48-1 behind. The final enhancer assay vector, p96E1ΔH, allows insertion of a desired sequence upstream of the basal promoter.

Pairs of complementary oligonucleotides were designed such that ligation would result in a head-to-tail arrangement of the individual GATA regions. After multimerization, the GATA regions were end-filled and subcloned into the EcoRV site of pBluescript SK+. The multimerized GATA regions were cleaved with PstI/HindIII and ligated into the PstI/HindIII sites of p96E1ΔH. The multimerized *Cb-ges-1* GATA sites (tandem, upstream, and downstream) were inserted into the p96E1ΔH enhancer assay vector. Final enhancer assay constructs were confirmed by sequencing, and are named as follows (written with the number of copies, the orientation, and which GATA site was inserted): p6xFGATAs, p6xRGATAs, p7xFUpGATA, p7xRDnGATA (F = forward sequence orientation, R = reverse sequence orientation). See table 2 for the list of primers used to create the multimerized GATA regions (specific primer sequences are listed in table 1).

Bandshifts

The *C. elegans* ELT-2 protein used to bind the DNA probe was produced using the T_NT Coupled Reticulocyte Lysate System (Promega). Lysate was programmed (as described by the manufacturer) to produce ELT-2 protein by adding full length *elt-2* cDNA plasmid template to the *in vitro*

transcription/translation reaction. An unprogrammed lysate reaction, which substituted water for the *elt-2* plasmid, was used in control bandshift experiments to identify any DNA binding proteins that were present in the rabbit reticulocyte lysate.

Sense oligonucleotides were labelled with γ-P^{32}-ATP using standard kinase reactions (Sambrook *et al.*, 1989). Spin columns containing Sephadex G-25 resin (Pharmacia) were used to remove unincorporated nucleotides (Sambrook *et al.*, 1989). Double stranded probes used for bandshifts were prepared by annealing sense oligonucleotides (containing either the *Ce-ges-1* or the *Cb-ges-1* GATA sites) to a 10-fold excess of the appropriate antisense oligonucleotide, as described in Stroeher *et al.* (1994). The *Ce-ges-1* GATA probe used oligonucleotides CeGATA-F and CeGATA-R, and the *Cb-ges-1* GATA probe used oligonucleotides CbGATA-F and CbGATA-R. Specific oligonucleotide sequences are listed in table 1.

Bandshift reactions consisted of 1-2 μL of lysate (programmed or unprogrammed), 50 000 cpm of labelled probe, 0.1-1 μg of poly dIdC:dIdC (Boehringer-Mannheim) as non-specific competitor, 0-200 ng of unlabelled probe as specific competitor, and 1x bandshift binding buffer (1x BSB: 10 mM HEPES, pH 7.6, 1 mM DTT, 5 mM Na_2EDTA pH 8.0, 5% glycerol). Bandshift reactions were incubated at 20°C for 20-30 minutes. The samples were loaded onto 1.5 mm thick, 1x TBE (89 mM Tris-HCl, 89 mM Boric acid, 2 mM Na_2EDTA) non-denaturing 5% polyacrylamide gels containing 5% glycerol. Gels were

electrophoresed at 20 V/cm at 4°C using a BioRad Mini-gel system, dried under vacuum, and exposed to Kodak BioMax film.

Sequencing

DNA template for sequencing was prepared via an alkaline lysis-PEG precipitation protocol supplied by Applied Biosystems, Incorporated. Sequencing reactions were performed using the Taq DyeDeoxy Terminator Cycle Sequencing Kit (ABI). Completed sequencing reactions were sent to The University of Calgary Sequencing Service for analysis. The integrity of all plasmids used for microinjection and *in vitro* protein production were confirmed by sequencing.

The 2.2 kb 3' non-coding region of the largest *Cb-ges-1* clone (pJM102) had not been sequenced at the beginning of this project. Sequencing was done by creating a series of nested deletions using the Exonuclease III-mung bean nuclease protocol described in Henikoff (1984). Briefly, target DNA was digested with two restriction enzymes, one that created an Exonuclease III sensitive 5' overhang in the direction desired for deletion, and the other created an Exonuclease III insensitive 3' overhang for protecting the DNA from Exonuclease III activity. The next step was to add Exonuclease III (Pharmacia) to the prepared target DNA. Aliquots of the Exonuclease III reaction were removed at regular intervals, with the Exonuclease III activity stopped immediately. This created a series of deletions of varying lengths (i.e. nested

deletions). The ends of the DNA were filled in using a combination of DNA Polymerase I Klenow Fragment (Boehringer-Mannheim) and Mung Bean Nuclease (Pharmacia) to create blunt ended DNA that was subsequently ligated together using T4 Ligase (Gibco-BRL). All enzymatic reactions were carried out according to manufacturers recommendations. pBluescript SK+ specific primers (e.g. M13 reverse) were used for sequencing of the deletions.

Sequence Manipulations

Assembly of the pJM102 3' non-coding sequence and all other sequence manipulations (comparisons, dot matrices, alignments, etc.) were performed using the Wisconsin Package software (version 9.1) produced by the Genetics Computer Group.

Results

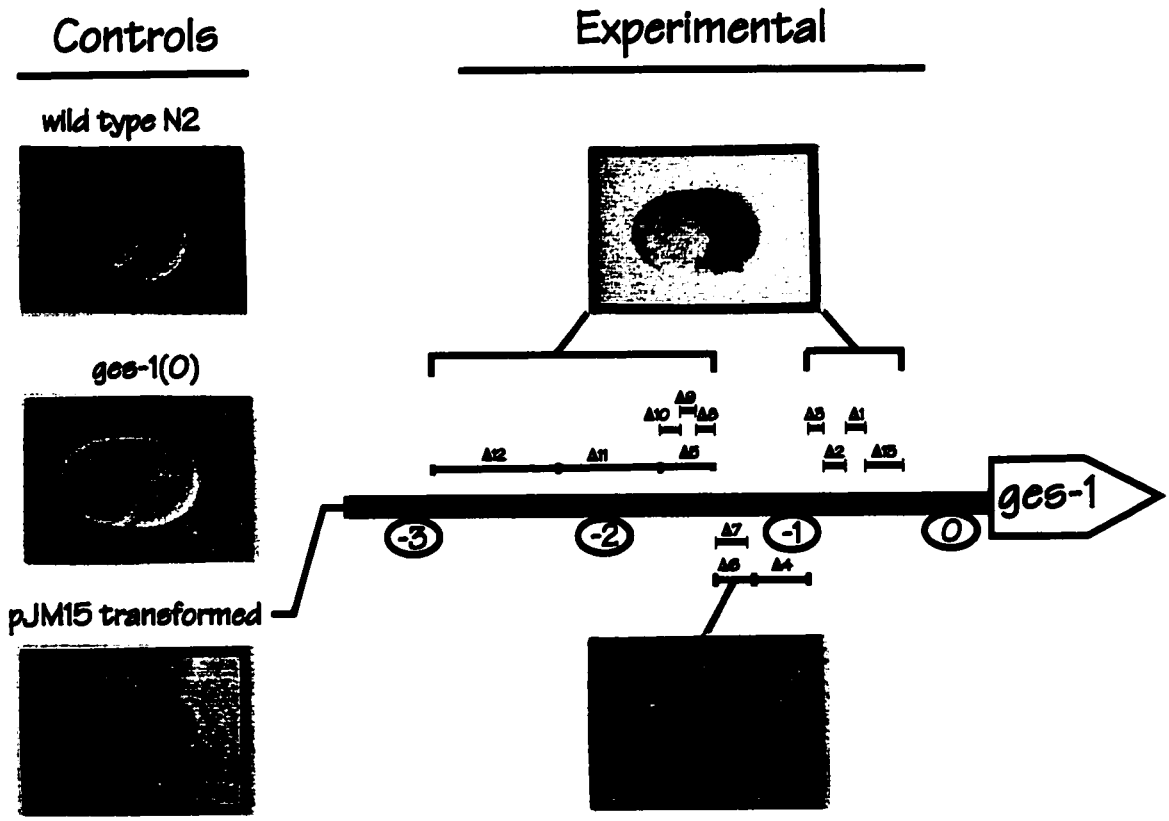
Overview of Previous Analysis of the C. elegans ges-1 Gene

Endogenous embryonic *C. elegans* GES-1 (Ce-GES-1) activity is present solely in the gut lineage, as shown by histochemical staining for esterase activity (Edgar and McGhee, 1986). Investigation of *Ce-ges-1* gene regulation was previously carried out by introducing both unidirectional and internal deletions into the 5' promoter region (Aamodt *et al.*, 1991; Egan *et al.*, 1995). Enhancer assays and sequence mutation analyses were done to characterize the identified regulatory elements (Egan *et al.*, 1995). Mapping and characterization of important regulatory elements was done by injection of various *Ce-ges-1* constructs into the *C. elegans ges-1* null mutant (JM1041), followed by esterase staining of embryos produced from transformed worms. Results from these experiments led to the identification of a tandem pair of GATA sites that could direct *Ce-ges-1* expression specifically to the gut (summarized in figure 10). Deletion of the tandem pair of GATA sites eliminated *Ce-ges-1* expression in the gut; however, *Ce-ges-1* expression was then observed in the pharynx and tail regions of the *C. elegans* digestive tract.

In order to simplify the comparison of the regulatory mechanisms between the two *ges-1* genes, results from previous analyses of *Ce-ges-1* regulation are presented with the current analyses I have done for *Cb-ges-1*.

Figure 10: Summary of the *C. elegans ges-1* deletion analysis.

"Control embryos stained for esterase activity are shown at the left: top, wildtype N2 embryo, stained for 60 min; middle, embryo from a *ges-1*(0) strain (JM1041) stained for 60 min; bottom, embryo from a *ges-1*(0) strain transformed with the plasmid pJM15 containing the intact *ges-1* gene with 3.3 kb of upstream sequence, stained for 3 min. At the top right is an embryo transformed with the Δ5 deletion and showing the wildtype expression pattern (i.e., in the gut). This embryos is representative of embryos transformed with the other deletion constructs indicated with brackets, all of which gave essentially wildtype staining patterns. At the bottom right is an embryo transformed with the Δ6 deletion construct of pJM15, showing *ges-1* expression in the region of the developing pharynx and in the tail. Unless indicated otherwise, all embryos from transformed lines were stained for 3 min. . . . Coordinates of the *ges-1* 5'-flanking region are shown (circled numbers) as kilobases upstream of the *ges-1* ATG codon." Figure and text taken from Egan *et al.* (1995).


Controls
Experimental
wild type N2
ges-1(0)
pJM15 transformed
ges-1
-3
-2
-1
0

Investigation of Cb-ges-1 Expression

Previous work demonstrated that endogenous Ce-GES-1 and Cb-GES-1 activities could only be detected in the gut of *C. elegans* and *C. briggsae* (Kennedy *et al.*, 1993). Since the endogenous expression patterns of the two *ges-1* genes were found to be similar, it was of interest to identify the *Cb-ges-1* regulatory elements for comparison with the *Ce-ges-1* regulatory elements (in terms of sequence, spatial arrangement, and function). Mapping and characterization of the *Cb-ges-1* regulatory elements was carried out through a series of deletions, mutations, and manipulations of spatial position.

Cb-ges-1 constructs that were used to identify and characterize the *Cb-ges-1* regulatory regions were injected into *C. elegans* and *C. briggsae*, with heritable lines examined for gene expression (esterase or reporter gene). While Kennedy *et al.* (1993) had previously demonstrated that *C. briggsae* could be transformed using *C. elegans* transformation methods, I found that stable transformation of *C. briggsae* was not as efficient as for *C. elegans* transformation. Inefficiency in producing heritably transformed *C. briggsae* lines was also noted during the study of another *C. elegans* gene, *dpy-7* (Gilleard *et al.*, 1997a). Given the difficulties in producing stable *C. briggsae* transformed lines, it was decided that all constructs would be injected into *C. elegans* first. When a particular deletion or mutation construct showed an expression pattern different from worms transformed with the longest *Cb-ges-1* clone (pJM102, see figure 7 in Materials and Methods), this construct was used to transform *C.*

briggsae. Although not all constructs were successfully introduced into *C. briggsae*, it appears that similar patterns of expression can be expected from transformations of both *C. elegans* and *C. briggsae*.

The parental plasmid used for promoter analysis of *Cb-ges-1* was pJM102, which contains the entire 2.00 kb open reading frame plus 1.69 kb of 5' flanking region upstream from the ATG and 2.22 kb of 3' flanking DNA downstream from the poly(A) addition site (shown in figure 7 of Materials and Methods). As will be discussed later in more detail, the endogenous *Cb-ges-1* expression pattern was observed in the gut (i.e. E lineage) of the worm. Transformation of pJM102 into JM1041 (a *C. elegans ges-1* null mutant) and AF16 (wildtype *C. briggsae*), with subsequent esterase staining, recreated the *C. briggsae* and *C. elegans* endogenous esterase gut staining pattern. A small amount of additional staining (~10-20%, compared to total esterase staining) was observed in the pharynx region. Esterase staining may also have been present in the rectum region, but any rectum staining was difficult to detect because the heavy gut staining often overlapped into the rectum region.

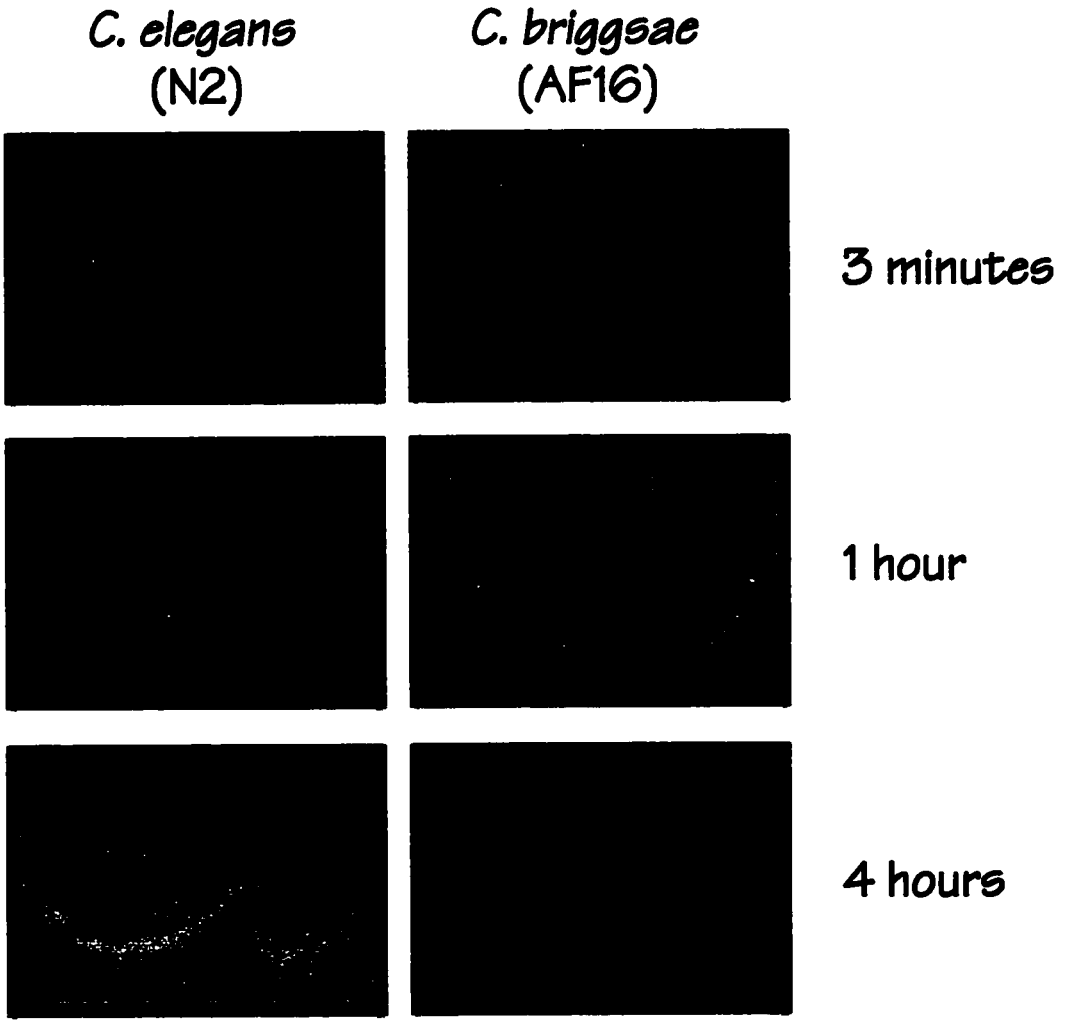
Both *C. elegans* and *C. briggsae* embryos containing the pJM102 transgene showed a slightly different staining pattern than was seen for the endogenous *Cb-ges-1* gene. This "extra" pharynx/rectum staining seen in pJM102 transformed worms could be explained in several ways: 1) transformation of *C. elegans* and *C. briggsae* does not accurately reflect a gene's endogenous expression pattern; 2) pJM102 does not contain the

complete set of regulatory elements required for proper *Cb-ges-1* expression; or, 3) weak endogenous expression of *Cb-ges-1* is present but is not detectable in *C. briggsae* at routine staining times (i.e. three minutes for transformed worms and one hour for endogenous *ges-1* staining), but is seen when multicopy arraysfrom the transformation overexpress the gene.

Of the three possibilities mentioned above, I showed that the third possibility is most likely. As shown in figure 11, neither *C. elegans* nor *C. briggsae* showed endogenous pharynx/rectum esterase activity in normal esterase staining times of three minutes and one hour (top and middle, respectively); three minute staining times were routinely used for transformed embryos, while one hour staining times were used to detect endogenous GES-1 activity. When the esterase staining reaction was carried out for an extended period of time (four hours), endogenous esterase activity was detected in the pharynx of *C. briggsae*, but not of *C. elegans* (figure 11). Endogenous expression of *Cb-ges-1* in the pharynx was first detected at four hours of staining and accounted for less than 1% (estimated) of the total esterase staining. This data suggested that most (if not all) of the major regulatory elements were indeed present in the *Cb-ges-1* parental plasmid, pJM102. The transformation assay faithfully recreated the expression pattern of the endogenous gene, though perhaps overstated the degree of *Cb-ges-1* pharynx/rectum expression.

Expression of *Ce-ges-1* ouside of the gut was originally reported as being in the pharynx and tail portions of the *C. elegans* digestive tract

Figure 11: Endogenous esterase staining of *C. elegans* (N2) and *C. briggsae* (AF16) embryos. Embryos were stained for three minutes, one hour, or four hours, as indicated in the figure. Arrow points to weak pharynx esterase staining in four hour esterase stained *C. briggsae* embryo (bottom right).


C. elegans
(N2)
C. briggsae
(AF16)
3 minutes
1 hour
4 hours

(Aamodt *et al.*, 1991). The esterase staining pattern outside of the gut has now been identified as staining in the pharynx and rectum. That GATA deleted *Ce-ges-1* constructs are expressed in the pharynx and rectum was shown by Fukushige *et al.* (1996). Elimination of the ABa, ABp, and MS founder cells through laser ablation specifically demonstrated that cells descendent from these three cells that make up the pharynx and rectum. Several gene mutations have been identified in *C. elegans* that can change the "identity" of cells from specific cell lineages. Embryos containing mutations within *skn-1*, *mex-1*, *pie-1*, *pop-1*, or *pha-4* were used to study the expression pattern of the pJM15 (*Ce-ges-1* clone) or the pJM15ΔGATA$_1$GATA$_2$ transgenes. This genetic analysis again revealed that *ges-1* was being expressed in cells descended from the ABa, ABp, and MS cell lineages. From this evidence it was concluded that *Ce-ges-1* can be expressed in cells of the pharynx (ABa and MS lineage) and rectum (ABp). Since *C. briggsae* is closely related to *C. elegans*, the assumption has been made that *Cb-ges-1* expression also has the ability to be expressed in cells of the pharynx and rectum.

Identification of the Regulatory Region(s) Required For Gut Expression of the C. briggsae ges-1 Gene

Earlier transformation experiments of *C. briggsae* with the *Ce-ges-1* gene showed that some of the *cis*-regulatory elements and *trans*-acting factors must be conserved, since the *Ce-ges-1* transforming gene was properly expressed in

C. briggsae embryos (Kennedy *et al.*, 1993). Furthermore, a *Cb-ges-1* construct that deleted 800 base pairs in the 5' end of the *Cb-ges-1* gene (pJM102ΔRI*), showed a switch in expression from the gut to the pharynx/rectum (M. A. Chung, personal communication). These observations suggested that the regulatory mechanisms between *Ce-ges-1* and *Cb-ges-1* may be conserved.

Previous promoter sequence alignments of the *ges-1* genes from *C. elegans* and *C. briggsae* did not reveal many significant regions of similarity (Kennedy *et al.*, 1993). The one exception was a region spanning approximately 100 bp in the 5' promoter of both *C. elegans* and *C. briggsae*. This region is roughly 70% identical over the entire length and contains a 17 of 17 bp match in the 5' promoter of both *C. elegans* and *C. briggsae* (see figure 6 in Introduction). Deletion of this 17 bp region from the *Ce-ges-1* gene showed no obvious change in the expression phenotype in embryos (Egan *et al.*, 1995).

As the initial computer assisted sequence comparisons did not suggest many promising regions with which to begin the promoter analysis, arbitrary deletions (using unique and rare restriction enzymes) were used to survey the *Cb-ges-1* gene for regulatory regions. Figure 12 provides a summary of the 5' and 3' deletion analysis of the *Cb-ges-1* flanking regions. These survey deletions revealed the presence of regulatory elements required for *Cb-ges-1* expression in both the gut and in the pharynx/rectum regions of the digestive tract. In particular, the $GATA_{Up}GATA_{Down}$ region appears to regulate gut

Figure 12: Summary of the expression pattern seen for the deletion analysis performed on the *Cb-ges-1* gene. Black lines above the schematic diagram of the *Cb-ges-1* gene (pJM102, in grey) indicate regions that were deleted (diagram is drawn to scale); abbreviated deletion construct names are indicated above black lines. Endogenous *C. elegans* and *C. briggsae ges-1* staining patterns at three minutes and four hours are shown at top left and middle left (arrow points to weak pharynx staining in *C. briggsae* embryo), respectively. pJM102 transformed worms are shown at bottom left. Deletions that had no effect on the expression pattern are shown above the line; deletions that altered the expression patterns are shown below the line. The staining pattern for deletions around a gut activator-pharynx/rectum repressor element ($GATA_{Up}GATA_{Dn}$) are at bottom middle; arrow shows endogenous *Cb-ges-1*gut activity. The staining pattern of deletions around the pharynx/rectum activator element (PRA) are shown at bottom right; a pJM102Δ3' transformed *C. briggsae* embryo is only one example of an apparently random staining pattern. All transformed embryos were stained for three minutes. *C. elegans* transformations were done using the *ges-1* null strain (JM1041); *C. briggsae* transformations were done using the *ges-1* wildtype strain (AF16).

Wildtype - 3 min.

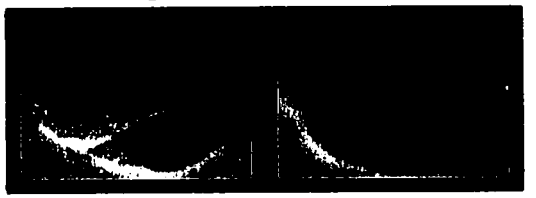
C. elegans N2 *C. briggsae* AF16

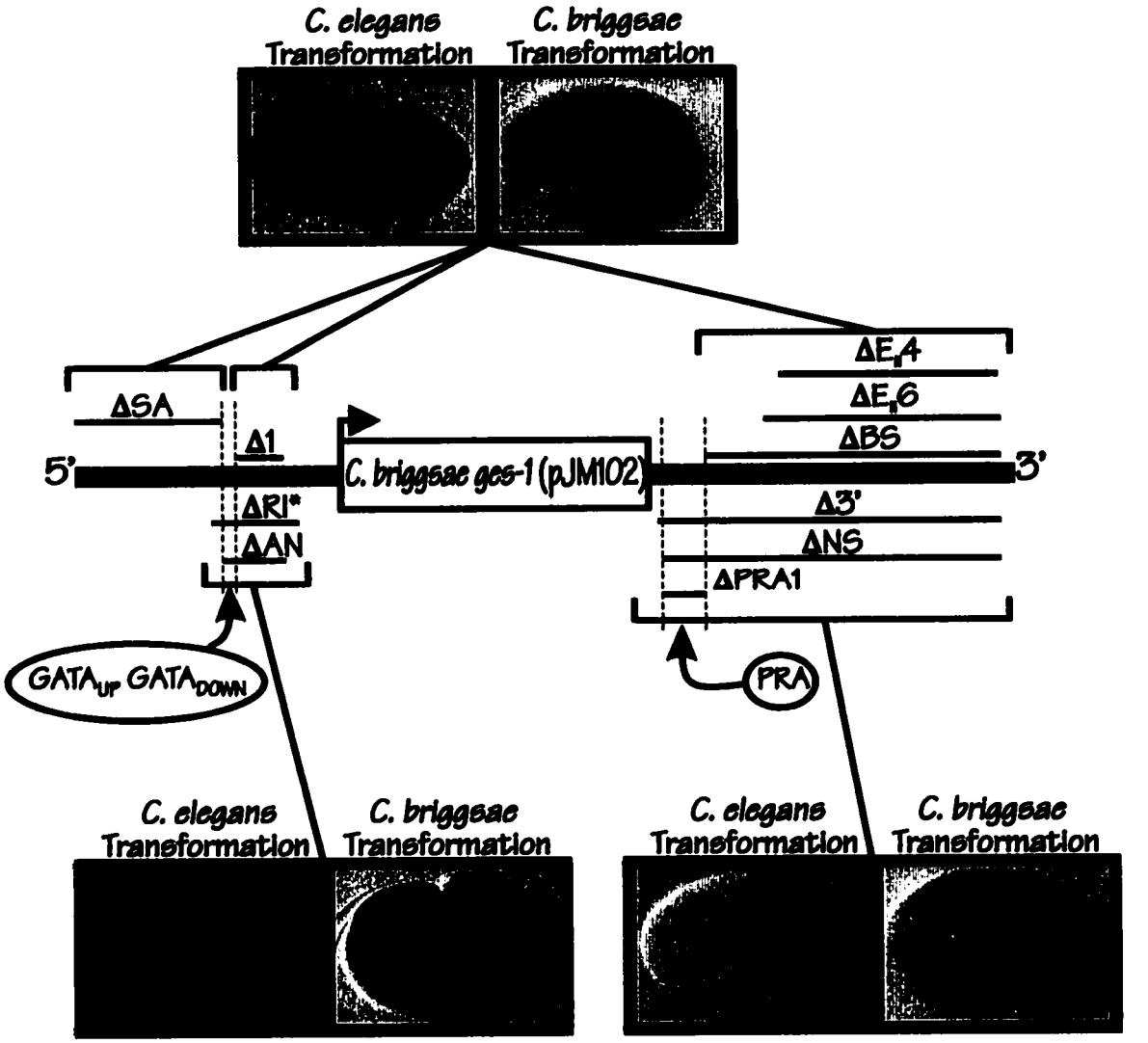
Wildtype - 4 hrs

C. elegans N2 *C. briggsae* AF16

pJM102 Transformation - 3 min.

C. elegans *C. briggsae*





activation-pharynx/rectum repression, while the PRA region seems to contain regulatory elements required for pharynx/rectum activation.

Since the original cloning of the *Cb-ges-1* gene was done, additional DNA sequence has been identified and sequenced for the *Cb-ges-1* gene. Figure 13 shows a dot matrix comparison of the longest *Ce-ges-1* clone (pJM15, 8.48 kb) and the *Cb-ges-1* clone (pJM102, 5.91 kb). The highest regions of similarity occur within the coding sequences (nucleotide positions 3316-7746 for *Ce-ges-1* and 1696-3662 for *Cb-ges-1*). Outside of the coding region, two other regions of sequence conservation were observed. At the 5' end of the *ges-1* gene is region 1, which is the same region noted earlier by Kennedy *et al.* (1993) (see figures 13 and 14). At the 3' end of the *ges-1* genes, an approximately 400 bp region (region 2) sharing a high degree of similarity was found. Approximately 100 bp of region 2 showed 76% identity (see figures 13 and 14). No other significant regions of conserved sequence were detected in the dot matrix; however, as will be described in the coming pages, two important regulatory regions were identified during the analysis of the *Cb-ges-1* promoter. The areas containing these sequences are marked in figure 13 (labelled as GATAs and PRAs) and a sequence alignment is shown in figure 15. It is interesting to note that neither the dot matrix nor the sequence alignments revealed the position of the GATAs region or the PRAs region.

Figure 13: Dot matrix sequence alignment of *C. elegans* and *C. briggsae ges-1* genes. Sequences from pJM15 (8487 bp; *Ce-ges-1*) and pJM102 (5915 bp; *Cb-ges-1*) clones were aligned using GCG "compare" and "dot plot" programs with the following parameters: window=21, stringency=14. Areas of identity are indicated by dots, with areas of high sequence conservation forming diagonal lines. The coding regions for *Ce-ges*-1 (3316-7746 bp) and *Cb-ges-1* (1696-3362 bp) are easily identified by the diagonal lines . Region 1 has an identity of 76%, and contains the previously identified 17 of 17 bp match reported in Kennedy *et al.* (1993). Region 2 has sequences showing approximately 76% identity, but this region remains to be characterized in *C. elegans*. GATAs contains a tandem pair of GATA sites that appear responsible for the gut activation-pharynx/rectum repression activity. The PRA region contains the pharynx/rectum activation activity. Neither the GATAs nor the PRA regions were detected using dot matrix analysis using a variety of parameters.

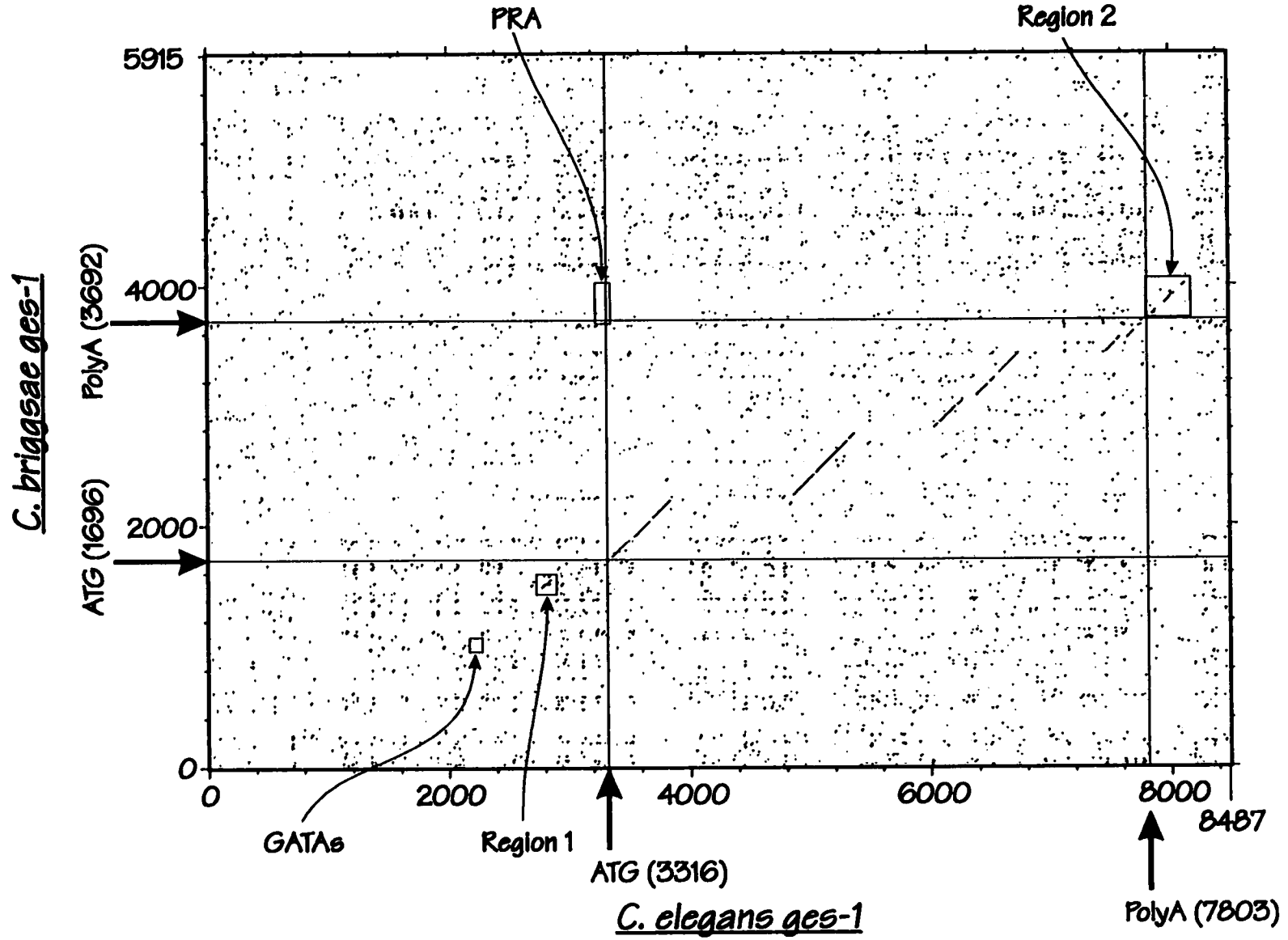

PRA
Region 2
5915
4000
2000
0
PolyA (3692)
ATG (1696)
C. briggsae ges-1
0
2000
4000
6000
8000
8487
GATAs
Region 1
ATG (3316)
C. elegans ges-1
PolyA (7803)

Figure 14: Sequence alignment of region 1 and 2 (A and B, respectively). Approximately 100 bp sharing 76% identity (both region 1 and region 2) were aligned using the GCG GAP program with the parameters set at: window=10, stringency=7. These sequence alignments were not altered after running the GCG GAP program.

A) Region 1

```
2750 AGTTGAATTTTGGCAAAATTGACCACACACCGATCAAAAACTATTGTGTG 2799
     ||    ||||| ||  |   ||    ||||||||||||||||
1460 AGACCGATTTTTGCCACGATGGGGGCACACCGATCAAAAAC..... 1500

2800 TGTGTGTATTTGTGTCTTCAATAA...TTCCTGTCTGTGTCTCGCACTAT 2846
       ||||| ||||||||   |||||   ||| |||||| ||||| ||||||
1501 ..TGTGTGTTTGTGTCACTAATAACTCTTCATGTCTGCGTCTCTCACTAT 1548

2847 CTATTGACTACAAGAACGTGATCG. 2870      C. elegans
     |||    ||   ||||||||||
1549 CTA...CCTGGCAGAACGTGATATG 1570      C. briggsae
```

B) Region 2

```
7950 ..CGGAGAGGACGACGCAAGCAAACATCGACGCACACGGCAAAAAGGCTC 7997
           |||||||| |||||||||| || ||||||||||||||  |||
3840 GTGCCGTTGGACGACGTAAGCAAACATGGAAGCACACGGCAAAAACTCTC 3889

7998 TC..CGGAAGTGGACAATAAGCCGAGTTTACAGTTGGATCGAGGGGGGTG 8045
     ||  ||||||||||||||     | | ||||||||||   || |||| |
3890 TCACCGGAAGTGGACAAT....GGCGGTTACAGTTGGCCAGAAGGGGATA 3935

8046 ACGAA 8050      C. elegans
     |   |
3936 AGCAG 3940      C. briggsae
```

Figure 15: Sequence alignment of GATAs and PRA regions (A and B, respectively). Approximately 100 bp sharing 40% and 42% identity (GATAs and PRA, respectively) were aligned using the GCG GAP program with the parameters set at: window=10, stringency=7. Underlined sequences correspond to the functional GATA sites. These sequence alignments were not altered after running the GCG GAP program. For an exact alignment of the GATA sites, see figure 16.

A) GATAs

```
2150 CAAGGTTTATGCAGATCAGCACTGATGCAACTGATAGCAAAACTGATAAG 2199
       |     |||  |  | |||    |  ||  |   ||  |          |
 950 ..AAACCAATGTACAACAGATTTCGTGAA...GAAAAACTGCTGTGATAA 994

2200 GGTCAAAATTTCAGTGGCCAGCACAAACACATACACATTCGTAGTTCGTT 2249
     |    |   ||  | ||   || ||  | |   |||  | | |||    | |
 995 GAAAAGTGTTATACTGATAAGAACGTAGAGTAACAAAGTGGTATGGTGGT 1044

2250 T.....  2250      C. elegans
     |
1045 TATTGA 1050       C. briggsae
```

B) PRAs

```
3220 TTTTAATTTTTAATTTTAAAAAATTTCATATAATTCGTCTGCGTCTTACA 3269
          | |  ||||||    ||        |  | ||   ||  | |
3765 ....CTTGTCAAATTTTCTTGAACAAACACTCTTCCGAGAGCCGCCTTAC 3810

3270 GTTTCAGGCTTCAAAAATCTATTACATATCTTATCTTTGAATTCAGATGC 3319
       |      ||||  | |    |  || | |    || ||   |  | ||
3811 CCTCTCCTATTCACCACTGGGCTCTATTTGTGCCGTTGGACGACGTAAGC 3860

3320 GGATAT....  3325      C. elegans
       | ||
3861 AAACATGGAA 3870       C. briggsae
```

Identification of the Cb-ges-1 Gut Regulatory Element(s)

The pJM102ΔRI* (i.e. ΔRI*) and pJM102ΔAN (i.e. ΔAN)deletions uncovered a region that caused a gut to pharynx/rectum switch in the expression pattern of *Cb-ges-1* (figure 12). This switch appeared similar to the switch observed for the *Ce-ges-1* gene when the GATA sites were deleted (Egan *et al.*, 1995). Inspection of the deleted *Cb-ges-1* sequence (from base pairs 937 to 1334) identified five consensus WGATAR sites; however, only one tandemly arranged pair of GATA sites was found within the deleted sequence. Located between base pairs 987 and 1021, this tandem pair of *Cb-ges-1* GATA sites showed similarity to the GATA sites found in the *Ce-ges-1* gene (figure 16). The major difference between the *Ce-ges-1* and *Cb-ges-1* tandem GATA sites is an eight base pair insertion between the *Cb-ges-1* tandem GATA sites.

PCR directed deletion was used to remove a 35 bp sequence encompassing both GATA sites to create the construct pJM102ΔGATAs. Figure 17 shows that the staining pattern of *C. elegans* and *C. briggsae* transformed with pJM102ΔGATAs was the same as the larger deletions of the "GATA site" region. In both nematode species, pJM102ΔGATAs expression was turned off in the gut, and simultaneously turned on in the pharynx/rectum regions. This result was reminiscent of the *Ce-ges-1* gene, when the *C. elegans* tandem GATA sites were deleted (figure 18; Egan *et al.*, 1995).

To determine the importance of each GATA site (in terms of contribution to *Cb-ges-1* expression), the upstream (pJM102ΔUpGATA) and downstream

Figure 16: Alignment of the *C. elegans* and *C. briggsae ges-1* tandem GATA sequences. Sequence position within the 5' DNA sequence is indicated and corresponds to the 5' most portion of the clone (pJM102 for *C. briggsae* and pJM15 for *C. elegans*). Vertical bars indicate sequence identity; boxes outline "large" conserved areas; underlined sequences indicate the consensus WGATAR sites.

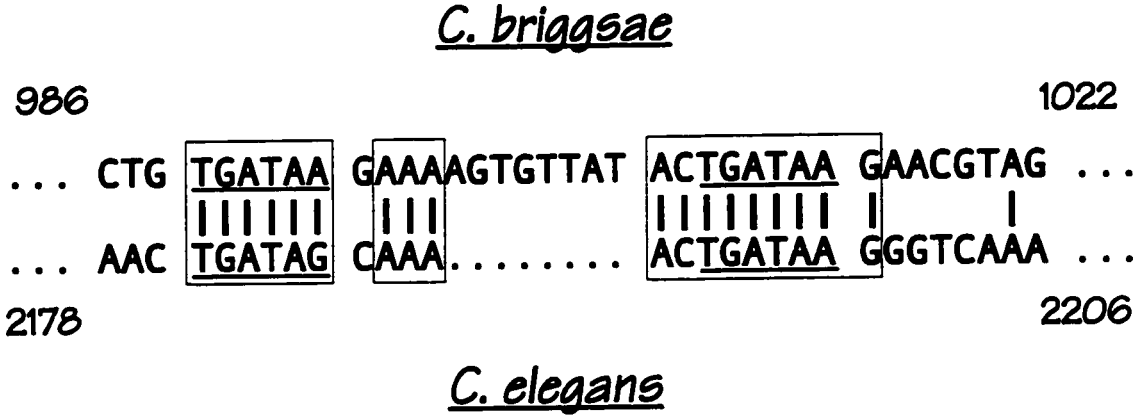

C. briggsae
986
1022
... CTG TGATAA GAAAAGTGTTAT ACTGATAA GAACGTAG ...
... AAC TGATAG CAAA..... ACTGATAA GGGTCAAA ...
2178
2206
C. elegans

Figure 17: Mapping of the *Cb-ges-1* gut activator-pharynx/rectum repressor regulatory element. Heavy gut staining with weak pharynx/tail staining was observed for pJM102 transformed embryos (top). Heavy pharynx/rectum staining is observed for all deletions: pJM102ΔRI*, pJM102ΔAN, pJM102ΔGATAs, pJM102ΔUpGATA, pJM102ΔDnGATA (middle and bottom). Arrows point out endogenous gut esterase staining in the transformed *C. briggsae* embryos. The thick grey line represents the pJM102 sequence, while the thick black lines indicate deleted sequences in the various "test" constructs. Schematic gene sequences are drawn to scale.

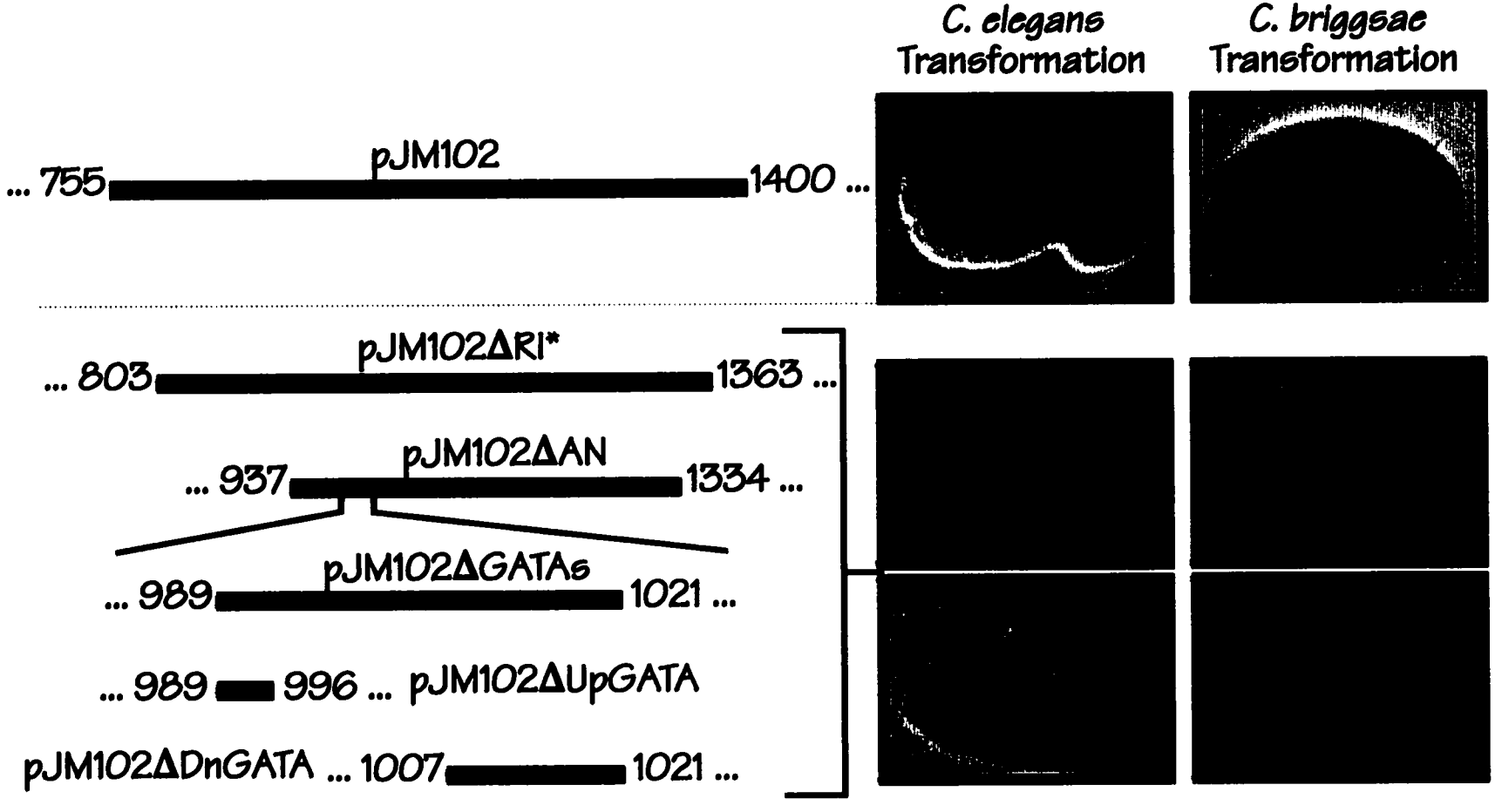
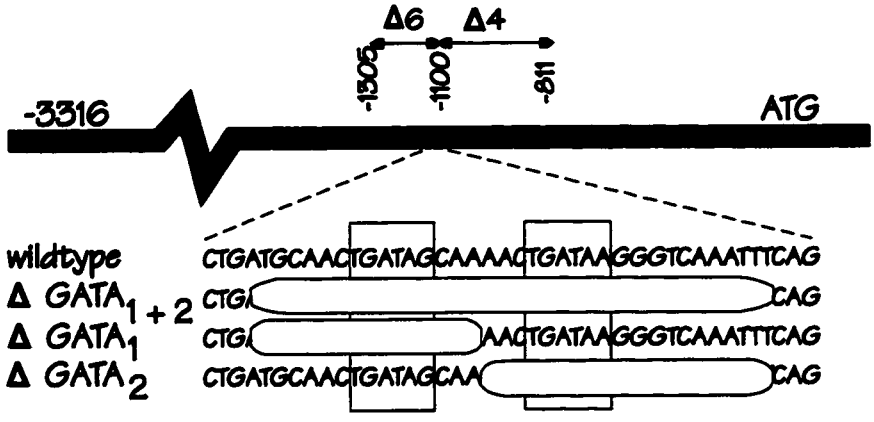
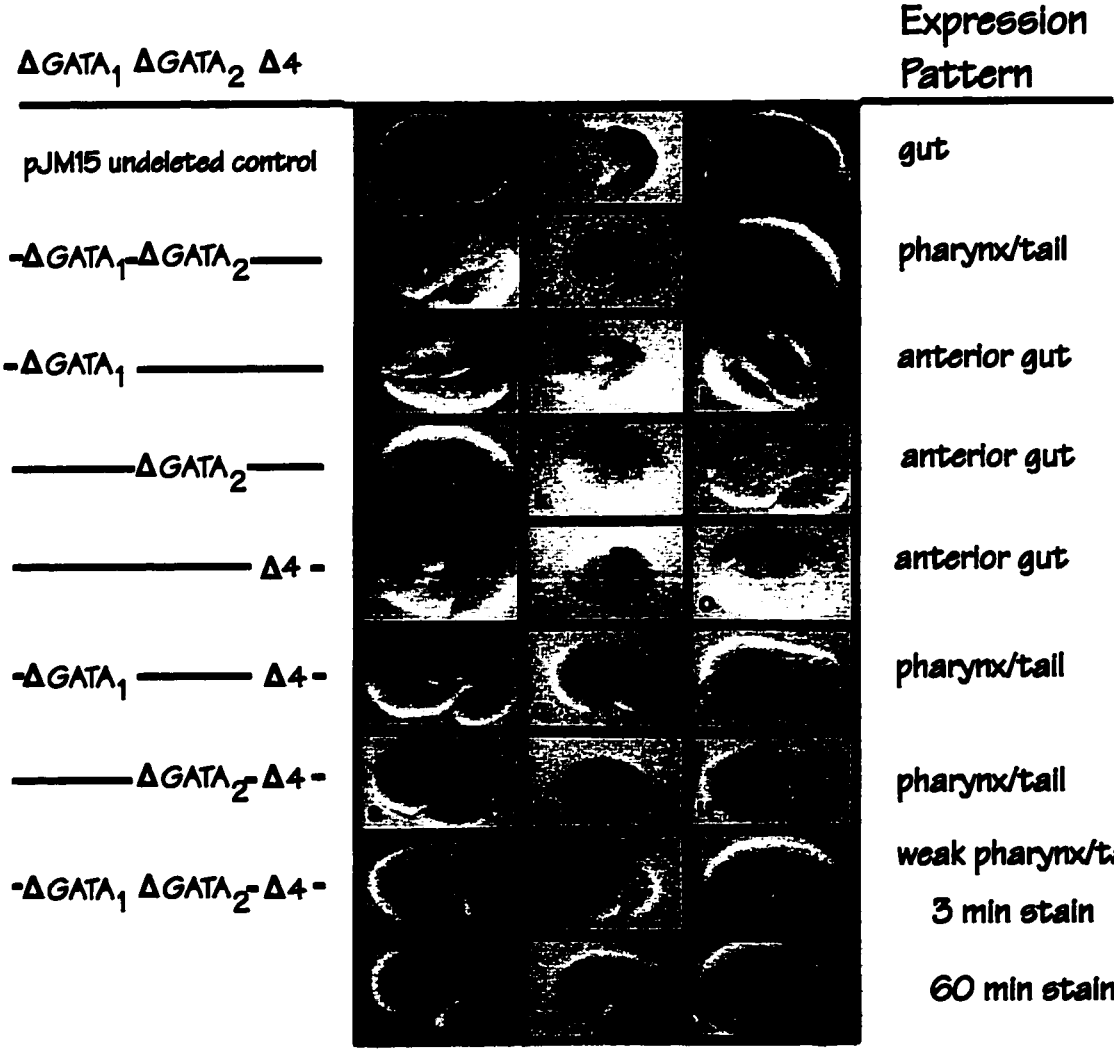

C. elegans Transformation
C. briggsae Transformation
... 755
pJM102
1400 ...
pJM102ΔRI*
... 803
1363 ...
pJM102ΔAN
... 937
1334 ...
pJM102ΔGATAs
... 989
1021 ...
... 989
996 ...
pJM102ΔUpGATA
pJM102ΔDnGATA
... 1007
1021 ...

Figure 18: Deletion analysis of the *Ce-ges-1* gut activator-pharynx/rectum repressor region.

"At the top is a schematic diagram of the 5'-region of the *ges-1* gene. The wildtype sequence from nucleotides -1105 to -1146 (inclusive) is shown, with the two tandem WGATAR motifs indicated in boxes. The Δ4 region (deleted for bp -811 to -1100 inclusive) is indicated at the top of the figure. The exact nucleotides removed along with each GATA site are indicated by the blank ovals on the sequence. The different combinations of the three regions ($\Delta GATA_1$, $\Delta GATA_2$, and Δ4) deleted in the various transforming constructs are indicated on the left; three esterase-stained embryos are shown for each transforming construct; the corresponding class of staining pattern is noted on the right. All embryos have been stained for 3 min (except v', w', and x', which were stained for 60 min)." Figure and text taken from Egan *et al.* (1995).


Δ6
Δ4
-1305
-1100
-811
-3316
ATG
wildtype
CTGATGCAACTGATAGCAAAACTGATAAGGGTCAAATTTCAG
Δ GATA1 + 2
CTG CAG
Δ GATA1
CTG AACTGATAAGGGTCAAATTTCAG
Δ GATA2
CTGATGCAACTGATAGCAA CAG


ΔGATA1 ΔGATA2 Δ4
Expression Pattern
pJM15 undeleted control
gut
-ΔGATA1-ΔGATA2-
pharynx/tail
-ΔGATA1
anterior gut
-ΔGATA2-
anterior gut
Δ4
anterior gut
-ΔGATA1 Δ4
pharynx/tail
-ΔGATA2-Δ4
pharynx/tail
-ΔGATA1 ΔGATA2-Δ4
weak pharynx/tail
3 min stain
60 min stain

(pJM102ΔDnGATA) GATA sites were deleted separately. Transformation of *C. elegans* with either pJM102ΔUpGATA or pJM102ΔDnGATA again resulted in esterase expression switching from the gut into the pharynx/rectum (figure 17). In the *Ce-ges-1* gene, removal of either the upstream GATA site or the downstream GATA site resulted in *Ce-ges-1* expression in the anterior portion of the gut (compare figure 17 with figure 18; Egan *et al.*, 1995). The anterior staining gut cells have now been identified as the int-1 and int-2 cells of the *C. elegans* gut (D. F. Schroeder, personal communication).

A third *Ce-ges-1* deletion (Δ4, just downstream from the GATA sites) also allowed anterior gut expression of *Ce-ges-1* (figure 18). It is interesting to note that deletion of any two of the upstream GATA, downstream GATA, or Δ4 region was required for *Ce-ges-1* to be expressed in the pharynx/rectum and not the gut (Egan *et al.*, 1995). Transformation of pJM102Δ1 (a *Cb-ges-1* deletion construct), which like the *Ce-ges-1* Δ4 deletion removes sequences immediately downstream from the GATA site region, into *C. elegans* had no effect on the *ges-1* expression pattern (figure 12).

Mutation of the Cb-ges-1 GATA Sites

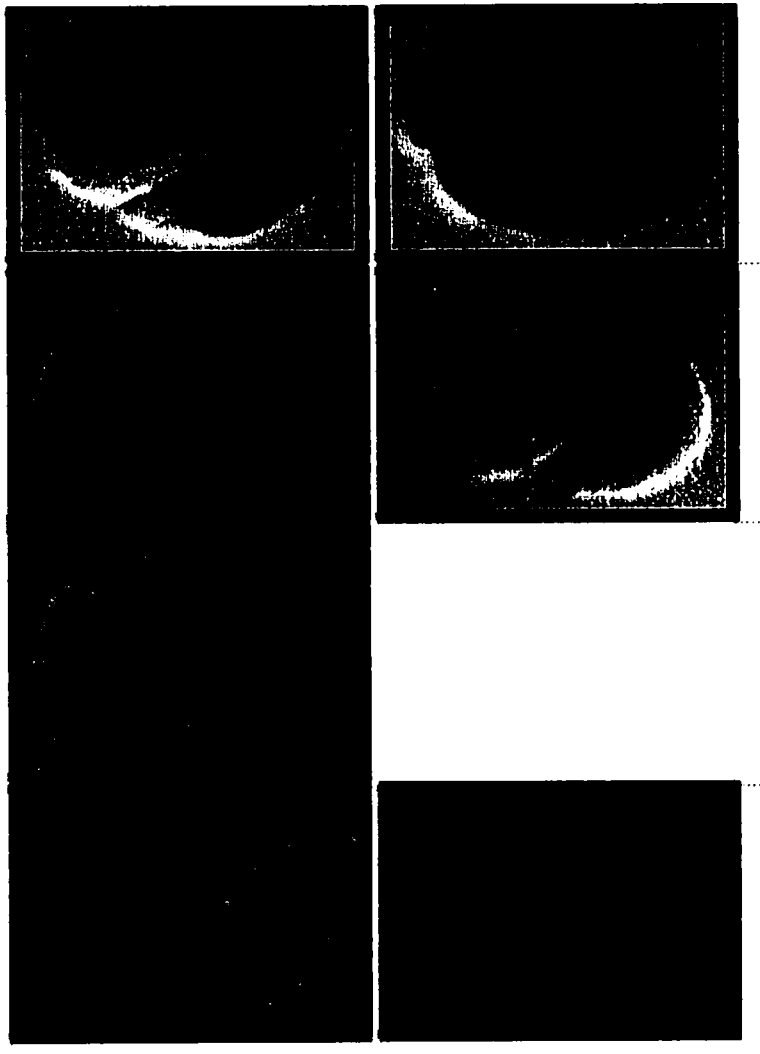
To understand the sequence requirements of the GATA sites for directing *Cb-ges-1* to the gut, the consensus WGATAR site sequences were specifically mutated (either separately or together). The tandem, upstream, and downstream *Cb-ges-1* GATA sites were altered from TGATAA to CCGCTG

(pJM102mGATAs, pJM102mUpGATA and pJM102mDnGATA, respectively; see figure 19). Mutation of the specific WGATAR sites allowed the importance of DNA spacing and the strict sequence requirements to be determined.

When both GATA sites were mutated simultaneously (pJM102mGATAs), the staining pattern of transformed *C. elegans* and *C. briggsae* embryos resembled the pJM102ΔGATAs transformation results; *Cb-ges-1* expression was turned off in the gut and was turned on in the pharynx/rectum (figure 19). The intensity of staining in the pharynx/rectum was comparable to that observed when both GATA sites were deleted (see figure 17). For *Ce-ges-1*, mutation of both GATA sites resulted in weak *Ce-ges-1* expression in the pharynx/rectum region, with a small percentage of embryos showing weak gut expression (figure 20; Egan *et al.*, 1995).

The upstream and downstream GATA sites were mutated separately to determine the effects that each GATA site may have in regulating *Ce-ges-1* expression. pJM102mUpGATA transformed *C. elegans* embryos stained weakly in the pharynx/rectum region and strongly in the gut, similar to pJM102 transformed embryos. This mutation result was different from the upstream GATA site deletion result, and may reflect the effect that spacing has on the ability of transcription factors to bind this region. Mutation of the *Cb-ges-1* downstream GATA site (pJM102mDnGATA) resulted in a switch from gut expression to pharynx/rectum expression (figure 19). Transformation using pJM102mDnGATA was successful for *C. elegans* and *C. briggsae*. Mutation of

Figure 19: Mutation of the *Cb-ges-1* GATA sites. The position of the consensus WGATAR sequences (wildtype and mutated sites) are surrounded by open boxes. Heavy gut staining with weak pharynx/rectum staining was observed for embryos transformed with pJM102 (top) or the mutated upstream GATA site construct (pJM102mUpGATA, middle bottom). Mutation of either both WGATAR sites (pJM102mGATAs) or the downstream WGATAR site (pJM102mDnGATA) results in heavy pharynx/rectum staining (middle top and bottom, respectively).



Transforming Construct

Wildtype WGATAR Sites
(pJM102)

...TGATAAGAAAAGTGTTATACTGATAAG...

Tandem WGATAR Sites Mutated
(pJM102mGATAs)

...CCGCTGGAAAAGTGTTATACGACGCCG...

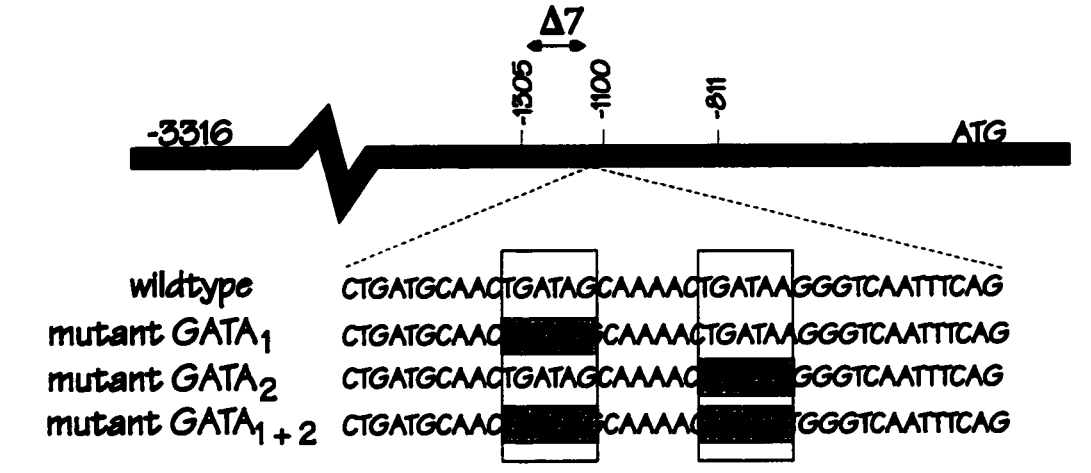
Upstream WGATAR Site Mutated
(pJM102mUpGATA)

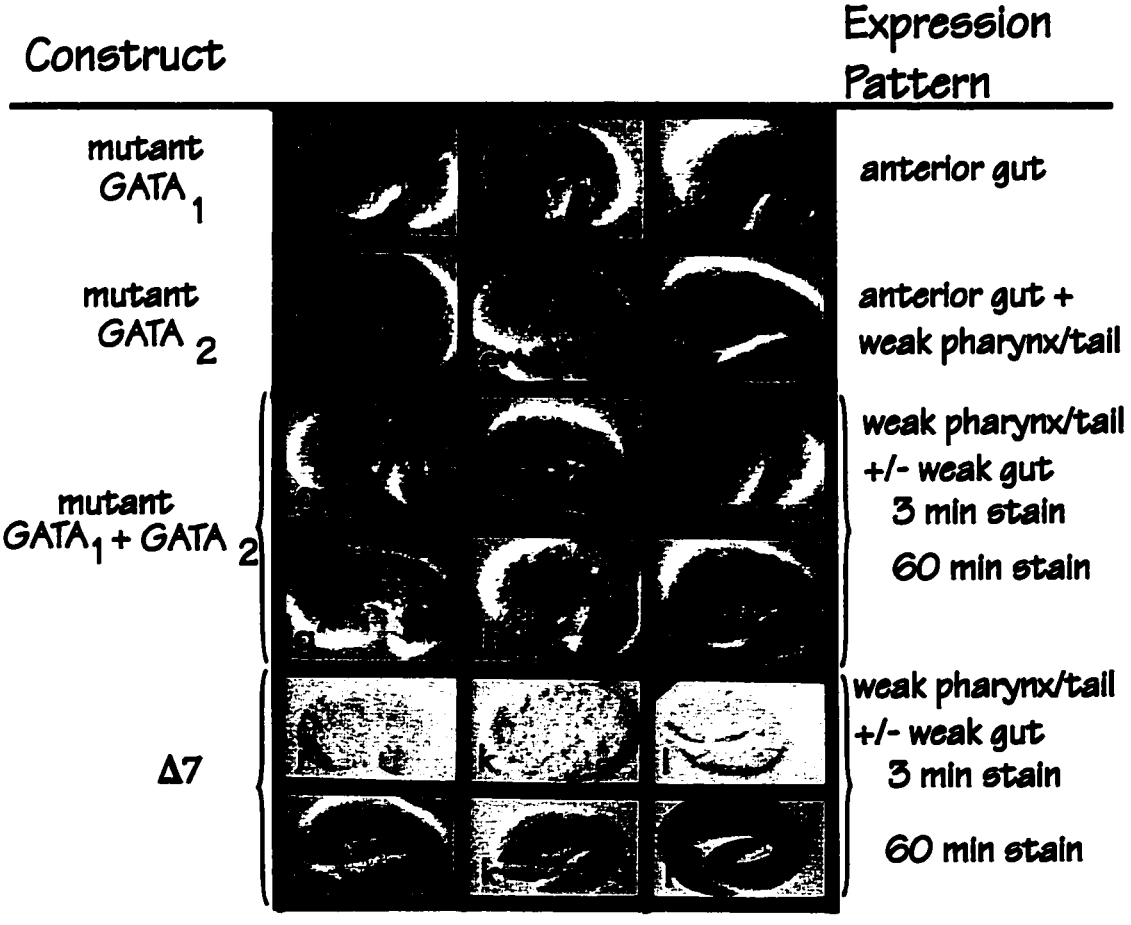
...CCGCTGGAAAAGTGTTATACTGATAAG...

Downstream WGATAR Site Mutated
(pJM102mDnGATA)

...TGATAAGAAAAGTGTTATACGACGCCG...

Figure 20: Mutation of the *Ce-ges-1* GATA sites and deletion of the Δ7 region. "At the top is a schematic diagram showing site-directed mutations introduced into the upstream WGATAR site (i.e. $GATA_1$ in figure), the downstream WGATAR site (i.e. $GATA_2$ in figure), or both. The Δ7 region (deleted for bp -1144 to -1305 inclusive) is also indicated at the top of the figure. Transforming constructs are shown at the left; interpretations of the corresponding staining patterns are noted on the right. All embryos have been stained for 3 min, except g'-l', which were stained for 60 min." Figure and text taken from Egan *et al.* (1995).


Δ7
-1505
-1100
-911
-3316
ATG
wildtype CTGATGCAACTGATAGCAAAACTGATAAGGGTCAATTTCAG
mutant GATA1 CTGATGCAAC CAAAACTGATAAGGGTCAATTTCAG
mutant GATA2 CTGATGCAACTGATAGCAAAAC GGGTCAATTTCAG
mutant GATA1+2 CTGATGCAAC CAAAAC GGGTCAATTTCAG


Construct
Expression Pattern
mutant GATA1
anterior gut
mutant GATA2
anterior gut + weak pharynx/tail
mutant GATA1 + GATA2
weak pharynx/tail +/- weak gut
3 min stain
60 min stain
Δ7
weak pharynx/tail +/- weak gut
3 min stain
60 min stain

the *Ce-ges-1* upstream GATA (a.k.a. $GATA_1$) or downstream GATA (a.k.a. $GATA_2$) sites resulted in anterior gut expression with some weak pharynx/rectum expression (figure 20; Egan *et al.*, 1995). These results indicate that the *Cb-ges-1* downstream GATA site sequence is required, while the upstream GATA site sequence is not specifically needed. Conversely, both GATA sites for *Ce-ges-1* appear to be required for proper *Ce-ges-1* gut expression.

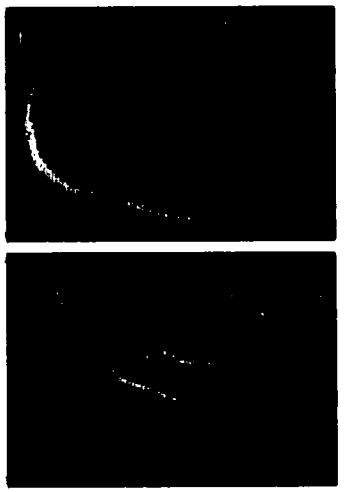
Repositioning of the Cb-ges-1 GATA Sites Within the Cb-ges-1 Promoter Still Directs Correct Cb-ges-1 Expression

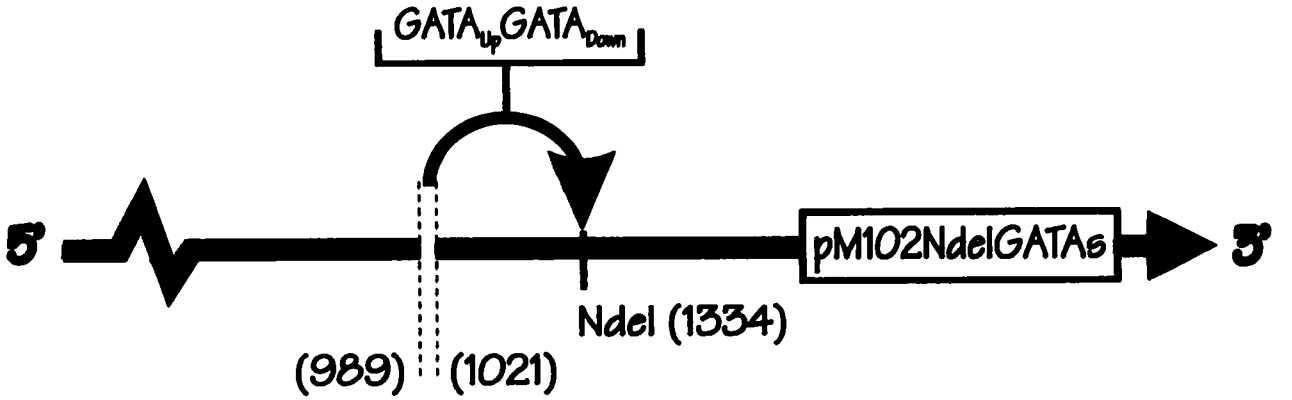
The *Cb-ges-1* tandem pair of GATA sites were inserted at an ectopic position within the *Cb-ges-1* promoter of the pJM102ΔGATAs construct to determine if contiguous sequence is required for proper *Cb-ges-1* expression. Re-insertion of the tandem pair of GATA sites (forward orientation) approximately 300 bp 3' from the normal position of the tandem GATA sites was sufficient to direct *Cb-ges-1* expression back to the gut (figure 21; pNdeIGATAs was not transformed into *C. briggsae*). Thus, it appears that the tandem GATA sites contain most (if not all) of the sequence information required to direct *Cb-ges-1* expression to the gut within the context of the *Cb-ges-1* gene, and this sequence is not required to be contiguous with other sequences.

When the *Ce-ges-1* tandem GATA sites were moved either upstream (four copies) or downstream (single copy) from their normal position, a wildtype expression pattern was observed, i.e. staining was only seen in the gut

Figure 21: Ectopic positioning of the *Cb-ges-1* tandem GATA sites directs proper *Cb-ges-1* expression. Heavy gut staining with weak pharynx/rectum staining was observed for pJM102 (bottom left) transformed embryos, while heavy pharynx/rectum staining was observed for pJM102ΔGATAs (bottom right) transformed embryos. Insertion of a single copy of the *Cb-ges-1* tandem GATA sites (in forward sequence orientation) into the NdeI restriction site within pJM102ΔGATAs (construct pNdeIGATAs) resulted in heavy gut staining with weak pharynx/rectum staining (top left).

C. elegans
Transformation





C. elegans
pJM102
Transformation

C. elegans
pJM102ΔGATAs
Transformation

(figure 22; Egan *et al.*, 1995). Therefore, for both *Ce-ges-1* and *Cb-ges-1*, the exact position of the tandem GATA sites does not appear to influence their ability to direct esterase activity to the gut (and exclude expression from the pharynx/rectum) within the context of the *Ce-ges-1* and *Cb-ges-1* promoters. As a cautionary note, insertion of a single copy of the tandem GATA sites upstream of the normal position was unable to restore gut activation-pharynx/rectum repression (Egan *et al.*, 1995). This last result suggests there may be some constraints for proper gut activation-pharynx/rectum repression.

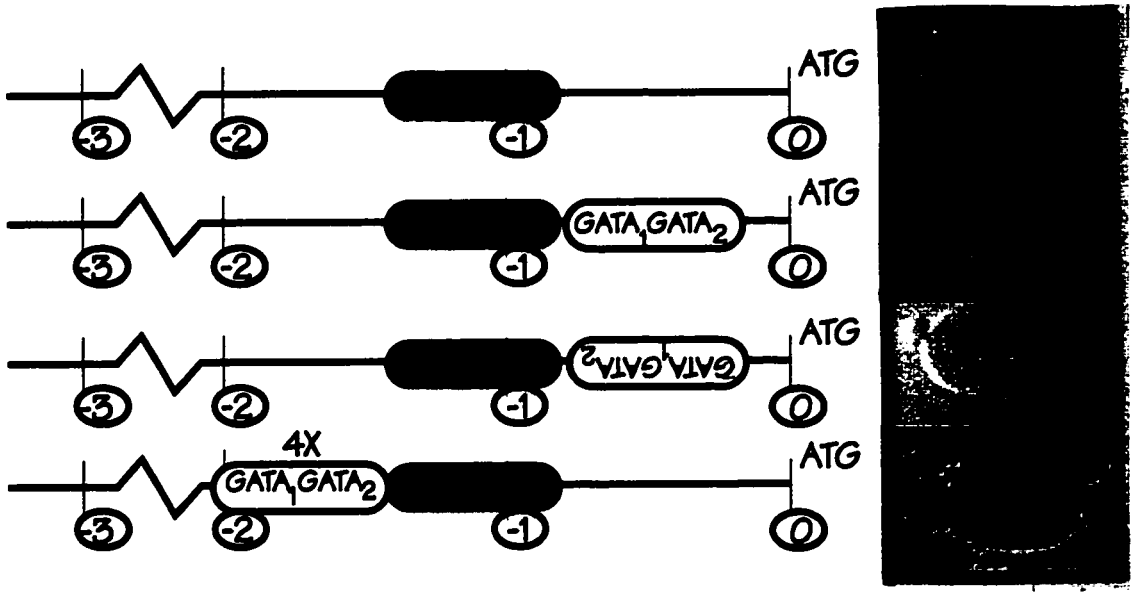
Cb-ges-1 GATA Sites Direct Gut Specific Expression Independently of the Cb-ges-1 Gene

A common method for determining the enhancer activity of a particular element is to determine if the sequence in question can drive proper expression of a reporter gene that contains a basal promoter. The enhancer assay vector p96E1ΔH (see Methods and Materials for details on vector construction) can be used to identify any transcriptional activation activity that may be present within a sequence of interest; transactivation can be detected via staining for the presence of β-galactosidase activity or visualization of GFP fluorescence.

The activity of the proposed gut regulatory elements was determined by placing six head-to-tail copies of the *Cb-ges-1* tandem GATA sites (in either a forward or a reverse orientation with respect to the basal promoter) into the p96E1ΔH vector to create p6xFGATAs and p6xRGATAs, respectively. After

Figure 22: Repositioning of the *Ce-ges-1* tandem GATA regions can direct proper *Ce-ges-1* gut expression.

"The tandem GATA regions can reestablish *ges-1* gut activation and pharynx/tail repression when inserted elsewhere into a pJM15-$\Delta GATA_1$-$\Delta GATA_2$ deletion construct. The transforming constructs are shown schematically on the left; circled numbers represent kilobases upstream of the *ges-1* ATG codon. A typical embryo from a strain transformed with these constructs is shown at the right . . . From top to bottom, the four constructs are: (i) pJM15-$\Delta GATA_1$-$\Delta GATA_2$ control, which directs pharynx/tail expression of the *ges-1* gene; (ii) pJM15-$\Delta GATA_1$-$\Delta GATA_2$ into which the deleted tandem GATA sites were reinserted in the forward orientation 534 bp downstream of their normal position; (iii) the same construct as (ii) but in which the deleted sites were reinserted in the reverse orientation; and (iv) pJM15-$\Delta GATA_1$-$\Delta GATA_2$ into which four head-to-tail copies of the deleted region were reinserted in the forward orientation 348 bp upstream of the normal location. The bottom three constructs all exhibit essentially wildtype *ges-1* expression, i.e., in the gut and not in the pharynx/tail. Figure and text taken from Egan *et al.* (1995).


ATG
-3
-2
-1
0
GATA1GATA2
4X

transformation of *C. elegans* and *C. briggsae*, expression from both p6xFGATAs and p6xRGATAs was observed only in the gut (see figure 23 for *C. elegans* transformations and figure 24 for *C. briggsae* transformations), as determined by both GFP fluorescence and β-galactosidase detection. In both species, expression was first seen at the four E cell stage, which is when *Cb-ges-1* esterase staining was first detected. Expression was detected up to the late L1/early L2 larval stages before disappearing.

To determine if the upstream and the downstream GATA sites are each alone competent to drive gut specific expression, seven head-to-tail copies of either the upstream or downstream GATA sites (in forward or reverse orientations, respectively) were inserted into the p96E1ΔH expression vector. The upstream GATA site construct is p7xFUpGATA, while the downstream GATA site construct is p7xDnGATA. Both enhancer assay constructs were injected into *C. elegans*; however, only p7xFUpGATA was successfully transformed into *C. briggsae*. p7xFUpGATA and p7xDnGATA were both able to direct gut specific expression as detected by GFP fluorescence and β-galactosidase staining (figures 23 and 24). When the *Ce-ges-1* GATA sites (tandem and upstream) were inserted into a similar enhancer assay vector (see Egan *et al.*, 1995 for vector details), the *Ce-ges-1* GATA sites were able to direct expression of β-galactosidase specifically to the gut (figure 25; Egan *et al.*, 1995). The GATA sites from both *Ce-ges-1* and *Cb-ges-1* appear competent to direct gut specific expression independently of the *ges-1* genes.

Figure 23: The *Cb-ges-1* GATA sites are able to direct gut specific expression independent of the *Cb-ges-1* gene (in *C. elegans*). A schematic diagram of the p96E1ΔH parental vector used in the enhancer assay is drawn at the top. Multimerized GATA sites were inserted into the multiple cloning site (MCS) of the enhancer assay vector p96E1ΔH. Copy number and insertion orientation of the multimerized GATA sites are indicated in the figure and correspond to the following constructs: 6x($GATA_{Up}GATA_{Dn}$) = p6xFGATAs; 6x($GATA_{Up}GATA_{Dn}$) - *written upside down* = p6xRGATAs; 7x($GATA_{Up}$) = p7xFGATAUp; 7x($GATA_{Dn}$) - *written upside down* = p7x$RGATA_{Dn}$. *C. elegans* N2 was used for transformation. The Nomarski image and GFP (i.e. fluorescent image) embryos correspond to the same embryo. β-gal stained embryos were stained for at least 15 hours at 37°C.

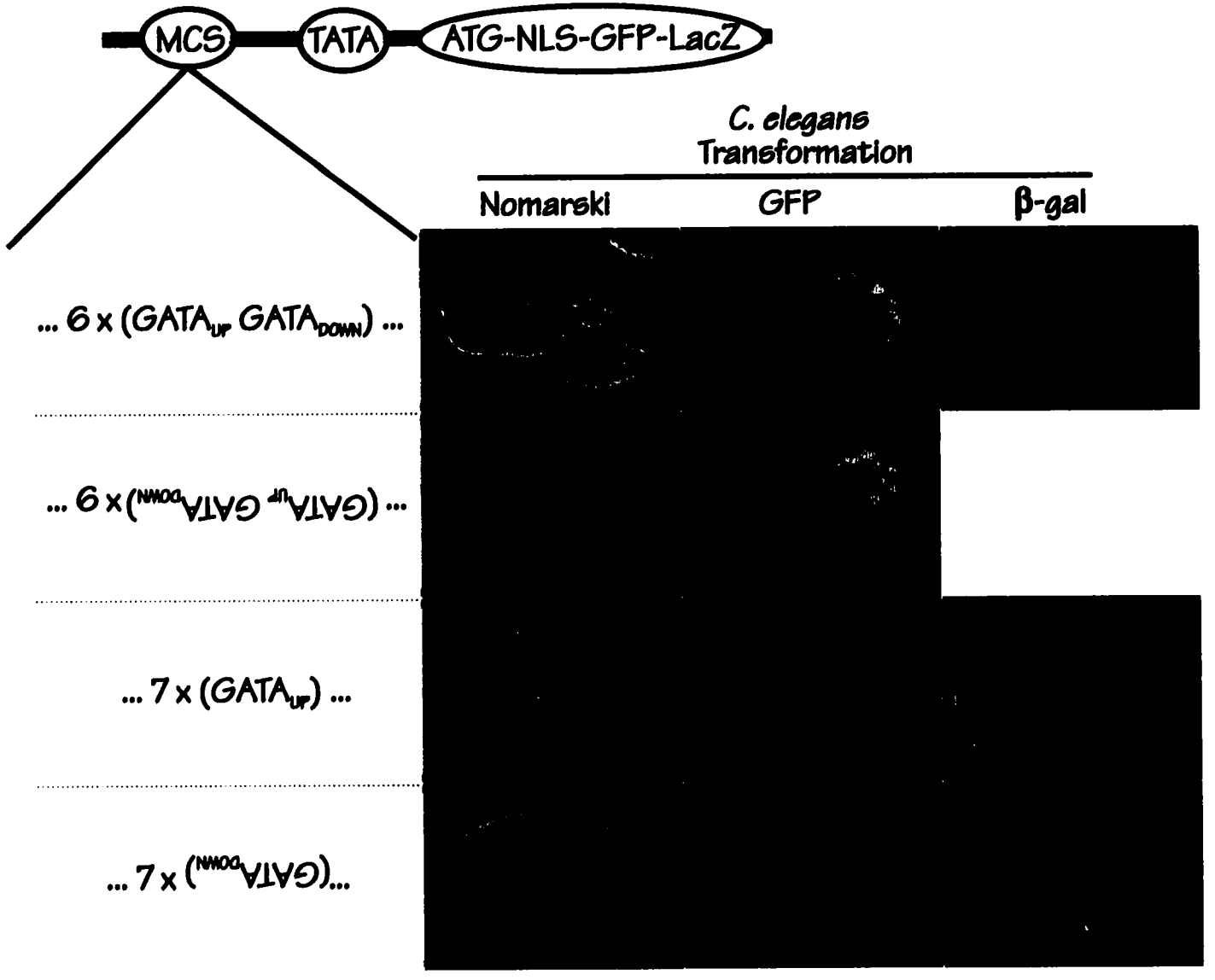

MCS
TATA
ATG-NLS-GFP-LacZ
C. elegans
Transformation
Nomarski
GFP
β-gal
... 6 x (GATA$_{UP}$ GATA$_{DOWN}$) ...
... 6 x (GATA$_{DOWN}$ GATA$_{UP}$) ...
... 7 x (GATA$_{UP}$) ...
... 7 x (GATA$_{DOWN}$) ...

Figure 24: The *Cb-ges-1* GATA sites are able to direct gut specific expression independent of the *Cb-ges-1* gene (in *C. briggsae*). A schematic diagram of the p96E1ΔH parental vector used in the enhancer assay is drawn at the top. Multimerized GATA sites were inserted into the multiple cloning site (MCS) of the enhancer assay vector p96E1ΔH. Copy number and insertion orientation of the multimerized GATA sites are indicated in the figure and correspond to the following constructs: 6x($GATA_{Up}GATA_{Dn}$) = p6xFGATAs; 7x($GATA_{Up}$) = p7xFGATAUp. *C. briggsae* AF16 was used for transformation. The Nomarski image and GFP (i.e. fluorescent image) embryos correspond to the same embryo. β-gal stained embryos were stained for at least 15 hours at 37°C.

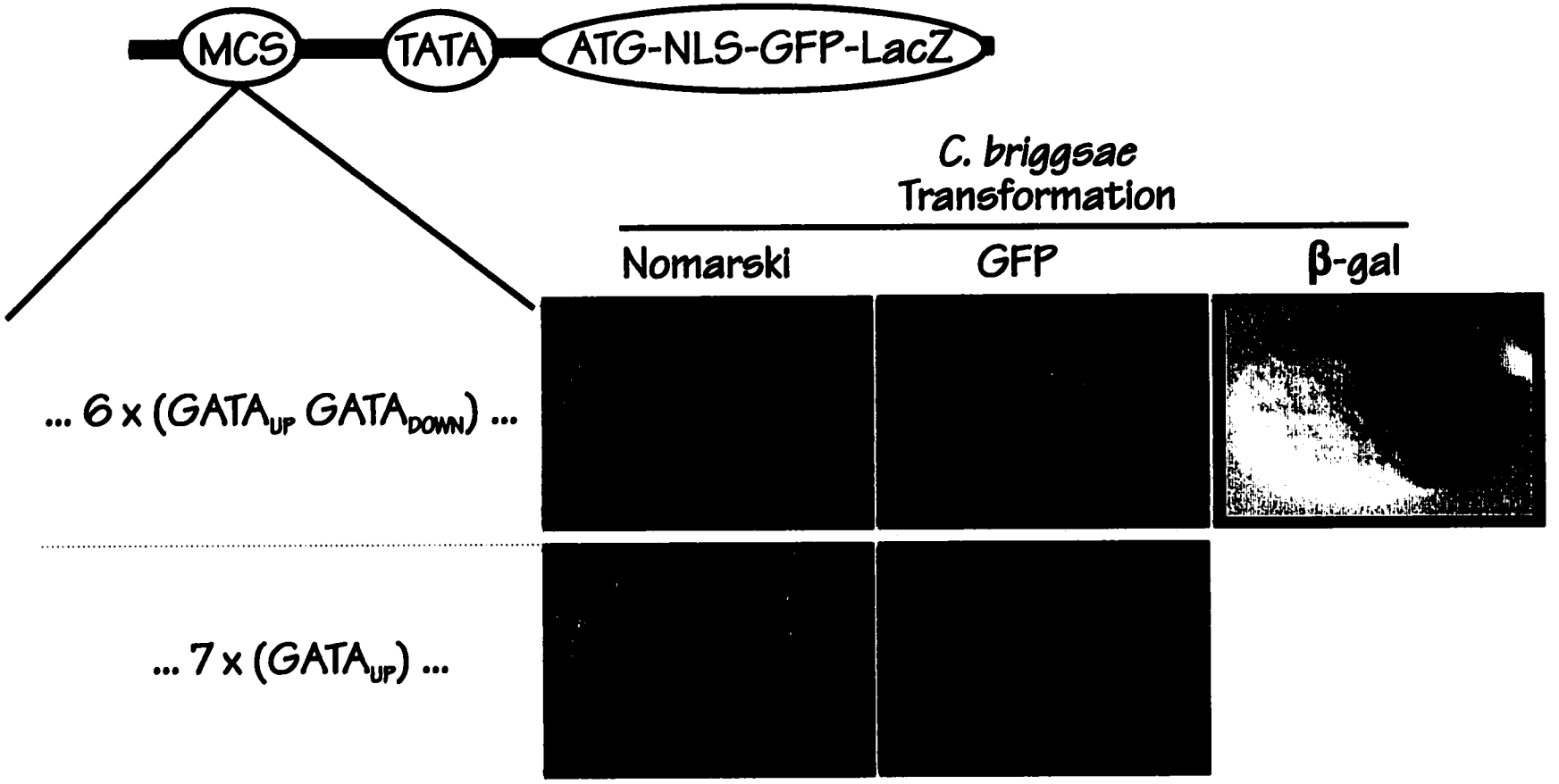
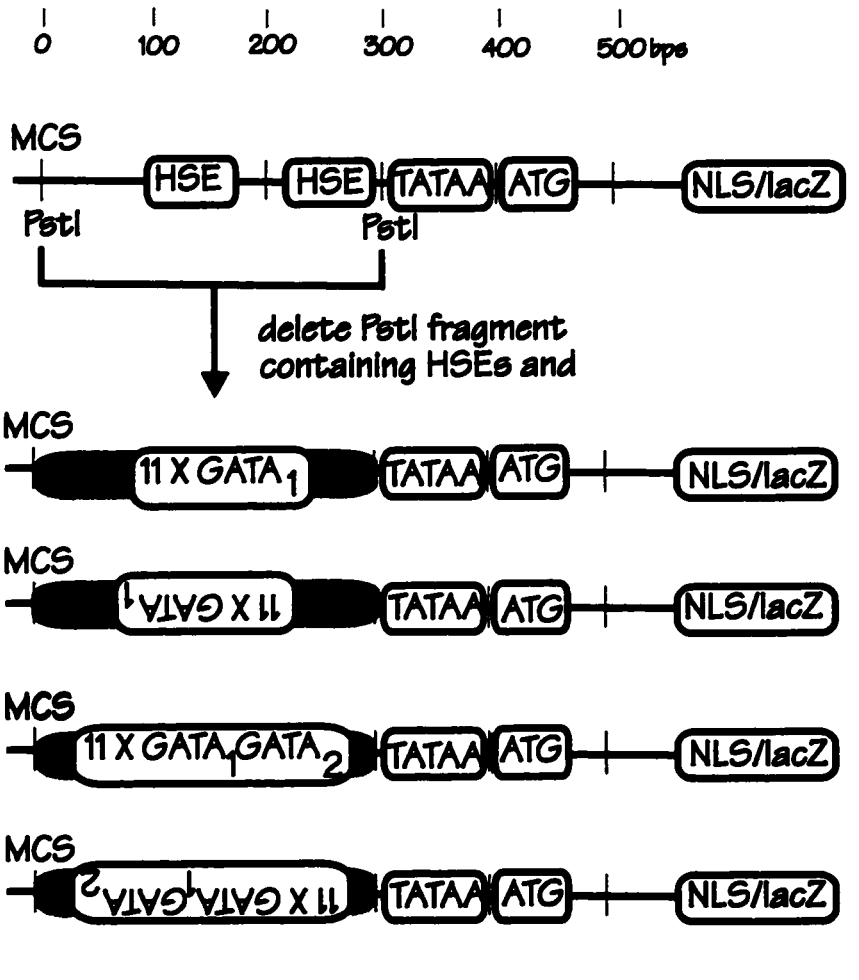

MCS
TATA
ATG-NLS-GFP-LacZ
C. briggsae
Transformation
Nomarski
GFP
β-gal
... 6 x (GATA$_{UP}$ GATA$_{DOWN}$) ...
... 7 x (GATA$_{UP}$) ...

Figure 25: Multimerization of the of the *Ce-ges-1* GATA regions can direct gut specific expression.

"At the top is a schematic diagram of the promoter region of the *C. elegans hsp16*-lacZ fusion construct pPC16.48-1 (Stringham *et al.*, 1992). The PstI fragment containing the heat-shock elements was removed and replaced with 11 head-to-tail copies of either the *ges-1* single $GATA_1$ region or the *ges-1* double GATA region, in either the forward or the reverse orientation as indicated. Transformed embryos were stained for β-galactosidase activity; expression is in the developing intestine." Figure and text taken from Egan *et al.* (1995).

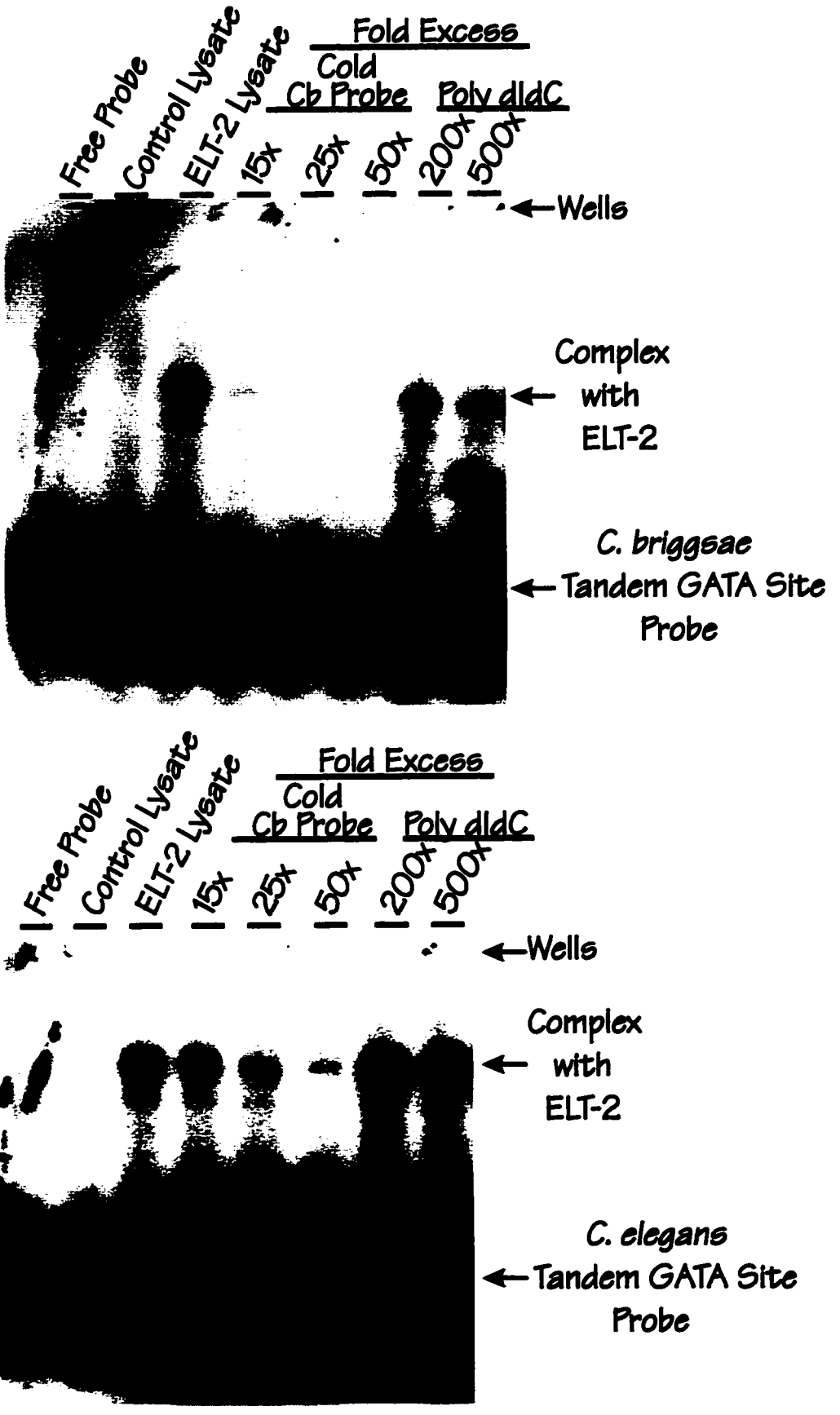

0
100
200
300
400
500 bps
MCS
HSE
HSE
TATAA
ATG
NLS/lacZ
PstI
PstI
delete PstI fragment
containing HSEs and
MCS
11 X GATA1
TATAA
ATG
NLS/lacZ
MCS
11 X GATA1
TATAA
ATG
NLS/lacZ
MCS
11 X GATA1GATA2
TATAA
ATG
NLS/lacZ
MCS
11 X GATA2GATA1
TATAA
ATG
NLS/lacZ

C. elegans ELT-2 Recognizes the C. briggsae GATA Sites

The *C. elegans elt-2* gene encodes a GATA factor that contains a single zinc finger. *C. elegans* ELT-2 protein was previously shown to bind the tandem pair of *C. elegans* GATA sites (Hawkins and McGhee, 1995). Current work suggests that *elt-2* may be involved in directly regulating *Ce-ges-1* expression (Fukushige *et al.*, in press; Hawkins and McGhee, 1995). Bandshift experiments were used to determine if ELT-2 could bind the *Cb-ges-1* tandem GATA sites, thereby suggesting that an ELT-2-like protein may also regulate *Cb-ges-1* expression.

Bandshifts were performed using the *C. briggsae* tandem pair of GATA sites as a probe. Ce-ELT-2 protein was produced *in vitro* from *elt-2* cDNA programmed rabbit reticulocyte lysate. Figure 26 shows that *in vitro* produced Ce-ELT-2 bound specifically to oligonucleotides containing the *C. briggsae* tandem GATA sites. The bandshift using the *C. elegans* tandem GATA sites was done as a control (figure 26). Binding was shown to be specific by adding increasing amounts of poly dIdC:dIdC (a non-specific DNA), and through successful competition using unlabelled *Cb-ges-1* tandem GATA site probe (figure 26). The competition assays also showed that Ce-ELT-2 appears to bind preferentially to the *Ce-ges-1* tandem GATA site probe relative to the *Cb-ges-1* tandem GATA site probe (figure 26).

Figure 26: *C. elegans* ELT-2 can bind the *Cb-ges-1* tandem GATA sites. Addition of ELT-2 lysate to either the *C. briggsae* (top) or the *C. elegans* (bottom) tandem GATA site probe resulted in a slow migrating band; no such band was observed in the control (i.e. unprogrammed) lysate lane. Addition of increasing amounts of unlabelled *C. briggsae* tandem GATA site probe (measured by mass) led to a gradual decrease in the intensity of the slower migrating "ELT-2" bandshift. Note that ELT-2 appeared to bind more readily to the *C. elegans* tandem GATA site probe over the *C. briggsae* probe. A vast excess (by mass) of poly dIdC:dIdC over the labelled probe did not appreciably reduce the intensity of the "ELT-2" bandshift. Approximately two nanograms of labelled probe was added to each lane. The minor complex under the major "ELT-2" bandshift is likely to be incompletely translated ELT-2 from the *in vitro* transcription/translation reaction.


Free Probe
Control Lysate
ELT-2 Lysate
Fold Excess
Cold
Cb Probe
Poly dIdC
15x
25x
50x
200x
500x
Wells
Complex with ELT-2
C. briggsae
Tandem GATA Site Probe
Free Probe
Control Lysate
ELT-2 Lysate
Fold Excess
Cold
Cb Probe
Poly dIdC
15x
25x
50x
200x
500x
Wells
Complex with ELT-2
C. elegans
Tandem GATA Site Probe

Identification of the Regulatory Region(s) Required For Pharynx/Rectum Expression of the C. briggsae ges-1 Gene

The potential for *Ce-ges-1* to switch from gut expression to pharynx/rectum expression, which was uncovered in the analysis of the *Ce-ges-1* promoter (see Egan *et al.*, 1995), leads to the question of how the pharynx/rectum regions of the digestive tract can support *Ce-ges-1* expression. Endogenous *Ce-ges-1* is normally only expressed in the gut, not in the pharynx/rectum of the digestive tract; however, under specific circumstances (i.e. removal of the tandem pair of GATA sites), expression of *Ce-ges-1* in the pharynx/rectum is possible. The specific regulatory site(s) and mechanism(s) allowing *Ce-ges-1* expression in the pharynx/rectum remains unknown.

A putative pharynx/rectum activator element was identified for *Cb-ges-1*. pJM102Δ3' removed the entire 3' end of pJM102, beginning 25 bp past the poly(A) signal site. After transformation of pJM102Δ3' into *C. elegans*, the esterase staining pattern showed only heavy gut staining; pharynx/rectum staining was not detected (figure 12). Disappearance of esterase activity in the pharynx/rectum suggested the presence of a pharynx/rectum activator element within the 3' flanking sequence of *Cb-ges-1*.

Isolation of a Pharynx/Rectum Activator Sequence for Cb-ges-1

Unidirectional deletion analysis mapped the putative pharynx/rectum activator region (PRA) to a 275 bp region beginning 54 bp downstream from the

poly(A) signal site (figure 27). PCR directed deletion was used to create an internal deletion that specifically removed this 275 bp region (construct pJM102ΔPRA1). *C. elegans* embryos containing pJM102ΔPRA1 showed no esterase staining in the pharynx/rectum area, but did show strong gut staining (figure 27). Transformation of *C. briggsae* using the pJM102Δ3', produced an unexpected result; figure 27 shows that *C. briggsae* embryos transformed with pJM102Δ3' stained for esterase in a "deregulated" manner (i.e. no obvious expression pattern was observed). While the particular embryo shown has heavy gut staining, this pattern was not observed for all pJM102Δ3' transformed *C. briggsae* embryos. Reasons for this deregulated pattern of esterase expression are discussed later.

A short pharynx/rectum activator sequence (e.g. 10-20 bp) has yet to be identified for *Ce-ges-1*. Unidirectional deletion analysis has tentatively mapped the putative *Ce-ges-1* PRA to a 68 bp region beginning 74 bp upstream from the translation start codon, as shown in figure 28 (Dr. T. Fukushige, personal communication). It appears that potential PRA regions have been identified for both *Ce-ges-1* and *Cb-ges-1*.

Figure 27: Mapping of the *Cb-ges-1* pharynx/rectum activator regulatory region. Heavy gut staining with weak pharynx/rectum staining was observed for pJM102 transformed embryos (top). Heavy gut staining with weak pharynx/rectum staining was observed for embryos transformed with the following deletion constructs (middle): pJM102$\Delta E_{II}6$, $\Delta E_{II}4$, and ΔBS. Heavy gut staining with no pharynx/rectum staining was observed for *C. elegans* embryos containing the following deletion constructs (bottom left embryo): pJM102Δ3', ΔNS, ΔPRA1. The bottom right *C. briggsae* embryo is one example of a pJM102Δ3' transformed *C. briggsae* embryo; a specific staining pattern was not observed for these embryos. The thick grey line represents the pJM102 sequence, while the thick black lines indicate deleted sequences in the various "test" constructs. The vertical dotted lines indicate position of the PRA in pJM102. Schematic gene sequences are drawn to scale.

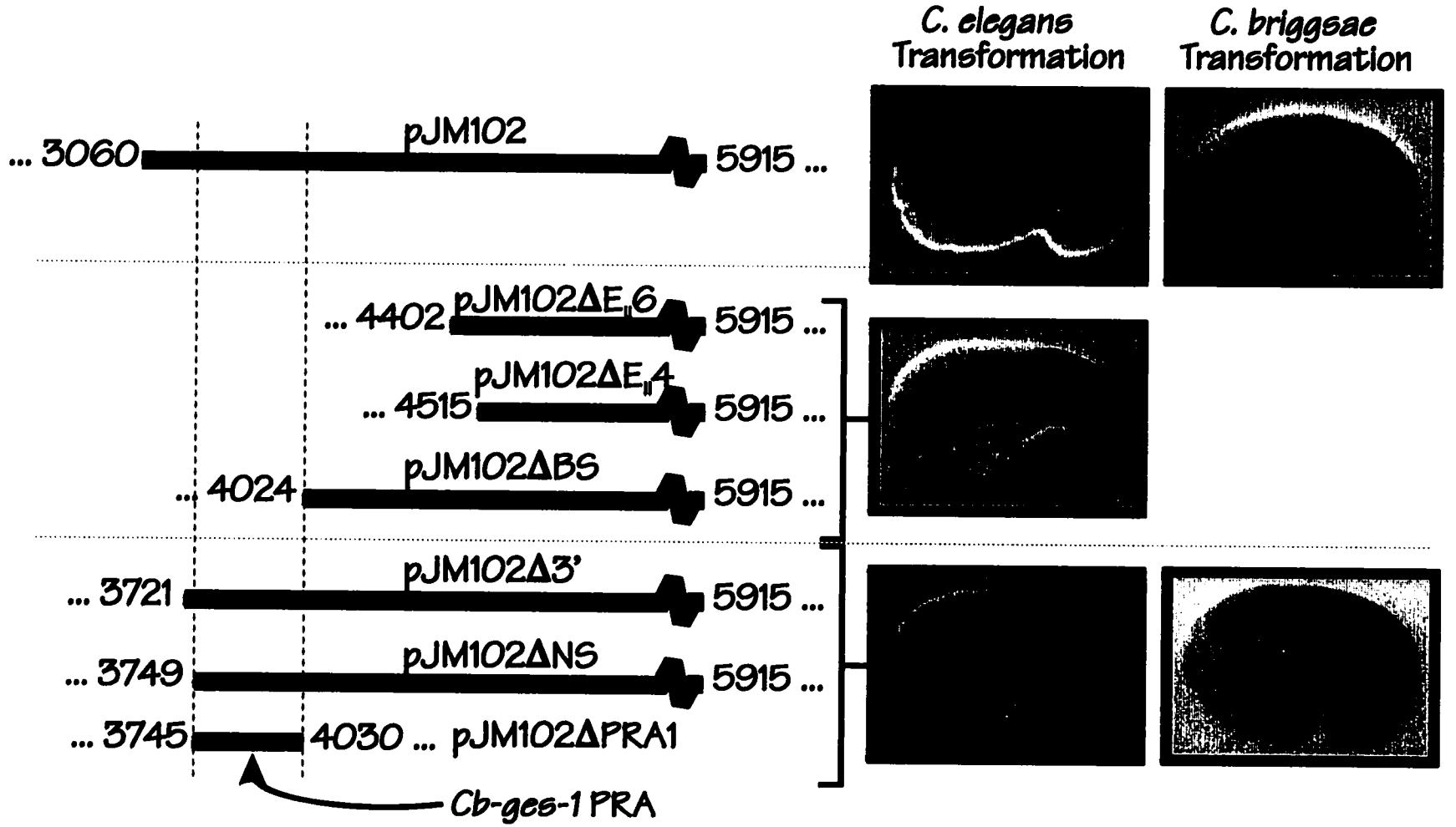
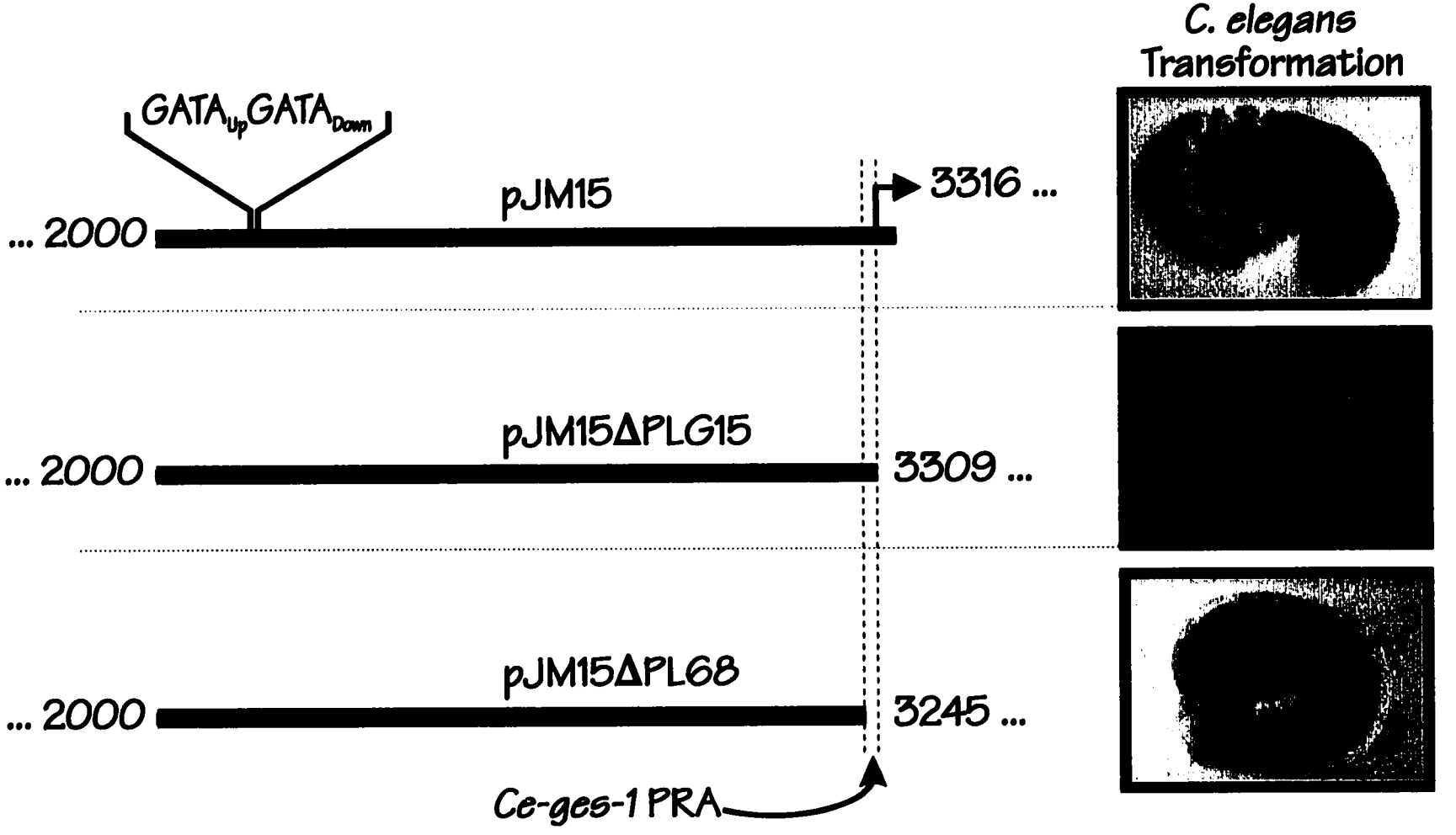

C. elegans Transformation
C. briggsae Transformation
... 3060
pJM102
5915 ...
... 4402
pJM102ΔE_II6
5915 ...
... 4515
pJM102ΔE_II4
5915 ...
... 4024
pJM102ΔBS
5915 ...
... 3721
pJM102Δ3'
5915 ...
... 3749
pJM102ΔNS
5915 ...
... 3745
4030 ... pJM102ΔPRA1
Cb-ges-1 PRA

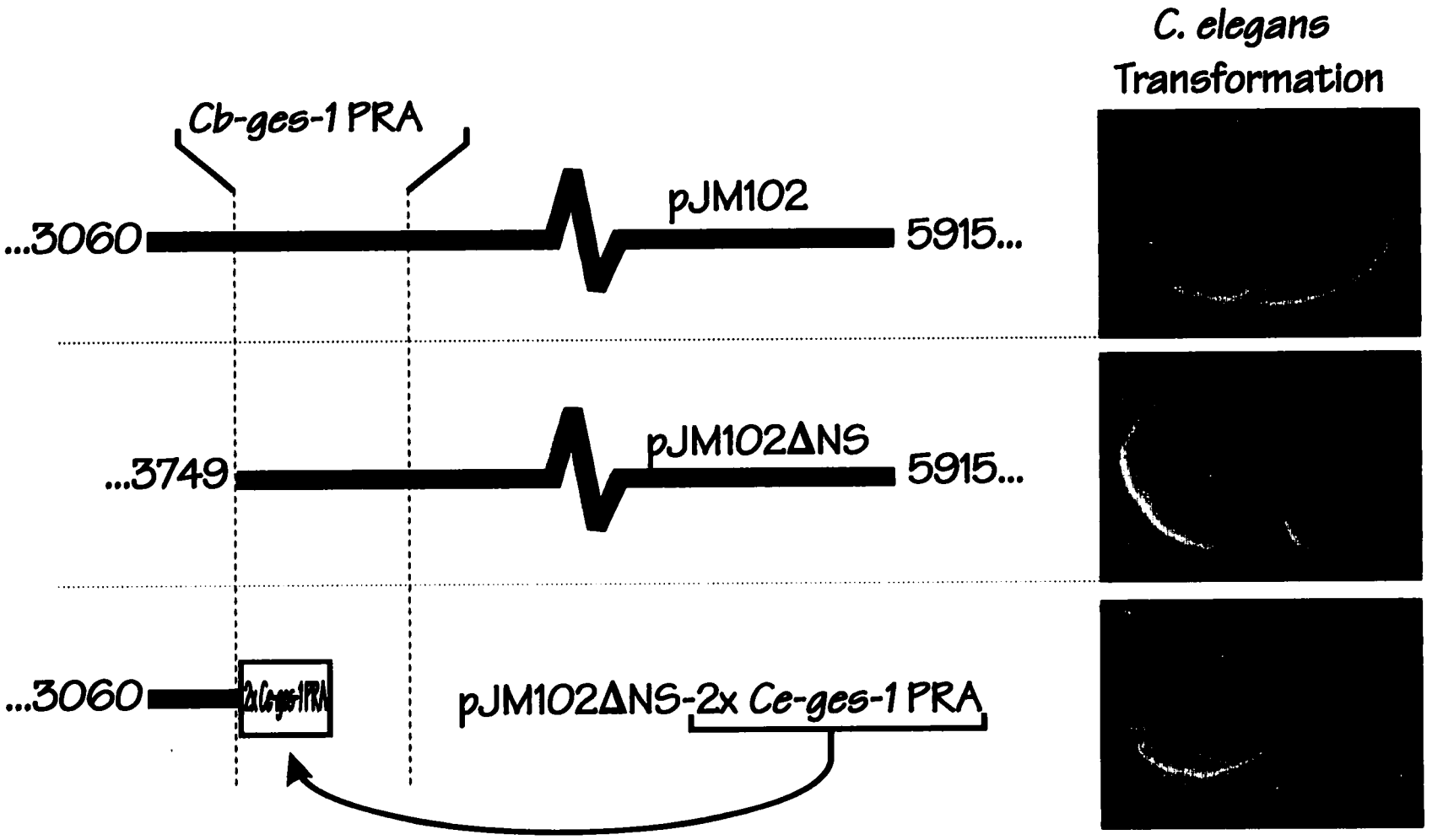
Figure 28: Isolation of the *Ce-ges-1* pharynx/rectum activator element. Heavy gut staining with weak pharynx/rectum staining was observed for pJM15 transformed embryos (top). No esterase staining was detected in pJM15ΔPLG15 transformed *C. elegans* embryos (middle). Heavy pharynx/rectum staining was observed for pJM15ΔPL68 (the GATA region was removed) transformed *C. elegans* embryos. The thick grey line represents the pJM15 sequence, while the thick black lines indicate deleted sequences in the various "test" constructs. The vertical dotted lines indicate position of the PRA in pJM15. Schematic gene sequences are drawn to scale. Experiment carried out by Dr. T. Fukushige.


C. elegans
Transformation
GATA$_{Up}$GATA$_{Down}$
... 2000
pJM15
3316 ...
pJM15ΔPLG15
... 2000
3309 ...
pJM15ΔPL68
... 2000
3245 ...
Ce-ges-1 PRA

A Potential Ce-ges-1 Pharynx/Rectum Activator Sequence Cannot Restore Pharynx Activity to a 3' end Deleted Cb-ges-1 Gene

To determine if the putative *Ce-ges-1* PRA could drive pharynx/rectum expression in a *Cb-ges-1* gene background, two copies of the 68 base pair *Ce-ges-1* PRA were placed at the 3' end of pJM102ΔNS (i.e. replacing the *Cb-ges-1* PRA). *C. elegans* transformed with pJM102ΔNS-2x*Ce-ges-1*PRA were stained for esterase activity; no pharynx/rectum expression was detected (figure 29). GAP alignments of the PRA sequences from *Ce-ges-1* and *Cb-ges-1* did not reveal any obvious sequence conservation.

Figure 29: The *Ce-ges-1* pharynx/rectum activator element is unable to replace the *Cb-ges-1* pharynx/rectum activator element. Heavy gut staining with weak pharynx/rectum staining was observed in pJM102 transformed *C. elegans* embryos (top). Heavy gut staining, but no detectable pharynx/rectum staining was observed for pJM102ΔNS and pJM102ΔNS-2x*Ce-ges-1*PRA transformed embryos (middle and bottom, respectively). The thick grey line represents the pJM102 sequence, while the thick black lines indicate deleted sequences in the various "test" constructs. The vertical dotted lines indicate position of the *C. briggsae* PRA in pJM102. The schematic gene sequences are drawn to scale, including the two copies of the *Ce-ges-1* PRA (in box).


C. elegans
Transformation
Cb-ges-1 PRA
...3060
pJM102
5915...
...3749
pJM102ΔNS
5915...
...3060
pJM102ΔNS-2x Ce-ges-1 PRA

Discussion

Manipulation of the "wildtype" DNA sequence through deletion or mutation often allows for the identification of specific regulatory sequences required for proper gene expression. Altering the DNA sequence is often assumed to affect the ability of DNA binding proteins to bind the DNA. The regulation of *ges-1* expression is believed to be controlled by a combination of activator and repressor molecules at the level of transcription. Other possible explanations for *ges-1* regulation include: nucleosome positioning, changes in mRNA stability, regulation of translation initiation and/or elongation; however, the idea of a combination of transcription factors regulating *ges-1* at the level of transcription requires a less complex explanation than those involving chromatin structure or post-transcriptional modifications. Furthermore, evidence from the enhancer assay experiments suggest that the *Ce-ges-1* and *Cb-ges-1* regulatory elements are modular and can function independently from the *ges-1* gene.

Spatial Arrangement of the Identified ges-1 Regulatory Elements

The spatial arrangement of the *ges-1* gut activator-pharynx/rectum repressor (i.e. tandem pair of GATA sites) regions are conserved between *C. elegans* and *C. briggsae*. The *Ce-ges-1* GATA sites are located 1.1 kb upstream from the ATG codon (figure 18), while the *Cb-ges-1* GATA sites are located 0.7 kb upstream from the ATG codon (figure 17).

Spatial arrangement of the *ges-1* pharynx/rectum activator is different in *C. elegans* and *C. briggsae*. The putative *Ce-ges-1* pharynx/rectum activator was mapped to the 5' end (-68 to -6 bp relative to the translation initiation codon; figure 28), while the putative *Cb-ges-1* pharynx/rectum activator was mapped to the 3' end (+57 to +332 bp relative to the poly-adenylation signal site; figure 27) of the respective *ges-1* genes. The presence of a *Cb-ges-1* PRA element was first discovered through transformation of *C. elegans* using the *Cb-ges-1* 3' deleted construct, pJM102Δ3'. When pJM102Δ3' was used to transform *C. briggsae*, an apparent deregulation of the esterase expression pattern was observed (i.e. no discernable pattern of expression), not the disappearance of the pharynx/rectum staining that was observed for pJM102Δ3' transformed *C. elegans* embryos.

Previous work done on the *Est-6* genes and *Adh* genes of Drosophila showed that these genes are regulated (in part) by sequences located at the 3' end of the gene. In the *D. melanogaster Est-6* gene, the regulatory element responsible for esterase expression in the ejaculatory duct was mapped to the 3' end of the *Est-6* gene (Healy *et al.*, 1996). The *D. pseudoobscura Est-5B* gene is not expressed in the ejaculatory duct of males, suggesting that *Est-6* has only recently acquired this tissue specific regulatory element (Healy *et al.*, 1996). Sequences in the 3' end of the *D. mulleri Adh-2* gene were found to regulate Malpighian tubule expression of *Adh*, while sequences in the 5' distal promoter of *D. melanogaster Adh* gene were found to regulate expression in the

Malpighian tubules (Fischer and Maniatis, 1986; Falb *et al.*, 1992). That the putative *Cb-ges-1* PRA region is located at the 3' end of the gene and the putative *Ce-ges-1* PRA region is located at the 5' end of the gene does not appear to be a unique feature of the *ges-1* genes.

Sequence Conservation of the Gut Activator-Pharynx/Rectum Repressor Elements

Endogenous Ce-GES-1 activity is found exclusively in the gut of *C. elegans* embryos, while the vast majority of *Cb-ges-1* activity is also found in the gut of *C. briggsae* embryos (less than 1% of endogenous esterase activity is seen in the pharynx/rectum; figure 11). For both *C. elegans* and *C. briggsae*, activation of *ges-1* expression in the gut centres around a conserved tandem pair of WGATAR sites. The most obvious sequence difference between the *ges-1* tandem GATA sites is the eight base pair insertion between the *Cb-ges-1* tandem pair of GATA sites versus the *Ce-ges-1* GATA sites (figure 16). Inspection of the sequences immediately flanking the GATA elements revealed a large degree of sequence variation. Furthermore, five consensus WGATAR sites were found within the deleted sequences of pJM102ΔAN, yet only the tandem GATA region demonstrated any obvious regulatory capacity. Three WGATAR sites were identified within the deleted sequences of the *Ce-ges-1* Δ6 construct, two of which account for the tandem GATA sites. Figure 30 lists the

Figure 30: Summary of the tandem GATA region sequence modifications for *Cb-ges-1* and *Ce-ges-1*. The GATA site positions are underlined; m = mutated GATA sequence; Δ = deleted GATA sequence.

Cb-ges-1

CTGCTGTGATAAGAAAAGTGTTATACTGATAAGAACGTAG	wildtype
CTGC GAAAAGTGTTATACTGATAAGAACGTAG	ΔUpstream
CTGCTGTGATAAGAAAAGTGTTAT G	ΔDownsteam
C G	ΔTandem
CTGCTGCCGCTGGAAAAGTGTTATACTGATAAGAACGTAG	mUpstream
CTGCTGTGATAAGAAAAGTGTTATACCCGCTGGAACGTAG	mDownstream
CTGCTGCCGCTGGAAAAGTGTTATACCCGCTGGAACGTAG	mTandem
TGCTGTGATAAGAAAAGTGTTATACTGATAAGAACG	Repositioned

Ce-ges-1

CTGATGCAACTGATAGCAAAACTGATAAGGGTCAAAATTTCAG	wildtype
CTG AACTGATAAGGGTCAAAATTTCAG	ΔUpstream
CTGATGCAACTGATAGCAA CAG	ΔDownstream
CTG CAG	ΔTandem
CTGATGCAACGTCGCTCAAAACTGATAAGGGTCAAAATTTCAG	mUpstream
CTGATGCAACTGATAGCAAAACGTCGCCGGGTCAAAATTTCAG	mDownstream
CTGATGCAACGTCGCTCAAAACGTCGCCGGGTCAAAATTTCAG	mTandem
ATGCAACTGATAGCAAAACTGATAAGGGTCAAAATTTCAG	Δ7 Region

sequence manipulations performed during the analyses of the *Ce-ges-1* and *Cb-ges-1* GATA regions.

The downstream *ges-1* GATA site sequences from *C. elegans* and *C. briggsae* are highly conserved; not only is the core consensus sequence conserved, but some of the flanking sequences are also conserved (i.e. ACTGATAAG; figure 16). As an aside, the sequence CTGATAAG was originally recognized in the promoters of the *C. elegans* vitellogenin genes; these GATA sites are important for gut specific expression of the vitellogenin genes (MacMorris *et al.*, 1992; MacMorris *et al.*, 1994). Perhaps the *Ce-ges-1* and *Cb-ges-1* downstream GATA sites have been more conserved than the upstream GATa sites because the downstream GATA site is more important than the upstream site for regulating *ges-1*. The *Ce-ges-1* and *Cb-ges-1* upstream GATA sites do vary somewhat, though the variation is still within the accepted WGATAR consensus sequence (figure 16).

For *Cb-ges-1*, the upstream core GATA site is TGATAA, while the *Ce-ges-1* upstream core GATA site is TGATAG (figure 16). Furthermore, three adenosine nucleotides on the 3' side of the upstream GATA site are conserved between *Ce-ges-1* and *Cb-ges-1*. Since the *Cb-ges-1* upstream GATA site does not appear to be strictly required for *Cb-ges-1* expression (as shown by mutation of the upstream GATA site), the sequence variation surrounding the *Ce-ges-1* and *Cb-ges-1* GATA sites could be attributed to mutations accumulated after speciation of *C. elegans* and *C. briggsae*. Differences in the *ges-1* upstream

GATA sequences (including surrounding sequences) are noteworthy because of the observed differences in the expression pattern of *ges-1* between *C. elegans* and *C. briggsae*. *Ce-ges-1* is endogenously expressed only in the gut of the *C. elegans* digestive tract. Endogenous *Cb-ges-1* is expressed mostly in the gut of *C. briggsae*, though faint esterase activity was observed in the *C. briggsae* pharynx/rectum regions at four hours of staining. Slight differences in the regulatory sequences may alter the binding efficiency of *ges-1* transcription factors, thereby allowing for the differences in esterase expression between *C. elegans* and *C. briggsae*.

ges-1 regulation is thought to be controlled (at least in part) by a GATA factor binding to the GATA region. At least four GATA factor proteins are known to exist in *C. elegans*, in addition to several potential GATA factor encoding genes recently identified from the *C. elegans* genome sequencing project (Dr. J. D. McGhee, personal communication). *elt-2* and *end-1* are expressed in the gut, while *elt-1* and *elt-3* are expressed in non-gut tissues (Hawkins and McGhee, 1995; Zhu *et al.*, 1997; Spieth *et al.*, 1991; Dr. J. S. Gilleard, personal communication). Over expression studies have shown that *elt-2*, but neither *end-1*, *elt-1* nor *elt-3* was capable of driving ectopic *ges-1* expression (Fukushige *et al.*, in press; unpublished data, Dr. T. Fukushige); *elt-2* has been proposed to regulate expression via the tandem GATA sites (Hawkins and McGhee, 1995). With GATA factors present in lineages outside of the gut, how are genes containing WGATAR regulatory regions properly expressed? It

seems reasonable to assume that the WGATAR sequence context within a given gene is important for biological function and not just the transcription factor binding sequence.

Study of the ELT-1 binding properties provides an example where GATA site context affects the DNA-protein interaction. As yet no downstream target genes have been identified, but *elt-1* contains (presumed) autoregulatory GATA sites (Shim *et al.*, 1995). Characterization of *elt-1* GATA sequence recognition was performed using a yeast *in vivo* expression system to map the functional domains of ELT-1 (Shim *et al.*, 1995). Transactivation of the β-galactosidase reporter gene by *elt-1* was dependent on the context of the *elt-1* GATA sites, i.e. flanking sequences, sequence copy number and the GATA sequence itself (Shim *et al.*, 1995). Extrapolation from the *elt-1* study suggests that the three additional GATA sites found in the *Cb-ges-1* promoter between base pairs 937 and 1334 (i.e. outside of the gut activator-pharynx/rectum repressor region) do not provide the proper sequence context for GATA factor binding. Apparently the *Cb-ges-1* tandem GATA sites are located in the correct context for regulation by a specific GATA factor.

With the possibility that the *ges-1* tandem GATA sites are in the correct sequence context, perhaps the *ges-1* genes can be grouped with other GATA regulated genes that possess a similar arrangement or context of GATA sites. Comparison of the *Ce-ges-1* and *Cb-ges-1* GATA binding sites with functionally relevant GATA sites from other genes reveals a large degree of variation in the

sequence arrangement (figure 31). There does not seem to be any pattern(s) in the orientation of the GATA sites or in the number of functional GATA sites present in these GATA controlled promoters. At this time it does not appear that the GATA sites from *Ce-ges-1* or *Cb-ges-1* can be placed within a family of similarly arranged GATA sites.

Sequence Conservation of the Pharynx/Rectum Activator Elements

The PRA regions from *C. elegans* and *C. briggsae* share a low degree of sequence identity, as determined from computer assisted sequence comparisons (figures 13 and 15). It is interesting that computer assisted searches of the 5' *Cb-ges-1* promoter were unable to identify any *Ce-ges-1*-like tandem GATA sites. The *ges-1* tandem GATA sites are, however, well conserved in spatial arrangement, general sequence, and overall regulatory function. The possibility remains for the PRA regions from *Ce-ges-1* and *Cb-ges-1* to have a conserved regulatory sequence; however, this conservation is not high enough to be detected using computer assisted methods.

While transformation of *C. elegans* and *C. briggsae* with pJM102Δ3' resulted in different expression patterns, elimination of the pharynx/rectum staining was extremely clean (figure 27). It is, therefore, difficult to believe that such a strikingly clean result is not due to the presence of a specific regulatory sequence. The deregulated pattern of pJM102Δ3' expression (in transformed *C. briggsae*) suggests that additional information important for *Cb-ges-1* repression

Figure 31: List of selected genes containing functionally relevant GATA regulatory regions. GATA site sequences are underlined; orientation of GATA site sequences are indicated by an arrow underneath the sequence. Functional GATA sites not conforming to the consensus WGATAR sequence are in bold.

Gene	GATA sequence and Orientation	Reference
C. elegans ges-1	gcaactgatagcaaaactgataagggtcaaa	(Egan *et al.*, 1995)
C. briggsae ges-1	aaaaacgctgtgataagaaaagtgttatactgataagaacgtag	(this thesis)
C. elegans elt-1	ctgataatctttatcttg	(Shim *et al.*, 1995)
C. elegans pha-4	agagcactgattgataagag	(Kalb, in press)
C.elegans vit-2	1)gtttctgataagggt; 2)aatgagcttatcaatgcgct	(MacMorris *et al.*, 1992)
C. elegans dpy-7	ttttcttatctccgataatt	(Gilleard *et al.*, 1997)
C. briggsae dpy-7	attccttatctccgacaatt	(Gilleard *et al.*, 1997)
D. melanogaster Adh (proximal promoter)	taacggtagataatga	(Heberlein *et al.*, 1985)
D. mulleri Adh-1 (boxA promoter)	agtggtattgataagaca	(Fischer and Maniatis, 1988)
B. mori HcA/B.12 (divergent gene pair)	ccttgagataagaaac	(Drevet *et al.*, 1994)
human α-*MyHC*	ggacatgataagga	(Huang *et al.*, 1995)
hamster α-*MyHC*	tccttatcatgtcct	(Huang and Liew, 1997)
human *APR-2*	1)cacaggcctatcaccctatcttg; 2)tatcaccctatcttggccttcac	(Uchida *et al.*, 1997)
human *Interleukin-5*	cattctctatctgattg	(Yamagata *et al.*, 1997)
human β-*globin* (3' enhancer)	ggacatgataagga	(Huang and Liew, 1997)
human A-γ-*globin*	cacactatctcaatgcaaatatctgtctg	(Huang *et al.*, 1995)
human *glycoprotein IIb*	1)ttttatcggg; 2)ggagattaga; 3)attgataggc; 4)tttgataaga	(Martin *et al.*, 1993)
mouse *epoR*	agcttatctgtccccag	(Zon *et al.*, 1991)
chicken *GATA-1*	tcgctatcagataaggccttatcagtgc	(Trainor *et al.*, 1996)
mouse *GATA-1*	agtccatctgataagacttatctgctgccc	(Trainor *et al.*, 1996)
chicken α-*D*	agcttgcggataagataaggccggaattca	(Trainor *et al.*, 1996)
N. crassa lao	acgccgatactgtatctcccct	(Feng *et al.*, 1993)

of non-digestive tract expression, but different from PRA1, may be encoded within the 3' 2 kb fragment that was removed. Refinement of the putative *Cb-ges-1* (and *Ce-ges-1*) PRA element may yet reveal a specific sequence.

Function of the Gut Activator-Pharynx/Rectum Repressor Elements

The *Cb-ges-1* and *Ce-ges-1* GATA sites perform similar functions. Deletion of the tandem GATA sites switches *ges-1* expression from the gut to the pharynx/rectum portions of the digestive tract (figures 17 and 18). This observation introduces the question of how a normally lineage restricted gene can be expressed in completely different cell lineages, long after the establishment of specific tissues and the genes expressed in them. Furthermore, why would a cryptic expression pattern be retained when this expression does not appear to be required for normal worm function?

It is possible that retention of the *Cb-ges-1* pharynx/rectum expression potential may have some importance that is not readily obvious under laboratory conditions. Alternatively, the idea that gene regulatory systems cannot be changed in just any manner has previously been discussed in the literature. In a gene regulatory model put forth by Dickinson (1988), it was suggested that regulatory systems are not so much based on an individual regulatory mechanism specific to each gene, but rather on a combination of several regulatory mechanisms that may have aspects common with other genes. This implies that regulatory networks are under constraints that may not be readily

appreciated; it may not be a trivial matter to alter the regulation of one gene without altering the control of another gene. Conservation of the *ges-1* pharynx/rectum expression pattern in *C. briggsae* and *C. elegans* may suggest that the *ges-1* regulatory network is constrained in some manner that is not yet understood.

Another regulatory model that could explain the pharynx/rectum expression potential of the *Ce-ges-1* and *Cb-ges-1* gene is the presence of another gene just upstream of the *Ce-ges-1* and *Cb-ges-1* loci. Syntenic gene clusters are common between *C. elegans* and *C. briggsae*, and this observation has been exploited as a method for cloning *C. briggsae* homologues of *C. elegans* genes (Kuwabara and Shah, 1994). Perhaps tampering with the tandem GATA region allows the regulatory elements from an upstream gene to control *ges-1* expression. Northern analysis of the sequence upstream from the *Ce-ges-1* gene does not support this hypothesis (Dr. J. D. McGhee, personal communication; Fukushige *et al.*, 1996). The *C. elegans* Genome Sequencing Consortium has not yet finished sequencing the chromosomal region where *Ce-ges-1* is located.

Deletion of the *Cb-ges-1* GATA sites (tandem, downstream, or upstream) switches esterase expression from the gut to the pharynx/rectum (figure 17). Relocation of the tandem GATA sites within the *Cb-ges-1* gene re-instated gut specific esterase staining (figure 21). Mutation of the tandem and downstream GATA sites turned the expression switch from the gut to the pharynx/rectum, but

mutation of the upstream GATA site (as opposed to deletion) had no obvious effect on the *Cb-ges-1* expression pattern (figure 19). It was also observed that multiple copies of either the tandem, downstream, or upstream GATA sites allowed gut specific expression of a reporter construct (i.e. GATA site activation of gut expression is independent of the remainder of the *Cb-ges-1* gene; figures 23 and 24).

The major difference between mutation and deletion of the *Cb-ges-1* upstream GATA site was that sequence mutation retained the sequence spacing (TGATAA was changed to CCGCTG; figure 19), while sequence deletion removed eight base pairs of the regulatory region (TGTGATAA was deleted; figure 17). The act of deletion altered the spacing within the gut activator-pharynx/rectum repressor element, thereby changing the orientation of the GATA sites with respect to rotational and transitional position along the double stranded DNA helix. Proper alignment of DNA binding proteins has been shown to be important for *trans*-activating some genes (van der Vliet and Verrijzer, 1993). Relocation of the *Cb-ges-1* tandem GATA sites (including five base pairs upstream from the upstream GATA site) to a position 400 bp downstream from the normal position fully restored gut expression, while simultaneously repressing pharynx/rectum expression (figure 21). Since relocation of the tandem GATA sites restored proper *Cb-ges-1* expression, any specific spacing requirement(s) between the two GATA sites must be contained within the relocated sequence. Relocation of the *Cb-ges-1* tandem pair of GATA sites

indicated that contiguous sequences were not strictly required for proper *Cb-ges-1* expression, implying a modular organization of regulatory elements (see Yuh *et al.*, 1998). Relocation experiments with the *Ce-ges-1* tandem GATA sites also indicated that sufficient information was present within the tandem GATA region for directing *Ce-ges-1* expression to the gut (Egan *et al.*, 1995; see figure 22).

It appears that the *ges-1* pharynx/rectum repressor elements are linked to the gut activator elements because simultaneous heavy gut and heavy pharynx/rectum expression was never observed during analysis of either *Ce-ges-1* or *Cb-ges-1* (Aamodt *et al.*, 1991; Egan *et al.*, 1995; Fukushige *et al.*, 1996; Kennedy *et al.*, 1993; this thesis). Examples of regulatory element linkage can be found in other genes. The *D. melanogaster Adh* gene is controlled by two promoters, ALE for larval expression and AAE for adult expression. Within the AAE sequence two overlapping sequences were identified that bound two different transcription factors (see figure 1 in Introduction). The consensus C/EBP binding site was shown to act in a positive manner, and was recognized by the mammalian C/EBP transcription factor (Abel *et al.*, 1992). The second binding site, which overlaps the C/EBP consensus site, was recognized by the transcriptional repressor, AEF-1 (Abel *et al.*, 1992). Binding experiments demonstrated that AEF-1 and C/EBP bound to the regulatory sequence in a mutually exclusive manner, with AEF-1 being dominant over C/EBP (Abel *et al.*, 1992). This example illustrates the potential of a short sequence to possess

both positive and negative regulatory functions. Closely positioned positive and negative regulatory elements appear to be present within the GATA site region for both *Ce-ges-1* and *Cb-ges-1*.

Function of the Pharynx/Rectum Activator Element

The *Ce-ges-1* PRA was mapped to the 5' end of the gene, while the *Cb-ges-1* PRA was located in the 3' end of the gene (figures 27 and 28). Even though the spatial arrangement of the putative PRAs are different for each *ges-1* homolog, function of the putative PRA elements appears to be conserved. Removal of either the *Ce-ges-1* or the *Cb-ges-1* PRA sequences eliminated all esterase activity in the pharynx/rectum. Both PRA elements were mapped to relatively small areas (68 bp for *Ce-ges-1* and 275 bp for *Cb-ges-1*), suggesting that the *ges-1* pharynx/rectum expression is controlled by a specific sequence, and thus has been specifically retained for an unknown reason. Specific differences in PRA function between *Ce-ges-1* and *Cb-ges-1* will be difficult to uncover until a short DNA sequence (e.g. 10-20 bp) has been identified for the *ges-1* PRA of either species.

While both *Ce-ges-1* and *Cb-ges-1* appear to have PRA elements, preliminary work suggests that the *Ce-ges-1* and *Cb-ges-1* PRA elements are partly species specific. Two copies of the *Ce-ges-1* were added to a *Cb-ges-1* 3' end deleted construct (pJM102ΔNS), but this was unable to restore pharynx/rectum expression (figure 29). Functional incompatibility of the *Ce-ges-*

1 PRA with *Cb-ges-1* suggests that the pharynx/rectum activator regions may not be under the same selective pressure as the GATA regulatory regions. The apparent divergence of the *Ce-ges-1* and *Cb-ges-1* PRA sequences suggests the PRA regulatory elements may be undergoing evolutionary change; however, some portion of the regulatory mechanism must be conserved because pharynx/rectum expression is observed when *Cb-ges-1* constructs are used to transform *C. elegans*. Further experiments will better characterize the *ges-1* PRA elements from each species.

Model For Cb-ges-1 Expression

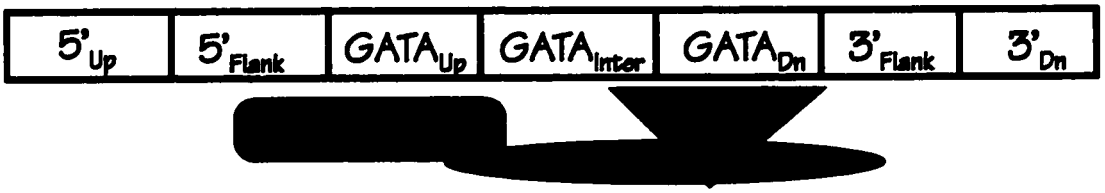
How can *Cb-ges-1* be expressed almost exclusively in the gut (only a small amount of pharynx/rectum expression exits), yet retain the ability to be expressed in other cell lineages of the digestive tract? Figure 32 (top) illustrates a possible model for *Cb-ges-1* regulation. For this model, the gut activator-pharynx/rectum repressor element has been arbitrarily divided into seven regions (figure 32): 5' upstream ($5'_{Up}$), 5' flanking ($5'_{Flank}$), upstream GATA ($GATA_{Up}$), inter-GATA ($GATA_{Inter}$), downstream GATA ($GATA_{Dn}$), 3' flanking ($3'_{Flank}$), and 3' downstream ($3'_{Dn}$) regions. Pharynx/rectum activation is not specifically discussed in this model because not enough evidence has been gathered at this time to suggest any particular mechanism.

Cb-ges-1 gut expression requires at least two transcription factors. Deletion or mutation of the downstream GATA site eliminated gut expression

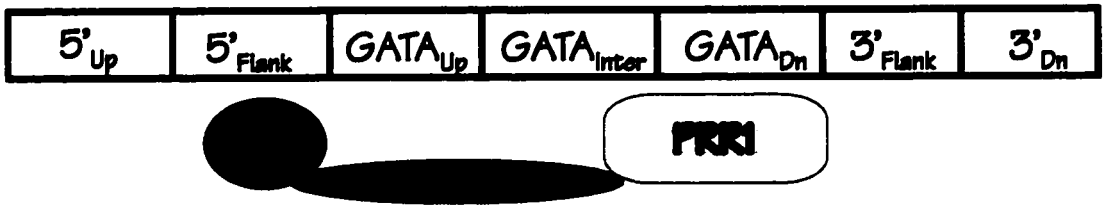
Figure 32: Possible models for *ges-1* gut activation and pharynx/rectum repression. The top panel illustrates the model for *C.b-ges-1* regulation, while the bottom panel illustrates the *Ce-ges-1* regulatory model. Six regulatory sequences containing protein binding sites are suggested. Proteins GA1 (a potential GATA factor) and GA2 are proposed to be central for gut activation, while PRR1 and PRR2 (a potential GATA factor) are required for full pharynx/rectum repression. The linker factors may not be required for proper *Cb-ges-1* expression, however, the linker protein is shown in the diagram. The linker factors are probably necessary for *Ce-ges-1* regulation. The GA3 and GA4 proteins are not required for *Cb-ges-1* regulation (top), but are needed for proper *Ce-ges-1* expression (bottom). A complete description of the proposed model is found in the text.

Model for *Cb-ges-1* Regulation

Gut Activation

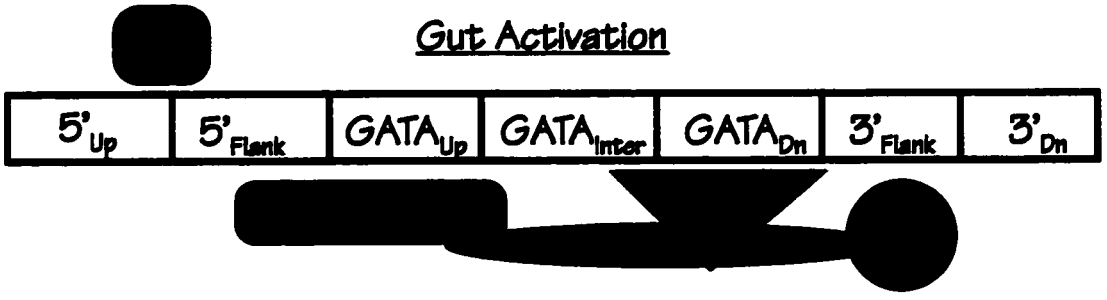


Pharynx/Rectum Repression

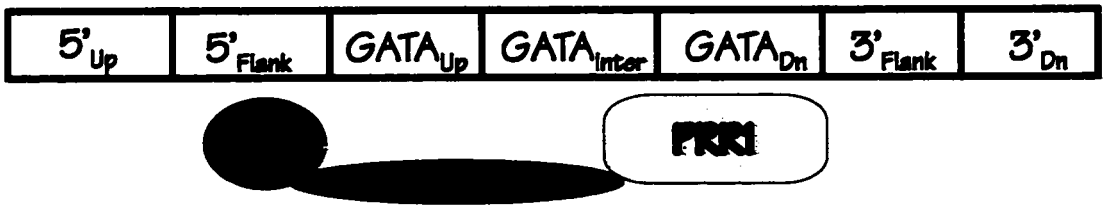


Model for *Ce-ges-1* Regulation

Gut Activation



Pharynx/Rectum Repression



(figures 17 and 19). Gut activator 1 (GA1) would specifically recognize the core GATA site, with a potential requirement for sequences within the inter-GATA region and the 3' flanking regions. GA1 is postulated to be a GATA factor expressed in the gut. Deletion of the upstream GATA site eliminated gut expression, while mutation had no effect on gut expression (figures 17 and 19). DNA binding of gut activator 2 (GA2) would specifically require only sequences in the 5' flanking region, though the A-rich inter-GATA region may also be required; the upstream GATA core sequence is not specifically required for *Cb-ges-1* expression, except that it contributes to the spacing between the GA1 and GA2 binding sites. After binding the regulatory elements, GA1 and GA2 would then interact to stabilize the activation complex. Alternatively, a stabilizing factor may be required to "link" GA1 and GA2 together at the gut activator element (labelled as linker in figure 32). Stabilization of GA1 and GA2 is required because deleting the upstream or the downstream GATA site sequences eliminated gut expression (figure 17). GA1 and GA2 are proposed to both function as transcriptional activators that cooperate with each other. Multimerization of the upstream and downstream GATA sites allowed gut specific expression (figures 23 and 24); sequence multimerization was apparently sufficient to overcome the need for stabilization. It is equally plausible that GA1 was able to recognize and bind (albeit weakly) the multiple copies of the upstream GATA site to activate transcription.

The observation that relocation of the tandem GATA sites restored proper *Cb-ges-1* expression argues that the regulatory elements function in a modular fashion. A DNA looping mechanism would allow interaction of the GATA regions, 5' regions and 3' regions to activate gut expression (and repress pharynx/rectum expression). Whether DNA looping or bending would be protein induced or a natural property of the DNA in the *ges-1* promoters remains to be tested.

For *Cb-ges-1* pharynx/rectum repression a minimum of two repressor factors are required. Deletion or mutation of the *Cb-ges-1* downstream GATA site allowed pharynx/rectum expression (figures 17 and 19). The pharynx/rectum repressor 1 (PRR1) would specifically recognize either the downstream GATA site, or an expanded version of the downstream GATA site. PRR1 is postulated to be a GATA factor expressed in the pharynx and rectum regions of the digestive tract. Deletion of the upstream GATA region allowed heavy pharynx/rectum expression, while mutation had no effect on the *Cb-ges-1* expression pattern (figures 17 and 19). The second repressor factor, pharynx/rectum repressor 2 (PRR2), would bind sequences encompassing the 5' flanking region, and possibly a portion of the inter-GATA region. PRR2 does not specifically recognize the upstream GATA site, but its binding may require the spacing provided by the upstream GATA site. Upon binding, PRR1 and PRR2 would interact (either directly or indirectly through a third "linker" protein) to form a stable repressor complex. Detection of weak endogenous pharynx/rectum

expression from the *Cb-ges-1* gene suggests that the sequences involved in *Cb-ges-1* pharynx/rectum repression do not allow strong repressor protein(s)-DNA interaction, thereby allowing low levels of transcription in the pharynx/rectum (figure 11).

The model presented for *Cb-ges-1* regulation must in some way also account for regulation of the *Ce-ges-1* gene. Transformation of *C. elegans* with the *Cb-ges-1* gene resulted in expression of *Cb-ges-1*, thereby suggesting conservation of some regulatory factors. The surrounding sequences of the *Ce-ges-1* have regulatory functions. Individually deleting the *Ce-ges-1* upstream GATA, downstream GATA, or Δ4 regions (corresponding to 3' flanking and/or 3' downstream regions) resulted in anterior gut expression (figure 18). To accommodate these results in the present model, a gut activator protein (GA3) is suggested to bind the Δ4 region. GA3 would not be able to activate transcription, but rather help to provide an environment conducive to GA1 and GA2 DNA binding. Deletion of any two of the three sequences was required to eliminate gut expression (Egan *et al.*, 1995; figure 20). These observations imply the presence of a factor in the anterior gut that forms a stable bridge between the Δ4 binding factor, GA1 and GA2 (labelled as linker in figure 32) . Deletion of the region located immediately 5' to the upstream GATA site virtually eliminated gut expression (Egan *et al.*, 1995). This observation requires the presence of another factor(s) (GA4) that can recognize sequences within the Δ7 region (denoted as 5' flanking and/or 5' upstream regions). GA4 may aid in

providing a DNA conformation conducive to transcription (e.g. introducing DNA bends, release of nucleosome proteins) that counteracts an upstream negative influence. With the observations collected for the regulation of *Cb-ges-1*, it appears that GA3 and GA4 are not required for *Cb-ges-1* expression; regions corresponding to the *Ce-ges-1*Δ7 and Δ4 regions were not found. Repression of *Ce-ges-1* pharynx/rectum expression is postulated to be essentially the same as for *Cb-ges-1*.

The present model is an extension of a model previously proposed by Egan *et al.* (1995). The major difference between the model proposed in this thesis and that by Egan *et al.* (1995) concerns the tandem GATA sites. In the previous model a GATA factor was able to simultaneously bind both GATA sites; I have suggested that a GATA factor only recognizes the downstream GATA site, while a second unknown factor recognizes sequences within the region of the upstream GATA site, but not necessarily the consensus upstream GATA site sequence. ELT-2 is a single zinc finger GATA factor that is currently the favoured protein involved for *Ce-ges-1* gut activation (discussed later on). With only a single zinc finger, two ELT-2 molecules would probably be required to bind the tandem GATA sites and mutation of the *Cb-ges-1* GATA site does not allow for a factor to specifically recognize the upstream GATA site.

There also appears to be a difference between the nematodes species in the ability of *Ce-ges-1* to be expressed in the anterior versus the posterior gut of *C. elegans* embryos. Separation of *Cb-ges-1* expression from the anterior to

posterior gut was not observed, therefore, the *Cb-ges-1* model does not require a difference between the anterior and posterior portion of the gut. Pharynx/rectum repression is similar to the previously proposed *Ce-ges-1* model. The exception being that the upstream GATA sequence is not specifically required for *Cb-ges-1*. Rather, specific sequences in the 5' flanking region are recognized by PRR2.

While this model is able to explain several of the experimental observations, it probably over simplifies *Cb-ges-1*/*Ce-ges-1* regulation. The central idea in this model is the cooperation of at least two binding factors within the tandem GATA site region; one of these factors is thought to be a GATA factor. Bandshift analysis of the tandem GATA sites using embryo extracts did not suggest two proteins were binding to the tandem GATA sequences; however, it is possible that the bandshift conditions were not optimal for DNA binding all factors (Stroeher *et al.*, 1994). Footprint analysis suggests that other proteins are protecting the sequences within the tandem GATA sites (Stroeher *et al.*, 1994). This model does not address the weak pharynx/rectum expression observed when the *Ce-ges-1* GATA sites are mutated. Nevertheless, the proposed model provides an experimental frame work for future investigation.

Potential ges-1 Regulating Factors

What are some factors that could be involved in regulating *ges-1* as postulated by the above model? A cDNA expression library was screened for

proteins that could specifically bind to a probe containing multiple copies of the *Ce-ges-1* downstream GATA site. A single zinc finger GATA factor named *elt-2* was discovered (Hawkins and McGhee, 1995). Competition bandshift experiments showed that *elt-2* could recognize both the upstream and the downstream GATA sites of *Ce-ges-1*, and ectopic expression of *elt-2* demonstrated that *elt-2* could specifically activate *ges-1* expression (Hawkins and McGhee, 1995; Fukushige *et al.*, in press). Furthermore, antibodies against *elt-2* revealed that *elt-2* is only present in the gut of *C. elegans* (Fukushige *et al.*, in press). Results from these experiments suggest that ELT-2 binds the downstream GATA site to facilitate expression of *Ce-ges-1* in the gut. However, *Ce-ges-1* is still expressed in an *elt-2* null allele strain of *C. elegans*, *ca15*, though the gut morphology of these mutant worms is severely defective (Fukushige *et al.*, in press). Together these observations suggest that *elt-2* may be sufficient for *Ce-ges-1* transactivation, but is not absolutely required.

Another GATA factor has recently been identified that could interact with the *Ce-ges-1* GATA sites. The *C. elegans end-1* gene is expressed in the gut, and appears to be important for specifying the E lineage, as shown by the gutless phenotype of a genomic deficiency that includes *end-1* (Zhu *et al.*, 1997). While *elt-2* is still considered to be the best candidate for controlling endogenous *Ce-ges-1* expression, the possibility of *end-1* regulating *Ce-ges-1* expression cannot be ruled out at this time. The deficiency that removes the

end-1 gene (i.e. *itDf2*) also eliminates *Ce-ges-1* staining; however, *itDf2* is estimated to remove <200 kb of DNA (Zhu *et al.*, 1997).

The *C. elegans pha-4* gene is a candidate for regulating *ges-1* expression in the pharynx/rectum regions. *pha-4* is a *forkhead*/HNF-3 homolog that is expressed in both the pharynx and rectum regions of the digestive tract, along with weak expression in the gut (Kalb *et al.*, in press). However, over expression of *pha-4* in all cells of *C. elegans* embryos was incapable of activating *ges-1* expression, suggesting that *pha-4* could function in suppressing *Ce-ges-1* expression in the pharynx/rectum regions (Dr. J. M. Kalb, personal communication). If *pha-4* does activate pharynx/rectum expression, a second factor would be required.

Summary

1) Identification and characterization of the *Cb-ges-1* regulatory elements was done.

   a) This work identified a gut activator-pharynx/rectum repressor element, which contains a tandem pair of GATA sites. The tandem GATA sites are located at the 5' end of the *Cb-ges-1* gene. Each GATA site was shown to possess *cis*-activating information.

   b) This work also identified a 275 bp region containing a potential pharynx/rectum activator element, which was mapped to the 3' end of the *Cb-ges-1* gene.

2) The *Ce-ges-1* and *Cb-ges-1* regulatory elements were compared.

   a) The tandem pair of GATA sites from *Ce-ges-1* and *Cb-ges-1* are similar in spatial position and sequence. The *Ce-ges-1* and *Cb-ges-1* GATA regions differ slightly in sequence and function: *Ce-ges-1* requires both GATA sites (plus additional sequence(s) in the Δ4 region) for intact gut expression-pharynx/rectum repression; *Cb-ges-1* requires the downstream GATA site and sequences immediately surrounding the upstream GATA site.

   b) Both *Ce-ges-1* and *Cb-ges-1* possess potential pharynx/rectum activator elements, with the 68 bp *Ce-ges-1* PRA found at the 5' end of the gene and the 275 bp *Cb-ges-1* PRA located at the 3'

end of the gene. The *Ce-ges-1* 68 bp PRA was functionally unable to replace *Cb-ges-1* 275 bp PRA function.

3) A model for *ges-1* regulation was presented. This model suggests a modular organization of the *ges-1* regulatory elements. The model centres around a tandem pair of GATA sites, with a GATA factor binding to the downstream GATA site and a second factor recognizing sequences surrounding the upstream GATA site (though not specifically the upstream GATA site).

Future Work

The work presented in this thesis provides a strong basis for future investigation of *ges-1* regulation. Below is a list of experiments that would be useful in furthering our knowledge of how *ges-1* is regulated:

1) Fine mapping of the *Ce-ges-1* and the *Cb-ges-1* PRA elements would be useful. Identification of a specific 10-20 bp sequence would provide a better understanding of how (and possibly why) pharynx/rectum esterase expression is activated in a normally lineage restricted gene. Insights into the evolution of a seemingly unnecessary regulatory elements may be possible through a comparison of the *Ce-ges-1* and *Cb-ges-1* PRA elements.

2) Investigation of *Ce-ges-1* at the 3' end would also be useful. The updated dot matrix analysis comparing *Cb-ges-1* with *Ce-ges-1* revealed a 100 bp region sharing approximately 76% identity. This result suggests another regulatory element may be present at the 3' end of *Ce-ges-1* and *Cb-ges-1*.

3) Completion of the *C. briggsae* transformations is required. Additionally, transformation of various *Ce-ges-1* constructs into *C. briggsae* would complete the characterization of the *ges-1* regulatory elements. These experiments would allow comparison of both conserved and divergent properties of *ges-1* regulation.

4) Investigation of the various *ges-1* regulatory regions outside of the normal gene context would provide information as to how these sequences interact with each other. The sum of the identified parts may not equal the whole of the observed expression pattern.

5) It is obvious that several transcription factors (and possibly other proteins) are involved in *ges-1* transcriptional activation/repression. Identification of these factors would provide a more complete understanding of *ges-1* regulation.

References

Aamodt, E. J., Chung, M. A., and McGhee, J. D. (1991). Spatial control of gut-specific gene expression during *Caenorhabditis elegans* development. *Science* **252,** 579-82.

Abel, T., Bhatt, R., and Maniatis, T. (1992). A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liver-specific regulatory elements. *Genes & Development* **6,** 466-80.

Abel, T., Michelson, A. M., and Maniatis, T. (1993). A Drosophila GATA family member that binds to *Adh* regulatory sequences is expressed in the developing fat body. *Development* **119,** 623-33.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987-present). Current protocols in molecular biology. *In* "Short protocols in molecular biology : a compendium of methods from Current protocols in molecular biology" (V. B. Chanda, Ed.). John Wiley & Sons, New York.

Beachy, P. A., Varkey, J., Young, K. E., von Kessler, D. P., Sun, B. I., and Ekker, S. C. (1993). Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites. *Molecular & Cellular Biology* **13,** 6941-56.

Beato, M., and Eisfeld, K. (1997). Transcription factor access to chromatin. *Nucleic Acids Research* **25,** 3559-63.

Brady, J. P., and Richmond, R. C. (1992). An evolutionary model for the duplication and divergence of *esterase* genes in Drosophila. *Journal of Molecular Evolution* **34,** 506-21.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71-94.

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

Corbin, V., and Maniatis, T. (1990). Identification of *cis*-regulatory elements required for larval expression of the *Drosophila melanogaster alcohol dehydrogenase* gene. *Genetics* **124,** 637-46.

Dickinson, W. J. (1988). On the architecture of regulatory systems: evolutionary insights and implications. *Bioessays* **8,** 204-8.

Drevet, J. R., Skeiky, Y. A., and Iatrou, K. (1994). GATA-type zinc finger motif-containing sequences and chorion gene transcription factors of the silkworm *Bombyx mori*. *Journal of Biological Chemistry* **269,** 10660-7.

Edgar, L. G., and McGhee, J. D. (1986). Embryonic expression of a gut-specific esterase in *Caenorhabditis elegans*. *Developmental Biology* **114,** 109-18.

Egan, C. R., Chung, M. A., Allen, F. L., Heschl, M. F., Van Buskirk, C. L., and McGhee, J. D. (1995). A gut-to-pharynx/tail switch in embryonic expression of the *Caenorhabditis elegans ges-1* gene centers on two GATA sequences. *Developmental Biology* **170,** 397-419.

Falb, D., Fischer, J., and Maniatis, T. (1992). Rearrangement of upstream regulatory elements leads to ectopic expression of the *Drosophila mulleri Adh-2* gene. *Genetics* **132,** 1071-9.

Feng, B., Xiao, X., and Marzluf, G. A. (1993). Recognition of specific nucleotide bases and cooperative DNA binding by the *trans*-acting nitrogen regulatory protein NIT2 of *Neurospora crassa*. *Nucleic Acids Research* **21,** 3989-96.

Fire, A., Harrison, S. W., and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93,** 189-98.

Fischer, J. A., and Maniatis, T. (1986). Regulatory elements involved in Drosophila *Adh* gene expression are conserved in divergent species and separate elements mediate expression in different tissues. *EMBO Journal* **5,** 1275-89.

Fischer, J. A., and Maniatis, T. (1988). Drosophila *Adh*: a promoter element expands the tissue specificity of an enhancer. *Cell* **53,** 451-61.

Fukushige, T., Hawkins, M. G., and McGhee, J. D. (in press). The GATA-factor *elt-2* is essential for development of the *C. elegans* intestine. *Developmental Biology* .

Fukushige, T., Schroeder, D. F., Allen, F. L., Goszczynski, B., and McGhee, J. D. (1996). Modulation of gene expression in the embryonic digestive tract of *C. elegans*. *Developmental Biology* **178,** 276-88.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of *lymphoid enhancer factor 1* bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69,** 185-95.

Gilleard, J. S., Barry, J. D., and Johnstone, I. L. (1997a). *cis* regulatory requirements for hypodermal cell-specific expression of the *Caenorhabditis elegans* cuticle collagen gene *dpy-7*. *Molecular & Cellular Biology* **17,** 2301-11.

Gilleard, J. S., Henderson, D. K., and Ulla, N. (1997b). Conservation of the *Caenorhabditis elegans* cuticle collagen gene *col-12* in *Caenorhabditis briggsae*. *Gene* **193,** 181-6.

Greenblatt, J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. *Current Opinion in Cell Biology* **9,** 310-9.

Grosschedl, R., Giese, K., and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends in Genetics* **10,** 94-100.

Hawkins, M. G., and McGhee, J. D. (1995). *elt-2*, a second GATA factor from the nematode *Caenorhabditis elegans*. *Journal of Biological Chemistry* **270,** 14666-71.

Healy, M. J., Dumancic, M. M., Cao, A., and Oakeshott, J. G. (1996). Localization of sequences regulating ancestral and acquired sites of *esterase6* activity in *Drosophila melanogaster*. *Molecular Biology & Evolution* **13,** 784-97.

Heberlein, U., England, B., and Tjian, R. (1985). Characterization of Drosophila transcription factors that activate the tandem promoters of the *alcohol dehydrogenase* gene. *Cell* **41,** 965-77.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28,** 351-9.

Heschl, M. F., and Baillie, D. L. (1990). Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. *Journal of Molecular Evolution* **31,** 3-9.

Hodgkin, J. (1994). Epigenetics and the maintenance of gene activity states in *Caenorhabditis elegans*. *Developmental Genetics* **15,** 471-7.

Hsieh, C. L. (1994). Dependence of transcriptional repression on CpG methylation density. *Molecular & Cellular Biology* **14,** 5487-94.

Huang, W. Y., Cukerman, E., and Liew, C. C. (1995). Identification of a GATA motif in the *cardiac alpha-myosin heavy-chain*-encoding gene and isolation of a human GATA-4 cDNA. *Gene* **155,** 219-23.

Huang, W. Y., and Liew, C. C. (1997). A conserved GATA motif in a tissue-specific DNase I hypersensitive site of the *cardiac alpha-myosin heavy chain* gene. *Biochemical Journal* **325,** 47-51.

Kalb, J. M., Lau, K. K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P. G., and McGhee, J. D. (in press). *pha-4* is *Ce-fkh-1*, a *fork head*/HNF-3 alpha, beta, gamma homolog that functions in organogenesis of the C. elegans pharynx. *Development*.

Kamakaka, R. T. (1997). Silencers and locus control regions: opposite sides of the same coin. *Trends in Biochemical Sciences* **22,** 124-8.

Kennedy, B. P., Aamodt, E. J., Allen, F. L., Chung, M. A., Heschl, M. F., and McGhee, J. D. (1993). The *gut esterase* gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae. Journal of Molecular Biology* **229,** 890-908.

Kuwabara, P. E., and Shah, S. (1994). Cloning by synteny: identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Research* **22,** 4414-8.

Lee, Y. H., Huang, X. Y., Hirsh, D., Fox, G. E., and Hecht, R. M. (1992). Conservation of gene organization and trans-splicing in the *glyceraldehyde-3-phosphate dehydrogenase*-encoding genes of *Caenorhabditis briggsae. Gene* **121,** 227-35.

Lu, Q., Wallrath, L. L., Granok, H., and Elgin, S. C. (1993). (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila *hsp26* gene. *Molecular & Cellular Biology* **13,** 2802-14.

MacMorris, M., Broverman, S., Greenspoon, S., Lea, K., Madej, C., Blumenthal, T., and Spieth, J. (1992). Regulation of *vitellogenin* gene expression in transgenic *Caenorhabditis elegans*: short sequences required for activation of the *vit-2* promoter. *Molecular & Cellular Biology* **12,** 1652-62.

MacMorris, M., Spieth, J., Madej, C., Lea, K., and Blumenthal, T. (1994). Analysis of the VPE sequences in the *Caenorhabditis elegans vit-2* promoter with extrachromosomal tandem array-containing transgenic strains. *Molecular & Cellular Biology* **14,** 484-91.

Maduro, M., and Pilgrim, D. (1996). Conservation of function and expression of *unc-119* from two *Caenorhabditis* species despite divergence of non-coding DNA. *Gene* **183,** 77-85.

Martin, F., Prandini, M. H., Thevenon, D., Marguerie, G., and Uzan, G. (1993). The transcription factor GATA-1 regulates the promoter activity of the platelet *glycoprotein IIb* gene. *Journal of Biological Chemistry* **268,** 21606-12.

McGhee, J. D. (1992). Gut Esterase Expression in the Nematode *Caenorhabditis elegans*. *In* "Advances in Developmental Biochemistry", Vol. volume 1, pp. 169-210. JAI Press Inc.

McGhee, J. D., Birchall, J. C., Chung, M. A., Cottrell, D. A., Edgar, L. G., Svendsen, P. C., and Ferrari, D. C. (1990). Production of null mutants in the major intestinal esterase gene (*ges-1*) of the nematode *Caenorhabditis elegans*. *Genetics* **125,** 505-14.

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO Journal* **10,** 3959-70.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. *Nature* **386,** 569-77.

Reuter, R. (1994). The gene *serpent* has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. *Development* **120,** 1123-35.

Rippe, K., von Hippel, P. H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. *Trends in Biochemical Sciences* **20,** 500-6.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences* **21,** 327-35.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular cloning : a laboratory manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Shen, N. L., Hotaling, E. C., Subrahmanyam, G., Martin, P. F., and Sofer, W. (1991). Analysis of sequences regulating larval expression of the *Adh* gene of *Drosophila melanogaster*. *Genetics* **129,** 763-71.

Shim, Y. H., Bonner, J. J., and Blumenthal, T. (1995). Activity of a *C. elegans* GATA transcription factor, ELT-1, expressed in yeast. *Journal of Molecular Biology* **253,** 665-76.

Siegfried, Z., and Cedar, H. (1997). DNA methylation: a molecular lock. *Current Biology* **7,** R305-7.

Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Research* **14,** 6711-9.

Spieth, J., Shim, Y. H., Lea, K., Conrad, R., and Blumenthal, T. (1991). elt-1, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Molecular & Cellular Biology* **11,** 4651-9.

Stringham, E. G., Dixon, D. K., Jones, D., and Candido, E. P. (1992). Temporal and spatial expression patterns of the small heat shock (*hsp16*) genes in transgenic *Caenorhabditis elegans*. *Molecular Biology of the Cell* **3,** 221-33.

Stroeher, V. L., Kennedy, B. P., Millen, K. J., Schroeder, D. F., Hawkins, M. G., Goszczynski, B., and McGhee, J. D. (1994). DNA-protein interactions in the *Caenorhabditis elegans* embryo: oocyte and embryonic factors that bind to the promoter of the gut-specific *ges-1* gene. *Developmental Biology* **163,** 367-80.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., and et al. (1992). The *C. elegans* genome sequencing project: a beginning. *Nature* **356,** 37-41.

Sulston, J. E., and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology* **56,** 110-56.

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology* **100,** 64-119.

Svaren, J., and Horz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends in Biochemical Sciences* **22,** 93-7.

Throckmorton, L. H. (1982). Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System (J. S. F. Barker and W. T. Starmer, Eds.), pp. pp. 33-47. Academic Press, Sydney.

Trainor, C. D., Omichinski, J. G., Vandergon, T. L., Gronenborn, A. M., Clore, G. M., and Felsenfeld, G. (1996). A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Molecular & Cellular Biology* **16,** 2238-47.

Uchida, S., Matsumura, Y., Rai, T., Sasaki, S., and Marumo, F. (1997). Regulation of *aquaporin-2* gene transcription by GATA-3. off. *Biochemical & Biophysical Research Communications* **232,** 65-8.

van der Vliet, P. C., and Verrijzer, C. P. (1993). Bending of DNA by transcription factors. *Bioessays* **15,** 25-32.

Verrijzer, C. P., and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. *Trends in Biochemical Sciences* **21,** 338-42.

White, B. A. (1993). PCR Protocols: Current Methods and Applications. *In* "Methods in molecular biology." (J. M. Walker, Ed.), Vol. volume 15, pp. 392. Humana Press, Totowa.

Wood, W. B. (1988). The Nematode Caenorhabditis elegans. *In* "Cold Spring Harbour Monograph Series" (W. B. Wood and C. o. C. e. Researchers, Eds.), pp. 667. Cold Spring Harbour Laboratory Press.

Yamagata, T., Nishida, J., Sakai, R., Tanaka, T., Yazaki, Y., and Hirai, H. (1997). Of the GATA-binding proteins, only GATA-4 selectively regulates the human IL-5 gene promoter in IL-5 producing cells which express multiple GATA-binding proteins. *Leukemia* **11,** 501-2.

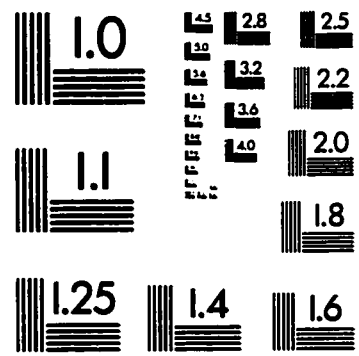
Yuh, C. H., Bolouri, H., and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* **279,** 1896-1902.

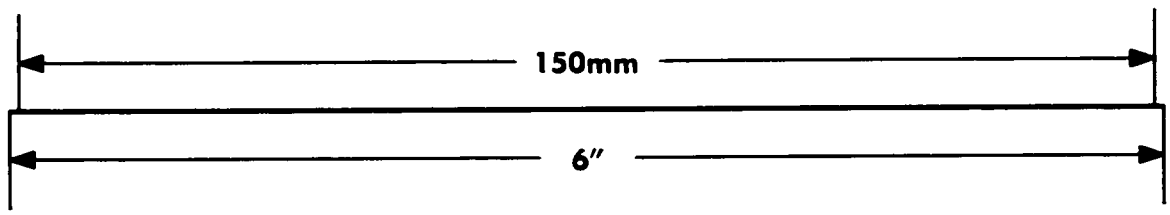
Zhu, J., Hill, R. J., Heid, P. J., Fukuyama, M., Sugimoto, A., Priess, J. R., and Rothman, J. H. (1997). *end-1* encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos. *Genes Dev* **11,** 2883-96.

Zon, L. I., Youssoufian, H., Mather, C., Lodish, H. F., and Orkin, S. H. (1991). Activation of the *erythropoietin receptor* promoter by transcription factor GATA-1. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 10638-41.

Zucker-Aprison, E., and Blumenthal, T. (1989). Potential regulatory elements of nematode *vitellogenin* genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution* **28,** 487-96.

IMAGE EVALUATION
TEST TARGET (QA-3)





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989