GENOME ANALYSIS IN *C. ELEGANS*: GENETIC AND MOLECULAR IDENTIFICATION OF GENES TIGHTLY LINKED TO *UNC-22(IV)*.

by

Marco Antonio Marra

B.Sc., Simon Fraser University, 1989

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Biological Sciences

SIMON FRASER UNIVERSITY

June 1994

APPROVAL

Name: **Marco Antonio Marra**

Degree: **Doctor of Philosophy**

Title of Thesis:

GENOME ANALYSIS IN C. ELEGANS: GENETIC AND MOLECULAR IDENTIFICATION OF GENES TIGHTLY LINKED TO UNC-22 (IV)

Examining Committee:

Chair: Dr. M. Smith

Dr. D. L. Baillie, Professor,Senior Supervisor
Institute for Molecular Biology and Biochemistry, S.F.U.

Dr. B. P. Brandhorst, Professor
Director, Institute for Molecular Biology and Biochemistry, S.F.U.

Dr. A. T. Beckenbach, Associate Professor
Institute for Molecular Biology and Biochemistry, S.F.U.

Dr. A. M. Rose, Professor
Department of Medical Genetics, U.B.C.

Dr. J. V. Price, Assistant Professor
Institute for Molecular Biology and Biochemistry, S.F.U.
Public Examiner

Dr. D. Riddle,Professor
Department of Biology, University of Missouri
External Examiner

Date Approved June 24- 1994

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Genome analysis in C. elegans: Genetic and molecular identification of genes tightly linked to unc-22(IV).

Author: _____
(signature)

MARCO ANTONIO MARRA
(name)

July 5, 1994
(date)

ABSTRACT

In this thesis the organization of genes within the *sDf10(IV)* region of the nematode *Caenorhabditis elegans* has been studied using classical genetic analysis and DNA sequencing.

In Section I, I describe the identification and analysis of developmentally required genes located within the region defined by the deficiency *sDf10(IV)*. The region spans approximately 0.5 map units or 500 kb, and contains the *unc-22* gene. Genetic screens employing 0.018 M EMS as the mutagen were conducted, resulting in the recovery of 17 mutations from 11,000 treated chromosomes. Four of these mutations each defined a new gene. There are now mutations defining eleven essential genes in this region, corresponding to a density of approximately one essential locus per 50 kb of DNA. With the exception of *let-65*, alleles of all previously identified essential genes in the region were recovered, including two alleles of *let-653*, a gene encoding a mucin-like protein. Analysis of alleles of *let-653* has shown they exhibit a complex complementation pattern.

In Section II, I present a new method to screen for duplications of essential genes located on the right half of chromosome IV. The method was used to recover two duplications of *let-56*, the first essential gene to the left of *unc-22*. One duplication is linked to the X chromosome and contains functional copies of other essential genes on chromosome IV. Analysis of this duplication has provided evidence that genes contained within the duplication may be subject to X chromosome dosage compensation.

In Section III, I describe the recovery and sequence analysis of cDNAs representing four genes located in the 50 kb to the left of *unc-22*. This has resulted in the identification of a putative Na^+ / H^+ exchanger gene, a putative histidine decarboxylase gene, a putative glutamate dehydrogenase gene, and a gene whose product possesses characteristics that suggest it may be localized to the nucleus. In

addition to these genes, four genes were predicted from nucleotide sequence 3' (left) of the *unc-22* gene. Analysis of the 50 kb interval has shown that six of eight genes are predicted to encode proteins which possess transmembrane helices, and three of eight may be involved in neurotransmitter transport or synthesis. These data are compatible with the notion that there exists an organizational theme among genes immediately to the left of *unc-22*.

ACKNOWLEDGEMENTS

There are a large number of people who have made significant positive contributions to the time I have enjoyed in the Baillie laboratory. Encouragement, guidance and superlative computing facilities were provided by my Senior Supervisor, Dr. David L. Baillie. I would like to express my gratitude to the members of my Supervisory Committee, Drs. B. P. Brandhorst, A. M. Rose and A. T. Beckenbach, for instruction and helpful advice. I have enjoyed many stimulating conversations with Dr. M. Smith, Dr. J. V. Price and Dr. B. M. Honda, who have always been available for consultation. Dr. Gabe Kalmar and the members of the I.M.B.B. Services Center, particularly Ewa Lis, Sarah Baldry and Stella Oh, have been incredibly helpful during the determination of much of the DNA sequence presented here. Thanks also to past and present members of the Smith, Price, Druehl, Brandhorst, Honda, Rose and Beckenbach labs for helpful discussions and friendship. Special thanks to Sharon Gorski, Karen Beckenbach and John Boom for early and continuing instruction on the finer points of molecular biology; to Francis Ouellette for friendship; and to Jeff Bryer for freely dispensed computing expertise. I am grateful to Dr. Erin Gilchrist, who has kindly read and commented on drafts of this thesis and other manuscripts. I would like to thank past and present members of the Baillie lab, including Dr. S. Prasad, Dr. D. Clark, Dr. R. Johnsen, Yousef Amaar, Helen Stewart, Diana Collins and Stacey Freeman for advice and support.

I have worked especially closely with Jacquie Schein, Steven Jones, Brad Barbazuk, Dinar Suleman, Kim McKim and Raja Rosenbluth. These individuals have been outstanding colleagues and friends. A finer bunch could not be found.

Special thanks to Sharon Gorski, who has graciously accepted the multiple roles of friend, partner, advisor and official Thesis Critic. She has contributed greatly to the intelligibility of this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

General Introduction.

The topic of this thesis is the identification, mapping and analysis of genes located within a small interval of the genome of the free living soil nematode *Caenorhabditis elegans*. This work has been conducted as a continuation of previous efforts to study gene expression, gene content and the organization of genes in the region of the *C. elegans* genome known as the *unc-22* region. This Introduction will first provide a consideration of reasons for studying genome organization. Next, a justification for studying genome organization in *C. elegans* will be provided. To place the work described in this thesis in the context of previous studies, descriptions of work conducted in the *unc-22* interval prior to this thesis will be provided in the Introductions to each of the Sections that follow.

Studies of genome organization attempt to describe and explain the arrangement of genes and other important sequences along the length of a chromosome. The ultimate goal of such studies is to elucidate the information content of the genome, and provide an understanding of how this information is deciphered in a temporally and spatially coordinated fashion during development. Studies of higher-order chromatin structure, summarized below, have provided an appealing model for how one level of genome organization might facilitate transcriptional regulation of gene expression.

The regulation of information flow from a eukaryotic genome is thought to result from the action of *trans*-acting factors on *cis*-linked regulatory elements, and the availability of the DNA template for transcription. It has been observed that structural changes in chromosomal organization correlate with transcriptional activation of genes. The generally higher sensitivity and site-restricted hypersensitivity to DNase I throughout the transcription unit (Butler 1983; Wu et al., 1979) indicate that transcriptional activation results from the "unpackaging" of DNA contained within the chromatin fiber. These observations are consistent with the notion that highly packaged chromatin may serve to hide the information

content of the genome. If so, there must exist mechanisms that regulate transcription by dictating the extent of DNA packaging in localized regions of the genome.

Another example of how higher order chromatin structure might be involved in the regulation of transcription has been provided by studies on the mechanisms of transcriptional enhancement. It is currently believed that transcriptional enhancers communicate with their target promoter(s) in an orientation and position-independent fashion via a mechanism that involves the outlooping of intervening DNA (Eissenberg and Elgin, 1991; Kennison, 1993). The factors preventing inappropriate outlooping and subsequent activation of heterologous promoters are unknown. One proposal has been that there exist so called "genetic boundary elements" which bracket transcription units and serve to insulate promoters from heterologous enhancers and restrict enhancer contact to only cognate promoter(s) (Eissenberg and Elgin, 1991).

The study of DNA packaging in the interphase and metaphase chromosomes of higher eukaryotes has provided evidence that chromosomes are organized into "loops", formed by the attachment of specific DNA sites [called **s**caffold **a**ssociated **r**egions (SARs) or **m**atrix **a**ssociated **r**egions (MARs)] to a proteinaceous nuclear "scaffold" (reviewed in Nelson et al., 1986). It has been proposed that scaffold binding regions located at the base of the chromosomal loops may play roles in chromosome segregation, transcriptional activation, and replication (Paulson and Laemmli, 1977; Smith et al., 1984; Mirkovitch et al., 1984; Van der Velden et al., 1984; Opstelten et al., 1989; Vaughn et al., 1990). These chromatin loops may contain only a single gene, as in the case of the *Drosophila Adh* gene (Gasser and Laemmli, 1986), or many genes, as in the case of the loop containing the *rosy* locus (Mirkovitch et al., 1986). Evidence has been obtained in chicken (Stief et al., 1989) and mouse (Bonifer et al., 1990) that SARs

can satisfy the criteria for "genetic boundary elements". That is, they can maintain the activity of existing enhancers by isolating them from chromosomal position effects and, when placed between an enhancer and reporter gene construct, can behave as functional barriers, preventing transcription from the reporter gene construct.

The examples presented above illustrate how higher order chromatin organization might play important roles in the regulation of gene expression. The linear arrangement of genes along the chromosome is another level at which genomes may exhibit organization. There are many descriptions of gene clusters in which the clustered genes exhibit a major organizational theme. For example, ultrastructural analysis by transmission electron microscopy of *Drosophila melanogaster* polytene chromatin has revealed clusters of tightly linked genes which are co-expressed during the white prepupal stage of development (Hager and Miller, 1991). These authors described a chromosomal region they named the "white prepupal locus" (or WPP). This locus consisted of a cluster of ten tightly linked co-expressed genes, spanning an estimated 42 kilobases of DNA. Nine of these genes were classified as being "highly transcribed". The arrangement of these genes, in terms of tight linkage and co-expression, parallels many developmentally regulated gene clusters in *Drosophila*. For example, there are three glue genes in the 68C glue gene cluster encoded within a 5 kb region (Meyerowitz and Hogness, 1982), five larval cuticle proteins encoded within a 9 kb region (Snyder et al., 1981; 1982) and seven closely linked heat shock genes found at the 67B locus (Ayme and Tissieres, 1985).

Descriptions of multigene families have provided examples of genes within clusters that are related structurally (i.e. at the level of DNA sequence similarity) or functionally. In general it is believed that these clusters have arisen as the result of gene duplication and inversion events during evolution, with the clustered

organization becoming fixed and maintained within the genome of a population by selective pressure (Sirotkin and Davidson, 1982; Eveleth and Marsh, 1986; Wright 1987). One of the many examples of a gene cluster containing functionally and structurally related genes is the DOPA decarboxylase (*Ddc*) gene cluster in *Drosophila melanogaster*. This cluster contains 18 genes, 14 of which are involved in the formation of cuticle. In *Caenorhabditis elegans*, genome organization of this sort has been observed in the case of certain of the actin genes (Krause et al., 1989), three of which form a tight cluster and appear to function only in myofilament containing tissues (R. H. Waterston, personal communication). Clustering of most of the characterized *C. elegans* major sperm protein genes has also been observed (Ward et al., 1988).

Recently, evidence for the existence of polycistronic RNAs in *Caenorhabditis elegans* has been obtained (Spieth et al., 1993). As in other higher eukaryotes, genes in *C. elegans* are usually monocistronic, with individual genes controlled by their own promoter elements. However, unlike most eukaryotes, the pre-mRNAs of some *C. elegans* genes are trans-spliced to one of two splice leader (SL) RNAs; SL1 (Krause and Hirsh, 1987) or SL2 (Huang et al., 1989). As part of their investigations into the mechanisms that result in trans-splicing of the SL2 splice leader to certain pre-mRNAs, Spieth et al. recovered polycistronic RNAs. Processing of the polycistronic RNA included trans-splicing of splice leader sequences to yield mature mRNA molecules. Several experiments conducted by Spieth et al. indicated that SL2 trans-splicing was the consequence of a gene's downstream location in a polycistronic transcription unit. Sufficient data has been generated to allow formation of two strong predictions. First, a cluster of closely spaced genes in the same orientation indicates the presence of a polycistronic transcription unit. Second, the first gene transcript in the transcription unit will receive SL1, while downstream gene transcripts usually receive SL2, but in some

instances will instead receive SL1. Recent unpublished data from the Blumenthal lab have borne out these predictions. The presence of polycistronic transcription units in *C. elegans* suggests a previously undescribed level of eukaryotic genome organization. It is tempting to speculate that polycistronic transcription units may provide a means of co-regulating the expression of certain genes that encode products that are functionally related.

One way to study genome organization involves the systematic identification of genes in defined regions of the genome. There are a number of ways to conduct such a study. In one method, genes residing in a particular interval are identified through mutational analysis. This method has been employed extensively in many regions of the *Drosophila melanogaster* genome, including the *zeste-white* region on the X chromosome (Judd et al., 1972), the *rosy* region on chromosome 3 (Hilliker et al., 1980), and the *dorsal - Bicaudal-D* region on chromosome 2 (Steward and Nusslein-Volhard, 1986).

A number of regions of the *C. elegans* genome have been characterized mutationally with regard to essential gene content. These include the *dpy-5* region (Rose and Baillie, 1980; Howell et al., 1987; Howell, 1989) and the *dpy-14* region (McKim et al., 1992), both on chromosome I; the region in the vicinity of the cross-over suppressor *mnC1* on chromosome II (Sigurdson et al., 1984); the region of chromosome V balanced by the reciprocal translocation *eT1(III;V)* (Rosenbluth, Cuddeford and Baillie, 1983 ; 1985; Rosenbluth et al., 1988; Johnsen, 1990; Johnsen and Baillie, 1991); the region near the *unc-22* gene on chromosome IV (Rogalski et al., 1982; Rogalski and Baillie, 1985; Clark et al., 1988; Clark 1990; Clark and Baillie, 1992); the region around the *ama-1* gene on chromosome IV (Rogalski and Riddle, 1988), and portions of the X chromosome (Meneely and Herman, 1979; 1981).

Another approach to the systematic identification of genes in defined regions of a genome has been the method of cross-species hybridization. For example, probes representing unique sequences in the human genome are often used to challenge "zoo blots", which are Southern blots containing DNA from closely related species (see McDonald et al., 1992 for an example). Hybridization of the probe to the blot indicates the evolutionary conservation of functionally important sequences within the probe DNA. The method, as applied to the detection of genes in the *C. elegans* genome, has involved hybridization of *C. elegans* DNA to the DNA of the closely related species *Caenorhabditis briggsae*. The work of Snutch (1984) demonstrated that in the region around the *hsp70* gene of *C. elegans* the only inter-species cross hybridizing restriction fragments were ones that contained *hsp70* gene sequences. This technique has been successfully employed in the detection of coding elements in other regions of the *C. elegans* genome, including a portion of the X chromosome (Heine and Blumenthal, 1986), the *dpy-14* region of chromosome I (Starr, 1989; McKim, Starr and Rose, 1992), and the *dpy-20* - *unc-22* interval on chromosome IV (Prasad, 1988; Prasad and Baillie 1989).

A recent approach to the systematic identification of genes is to determine the complete nucleotide sequence of entire eukaryotic genomes. This approach is currently being employed by groups sequencing the genomic DNA of *Saccharomyces cerevisieae* (Ouellete et al., 1993; Oliver et al., 1992) and *C. elegans* (Sulston et al., 1992, Wilson et al., 1994). Large scale sequencing projects will determine the nucleotide sequences of all of the genes contained within the genomes of these organisms; however, some method is required to relate the molecular structure of the DNA, RNA and predicted proteins to their biological functions.

Caenorhabditis elegans presents a unique system in which to pursue studies of genome organization. The most attractive feature of the worm, earmarking it for studies of this sort, is its anatomical simplicity (Brenner, 1974). Although composed of only 959 somatic cells, it possesses all the basic differentiated tissues of multicellular eukaryotes. It is the only metazoan for which the complete cell lineage from zygote to adult is known (Sulston and Horvitz, 1977; 1981; Sulston et al., 1983). It has the smallest genome known for any metazoan, with a haploid genome size estimated at 10^8 base pairs (Sulston and Brenner, 1974; Coulson et al., 1991). Virtually the entire genome has been reconstructed in overlapping yeast artificial chromosomes and cosmid clones (Coulson et al., 1986; Coulson et al., 1988; A. Coulson, personal communication) and systematic sequencing has commenced (Sulston et al., 1992; Wilson et al., 1994). Procedures for the introduction of cloned DNA fragments into the genome of individual animals have been developed (Fire, 1986; Mello et al., 1991, T. Fukushige and S. S. Siddiqui, personal communication). In addition, a system for the recovery of mutations induced by the integration and subsequent excision of the transposon Tc1 has been devised (Plasterk and Groenen, 1992). Although the primary mode of reproduction is that of a hermaphrodite, males do exist, so that genetic crosses can be performed. The generation time of three and one-half days at 20° C is short compared to those of other genetically tractable eukaryotes and in excess of 300 progeny can be obtained from a single hermaphrodite, enabling the rapid recovery and analysis of mutations. It is possible to retain mutant strains indefinitely in liquid nitrogen. Examination of the latest available release of the *C. elegans* genome mapping information database **ACeDB** (V. 2.9; Sulston et al., 1992) reveals the existence of 1412 identified loci that have been placed on the genome map, at least 980 of which have been defined by mutation (D.L. Baillie, personal communication). Recently, an additional 100 loci have been defined by mutation

on chromosome III (H. I. Stewart, D. L. Collins and D. L. Baillie, personal communication). These features combine to make *C. elegans* a near-ideal system in which to conduct studies of genome organization.

In this thesis, I have examined the organization of genes in a region of the *C. elegans* genome using both classical genetic analysis and DNA sequence analysis. In Section I, I describe the identification and genetic analysis of developmentally required genes located within the interval delimited by the deficiency *sDf10(IV)*. The region is bounded on the left by the *let-65* gene and on the right by the *let-93* gene, and spans a recombinational distance of approximately 0.5 map units, which corresponds to approximately 500 kb. Central to this region is the *unc-22* gene, which encodes a component of the worm's muscle. All mutant alleles of *unc-22* result in a constant twitch of the body surface that originates in the underlying muscle (Waterston, Thomson and Brenner, 1980; Moerman 1980). Genetic analysis has provided evidence that the *unc-22* gene product interacts with myosin (Moerman et al., 1982). *unc-22* has been cloned (Moerman, Benian and Waterston, 1986) and sequenced (Benian et al., 1989; Benian, L'Hernault and Morris, 1993). The *unc-22* transcription unit is 38,308 bp, and produces an Unc-22 protein predicted to be 6839 amino acids in length, a size that is supported by western blot analysis (Moerman et al., 1988; Benian, L'Hernault and Morris, 1993).

Previous efforts towards identifying all of the essential loci within this region of the genome have been described (Clark et al., 1988; Clark and Baillie 1992). In those studies 9889 chromosomes were treated with 0.012 M EMS and seven lethal mutations were recovered, only one of which identified a previously uncharacterized gene. In continuation of that work, I have screened over 11,000 chromosomes treated with a higher concentration of EMS (0.018 M) and recovered 17 lethal mutations, four of which each define a new complementation

group. These complementation groups have been positioned on the genetic map using a number of deficiency breakpoints. Together with previous work, a density of approximately one mutation per cosmid length of DNA has been achieved within the interval. The phenotypes of individuals homozygous for each recovered mutation have been characterized with respect to developmental arrest stage. My results demonstrate that the *let-65 - let-93* interval is not yet saturated for mutations in essential genes, and that lethal mutations provide a high density of genetic map landmarks at which the physical and genetic maps may be correlated.

In Section II, I describe a scheme devised to facilitate rapid screening for duplications of essential loci located within a large region of the genome including the *let-65 - let-93* interval. This scheme takes advantage of the potent anthelmintic levamisole. I have used this methodology to recover two duplications that contain functional copies of *let-56*, the first essential gene to the left of *unc-22*. Analysis of these duplications has revealed that one of the duplications is attached to chromosome IV. The other duplication is attached to the X chromosome and contains a functional copy of the *unc-22* gene, as well as functional copies of several essential loci adjacent to *unc-22*. Results I have obtained during analysis of the latter duplication are compatible with the notion that the copy of the *unc-22* gene located on the duplication is subject to X chromosome dosage compensation.

In Section III, I describe the recovery and DNA sequence analysis of cDNAs for four genes located on the cosmid C11F2. Three of the four predicted protein products have informative matches to other sequences contained in Genbank. One of the genes encodes a protein predicted to possess multiple transmembrane helices, and exhibits similarity to vertebrate Na^+/H^+ exchangers. Another of the genes encodes a protein with a high degree of similarity to amino acid decarboxylase enzymes. The third gene encodes a protein that exhibits similarity to vertebrate glutamate dehydrogenase enzymes. The fourth gene

encodes a protein which possesses characteristics that indicate it may bind DNA. In a collaborative effort with J. E. Schein, I have analyzed DNA sequence determined by G. Benian et al. (1989). This sequence is contained within the cosmid C18D3, and extends from *unc-22* approximately 13 kb to the left. Within this sequence we predicted the existence of four genes. In collaboration with G. Benian, we obtained evidence that three of the four predicted genes produced transcripts, and have shown that all of the putative genes encode proteins that are predicted to contain membrane spanning helices. Analysis of two deficiencies that affect all of the genes between and including *let-56* and *unc-22* (a distance of approximately 30 kb) has shown that, under laboratory conditions, none of these genes are likely to be required for any vital process during the worm's larval stages, either singly or in combination with each other.

Materials and Methods.

1. Genetics.

The genetic nomenclature used in this work follows the uniform system proposed by Horvitz et al. (1979).

Nematode strains and culture conditions: Nematodes were maintained on petri plates containing nematode growth medium (NGM) streaked with *Escherichia coli* OP50 (Brenner 1974). Mutations used in this work are listed in Table 1. The wild-type *C. elegans* strain N2 (var. Bristol) was obtained from the stock collection at the Medical Research Council, Cambridge, England. The strain DR789 was obtained from the laboratory of D. Riddle. Throughout this work, *unc-22(s7)* (Moerman and Baillie 1979) was used as a marker. All *unc-22* mutations have a recessive "twitcher" phenotype (Waterston, Thomson and Brenner 1980; Moerman 1980) with the exception of *unc-22(m52)* (D. Riddle, personal communication) and the deficiency *sDf19* (Rogalski and Baillie 1985), which have dominant twitcher phenotypes. Point mutations in *unc-22* and deficiencies that affect *unc-22* result in the production of a conditionally dominant phenotype, such that individuals heterozygous for an *unc-22* mutation twitch in a 1% nicotine solution (Sigma). Strains heterozygous for *unc-22* mutations are otherwise phenotypically wild-type (Moerman and Baillie 1979). Genotypically wild-type individuals are contracted and paralyzed in 1 % nicotine.

Construction of BC 3687: The *unc-22(s7) lev-1(x22)* chromosome on which lethal mutations were recovered following EMS treatment was constructed as follows: *unc-22(s7) unc-31(e169)* hermaphrodites were isolated from BC 2917 [*unc-22(s7) unc-31(e169)/nT1(IV); +/nT1(V)*] stock plates. These hermaphrodites were mated with *lev-1(x22)* males. Wildtype hermaphrodite progeny resulting from this cross [*unc-22(s7) unc-31(e169) + / + + lev-1(x22)*] were picked individually to separate plates and allowed to produce self progeny. Unc-22 non-Unc-31 individuals were selected from among these progeny, picked to separate

TABLE 1.

Mutations used in this study.

Mutation	Source	Reference
dpy-4(e1166)	M.R.C.[a]	D. Riddle
dpy-20(e1282)	M.R.C.	B. Edgar
lev-1(x22)	J. Lewis	Lewis *et al.*, 1980
mDf7[e]	D. Riddle	Rogalski and Riddle 1988
nDf27	R. Horvitz	Clark *et al.*, 1988
nT1(IV;V)	M.R.C.	Ferguson and Horvitz 1985
nT1(IV;V)-[let (m435)]	C.G.C.[b]	D. Riddle
par-5(it121)	K. Kemphues	unpublished
unc-22(s7)	D. Moerman	Moerman and Baillie 1979
unc-26(e345)	M.R.C.	Brenner 1974
unc-30(e191)	M.R.C.	Brenner 1974
unc-31(e169)	M.R.C.	Brenner 1974
sDf7	D. Moerman	Moerman and Baillie 1982
sDf8[l]	D. Moerman	Moerman and Baillie 1982
sDf9	D. Moerman	Moerman and Baillie 1982
sDf10	D. Moerman	Moerman and Baillie 1982
sDf19	D. Moerman	Rogalski *et al.*, 1982
sDf60	D. Clark	Clark and Baillie 1992
sDf61	D. Clark	Clark and Baillie 1992
sDf63[e]	D. Clark	Clark and Baillie 1992
sDf64[e]	D. Clark	Clark and Baillie 1992
sDf65	D. Clark	Clark and Baillie 1992

TABLE 1 (continued).

Mutations used in this study.

Mutation	Source	Reference
sDf66	J. Schein	unpublished
sDf68^{e}	J. Schein	unpublished
sDf80^{e}	J. Schein	unpublished
sDf81	J. Schein	unpublished
sDf82^{l}	J. Schein	unpublished
sDf83	J. Schein	Schein *et al.*, 1993
sDf84	J. Schein	unpublished
sDf85^{l}	J. Schein	unpublished
sDf86^{l}	J. Schein	unpublished
sDf87	J. Schein	unpublished
let-52(s42)	D. Moerman	Rogalski *et al.*, 1982
let-56(s46)	D. Moerman	Rogalski *et al.*, 1982
let-56(s173)	T. Rogalski	Rogalski *et al.*, 1982
let-59(s1087)	D. Clark	Clark and Baillie 1992
let-60(s1124)	D. Clark	Clark *et al.*, 1988
let-64(s216)	T. Rogalski	Rogalski *et al.*, 1982
let-65(s174)	T. Rogalski	Rogalski *et al.*, 1982
let-70(s1132)	D. Clark	Clark *et al.*, 1988
let-71(s692)	T. Rogalski	Rogalski and Baillie 1985
let-73(s685)	T. Rogalski	Rogalski and Baillie 1985
let-91(s753)	L. Donati	Donati 1985
let-92(s504)	T. Rogalski	Rogalski and Baillie 1985

TABLE 1 (continued).

Mutations used in this study.

Mutation	Source	Reference
let-92(s677)	T. Rogalski	Rogalski and Baillie 1985
let-93(s734)	L. Donati	Donati 1985
let-97(s1121)	D. Clark	Clark *et al.*, 1988
let-98(s1117)	D. Clark	Clark *et al.*, 1988
let-99(s1201)	D. Clark	Clark *et al.*, 1988
let-100(s1160)	D. Clark	Clark *et al.*, 1988
let-311(s1195)	D. Clark	Clark *et al.*, 1988
let-312(s1234)	D. Clark	Clark *et al.*, 1988
let-315(s1101)	D. Clark	Charest *et al.*, 1990
let-316(s1227)	D. Clark	Charest *et al.*, 1990
let-323(s1719)	D. Clark	Charest *et al.*, 1990
let-325(s1738)	D. Clark	Charest *et al.*, 1990
let-653(s1733)	D. Clark	Clark and Baillie 1992
let-654(s1734)	D. Clark	Clark and Baillie 1992
let-655(s1748)	D. Clark	Clark and Baillie 1992
lin-3(s1750)	D. Clark	Clark and Baillie 1992

[a]M.R.C. is the Medical Research Council, Cambridge, England.

[b]C.G.C. is the *Caenorhabditis* Genetics Centre.

[l] Indicates a deficiency used in PCR. Individuals homozygous for the deficiency may hatch, and arrest at a larval stage.

[e] Indicates a deficiency used in PCR. Individuals homozygous for the deficiency are arrested as embryos.

plates and allowed to produce self progeny. These individuals were exposed to 1mM levamisole. The desired recombinants were identified as Unc-22 individuals exhibiting strong resistance to levamisole. The genotype of these animals was expected to be *unc-22(s7) lev-1(x22)*. The absence of *unc-31(e169)* was confirmed by complementation tests. A single *unc-22(s7) lev-1(x22)* animal was selected and used to establish the strain BC 3687. Individuals from this strain were treated with EMS in the MAM1 screen for lethal mutations, as described in the next section.

Recovery of recessive lethal mutations tightly linked to *unc-22*:

There were two sources of previously unanalyzed lethal mutations that were used in this study. One of these sources was a screen for lethal mutations conducted by M. E. Green and D. V. Clark (screen MEG1). The other of these sources was a screen for lethal mutations conducted principally by myself (screen MAM1). The screen conducted by Green and Clark and the screen performed by myself differed in the genotype of the EMS treated individuals (below) and the number of chromosomes screened (Results). Green and Clark treated young gravid hermaphrodites of the genotype *unc-22(s7) unc-31(e169)* with 0.018 M EMS, essentially as described by Sulston and Hodgkin (1988). In the MAM1 screen, young gravid hermaphrodites of the genotype *unc-22(s7) lev-1(x22)* (from the strain BC3687) were treated with 0.018 M EMS. After treatment, animals were allowed to recover for two hours before they were mated with N2 males. Single wild-type fourth larval (L4) stage progeny (the F_1 generation) resulting from this mating were picked to individual plates and allowed to produce self progeny. The F_2 generation was screened for the absence of gravid Unc-22 hermaphrodites. F_1 individuals that gave more than three gravid Unc-22 progeny per approximately 150 progeny screened were discarded. F_1 individuals that produced three or less gravid Unc-22 progeny per approximately 150 progeny screened were considered to carry a lethal mutation, and were retained. Therefore,

for the purposes of this study, a "lethal" mutation was defined as any lesion which prevented F_2 Unc-22 hermaphrodites from either hatching or developing or accumulating eggs in the time required for the F1 parent to produce approximately 150 adult progeny. Using these criteria, a set of lethal mutations was obtained that mapped within approximately three map units of *unc-22*. Each newly isolated lethal bearing strain was maintained by picking three phenotypically wild-type individuals that twitched in 1% nicotine [genotype *let-(sx) unc-22 unc-31/+++* or *let-(sx) unc-22 lev-1/+++*], and selecting a line that did not segregate more than three Unc-22 individuals per approximately 150 progeny each generation. Lethal-bearing chromosomes were balanced over the rearrangement *nT1(IV;V)* (Ferguson and Horvitz 1985). *nT1(IV;V)* is a complex rearrangement that causes pseudo-linkage of chromosome *IV* and chromosome *V* and serves to suppress recombination over portions of these chromosomes, including that portion of chromosome *IV* containing the *unc-22* gene (Clark et al.1988; see Figure 1). Balancing of lethal mutations was accomplished in most cases using *unc-31(e169)/nT1(IV);+/nT1(V)* males as described by Clark and Baillie (1992). Lethal mutations that were recovered in the MEG1 screen were balanced by M. E. Green or D. V. Clark. Lethal mutations that were recovered in the MAM1 screen were balanced by myself. Some of the lethal mutations that originated in the MAM1 screen were balanced using the strain DR789, which has the genotype *dpy-13(IV)/[let-?(m435)]nT1(IV); unc-42(V)/[let-?(m435)]nT1(V)*. *let-?(m435)* is a recessive lethal mutation linked to either *nT1(IV)* or to *nT1(V)* that results in the early larval arrest of individuals homozygous for the *nT1(IV;V)* rearrangement. In the cases where DR789 was used to balance lethal-bearing chromosomes, lethal bearing males of the genotype *let-(sx) unc-22 lev-1/+++* were mated to DR789 hermaphrodites. Wild-type L4 hermaphrodite progeny that twitched in nicotine

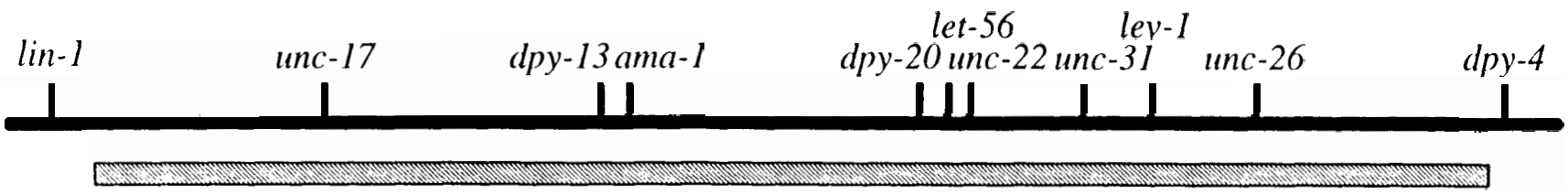


Figure 1:

A partial genetic map of the right-most portion of chromosome IV. The region of chromosome IV recombinationally balanced by the *nT1(IV;V)* reciprocal translocation is indicated by the hatched line beneath the genetic map (Ferguson and Horvitz, 1985; Clark et al., 1988).

[either *let-(sx) unc-22 lev-1/dpy-13(IV) ; +/unc-42(V)* or *let-(sx) unc-22 lev-1/[let-?(m435)]nT1(IV) ; +/[let-?(m435)]nT1(V)*] were picked to single plates, and allowed to produce self progeny. A single line failing to segregate Dpy-13 Unc-42 progeny in the next generation was retained for each lethal mutation balanced in this way.

Complementation tests: Complementation tests were conducted at room temperature (approximately 20 ^{o}C). Lethal bearing males of the genotype *let-(sx) unc-22 lev-1/+++* or *let-(sx) unc-22 unc-31/+++* were generated by crossing the *nT1(IV;V)* balanced lethal bearing strains to N2 males, and selecting F_1 males that twitched in nicotine. These males were mated to hermaphrodites of the genotype *Dfx / nT1(IV);+ / nT1(V)*, or of the genotype *Dfx / +*, or of the genotype *Dfx / [let-?m435] nT1(IV) ; + / [let-?m435] nT1(V)* where *Dfx* was one of the deficiencies in Table 1. The F_1 progeny resulting from this cross were scored for the presence of Unc-22 animals [genotype *let-(sx) unc-22 lev-1/Dfx* or *let-(sx) unc-22 unc-31/Dfx*]. If Unc-22 hermaphrodites were seen, at least four were picked to individual plates, and examined for fertility. Unc-22 animals that failed to produce progeny within five days were considered to carry a mutation that failed to complement the deficiency. Once a mutation had been mapped to the highest resolution possible with available deficiencies, it was complementation tested against alleles of the gene (or genes) residing in its immediate vicinity. In this way, mutations were assigned to pre-existing complementation groups, or were considered to define new complementation groups.

Determination of lethal arrest stage: In order to determine the stage at which each of the mutations arrested development, twitching larvae [either *let-(sx) unc-22 lev-1/let-(sx) unc-22 lev-1* or *let-(sx) unc-22 unc-31/let-(sx) unc-22 unc-31*] were picked from stock plates to individual plates at room temperature. Three days later, the general appearance of the animal on the plate was noted. After

seven days, animals were again examined. Only animals that were alive after the seven day interval were chosen for examination under an Olympus Vanox AHBS3 microscope fitted with Nomarski optics. Living animals were placed in a drop of S Buffer (Brenner 1974) on 5% agar pads, and immobilized by placing a coverslip over them, as described by Sulston and Horvitz (1977). In some cases, animals were photographed using Kodak Technical Pan film. The stage of larval arrest was determined by examining the extent of gonadal development (Kimble and Hirsh1979) in five to seven animals homozygous for each of the lethal mutations. In some instances, no living animals were observed on plates after the seven day interval. In these cases, lethal homozygotes were re-isolated from stock plates and left for a shorter amount of time (usually three days) before the extent of their gonadal development was determined. A list of all lethal mutations examined in this way and their stage of developmental arrest is presented in Table 3 (Results).

The arrest stage of relevant lethal mutations, recovered during the MAM1 or MEG1 mutagenesis screens, in *trans* to *sDf10* was also determined (Table 3). This was accomplished by mating males of the genotypes *let-(sx) unc-22 lev-1/+++* or *let-(sx) unc-22 unc-31/+++* with hermaphrodites of the genotype *sDf10(IV)/[let-?(m435)]nT1(IV) ; +/[let-?(m435)]nT1(V).* Single mated hermaphrodites were transferred to individual plates and allowed to lay eggs in two 24 hour broods. When twitching F_1 larvae were seen, between five and ten were transferred to fresh plates and observed over a period of seven days. Animals alive after seven days were chosen for examination under an Olympus Vanox microscope equipped with Nomarski optics, as described above.

Screen for duplications of *let-56(s2321)* following γ ray treatment:

A diagrammatic representation of the methodology used to screen for duplications is presented in Figure 2. In the screen for duplications of *let-56(s2321)*, young gravid hermaphrodites of the genotype *let-56(s2321) unc-22(s7)*

lev-1(x22)/nT1(IV);+/nT1(V) were treated with 1500 R of γ radiation as recommended by Rosenbluth, Cuddeford and Baillie (1985). P_0 individuals were placed on plates and allowed to lay eggs for two 24 hour broods. Three days after the brood had been set, F_1 animals were collected in M9 buffer (Sulston and Hodgkin 1988) and spotted onto NGM plates containing 1mM levamisole (Lewis et al. 1980). Twenty-four hours later, these plates were inspected. No recombination was expected between *let-56* and *lev-1* due to the presence of *nT1(IV;V),* which suppresses recombination in this interval (Ferguson and Horvitz 1985, Clark et al. 1988; see Figure 1). Therefore, individuals homozygous for *lev-1(x22)* (and thereby resistant to levamisole) would also be homozygous for *let-56(s2321)* and, in the absence of a duplication carrying a wild-type copy of *let-56*, would appear as developmentally retarded larvae at the time of inspection. I screened the levamisole-containing plates for exceptional adult mobile hermaphrodites in a background of developmentally retarded larvae or paralyzed hypercontracted individuals. These exceptional adult mobile individuals would be candidate duplication-bearing animals.

Construction of *let-664(s2374)* homozygous animals bearing *sEx26:* To determine whether the cosmid B0033 was capable of complementing the *let-664(s2374)* mutation, animals that were homozygous for *let-664(s2374)* and that carried the *sEx26* extra-chromosomal array were constructed. The *sEx26* extra-chromosomal array consists of the cosmid B0033, and the plasmid pRf4 (S. Jones, personal communication). pRf4 contains the dominant "roller" mutation *rol-6(su1006)* (Kramer et al. 1990), so that an array containing pRf4 confers a dominant "Roller" phenotype on individuals carrying the array (Mello et al. 1990). This distinctive phenotype consists of the animal rotating about its long axis.

Figure 2:

A generalized flow diagram illustrating the methodology used to recover duplications of *let-56(s2321)*. P_0 hermaphrodites of the indicated genotype were mutagenized, and F_1 animals were subjected to treatment with 1 mM levamisole after the majority of F_1 animals had reached a late larval / early adult stage. In this way I avoided selecting *let-56(s2321) unc-22(s7) lev-1(x22)* arrested homozygotes as candidate duplication bearing animals. After levamisole selection, candidate duplication-bearing animals are recognized as individuals which are motile, and not hypercontracted. Levamisole sensitive individuals are severely paralyzed, and in many cases die following exposure to 1mM levamisole.

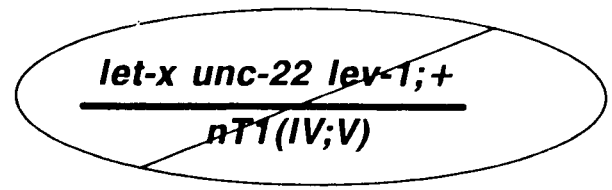
Mutagen Treatment

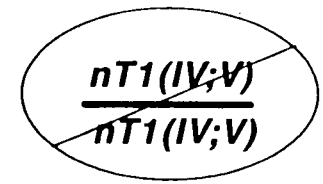
P0 Genotype: $\frac{\textit{let-x unc-22(s7) lev-1(x22)}}{\textit{nT1 (IV)}}$; $\frac{+}{\textit{nT1 (V)}}$ ♂

1. Allow mutagenized animals to produce self progeny.
2. Subject self progeny to 1mM levamisole treatment.
3. Select fertile resistant animal(s).

F1 Genotypes:

$\frac{\textit{let-x unc-22 lev-1}}{\textit{let-x unc-22 lev-1}}$; *Dup*





levamisole resistant:
-do not hypercontract
-motile

levamisole sensitive:
-hypercontracted
-paralyzed

levamisole sensitive:
-hypercontracted
-paralyzed

Because roller males did not mate, it was not possible to transmit extra-chromosomal arrays through the male in crosses. Therefore, construction of a strain carrying *sEx26* in a *let-664* background was performed as follows. Hermaphrodites of the genotype *let-664(s2374) unc-22(s7) lev-1(x22)/nT1(IV) ; +/nT1(V)* were mated with N2 males. Wild-type males that twitched in nicotine [genotype *let-664 unc-22 lev-1/+++*] were selected from among the progeny of this cross. These males were then mated with hermaphrodites carrying the *sEx26* array (*+ ; sEx26*; constructed by S. Jones). Roller progeny from this cross were removed to another plate and screened in nicotine for animals that twitched [genotype *let-664 unc-22 lev-1/+++ ; sEx26*]. When twitching animals were found, they were picked to a separate plate and allowed to produce self progeny. Animals that twitched constitutively and rolled [genotype *let-664 unc-22 lev-1/let-664 unc-22 lev-1 ; sEx26*] were selected from among the self progeny and picked to separate plates for further observation.

Construction of *sDf83* homozygous animals bearing *sEx25:* To determine whether there existed an essential gene or genes in the interval between *let-56* and *unc-22*, I constructed a strain that was homozygous for the deficiency *sDf83* and that carried the extra-chromosomal array *sEx25*. The extra-chromosomal array *sEx25* consists of the plasmid pRf4 (see above) and two plasmids (pCes122 and pCes 891) that contain overlapping DNA subcloned from the cosmid C11F2. These two plasmids represent the right-most 12 kb of the cosmid C11F2, and have been shown to complement the *let-56(s173)* mutation (S. Jones, personal communication). I obtained a strain carrying *sEx25* from S. Jones. The genotype of this strain (BC 4434) was *let-56(s173) unc-22(s7) / let-56(s173) unc-22(s7) ; sEx25*. Animals of this genotype twitch and roll weakly. BC 4434 hermaphrodites were mated with N2 males, and the progeny of this cross were scored for the presence of non-twitching roller animals. The genotype of these

individuals was *let-56 unc-22/++ ; sEx25*. When these animals were found, they were picked to separate plates and allowed to produce self progeny. From among these self progeny, rolling animals that failed to twitch in nicotine were selected. These animals were of the genotype + ; *sEx25*. Animals were again picked to separate plates and allowed to self. The absence of rolling twitchers and rollers that twitched in nicotine among the progeny of these animals served to confirm that the *unc-22* marked chromosome had been removed from the stock. A single roller was selected, and used to establish the strain BC 4537.

I constructed a strain carrying *sEx25* and the deficiency *sDf83* as follows. N2 males were mated to hermaphrodites of the genotype *sDf83/+*. Males that twitched in nicotine (genotype *sDf83/+*) were selected from among the progeny resulting from this cross, and mated with roller hermaphrodites from the strain BC 4537 (genotype + ; *sEx25*; see above). Mated hermaphrodites were set individually and allowed to lay eggs. The progeny of this cross were screened for the presence of rolling animals that twitched in nicotine. Progeny exhibiting this phenotype were of the genotype *sDf83/+*; *sEx25*. These animals were picked individually to separate plates and allowed to lay eggs. Their progeny were screened for the presence of constitutive twitcher animals that rolled. These animals are of the genotype *sDf83/sDf83*; *sEx25*. When animals exhibiting this phenotype were found they were removed singly to individual plates for observation.

2. Molecular Biology.

Isolation of *C. elegans* genomic DNA: Genomic DNA was prepared using the protocol of Emmons et al. (1979), with the modifications of J. Curran and D. L. Baillie (personal communication). Details of the procedure are described in Starr (1989).

Restriction enzyme digestions: Restriction enzymes were obtained from BRL, Pharmacia or New England Biolabs. Restriction enzyme buffers were obtained from BRL. Restriction digests of both genomic DNA and cloned DNA were performed using a minimum of two units of enzyme per μg of DNA for at least two hours at the temperature specified by the enzyme manufacturer.

Agarose gel electrophoresis: DNA was electrophoresed through agarose gels containing 0.5 μg/ml ethidium bromide. Gels were cast in either 1 X TAE or 0.5 X TBE (Sambrook et al. 1989), using 1 X TAE or 0.5 X TBE, respectively, as the electrophoresis buffer. Gels were run at 1 - 5 V/cm, as recommended by Sambrook et al. (1989). Typically, digested genomic DNA (approximately 3 to 5 μg of DNA per lane) was electrophoresed through 0.8 - 1 % (w:v) gels, PCR products were electrophoresed through 0.8 - 1.2 % gels, and digests of cloned DNA were electrophoresed through 0.8 % gels. DNA was made visible by exposure to 300 nm U.V. light. Gels were photographed using either Kodak PLUS-X pan film, or a UVP Image Store 5000 video documentation system.

Isolation of DNA fragments from agarose gels: DNA fragments were isolated from agarose gels for the purpose of probe construction, or in subcloning experiments designed to clone specific restriction fragments. DNA was electrophoresed in 0.7% Seaplaque (FMG) low melting-point agarose gels containing 0.5 μg/ml ethidium bromide. Specific fragments were then excised over a 360nm ultraviolet light transilluminator. Both labeling reactions and ligation reactions were conducted in Seaplaque agarose. Alternatively, DNA was

electrophoresed in a gel cast from Multi Purpose agarose (Boehringer Mannheim) and excised over a 360nm ultraviolet light transilluminator. The excised gel fragment containing the DNA of interest was then placed onto a silanized glass wool plug in a cut-off P-1000 pipette tip, and spun in a micro-centrifuge at 6000 rpm for 10 - 30 minutes as described (Heery, Gannon and Powell 1990). DNA in electrophoresis buffer was deposited into an Eppendorf tube as a result of the centrifugation step, but residual agarose was retained by the glass wool plug. DNA isolated in this fashion could be used without further purification in both labeling and ligation reactions. In hybridization experiments involving the ECL (Enhanced Chemi-Luminescent) labeling and detection system (Amersham), double gel isolation of probe DNA was required. Double gel isolation was achieved by inserting an excised gel fragment containing the desired DNA into a well of a freshly cast gel and then repeating the electrophoresis.

Labeling of DNA probes: Approximately 50 ng of gel isolated DNA were labeled with ^{32}P-dATP using the method of Feinberg and Vogelstein (1983). Random hexamer oligonucleotide primers were purchased from Pharmacia. Labeling was carried out for two to three hours at room temperature, or overnight at 15 ^{o}C. Reactions were terminated by the addition of "stop solution" (10 X stop solution was 200 mM EDTA, 5% SDS) and then centrifuged through a G-25 Sephadex (Pharmacia) spin column. Probes with a specific activity of 1.0 x 10^8 to 2.0 x 10^9 cpm/µg were synthesized by this procedure. Prior to denaturation, probes were added to hybridization solution (see below) and filtered through 0.45 µm nitrocellulose filters (Schleicher and Schuell). Alternatively, probes were constructed using the ECL random prime kit, using protocols and reagents supplied by the manufacturer (Amersham). For this procedure, a minimum of 50 ng of DNA were required, and labeling was with fluorescein-11-dUTP using the random prime method of Feinberg and Vogelstein (1983). Probe synthesis was

complete in two to three hours at room temperature. Probes were denatured by immersion in a boiling water bath for five to ten minutes and snap cooled on ice prior to addition to the hybridization solution.

Southern blotting: Southern blotting of agarose gels containing restriction enzyme digested genomic DNA or restriction enzyme digested cloned DNA was performed using the following protocol. Gels were soaked in 0.25 M HCl for 10 minutes, thoroughly rinsed two times in distilled water, and then soaked for 30 minutes in a solution that was 0.5 M NaOH and 1.5 M NaCl. In order to prevent the gel from floating in this solution, it was necessary to weight the gel by placing Pasteur pipettes onto its surface. The gel was then rinsed twice in distilled water, and soaked for one hour in "transfer buffer" (1 M NH_4CH_3COO, 0.02 M NaOH). The DNA was then transferred to Genebind (Pharmacia) nylon membrane that had been soaked in transfer buffer for 30 minutes. The preferred method of DNA transfer for blots of gels containing genomic DNA was the unidirectional capillary reservoir technique of Southern (1975) as described in Sambrook et al. (1989). For blots of cloned DNA, the bidirectional transfer method of Smith and Summers (1980) as described in Sambrook et al. (1989) was occasionally used. Transfers were normally allowed to proceed for 10 - 12 hours. After the transfer, the membrane was carefully rinsed in 5 X SSC (Sambrook et al. 1989) to remove residual agarose and other undesirable adhering material, air dried on Whatman paper, and then baked at 80°C for at least two hours to fix the DNA to the membrane.

Subcloning: The "shotgun" subcloning method (Sambrook et al., 1989) was used to clone *Bam*HI and *Hin*dIII restriction fragments of C11F2. In this method, restriction enzyme digested cosmid DNA was mixed with plasmid vector DNA digested with the same enzyme. After heat inactivation of the restriction enzyme, the mixture was ligated as described by Snutch (1984). Alternatively, the

desired DNA fragments were isolated from agarose gels as previously described, and ligated to plasmid vector DNA digested with the appropriate enzyme. The plasmid vectors used in this study were pUC19 (Messing 1983), pBluescript SK+ and SK- (Stratagene), or pVZ1 (Henikoff and Eghtedarzadeh 1987). Ligations were conducted overnight at 15°C. Completed ligations were transformed into *E. coli* DH5α subcloning efficiency competent cells obtained from BRL. Colonies of cells bearing recombinant plasmids were identified as white colonies on agar plates containing ampicillin, X-gal and IPTG (Sambrook et al. 1989). Mini-preps of plasmid DNA were performed using the Pharmacia Mini-Prep Kit Plus.

Hybridization of probes to DNA filters: For hybridization experiments involving ^{32}P-labeled DNA probes, membranes carrying DNA were prehybridized in "hybridization buffer", which was 5 X SSPE (Sambrook et al. 1989), 0.2% SDS and 5 X Denhardt's (Sambrook et al. 1989), at 68°C for at least three hours. Hybridization was carried out in a fresh aliquot of hybridization buffer containing heat denatured probe at 68°C overnight. Filters were washed with constant agitation at the hybridization temperature for at least one hour in 0.1 X SSPE, 0.2% SDS with at least three changes of wash solution. Autoradiography was accomplished by exposing filters to preflashed Kodak X-OMAT K film at -70°C for the required duration (4 hours to 3 days) using a Dupont Lightning Plus intensifying screen.

For hybridization experiments involving DNA probes labeled using the ECL random prime labeling kit (Amersham), membranes carrying DNA were prehybridized in 5 X SSC (Sambrook et al. 1989), 0.3% SDS and 5 X Denhardt's (Sambrook et al. 1989) for at least three hours at 60°C. An aliquot of denatured probe was added directly to the hybridization bag, and hybridization was allowed to proceed overnight at 60°C. Washes were conducted as recommended (ECL random prime labeling and detection protocol book), except that the filter was

washed for one hour in 0.1% Tween 20 in antibody wash buffer with four changes of solution instead of two washes of ten minutes duration each, followed by two washes of five minutes duration each. In general, it was found that increasing the number and the duration of washes in the 0.1% Tween 20 solution resulted in decreased background. Adequate signal intensity for membranes carrying genomic DNA was usually achieved after exposure of membranes to Kodak X-OMAT K film for 20 to 30 minutes, while membranes carrying cloned DNA produced adequate signal in three seconds to two minutes, depending on the experiment.

For low stringency hybridization conditions, hybridization was carried out overnight at 57 C in hybridization buffer containing 5 X SSC. Two twenty-minute stringency washes were carried out in 2 X SSC at 45 C prior to detection and chemi-luminography.

Screening a *C. elegans* cDNA library: A *C. elegans* cDNA library was obtained from B. Barstead (Barstead and Waterston, 1989). This library had been constructed in λZAP (Stratagene) from poly-A^+ RNA isolated from a population of worms that contained animals at all stages of the worm's life-cycle. The titer of the library was determined, and found to be approximately $2x10^{10}$ pfu/ml. Manipulation and screening of the library were performed according to the manufacturer's instructions (Stratagene), except that the library was propagated in BB4 host cells. Phage were adhered to pre-cut 82 mm nitrocellulose discs (Schleicher and Schuell) or 150 mm Genebind nylon discs (Pharmacia). The subsequent denaturation of phage particles and fixing of phage DNA to the membrane was accomplished using either a method modified from that of Benton and Davis (1977), as described in Sambrook et al. (1989), or according to the ECL random prime labeling and detection systems protocol book (Amersham). Preparation of probe DNA and hybridization of labeled probe to membranes was

performed as previously described. Plaques were picked using a cut-off P-1000 pipette tip, and deposited into 500 µl of SM buffer (Sambrook et al. 1989) that contained a drop of chloroform. Picked plaques were found to retain their titer for at least 1 month at 4°C. After three rounds of screening, cDNAs were recovered in the *Eco*RI site of the pBluescript SK- plasmid following the excision protocols obtained from Stratagene.

Oligonucleotide primers designed and/or used in this study: A complete list of all "custom" oligonucleotide primers designed and/or used in this thesis and the appropriate references has been included in Table 2. The program OLIGO (Rychlik and Rhoads 1989) was used to select and evaluate all primers used in either DNA sequence determination or in PCR experiments.

PCR techniques: Reagents for some PCR experiments were obtained from Perkin-Elmer Cetus (Gene-Amp PCR Reagent Kit). PCR reaction conditions were as described in the protocol book supplied with the Gene-Amp Kit. Alternatively, *Taq* polymerase and *Taq* buffers were obtained from Promega. In some experiments *Pfu* polymerase (Stratagene) was used. The deoxynucleoside triphosphates used in some experiments were obtained from Pharmacia.

(A) Worm and egg PCR: PCR amplifications were conducted in 25 µl volumes using the procedure of Barstead, Kleiman and Waterston (1991), with the modifications of Williams et al. (1992). Temperature cycling was performed using standard 0.5 µl microfuge tubes in the Ericomp Twin Block System, or in glass capillaries in an Idaho Technology model 1605 Air Thermo-Cycler. N2 worms or arrested twitching larvae homozygous for one of the deficiencies indicated by a "*l*" in Table 1 were placed into 2.5 µl of lysis buffer (Williams et al. 1992). N2 embryos or arrested embryos homozygous for one of the deficiencies indicated by an "*e*" in Table 1 were exposed to chitinase solution (Williams et al.

TABLE 2.

Oligonucleotide primers used in this study.

PRIMER	SEQUENCE (5' - 3')	APPLICATION	REFERENCE [a]
5P2	GCCTAAAACCAAATACAGC	SEQUENCE - *nhe-1*	2
C6-1	GCCAGCAACAAACCGAAG	PCR - *adl-1*	1
C6-2	AAGAACAGCAGACAACGC	PCR - *adl-1*	1
DB-19	GCTATGATGGAGGAAAGTG	SEQUENCE - *nhe-1*	2
DB-20	CTCATTATTGTCGGTTTGGC	SEQUENCE - *nhe-1*	2
DB-21	TCGGAGAGCACTTTTCAATG	SEQUENCE - *nhe-1*	2
DUP-1	ATTCACAAAACACATCCC	PCR - dup	1
DUP-2	GTTATGTGACCGATGAGC	PCR - dup	1
DUP-3	GCTCATCGGTCACATAAC	PCR - dup	1
DUP-4	TAATCTTTTCACTGCTCC	PCR - dup	1
K08-3	TCAAAACGCATCTAAACTGG	PCR - (LGV)	3
K08-7	GCCACACTAAAACAGGATTC	PCR - (LGV)	3
KT01	CTCCAAGTAACACTACGTCGTGG	PCR - *spe-17*	1
KTO3	ATGCACTATATTTTCCAATATTTTTCAT	PCR - *spe-17*	1
LUNC-1	CCGTTTGAGAGTTGAGCG	PCR - *gtl-1*	1
LUNC-2	CGACGAGACAGCCACAAC	PCR - *gtl-1*	1
MAM-1	CATCCAAGTGCCAAACCG	SEQUENCE - *nhe-1*	2
MAM-2	AGGTTTCTCAGTTTTTGG	SEQUENCE - *nhe-1*	2
MAM-3	TGTCACTGTTGTTCTGTATC	SEQUENCE - *nhe-1*	2
MYS-1	AAGTTCGTAAAGATGCCC	PCR - *ptl-1*	1
MYS-2	AGATTTGGGCATTTTTCG	PCR - *ptl-1*	1
MYS-3	AACAACTCTGCCCATTCG	PCR - *ptl-1*	1
179	TGATACTTCCCTTTTTCG	PCR - (LGV)	4
180	CATTACACGGAGAAGACG	PCR - (LGV)	4
QFOR1	CTCAACGTCATCAGCAGATC	PCR - (frag. Q)	2
QREV1	TTTATCCCATTTTTCGCTAC	PCR - (frag. Q)	2
R1	CGGCGGAGAAGAACAAGC	SEQUENCE - *hdl-1*	2
R2D	GCTACTTGAGGCGATTTG	SEQUENCE - *hdl-1*	2
R3	CCAGGACTTTCTCACTCG	SEQUENCE - *hdl-1*	2
R4	CGAGTGGATTTTTGCTGG	SEQUENCE - *hdl-1*	2
R5	TGATGATGGATTGGCTCGGC	SEQUENCE - *hdl-1*	2

TABLE 2. (continued)

Oligonucleotide primers used in this study.

PRIMER	SEQUENCE (5' - 3')	APPLICATION	REFERENCE [a]
R6	GTTTTGTAAGACGCACGATG	SEQUENCE - *hdl-1*	2
R7	GTTGCGATGGTAAAGATGCG	SEQUENCE - *hdl-1*	2
R8	GCACGGAAGCAAATCAACCC	SEQUENCE - *hdl-1*	2
R9	ATAAGAGGGCTGATGAGAGG	SEQUENCE - *hdl-1*	2
R10	CGAACCCAGAGACAACAAAC	SEQUENCE - *hdl-1*	2
R11	TACTCGCATCTTTACCATCG	SEQUENCE - *hdl-1*	2
RF1	AATCAGAGGAGACCGAAG	PCR - (frag. R)	2
RR1	TGACTTTTTGATGCCTGC	PCR - (frag. R)	2
265-F	CAAAATGGATGATGAGAAGTGC	PCR - (LGV)	3
265-R	CAACTGCCCTGACAACTCCGAC	PCR - (LGV)	3

[a] Reference 1 is Schein *et al.* 1993. Reference 2 is Marra *et al.* 1993. Reference 3 is K. S. McKim, personal communication. Reference 4 is W. B. Barbazuk, personal communication.

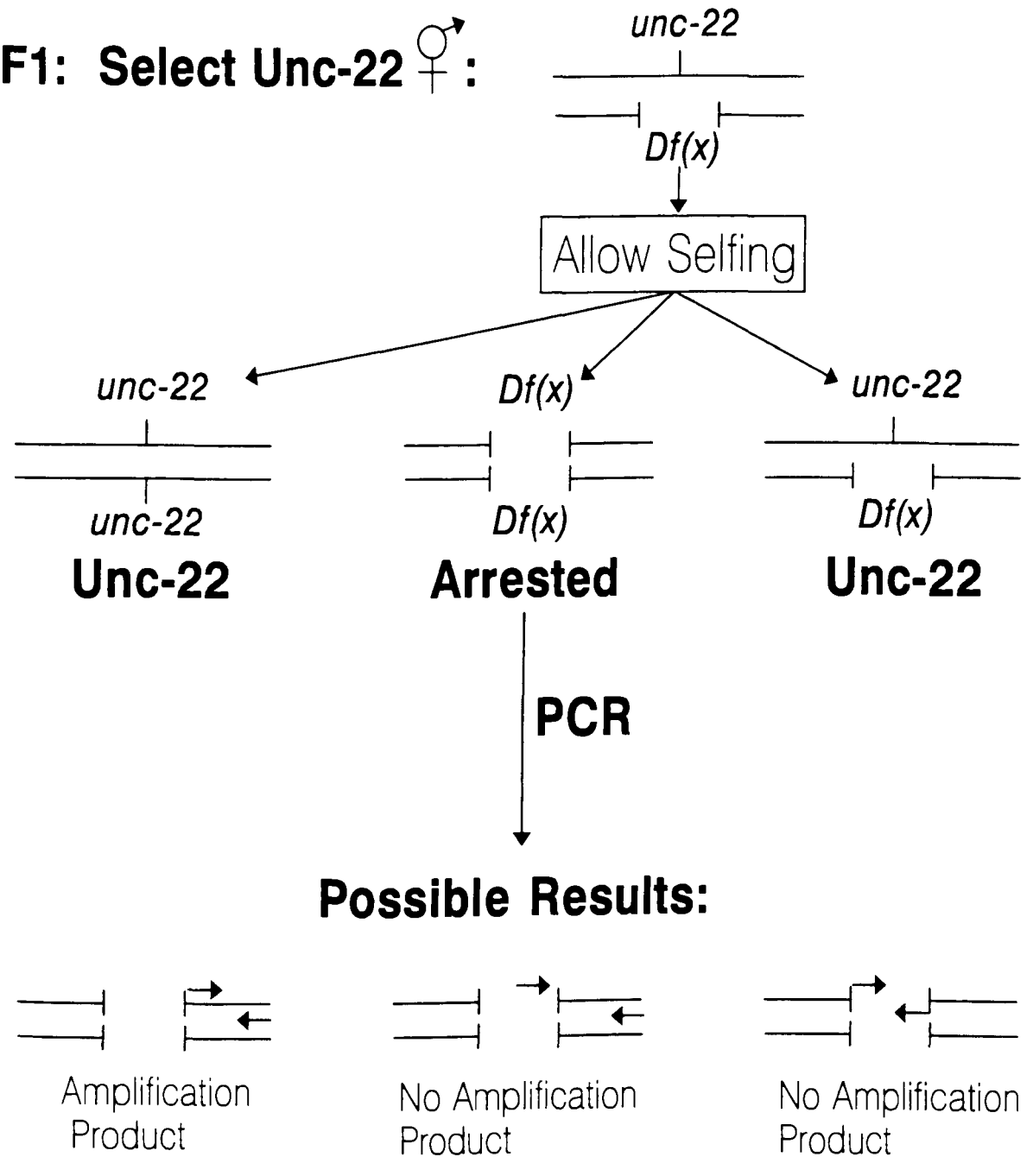
1992) prior to being placed in lysis buffer. In the cases of deficiencies balanced over *nT1(IV;V)*, it was necessary to remove the balancer prior to selecting arrested embryos or larvae for PCR to avoid selection of arrested *nT1* aneuploid individuals. Removal of the *nT1(IV;V)* background was accomplished by crossing *unc-22(s7)/++* males with *Dfx/nTI(IV) ; +/nTI(V)* hermaphrodites. Unc-22 progeny resulting from this cross are *unc-22/Dfx,* and therefore do not carry the *nT1* rearrangement (Figure 3). These individuals were set on plates and allowed to lay eggs for 12-24 hours, and were then removed. Two to three days later arrested progeny were collected for PCR. PCR amplifications were repeated at least in duplicate, and control primers annealing to sequences on chromosome *V* were incorporated into each reaction. The presence of a control amplification product allowed me to distinguish between failed PCR reactions and cases in which the deficiency under examination deleted at least one of the primer annealing sites (Figure 3). Typical thermal profiles for the Idaho Technologies Air Thermo-Cycler were 1 minute @ 94°C followed by 35 cycles of 10 seconds @ 94°C, 25 seconds @ 53°C, 40 seconds @ 72°C, followed by 2 minutes at 72°C. Typical thermal profiles for reactions cycled in the Ericomp Twin Block System were: 5 minutes @ 94°C, 1 minute @ 58°C, 3 minutes @ 72°C for one cycle, followed by 25-30 cycles of 1 minute @ 94°C, 1 minute @ 58°C, 3 minutes @ 72°C, and one cycle of 1 minute @ 94°C, 1 minute @ 58°C and 10 minutes @ 72°C.

(B) PCR amplification of genomic and cloned DNA: The thermal profiles listed above were found to be satisfactory during PCR amplification of both genomic and cloned DNAs, except that adequate amplification was usually achieved after 25 cycles on both the Ericomp and Idaho cyclers. For amplification of cloned DNA, approximately 1 ng of template DNA was used. For amplification of genomic DNA, approximately 50 ng of template DNA were used.

Figure 3:

A flow diagram illustrating the PCR using individuals homozygous for an *unc-22* deficiency as amplification "templates". *unc-22(s7) / +* males are mated with hermaphrodites bearing a deficiency of *unc-22*. Unc-22 hermaphrodite progeny resulting from the cross do not carry the *nT1(IV;V)* rearrangement, and therefore will not produce *nT1* aneuploid progeny, which, when arrested, could be confused with individuals homozygous for the deficiency. Unc-22 hermaphrodites are picked to separate plates and allowed to produce self progeny. Arrested self progeny are removed to separate plates and observed after 24 hours to ensure they are homozygous for the deficiency. Amplification products are seen when PCR is performed using individuals homozygous for deficiencies in which both primer annealing sites are present.

P0: *unc-22/+* ♂ X *Df(x)/nT1(IV) ; +/nT1(V)*♀



(C) PCR amplification of recombinant λ ZAP phage particles: DNA was isolated from 20 ml of the previously described λ ZAP cDNA library following the polyethylene glycol (PEG) extraction protocol described in Sambrook et al. (1989). The resulting phage DNA solution was diluted 10-fold and 1 μl was used as the DNA template in subsequent PCR reactions.

DNA sequence determination: DNA sequence was determined using the dideoxy chain terminator method of Sanger et al.(1977), with either custom designed or commercially available oligonucleotide primers. Double-stranded DNA was prepared using a modified alkaline lysis procedure as described in the Pharmacia Mini-prep Kit Plus Protocol booklet. Further purification of DNA templates by passage of the crude Mini-prep through Pharmacia Mini-prep spun columns was required for DNA to be manually sequenced, but not for DNA sequenced using an ABI model 373A automated sequencer. Double-stranded DNA templates were sequenced "manually" following a protocol obtained from T. Snutch using α^{35}S-dATP, and reagents supplied in the Sequenase Version 2.0 Sequencing Kit (USB). "Manual" sequencing reactions were electrophoresed through 8% polyacrylamide gels containing urea. Autoradiographs of sequencing gels were obtained by exposure of the dried gel to Kodak X-OMAT K film for 24 to 48 hours. Automated DNA sequence was collected on an ABI model 373A sequencer using reagents and the *Taq* cycle sequencing protocol supplied by ABI. The only modification to the ABI protocol was to roll the completed reactions down a strip of parafilm to remove residual traces of mineral oil (K. Peters, personal communication) prior to ethanol precipitation. Cycle sequencing reactions were performed using dye-labeled -21M13 or dye-labeled M13 reverse primers in an Ericomp Twin Block System thermal cycler. Exonuclease III generated unidirectional deletions (Henikoff 1987) of cDNAs or genomic clones were prepared using the Erase-a-base system (Promega).

DNA sequence analysis: Nucleotide sequences were aligned, formatted, and translated using the Eyeball Sequence Editor (ESEE) program (Cabot and Beckenbach 1989). Amino acid sequences predicted using ESEE were used to search the latest available releases of the SWISS and translated EMBL databanks. Searches were done with the FASTA and TFASTA programs of Pearson and Lipman (1988). Alternatively, amino acid sequences were searched against entries in the SwissProt, PIR, GenPept, and GPUpdate peptide sequence databases (collectively known as the "non-redundant" database), using the BLAST algorithm (Altschul et al. 1990). Some searches were performed against **dbest** (**d**ata**b**ase of **e**xpressed **s**equence **t**ags) which contains the cDNA sequences determined by the *C. elegans* genome sequencing consortium. These computations were performed at the National Center for Biotechnology Information (NCBI) using the BLAST network service. Amino acid sequences were also analyzed using the BLOCKS utility server (Henikoff and Henikoff, 1991), and software available in the PCGENE (Intelligenetics; release 6.8) suite of programs.

SECTION I.

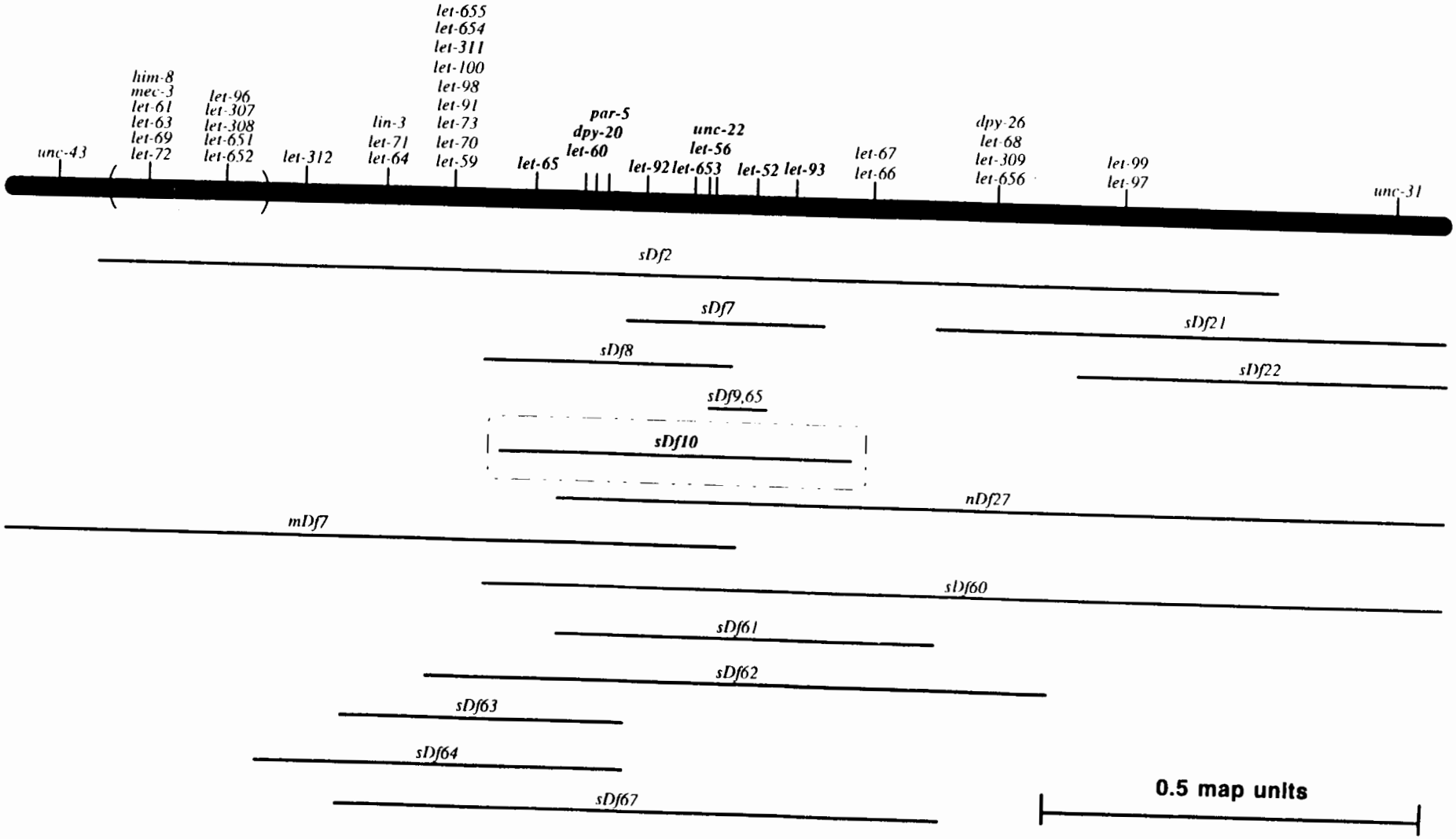
Recovery and analysis of lethal mutations in the *let-65* - *let-93* region.

INTRODUCTION

Review of the *unc-22* region: The *let-65 - let-93* interval is contained entirely within the region of the genome known as the *unc-22* region. The *unc-22* region is near the right end of chromosome IV in the nematode *Caenorhabditis elegans* and has previously been defined as the interval spanned by the deficiency *sDf2* (Figure 4), a distance of approximately two map units (Clark et al., 1988). The *unc-22* region has been investigated in our laboratory at both the genetic and molecular levels to determine if constituent genes exhibit a major organizational theme. Investigations at the genetic level have involved EMS mutagenesis screens designed to recover recessive lethal mutations either tightly linked to *unc-22*, or within the *nT1* balanced region of chromosome IV (Figure 1; Rogalski, Moerman and Baillie 1982; Rogalski and Baillie 1985; Clark et al. 1988; Rogalski and Riddle 1988; Clark and Baillie 1992). These studies have resulted in the identification of 84 lethal or sterile mutations that define 36 complementation groups mapping within the breakpoints of *sDf2* (Clark and Baillie, 1992). Based on these data, Clark and Baillie (1992) predicted the *unc-22* region was at most 73% saturated for mutations in essential genes. Therefore, these authors predicted a minimum of 49 essential genes in the *sDf2* region. To facilitate the mapping and characterization of these mutations, a number of deficiencies had been recovered (Moerman and Baillie 1981; Clark and Baillie 1992; Schein et al. 1993; J. E. Schein, personal communication). The result of these studies has been the construction of a genetic map populated primarily with essential gene markers, some of which are separated from neighboring genes by the breakpoints of deficiencies (Edgley and Riddle, 1990; Figure 4).

Figure 4:

A genetic map of the *sDf2* region of chromosome IV prior to this thesis (adapted from Clark, 1990). The heavy line represents the genetic map. The parentheses on the heavy line indicate that these loci had not been positioned with respect to one another. The deficiencies with breakpoints in the region are indicated below the genetic map line. The deficiency *sDf19*, not shown on this map, disrupts the *unc-22* locus and extends left, separating the *let-56* complementation group from the *let-653* complementation group. Each of the thin vertical lines intersecting the genetic map line represents a "zone", or region of the genetic map between two adjacent deficiency breakpoints. Exceptions are the thin vertical lines under *let-60*, which was initially positioned to the left of *dpy-20* by Min Han, and *par-5*, which was positioned to the right of *dpy-20* by Diane Morton (Edgley and Riddle, 1987). *dpy-26* was mapped by P. Meneely (personal communication). *let-654* was renamed *sem-3*, as some alleles affect sex muscle development in the hermaphrodite (M. Stern, personal communication). The work in this thesis has focused on the region defined by the deficiency *sDf10*, which is boxed and indicated in bold type. The genes that were within this deficiency are also indicated in bold type above the genetic map line.


unc-43
him-8
mec-3
let-61
let-63
let-69
let-72
let-96
let-307
let-308
let-651
let-652
let-312
lin-3
let-71
let-64
let-655
let-654
let-311
let-100
let-98
let-91
let-73
let-70
let-59
let-65
par-5
dpy-20
let-60
let-92
unc-22
let-56
let-653
let-52
let-93
let-67
let-66
dpy-26
let-68
let-309
let-656
let-99
let-97
unc-31
sDf2
sDf7
sDf21
sDf8
sDf22
sDf9,65
sDf10
nDf27
mDf7
sDf60
sDf61
sDf62
sDf63
sDf64
sDf67
0.5 map units

40 (b)

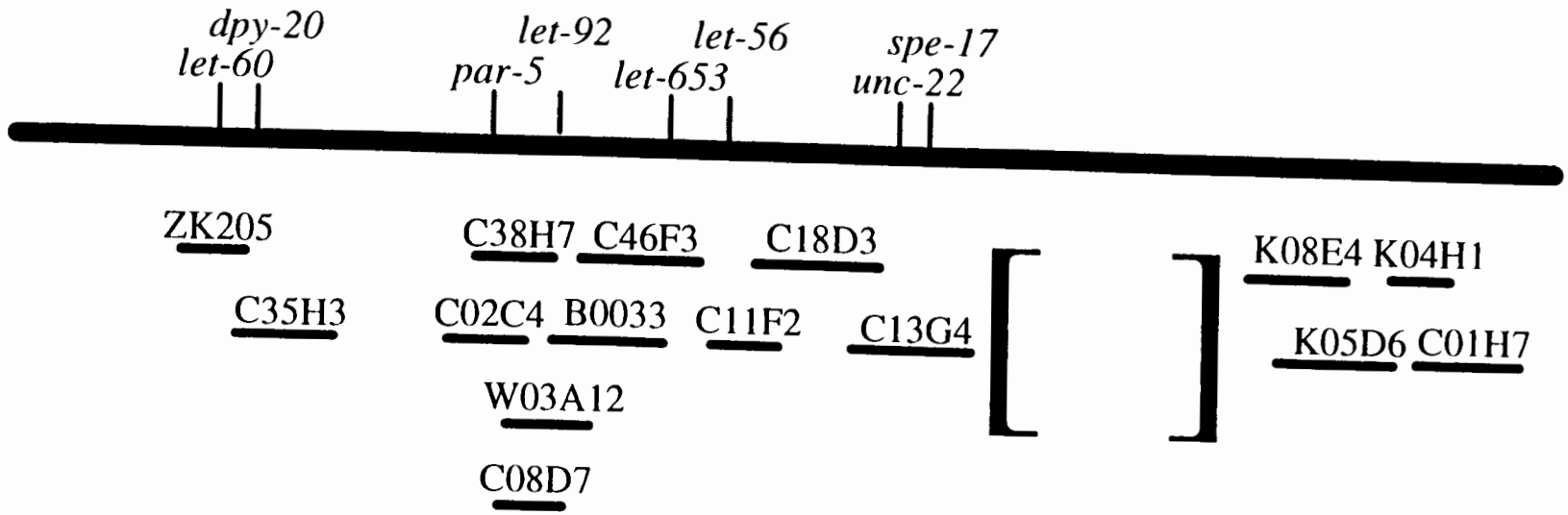
Molecular characterization of the *let-65 - let-93* interval had revealed genes not identified by mutation: The intervals immediately to the right and left of *unc-22* had been examined at the molecular level. Immediately to the right of *unc-22*, the spermatogenesis-defective *spe-17* gene had been examined in detail (L'Hernault, Benian and Emmons, 1993). Examination of the physical map (Coulson et al., 1986; Coulson et al., 1988) of the region to the right of *spe-17* (cosmid C13G4; L'Hernault, Benian And Emmons, 1993) revealed an island of cosmids, separated from the *unc-22* contig by a cosmid "gap" (Figure 5). This gap was bridged by overlapping yeast artificial chromosomes (Coulson et al., 1988).

To the left of *unc-22*, the physical map between *let-60* and *unc-22* was composed of contiguous cosmid clones (Coulson et al., 1986) and all markers on the genetic map had been localized on the physical map (Figure 5). These included *let-60,* which has been found to encode a Ras-like protein, (Han and Sternberg 1990), *dpy-20* (Clark 1990; Suleman, Clark, Beckenbach and Baillie, submitted), *par-5* (tentatively assigned to cosmid C38H7; D. Shakes, personal communication), *let-92* (S. Jones, 1994), *let-653* (Clark and Baillie 1992; S. Jones, 1994), *let-56* (Clark and Baillie, 1992) and *unc-22,* which had been shown to encode a large protein named "twitchin" that has been implicated in the regulation of myosin activity (Moerman et al., 1986; Benian et al., 1989).

Other genes, not defined by mutation, were demonstrated to reside in this interval. Prasad and Baillie (1989) hybridized restriction fragments of cloned *C. elegans* genomic DNA to genomic DNA obtained from *Caenorhabditis briggsae* to identify and map the location of sequences conserved between these closely related species. These authors found nine regions of conservation localized to four cosmids (C02C4, C08D7, C11F2 and C18D3) that contained approximately 150 kb of DNA, or one half of the DNA separating *dpy-20* from *unc-22* (see

Figure 5:

A portion of the physical map in the *let-60* - *unc-22* interval, adapted from **ACeDB**. Only selected cosmid clones are shown. The heavy horizontal line represents the correlated physical and genetic maps. The following genes were positioned on cosmids by germline transformation rescue. *let-60* was rescued by ZK205 (Han and Sternberg, 1990), *dpy-20* was rescued by C35H3 (D. V. Clark, 1990), *let-92* was rescued by B0033 (S. J. M. Jones 1994), *let-653* was rescued by C46F3 (S. J. M. Jones, 1994), *let-56* was rescued by C11F2 (D. V. Clark, 1992) and both *unc-22* and *spe-17* sequences are on C13G4 (Benian et al., 1989). Diane Shakes (personal communication) has tentatively assigned *par-5* to C38H7. The cosmid "gap" to the right of C13G4 is indicated by the brackets. The cosmid clones studied by Prasad (1988) are included on the map. They are C08D7, C02C4, C11F2 and C18D3.


dpy-20
let-60
let-92
par-5
let-56
let-653
spe-17
unc-22
ZK205
C38H7
C46F3
C18D3
K08E4
K04H1
C35H3
C02C4
B0033
C11F2
C13G4
K05D6
C01H7
W03A12
C08D7

42 (b)

Figure 5). These regions of conservation represent gene sequences, as non-coding regions of *C. elegans* DNA do not hybridize to *C. briggsae* genomic DNA even at low stringency (Snutch, 1984). Marra, Prasad and Baillie (1993) demonstrated that one of these conserved regions encoded a *C. elegans* Na^+/H^+ antiporter-like protein, and another encoded a *C. elegans* amino acid decarboxylase-like protein (see Section III). An L-proline transporter-like gene (Liu et al. 1992; Schein et al. 1993) and a glucose transporter-like gene (Schein et al. 1993) had been identified immediately to the left of *unc-22* by analyzing DNA sequence obtained by Benian *et al.* (1989) during the sequencing of *unc-22* (see Section III).

No mutations were known to correspond to the loci described in the preceeding paragraph. As a basis for the continued genetic characterization of these genes and other functionally important sequences residing in the *let-65 - let-93* interval, I undertook a screen for mutations that slowed or arrested development. Such a screen would identify loci required for normal growth and reproduction and would, in addition, identify mutations with more subtle effects. A screen of sufficient generality was required because the phenotypes of mutations in the aforementioned genes, or other genes in the interval, could not be predicted beforehand. In spite of the extensive genetic analysis that had previously been conducted in the *sDf2* region, I suspected that there were additional loci in the interval that could be defined by mutation for the reasons given below.

The observation that nine regions of conservation were localized to one half of the DNA separating *dpy-20* from *unc-22* (Prasad and Baillie, 1989) was compatible with the notion that there were additional genes in the unexamined DNA separating *dpy-20* from *unc-22* (Figure 5). Genetic analysis of the entire *dpy-20 - unc-22* interval had only identified six genes with mutant phenotypes, including *dpy-20* and *unc-22*. An explanation for the apparent paucity of mutationally defined loci was that there were other genes residing in the interval

which would yield phenotypes when mutant, and mutations in these genes could be recovered given the application of a suitably designed mutagenesis screen of sufficient size. I felt this explanation to be likely for the following reasons. First, *let-653* was represented by a single allele. If the region had been saturated for *let* mutations, multiple alleles of *let-653* would be expected. Second, Clark and Baillie (1992) had estimated the *sDf2* region, which encompasses the *let-65* - *let-93* interval, to be at most 73% saturated for mutations in essential genes. Third, *let-52* and *let-93,* to the right of *unc-22*, were each represented by single alleles. These lines of evidence supported the prediction that mutations in as yet undiscovered genes in the *let-65* - *let-93* interval could be recovered.

Identification of the already discovered *let-* genes in the *let-65* - *let-93* interval had been the result of a number of previous lethal screens in the Baillie laboratory. The combined data from the most recent screens (Clark et al., 1988; Clark and Baillie 1992) shows that from 9889 chromosomes treated with 0.012 M of the alkylating agent EMS, seven mutations were recovered in the interval, only one of which identified a gene that had not been discovered previously. To ensure the recovery of a larger size sample of mutations, over 11,000 chromosomes treated with a higher concentration of EMS (0.018 M) were screened. In this section of the thesis, I describe the isolation, mapping, and phenotypic characterization of 17 new recessive lethal mutations, tightly linked to *unc-22,* that resulted from this screen. Four of these lethal mutations each define a new complementation group. Another lethal mutation was found to be a deficiency. My analysis of two others has revealed that the *let-653* locus exhibits a complex complementation pattern. My results suggest that the *let-65* - *let-93* region contains additional essential genes not yet identified by mutation. In conjunction with previous studies, a density of approximately one mutation per cosmid length of DNA has been achieved over the majority of the *let-65* - *let-93* interval.

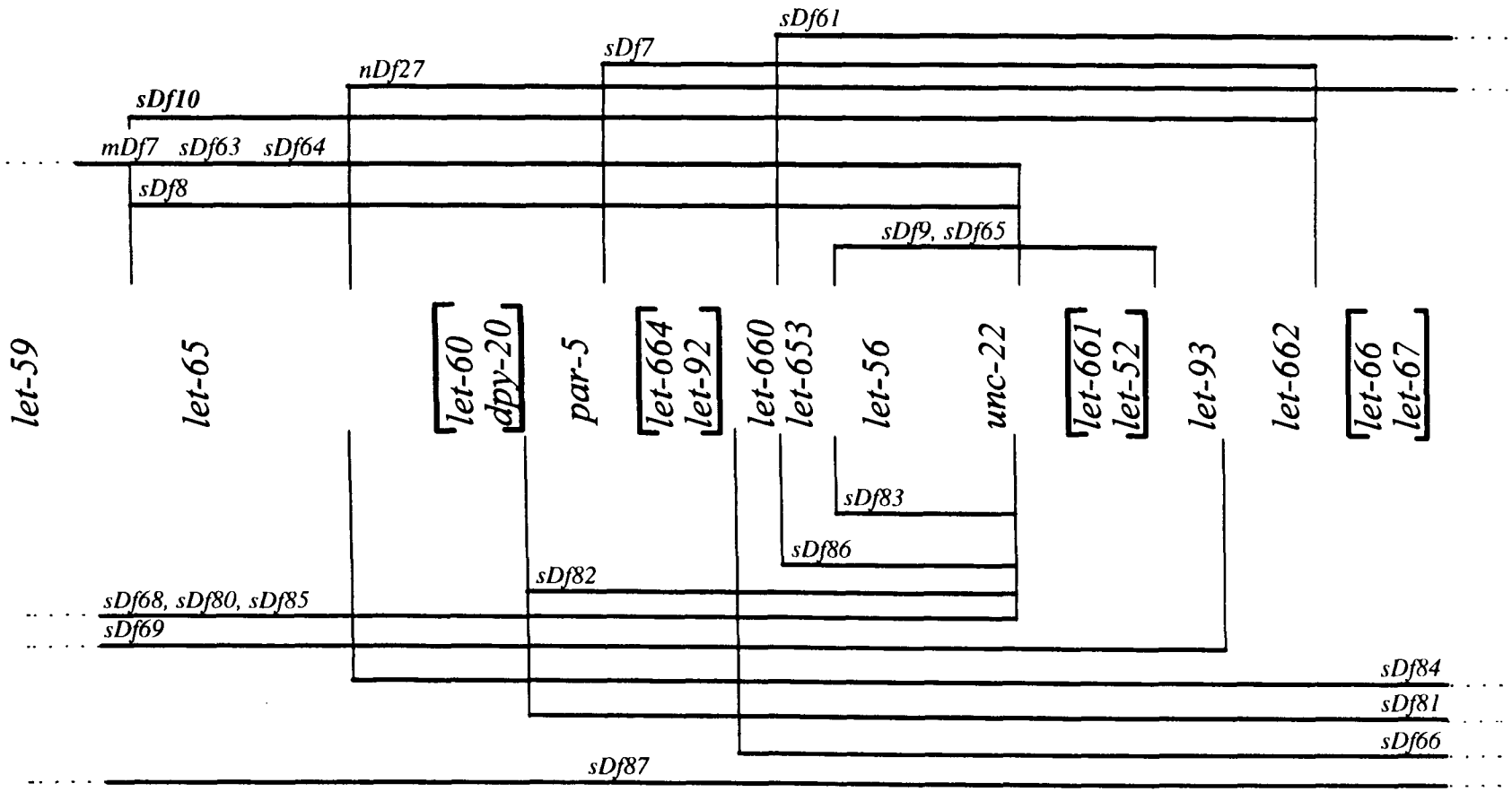
RESULTS

Identification of recessive lethal mutations in the *let-65* - *let-93* interval: The MEG1 and MAM1 0.018 M EMS screens were designed to recover mutations, mapping to within approximately 3 map units of *unc-22,* that arrested or substantially slowed development. I have chosen to call these mutations "lethals" (see Materials and Methods for the definition of a "lethal"). The progeny of 8918 *unc-22(s7) lev-1(x22) (?let)/++* individuals were examined in the MAM1 screen as described in Materials and Methods. This resulted in the recovery of 175 lethal mutations. Therefore, 1.96% of the treated chromosomes were found to carry a lethal. The progeny of 2158 *unc-22(s7)unc-31(e169) (?let)/++* individuals were examined in the MEG1 screen, resulting in the recovery of 41 lethals (M. E. Green and D. V. Clark, personal communication). In this screen, 1.90% of the treated chromosomes were found to carry a lethal mutation. To identify lethals that fell within the *let-65* - *let-93* interval, I tested all of the lethal mutations recovered from both mutagenesis screens for complementation with the deficiencies *sDf7* and *sDf8*. These deficiencies define the region pertinent to this study (Figure 6). Of the 175 lethals recovered in the MAM1 screen, 15 were found to fail to complement *sDf7* and/or *sDf8*. Of the 41 lethals recovered in the MEG1 screen, I found that two failed to complement both *sDf7* and *sDf8*.

Complementation tests: In order to further refine the map positions of each of the lethals that were identified in the region of interest, I tested the 17 lethals for complementation with a set of deficiencies that had breakpoints in the *sDf10* region. The breakpoints of certain of these deficiencies define what has been termed a "zone" (see, for example, Clark and Baillie, 1992). A zone is a region of the genetic map that is flanked on each side by at least one deficiency breakpoint. The strategy that was used to select the appropriate deficiencies for complementation testing was as follows. Lethals that failed to complement *sDf7*

Figure 6:

A genetic map of the *sDf10* interval. The deficiencies above the gene names are those that were available at the start of this thesis. *sDf10* is indicated in bold faced type. *sDf10* and *sDf8* share the same left breakpoint on the genetic map. Deficiencies below the genes were kindly provided by J. Schein. The deficiency breakpoints are indicated by the vertical lines extending toward the gene names. Genes that are bracketed are not separated from each other by breakpoints. The genes identified in this Section of the thesis are *let-660*, *let-661*, *let-662*, and *let-664*. The deficiency *sDf69* was recovered in the MAM1 mutagenesis screen. It is presumed to be of spontaneous origin. The right breakpoint of *sDf69* serves to order *let-93* and *let-662* with respect to each other.


sDf61
sDf7
nDf27
sDf10
mDf7 sDf63 sDf64
sDf8
sDf9, sDf65
let-59
let-65
let-60
dpy-20
par-5
let-664
let-92
let-660
let-653
let-56
unc-22
let-661
let-52
let-93
let-662
let-66
let-67
sDf83
sDf86
sDf82
sDf68, sDf80, sDf85
sDf69
sDf84
sDf81
sDf66
sDf87

0.1 map unit

but complemented *sDf8* were tested for complementation with *mDf7, sDf9, sDf10, sDf63, sDf64* and *sDf65*. Lethals that failed to complement both *sDf7* and *sDf8* were tested for complementation with *mDf7, sDf9, sDf61, sDf63, sDf64* and *sDf65*. Lethals that failed to complement *sDf8* but complemented *sDf7* were tested for complementation with *nDf27*. In addition to these deficiencies, a number of formaldehyde induced deficiencies constructed by J. Schein were made available to me. These deficiencies *(sDf66, sDf68, sDf80, sDf81, sDf82, sDf83, sDf84, sDf85, sDf86 and sDf87)* are indicated in Table 1, and are included on the genetic map in Figure 6. Lethal mutations were positioned on the genetic map to the highest possible resolution with the available deficiencies, placing them within particular zones. Next, complementation tests against alleles of genes resident in the same zone were performed.

As a result of the analysis described above, I have identified four new complementation groups. To the left of *unc-22*, two of these *(let-664 and let-660)* fall between *dpy-20* and *unc-22*. One of these, *let-660*, is located in a zone in which it is the sole occupant. The other new complementation group in this interval, *let-664*, falls into the same zone as *let-92* (Figure 6).

I have identified two new complementation groups to the right of *unc-22*. These complementation groups, *let-661* and *let-662*, map to the same zones as *let-52* and *let-93*, respectively. None of the previously available deficiencies allowed me to separate *let-661* from *let-52*. However, I have shown that one of the lethal mutations recovered in the MAM1 screen is a deficiency, and the right breakpoint of this deficiency, *sDf69*, has separated the *let-93* complementation group from the *let-662* complementation group (Figure 6).

In addition to identifying new complementation groups, I recovered alleles of all previously existing lethal complementation groups between (and including) *let-60* and *let-93* (Table 3). The two alleles of *let-653* I recovered reveal an

interesting property of the *let-653* gene. One of the alleles, *s2270*, fails to complement the single previously existing *let-653* allele, *s1733*, recovered by D. V. Clark (Clark and Baillie, 1992). *s2270* also fails to complement *s2377*, the other *let-653* allele I recovered. However, *s2377* complements *s1733* (Table 4). Therefore, it appears that alleles of *let-653* exhibit a complex interallelic complementation pattern (see Discussion).

Determination of developmental arrest phase phenotype: In order to establish the approximate stage at which animals homozygous for each of the lethal mutations were arrested in development, Unc-22 animals [genotype *let-(sx) unc-22(s7) lev-1(x22)* or *let-(sx) unc-22(s7) unc-31(e169)*] were picked from stock plates and set on separate plates, where they were kept at room temperature (approximately 20°C). The approximate developmental stage at which these animals were arrested was determined by examining the extent of their gonadal development (Kimble and Hirsh 1979) three or seven days after they were removed from the stock plate (see Materials and Methods). In a similar fashion, I determined the approximate arrest stage for each lethal mutation in *trans* to *sDf10*. The arrest stages are presented in Table 3. Additional comments regarding some of these results are included below.

let-660(s1996): In my initial experiments with *let-660(s1996)*, I found that individuals homozygous for *let-660(s1996)* were arrested in development at an early larval stage. However, animals of the genotype *let-660(s1996) unc-22(s7) unc-31(e169)/sDf7* exhibited a sterile phenotype. These observations could be explained if *s1996* was a hypermorphic mutation, or if there existed a second early blocking lethal mutation on the *s1996* chromosome that was not uncovered by *sDf7*. The latter explanation was shown to be the correct one, as the early blocking phenotype was recombinationally separable from the sterile phenotype, and a

TABLE 3.

Lethal mutations isolated in this study, and their arrest stages.

Gene	Allele	Developmental arrest stage		Total number of alleles
		let-(sx)/let-(sx)	*let-(sx)/sDf10*	
let-52	*s2346*	L2	L2	3
	s2381	L3	L2	
let-56	*s2217*	Sterile	Sterile	11
	s2230	Sterile	Sterile	
	s2321	Sterile	Sterile	
let-60	*s2336*	L3	L2	4 [a]
let-92	*s2021*	L2	L2	4
	s2209	L2	L2	
let-93	*s2254*	L2	L1	3
	s2357	L2	L1	
let-653	*s2270*	L2 (vacuole)	L1 (vacuole)	3
	s2377	Sterile	L2 (vacuole)	
let-660	*s1996*	Sterile (leaky)	Sterile (leaky)	1
let-661	*s2203*	L2	L2	1
let-662	*s2219*	Sterile (leaky)	Sterile (leaky)	1
let-664	*s2374*	Sterile	Sterile	1

[a] This is the number of *let-60* alleles isolated in lethal screens only.

TABLE 4.

let-653 alleles exhibit heteroallelic complementation.

	s1733	*s2270*	*s2377*
s1733	Death at L1-L2.§ Vacuole visible.*	Death at L1-L2. Vacuole visible.	Fertile adult. No vacuole.
s2270		Death at L1-L2. Vacuole visible.	Death at L1-L2. Vacuole visible.
s2377			Sterile adult. No vacuole.

§ The lethal arrest stage of *s1733* homozygotes was initially estimated from comparison of the lengths of these animals to a growth curve generated from homozygous *unc-22(s7) unc-31(e169)* individuals (the genetic background in which *s1733* is contained). In some cases, L1 moults were observed, indicating progression of *s1733* homozygotes to the L2 stage (S. J. M. Jones, 1994). The stage of arrest was confirmed by examination of the extent of gonadal development (Materials and Methods).

* The vacuole present in *let-653(s1733)* homozygous individuals was first noted and described by S. J. M. Jones (1994).

strain segregating sterile Unc-22 Unc-31 animals was recovered (S. J. M. Jones, 1994). I verified that the sterile mutation remaining on the *unc-22* - marked chromosome was *s1996* by complementation tests with appropriate deficiencies.

In order to examine the phenotype of *s1996* homozygous individuals in a more precise fashion, I picked L2 - L3 Unc-22 individuals from stock plates, set them singly on individual plates, and observed them over a period of nine days. Of 11 *s1996* homozygous individuals, three produced no progeny. The remaining eight animals produced an average of five progeny each. All of the animals appeared clear and sickly, and grew slowly.

I examined 15 six-day old adult homozygous *s1996* animals with a Nomarski equipped microscope (Materials and Methods). All of the examined animals contained embryos and/or internal hatchlings. All animals possessed an apparently normal vulva, suggesting that retention of the embryos was not due to the absence of a vulva (i.e. a Vul phenotype). Six of these animals exhibited varying degrees of gonadal disorganization.

In complementation tests involving *s1996* and various deficiencies, I recognized deficiencies that failed to complement *s1996* by the extremely slow growth, sickly appearance, and near-complete sterility of the hemizygous individuals. Hemizygous animals occasionally produced up to three progeny 5-7 days after they were removed to separate plates for progeny testing (Materials and Methods). Hemizygous animals were examined under a Nomarski equipped microscope. These individuals did not display a noticeably more severe phenotype than animals homozygous for *s1996*.

***let-662(s2219)*:** To determine the developmental arrest phase of *s2219* homozygous individuals, I picked L2 - L3 Unc-22 individuals from stock plates, set them singly on individual plates, and observed them over a period of nine days. Of seven individuals, one produced no progeny. This individual was arrested in

development at the fourth larval stage (L4). The six remaining animals produced an average of five progeny each. Similar to *s1996* animals, *s2219* animals were clear and sickly in appearance, and developed slowly.

I examined nine six-day old adult *s2219* animals with a Nomarski equipped microscope. In all cases except one, embryos and internal hatchlings were seen, although the vulva appeared normal. Hemizygous animals carrying *s2219* in *trans* to *sDf10* did not appear to differ in morphology from *s2219* homozygous individuals.

let-56(s2217), let-56(s2230), let-56(s2321): I found the developmental arrest stage of the three alleles of *let-56* that I recovered (Table 3) to be later than the arrest stages noted for pre-existing *let-56* alleles (Clark et al. 1988, Clark 1990, Clark and Baillie 1992). These prior studies report arrest stages ranging from mid to late larvae for *let-56* homozygous individuals, and arrest phases ranging from early to mid larval stages for *let-56* hemizygous individuals. Under my assay conditions (Materials and Methods), I found that animals homozygous or hemizygous for the *let-56* alleles *s2217*, *s2230* and *s2321* can attain adulthood, but are infertile. The assay employed by Clark et al. (1988) used length as an indicator of development. That assay involved selecting Unc-22 or Unc-22 Unc-31 animals [genotype *let-(sx) unc-22* or *let-(sx) unc-22(s7) unc-31 (e169)*] from stock plates and, three days later, calculating the average length of these animals using an ocular micrometer fitted to a dissecting microscope. The stage of arrest was then determined by comparing the length of mutant individuals to an *unc-22(s7)* growth curve (Rogalski, Moerman and Baillie 1982). In the assay I used, homozygous *let-56* animals were left for seven days prior to examination instead of for three days, and determination of the lethal arrest phase was accomplished by examining the extent of gonadal development and using this feature as an indicator of the extent of development (Kimble and Hirsh 1979).

I used this assay to determine the developmental arrest stage of *let-56(s173)*, the earliest arresting *let-56* allele analyzed by Clark et al. (1988). When homozygous, this allele was reported to arrest development in a mid-larval stage. In the *s173* hemizygote, development was arrested in an early larval stage (Clark et al. 1988). I found that *s173* homozygous individuals attained adulthood, as did animals carrying *s173* in *trans* to *sDf10*. As in the cases of animals homozygous for the other *let-56* alleles that I examined, *let-56(s173)* homozygous individuals occasionally appeared to carry a single oocyte. I conclude that the primary phenotypes of animals homozygous for *s173* are extremely slow growth and adult sterility. These phenotypes are likely to represent the phenotypes of null or near-null mutations in the *let-56* gene, as the phenotypes of *let-56(s173)* hemizygotes were not noticeably different from that of *s173* homozygotes. The developmental arrest phenotypes of the *let-56* alleles I isolated in this study all resembled those of *let-56(s173)*, and are therefore likely complete or near-complete loss-of-function mutations in the *let-56* gene.

***let-664(s2374)*:** I found that individuals homozygous for *let-664(s2374)* were arrested in development at the L3 larval stage. However, animals hemizygous for *s2374* exhibited a sterile adult phenotype in *trans* to all of the deficiencies I had used to position *s2374* on the genetic map. These results could be explained if *s2374* was a hypermorphic mutation, or if there was a second earlier-arresting lethal mutation on the chromosome that carried *s2374*. This second mutation would have to be in a region of the chromosome not affected by any of the deficiencies used to position *s2374*. In order to test whether there existed a second mutation on the chromosome bearing *s2374*, I placed *let-664(s2374)* in *trans* to both *mDf7* and *nDf27* (Figure 6), and observed the approximate stage at which hemizygous individuals arrested development. I found that *s2374/mDf7* individuals exhibited a sterile adult phenotype, whereas

s2374/nDf27 individuals were arrested at a mid-larval stage. Therefore, it seemed that *s2374* was not a hypermorph, but rather there existed a second earlier arresting lethal mutation on the *s2374* chromosome. Furthermore, this second earlier arresting mutation was to the right of the right-most breakpoint of *mDf7*, as it was deleted by *nDf27* but not *mDf7*.

I allowed the second lethal mutation to recombine away from *s2374* as follows. *let-664(s2374) unc-22(s7) [lev-1(x22) let-(sx)]/nT1(IV); +/nT1(V)* hermaphrodites were mated with N2 males. L4 hermaphrodite progeny resulting from this cross were examined in 1% nicotine, and animals that twitched in nicotine were picked to separate plates and allowed to produce self progeny. Animals that twitched in nicotine were selected from among these self progeny, picked to separate plates, and allowed to produce self progeny. A plate containing multiple sterile adult Unc-22s identified the desired recombinant. I verified that the sterile mutation remaining on the *unc-22*-marked chromosome was *s2374* by complementation tests with appropriate deficiencies. Wild-type male sperm was not capable of rescuing the sterility of *s2374* homozygous hermaphrodites, which indicated that the sterility of these animals is not due to the production of defective sperm, or to the absence of some factor required in the hermaphrodite sperm.

Complementation tests used to map the location of *s2374* placed the mutation in the same zone as *let-92*, between *par-5* and *let-653*. S. J. M. Jones had shown that the cosmid B0033 (Figure 5) was capable of restoring a portion of the missing *let-92*(+) function, by demonstrating that *let-92(s504) unc-22(s7)* and *let-92(s677) unc-22(s7)* homozygous hermaphrodites, which normally do not proceed in development past an early larval stage, could attain adulthood when carrying an extra-chromosomal array containing B0033. Although these individuals were not fertile, this was good evidence that at least a portion of the *let-92* gene was contained within B0033 (S. J. M. Jones, 1994). Because *let-92* and *s2374* mapped

to the same zone, I tested whether B0033 was capable of rescuing *s2374* (Materials and Methods). I found that *s2374* was not rescued by B0033.

PCR analysis of deficiencies with a genetic breakpoint between *unc-22* and *let-52*: In addition to their utility in genetic mapping experiments and dosage studies, chromosomal deficiencies are often valuable in gene cloning experiments. The placement of deficiency breakpoints on the physical map of a region of the genome can greatly accelerate efforts to molecularly characterize genes that have been identified by mutation. J. Schein and I had observed that six of ten formaldehyde induced deficiencies recovered by J. Schein had genetically inseparable right breakpoints between *unc-22* and *let-52* (*sDf68*, *sDf80*, *sDf82*, *sDf83*, *sDf85* and *sDf86,* see Figure 6). Other previously identified deficiencies also had genetic breakpoints in this interval (*mDf7, sDf8, sDf63 and sDf64;*see Figure 6). All of these deficiencies, with the exception of *mDf7* (Rogalski and Riddle, 1988), were selected, following treatment with formaldehyde, as heterozygotes on the basis of an Unc-22 phenotype in 1% nicotine (Materials and Methods).

Because all of the deficiencies, with the exception of *mDf7*, were selected on the basis of an Unc-22 phenotype in nicotine, it was possible that these deficiencies did not delete the entire *unc-22* locus but rather broke within it. This disruption of the *unc-22* gene would result in the production of the Unc-22 phenotype. If the right breakpoints of certain of these deficiencies were located within the *unc-22* locus, these deficiencies would be of no use in future attempts to clone the genes immediately to the right of *unc-22* (i.e. *let-52* and *let-661*). In order to determine whether these deficiencies disrupted the *unc-22* locus, or completely deleted it, I performed PCR on deficiency homozygotes (Materials and Methods) with primers that amplify a region approximately 2 kb upstream (i.e. to

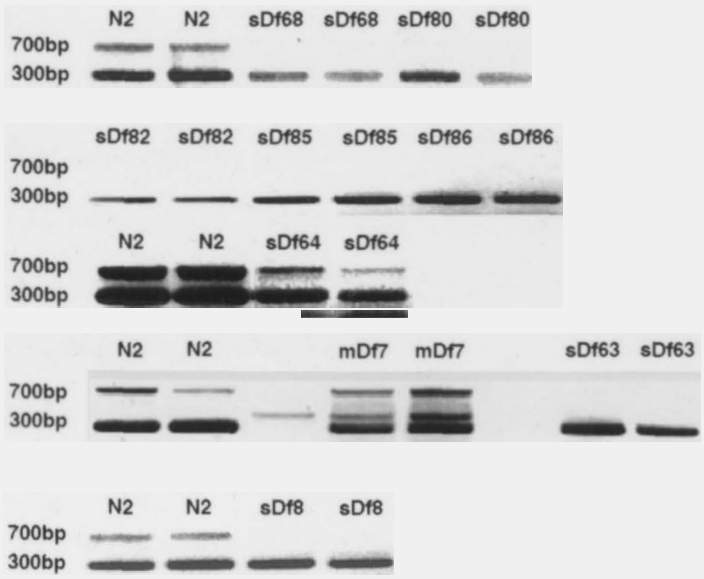
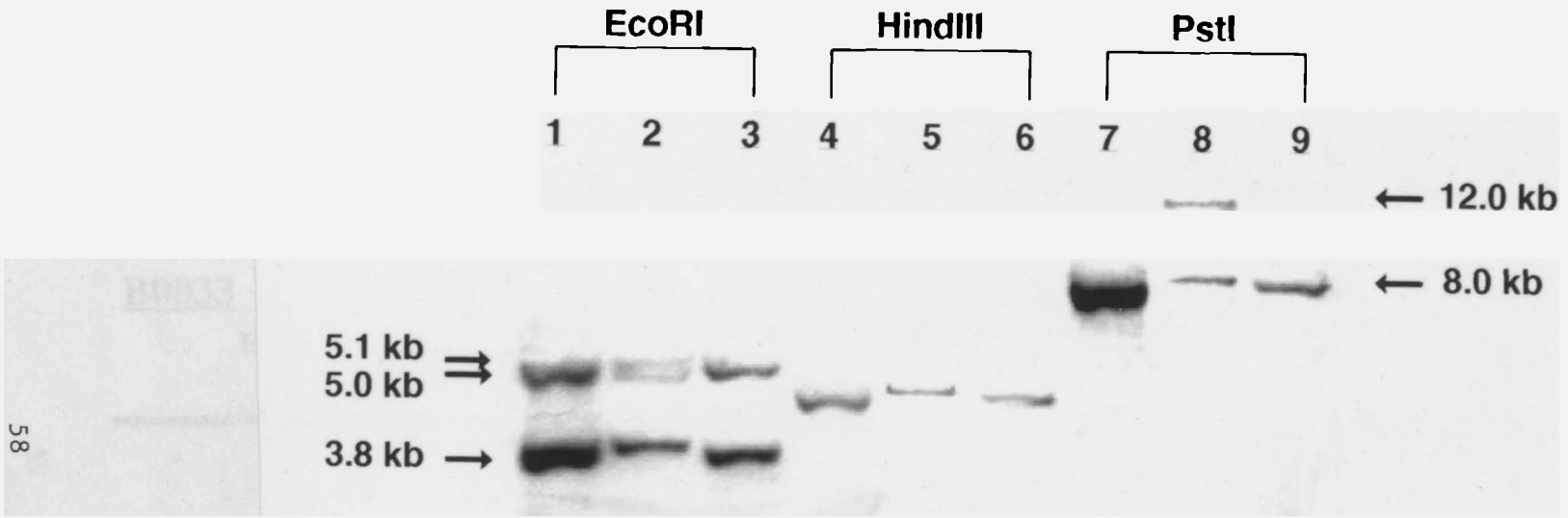


Figure 7:

PCR from deficiency homozygotes. The 700 bp band corresponds to the KT01 - KT03 amplification product. The 300 bp band corresponds to an amplification product produced by chromosome V-specific primers incorporated into all reactions as a control (primers 179 - 180; Table 2). With the exception of *sDf64* and *mDf7*, all of the deficiencies failed to give the 700 bp amplification product. This indicates that they disrupt at least the KTO3 primer annealing site, and therefore completely delete the *unc-22* gene. The unlabeled lanes contain failed reactions.

the right of *unc-22* on the genetic map) of the *unc-22* gene (Figure 6; see Section III). These primers (KTO1 and KTO3; see Table 2 and Section III) were obtained from G. Benian (Schein et al. 1993), and are specific for the spermatogenesis defective gene *spe-17* (L'Hernault, Benian and Emmons 1993). The absence of an amplification product indicated the deficiency deleted at least one of the priming sites and therefore broke at least 2 kb upstream of the *unc-22* initiating methionine, and not within the *unc-22* gene. I found that individuals homozygous for all of the deficiencies, with the exceptions of *mDf7* and *sDf64,* failed to produce an amplification product with the KT01/KT03 primer pair (Figure 7; *sDf83* not shown). These results indicated that all of the deficiencies, with the exceptions noted above, had right breakpoints in the interval separating *unc-22* from *let-52.*

Positioning the left breakpoint of *sDf9* on the physical map: Three deficiencies, *sDf9*, *sDf19* and *sDf65*, had left genetic breakpoints between *let-56* and *let-653*. D. Clark had previously obtained evidence showing that the cosmid B0033 hybridized to a polymorphic *Eco*RI fragment on Southern blots containing *sDf19 / nTI(IV;V)* genomic DNA (Clark 1990). This result indicated that the left breakpoint of *sDf19* was within DNA contained in B0033. To position the breakpoints of the other deficiencies on the physical map, I undertook a similar experiment using *Eco*RI, *Hin*dIII, and *Pst*I digested genomic DNA prepared from *sDf9* (prepared by S. Prasad) and *sDf65* (prepared by D. Clark). Southern blots containing these DNAs were probed with a *Pst*I subclone of the cosmid C46F3, pCes1940 (S. Jones, 1994). This subclone contains most of the *let-653* gene, and does not contain B0033 DNA. As a result of this hybridization, polymorphic bands were seen in the lanes containing *sDf9 / nT1(IV;V)* DNA digested with *Eco*RI and *Pst*I. No polymorphic bands were seen in lanes containing *sDf65 / nT1(IV;V)* DNA, or in any of the the lanes containing *Hin*dIII digested DNA. These results are shown and interpreted in Figures 8 and 9, respectively.



58

Figure 8:

Portion of a chemi-lumigraph of a Southern blot containing *Eco*RI, *Hin*dIII and *Pst*I digested DNA prepared from N2, *sDf9 / nT1(IV;V)* and *sDf65 / nT1(IV;V)* animals. Lanes 1, 4 and 7 contain N2 DNA. Lanes 2, 5 and 8 contain *sDf9 / nT1(IV;V)* DNA. Lanes 3, 6 and 9 contain *sDf65 / nT1(IV;V)* DNA. The N2 lanes are overloaded with respect to the *sDf9* and *sDf65* lanes. Arrows indicate the sizes of the wild-type bands and the *sDf9* "fusion" fragments in the *Eco*RI and *Pst*I lanes. *Hin*dIII restriction fragments showed no polymorphic bands. There are at least three additional hybridizing *Hin*dIII fragments, too small to be visible in this photograph.

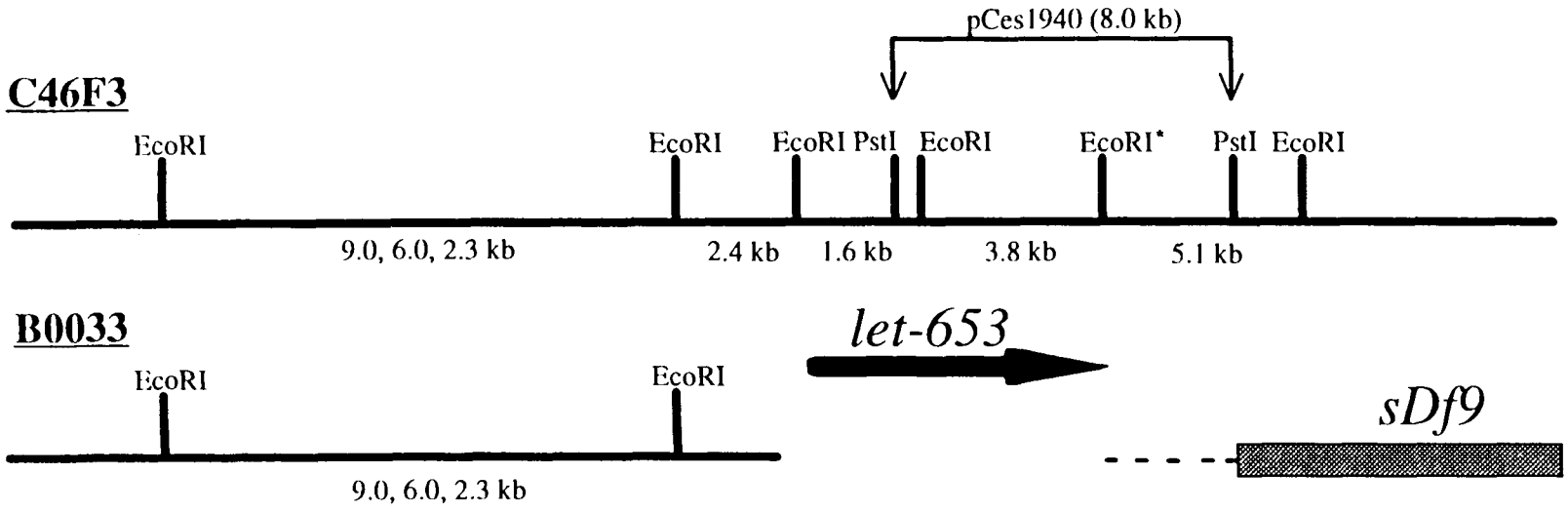


Figure 9:

An interpretation of the results shown in **Figure 8**. The restriction maps of the cosmids C46F3 and B0033, in addition to the direction and extent of the *let-653* open reading frame, were determined by S. J. M. Jones (1994). The sizes of the *Eco*RI fragments are given below the restriction maps of the cosmids. The arrow indicates the *let-653* gene and the direction of the *let-653* open reading frame. The probe used in the experiment shown in **Figure 8** is contained within the plasmid pCes1940. *sDf9* extends at least as far as the right-most *Pst*I site of pCes1940. *sDf9* is unlikely to extend further to the left than the *Eco*RI site indicated by the asterisk, as *let-653* coding sequences span this *Eco*RI site and *sDf9* genetically complements *let-653*.

DISCUSSION

This section of my thesis has described efforts to identify and characterize developmentally required genes in the *sDf10(IV)* region. Ultimately, this work is directed toward the long term goal of understanding the organization and molecular nature of the genes in this interval that, when mutated, give rise to death, arrested development or slow growth.

My strategy involved screening for recessive mutations following treatment with EMS. Out of a total of 216 mutations recovered in two screens, I identified 17 lethal mutations that mapped to the region of interest. Complementation analysis of these mutations showed that I had identified four new lethal complementation groups. Each of these new complementation groups is represented by a single allele. The identification of these new complementation groups increases to eleven the number of essential genes identified in the *sDf10* interval (Figure 6). The observation that four of the eleven essential genes identified to date in this region are represented by single alleles indicates that the region has not been saturated for mutations in essential genes.

Other studies have generated estimates of the minimum number of essential genes in the *C. elegans* genome. Johnsen and Baillie (1991) estimated a minimum of 2850 essential genes in the genome. Howell and Rose (1990) estimated a minimum of 3300 essential genes, and Clark (1990) estimated a minimum of 3500 essential genes in the genome. Assuming a genome size of 100 000 kb (Sulston et al. 1992) and an average cosmid size of 40 kb, 2500 cosmids would be required to reconstruct the haploid genome. This figure, taken together with the estimated minimum number of essential genes from the studies cited above, predicts a minimum of approximately one essential gene per cosmid on average, over the entire genome. The physical map of the region between *let-60* and *unc-22* is composed of contiguous cosmid clones, and these markers have been placed on the

physical map (see Introduction). Approximately 8 cosmids span this region, and cumulative efforts have resulted in the identification of six essential genes here (including *let-60*). Therefore, a density of approximately 0.75 essential genes per cosmid has been achieved in this interval, below the minimum average estimate calculated above. This could be due to a lower than average number of essential genes in this region. An alternative possibility is that there are as-yet unidentified essential genes in the interval. My identification of additional essential genes represented by single alleles supports the notion that the region is not yet saturated for mutations in essential genes. Therefore, an even higher resolution genetic map of this interval based upon recovery of similar mutations seems possible, albeit at considerable additional effort.

As well as identifying four new essential genes in the *sDf10* region, I have identified additional alleles of all previously identified essential genes in the interval, with the exception of *let-65* (Table 3). Prior to this work, *let-52*, *let-93* and *let-653* were each defined by a single allele. I have identified two additional alleles for each of these complementation groups. In addition, I have isolated three new alleles of *let-56*. The *let-56* locus is currently the largest developmentally required mutagenic target in the region under study, with 11 alleles recovered following treatment with EMS. An additional loss-of-function allele of *let-60* has also been recovered, as have two alleles of *let-92*. These new alleles will aid in our long term goal of understanding the function of these genes at the molecular level. For example, a collection of alleles at each locus may define domains that are required for different functions during development, as may be the case with the *let-653* gene.

Prior to this study, a single allele of *let-653* had been identified (Clark and Baillie, 1992). A phenotypic characterization of this allele, *s1733* , revealed that homozygous individuals developed a vacuole near the posterior pharyngeal bulb

soon after hatching, and this vacuole gradually expanded to fill a large fraction of the body cavity. Death occurred late in the L1 stage or early in the L2 stage (S. J. M. Jones, 1994).

I have identified two additional alleles of *let-653*. One of these, *s2270*, results in a phenotype that is virtually identical to that produced in *s1733* animals. The other newly identified allele, *s2377*, produces a different phenotype. Individuals homozygous for this allele survive to adulthood, with no indication of vacuole formation. I have found that, whereas *s2377* complements *s1733*, it fails to complement *s2270*. *s2270* fails to complement *s1733*. Therefore, alleles of *let-653* exhibit a complex complementation pattern (Table 4). Complex complementation patterns have also been seen with alleles of *unc-15* (Rose and Baillie, 1980), *cha-1 - unc-17* (Rand, 1989), *bli-4* (Peters, McDowall and Rose, 1991), *lin-40* (R. C. Johnsen, W. B. Barbazuk and D. L. Baillie, in preparation), and *unc-52* (Gilchrist and Moerman, 1992).

An alternative scenario that would explain the complementation results is one in which *s2377* is a mutation in an essential gene adjacent to *let-653*, and the non-complementing allele, *s2270*, is a deficiency that affects both *let-653* and the adjacent gene defined by *s2377*. This scenario is supported by the observation that *s2377* homozygotes do not exhibit the distinctive vacuole phenotype displayed by animals homozygous for either *s1733* or *s2270*. However, this scenario is unlikely since animals hemizygous for *s2377* arrest at the second larval stage (L2) and exhibit the characteristic *s1733* "vacuole" phenotype. Furthermore, the vacuole phenotype is also seen in *s2270/s2377* hetero-allelic animals. Based on these data, I believe that *s2270* is a null or near-null allele of *let-653*, that *s2377* is a hypomorphic mutation affecting the *let-653* gene, and that the *let-653* locus exhibits a complex complementation pattern. This sort of complementation pattern has been explained by a model proposing the locus encodes a

multifunctional polypeptide with two (or more) functionally distinct domains which are required for normal development. This multifunctional polypeptide could function *in vivo* as a homodimer (Rand, 1989). In the case of *let-653*, this model would predict that the *s2377* mutation disables one discrete domain, while *s1733* disables another. The homodimeric Let-653 protein produced in the *s2377/s1733* heterozygote would retain one of each of the required functional domains, and would therefore have sufficient *in vivo* activity to produce the observed complementation result. The non-complementing allele, *s2270*, would affect both the domain affected by *s2377*, and the domain affected by *s1733*, and would therefore be unable to provide the missing function in *s2377/s2270* or *s1733/s2270* heterozygotes.

Another model explaining intragenic complementation has been proposed. This model invokes alternative splicing, in which multiple products from a single gene are produced. In this model, alleles would be found to complement one another if the complementing mutations affected different transcripts from the same gene. Evidence for this model has been obtained in the case of *bli-4* (C. Thacker and M. Srayko, personal communication), *unc-52* (Gilchrist, 1993; Gilchrist and Moerman, 1992, Rogalski et al., 1993) and *unc-60* (McKim et al., 1994).

The genomic DNA corresponding to the *let-653* locus has been cloned and sequenced, as has a single *let-653* cDNA (Jones, 1994). The predicted LET-653 protein has some structural similarities to a family of glycoproteins known as "mucins" (Jones, 1994). Some of these similarities include a hydrophobic leader sequence, a central serine-threonine rich domain which may be glycosylated, and the presence of cysteine rich regions flanking the serine-threonine rich domain. These flanking regions contain sites at which asparagine residues may be glycosylated. No evidence that the *let-653* gene produces multiple transcripts has

been obtained, but there exists evidence that mucins may function as homomultimers in human and rat (Carlstedt and Sheenan, 1989 and Thornton et al., 1990; cited in Jones, 1994). In order to better understand the role the *let-653* gene assumes in development and the functional importance of the various domains noted in the predicted LET-653 protein, additional molecular and genetic characterization of the *let-653* locus and the lesions affecting it are required.

An attempt was made to rescue *let-664(s2374)* with an extrachromosomal array, constructed by Jones (1994), that contained the cosmid B0033. The inability of this array to rescue *s2374* indicates that the *s2374* rescuing cosmid will likely lie to the left of B0033 for the following reasons. The genetic marker immediately to the left of *let-92*, *par-5*, has been tentatively assigned to the cosmid C38H7 (D. Shakes, personal communication). The closest genetic marker to the right, *let-653*, has been rescued by C29E6 (Clark and Baillie, 1992) and C46F3 (Jones, 1994). The overlap between B0033 and C46F3 is extensive; only the rightmost 15 kb of C46F3 is not contained within B0033. DNA sequence analysis of this region has shown that the *let-653* gene lies immediately to the right of B0033 on C46F3 (Jones, 1994). In addition, C46F3 does not rescue *let-92* (Jones, 1994). This implies that *let-92* resides at the extreme left end of B0033, in the region not contained within C46F3. Because B0033 does not rescue *s2374*, the *let-664* gene must either traverse the leftmost breakpoint of B0033, or reside on an adjacent cosmid. Based on the map position of *let-664(s2374)* with respect to *par-5*, it is improbable that any cosmid to the left of C38H7 (the cosmid tentatively containing *par-5* rescuing activity) would rescue *let-664(s2374)*. Therefore, suitable choices with which to attempt cosmid rescue of *s2374* in future studies would include C08D7 and W03A12, in addition to C38H7. C08D7 and the cosmid C02C4, which share extensive overlap with C38H7, have been shown to contain several molecularly defined genes which may correspond to *let-664*. Using the method of

interspecies hybridization, Prasad (1988) found evidence for the existence of at least 2 genes on C08D7, and detected a transcript for one of these genes. A developmental expression profile of this gene, contained on restriction fragment "k", showed maximal expression in the adult. Prasad identified another gene on C08D7, on restriction fragment "m", but did not detect a transcript for this gene. In addition, he identified a restriction fragment ("g") from the cosmid C02C4 that hybridized to two transcripts on Northern blots (transcripts "g1" and "g2"). Both g1 and g2 show maximal expression at the L2 stage (Prasad, 1988; Prasad and Baillie, 1989).

The systematic recovery and analysis of mutations in well balanced regions of the *C. elegans* genome provides a number of useful tools to the *C. elegans* community in general, and to the study of genome organization in particular. A large fraction of *C. elegans* genes can be identified by "lethal" mutation; therefore, screens for lethals will provide mutant alleles of genes involved in many different biological processes. In addition, because essential genes are by far the largest of the classes of mutable loci yielding easily detectable phenotypes, genetic maps constructed using lethals offer very high resolution, on the order of one mutation per cosmid. This very high resolution at the genetic level makes possible the attainment of high resolution correlations between the physical and the genetic maps through the use of germline transformation rescue technology (Fire 1986; Mello et al., 1991). In turn, a high resolution correlation between the two map types accelerates the mapping and cloning of interesting mutations isolated in lethal and other mutagenesis screens. Ultimately, high resolution genetic maps of the genome and collections of alleles at each locus will aid in the functional analysis of the genomic DNA sequence as it is generated by the *C. elegans* genome sequencing consortium.

SECTION II.

Recovery of duplications by drug resistance selection.

INTRODUCTION

In Section I, I described the recovery and analysis of recessive mutations that arrested or slowed development and/or reproduction. These mutations were tightly linked to *unc-22*, within the region under study in this thesis. In Section II, I describe a method I have devised, following a suggestion made by D. Riddle, to recover duplications of essential loci. In theory, this method can be used to screen for chromosomal duplications of vital loci that lie within the region of chromosome *IV* genetically balanced by the reciprocal translocation *nT1(IV;V)* (Ferguson and Horvitz 1985; Clark et al. 1988; see Figure 1). On chromosome *IV*, *nT1* has been shown to effectively suppress recombination over the 21 map unit interval from *lin-1* to *dpy-4*. This interval represents about 7% of the *C. elegans* genome, and based on recent estimates of approximately 17000 genes in the genome (Sulston et al. 1992; Waterston et al. 1992, Wilson et al., 1994), would contain more than 1200 genes. I have demonstrated the effectiveness of the methodology by employing it to recover duplications of *let-56*, the first essential gene to the left of *unc-22* (Figure 6).

In *Caenorhabditis elegans*, chromosomal duplications have been shown to be valuable tools in studies of gene dosage effects (for an example see Dalley et al. 1993), mosaic analysis (Herman 1987), and studies of the meiotic aspects of chromosome behaviour (Herman, Madl and Kari 1979; Rose, Baillie and Curran 1984; Herman and Kari 1989; McKim and Rose 1990). In addition, duplications are of demonstrated utility in gene mapping studies (McKim and Rose 1990; McKim, Howell and Rose 1988), and have been used as genetic balancers for the induction and maintenance of lethal mutations (Howell et al. 1987; H. I. Stewart, D. L. Collins, A.M. Howell, A.M. Rose and D.L. Baillie, in preparation). Given the many useful features of duplications, a rapid, easily executed scheme for their recovery is desirable.

The duplication selection scheme I have implemented makes use of the *lev-1(x22)(IV)* mutation, which confers resistance in a recessive fashion to the potent anthelmintic levamisole (Lewis et al. 1980). Candidate duplication bearing individuals are selected based on their resistance to levamisole. According to the scheme, hermaphrodites of the genotype *let-x unc-22(s7) lev-1(x22)/nT1(IV) ; +/nT1(V)* are treated with a mutagen causing chromosome breakage (e.g. X-radiation or gamma radiation). Treated animals are then allowed to produce self progeny, and the self progeny are exposed to levamisole. Levamisole-sensitive euploid progeny are of the following genotypes: *let-x unc-22(s7) lev-1(x22)/nT1(IV) ; +/nT1(V)* or *nT1(IV)/nT1(IV) ; nT1(V)/nT1(V).* These animals are hypercontracted and paralyzed in the presence of levamisole. Levamisole-resistant progeny are of the genotype *let-x unc-22(s7) lev-1(x22)/let-x unc-22(s7) lev-1(x22).* These individuals are homozygous for the *unc-22* marker, and consequently twitch in a constitutive fashion. In addition, they are homozygous for the lethal mutation, and are therefore developmentally arrested or dead. In contrast, progeny of the above genotype that, in addition, carry a duplication containing a wild type copy of the mutant essential gene are not developmentally arrested. A flow diagram illustrating the scheme is shown in Figure 2. In order to test the scheme, I screened for chromosomal duplications carrying a wild-type copy of *let-56,* the first essential locus to the left of *unc-22(IV)* (Schein et al. 1993; Figure 6). The allele of *let-56* I chose for this test was *s2321.* The recovery and analysis of *let-56(s2321)* is described in Section I. Briefly, animals homozygous for the *let-56(s2321)* mutation grow extremely slowly compared to wild-type individuals or individuals heterozygous for the mutation. *let-56(s2321)* homozygous individuals attain adulthood after approximately two weeks at 20^{o}C, and are sterile. Three days after hatching, *s2321* homozygous animals are in the early larval stages of development. I chose for my duplication screen a mutagen

that causes chromosome breakage (i.e. gamma radiation) expecting to recover duplications containing a wild-type copy of *let-56*. I presumed these duplications would be derived from *nT1(IV)*.

A paper describing some of the results presented in this Section of the thesis has been accepted for publication (M. A. Marra and D. L. Baillie, **Genome**, in press).

RESULTS

1050 young gravid hermaphrodites of the genotype *let-56(s2321) unc-22(s7) lev-1(x22)/nT1(IV);+/nT1(V)* were treated with 1500 R of gamma radiation, following the dose recommended by Rosenbluth, Cuddeford and Baillie (1985; see Figure 2). Fifty treated individuals (the P_0 generation) were placed on each of 21 plates and allowed to lay eggs for two 24 hour broods. Three days after the brood had been set, F_1 animals were collected in M9 buffer and treated as described in Materials and Methods (**Genetics**).

Six fertile animals displaying the characteristic Unc-22 "twitching" phenotype (Moerman and Baillie 1979; Waterston et al. 1980) and exhibiting varying degrees of resistance to levamisole were recovered from a population of approximately 24000 F_1's. Four of these individuals appeared only partially resistant to the effects of levamisole. Progeny testing of these four animals revealed that, in each case, the majority of the eggs that were laid failed to hatch. In addition, each of the four animals produced some progeny that lacked vulvae. Both of these characteristics are indicative of the presence of *nT1(IV;V)* (Ferguson and Horvitz 1985). These animals were presumed to be resistant to levamisole due to the induction of an *unc-22* mutation on the *nT1(IV)* chromosome that conferred partial levamisole resistance, and were not analysed further. Unc-22 individuals are weakly resistant to the effects of levamisole, probably due to their inability to hypercontract upon exposure to the drug (Brenner 1974, Lewis et al. 1980). The two remaining animals exhibited strong resistance to levamisole. In addition, neither of these individuals produced progeny that indicated the presence of *nT1(IV;V)*. One of these animals displayed a considerably more severe "twitch" than did the other, and more closely resembled *unc-22(s7)* homozygous individuals.

The phenotypes of the self progeny produced by the "severe twitcher" were examined. In the F_1, fertile twitchers and sterile twitchers were found. I proposed that the fertile twitchers carried a duplication that covered the *let-56* locus. In order to determine whether the putative duplication carried in this strain could be present in two copies without detriment to the organism, I picked 20 animals to separate plates and observed their progeny. Several of the animals failed to segregate any sterile twitchers, producing only fertile twitcher progeny. This indicated to me that the duplication had become linked to a chromosome, and in these animals was present on both homologues. The duplication was designated *sDp11*. A single animal derived from a parent that had failed to segregate sterile twitcher progeny was selected, and used to establish a strain (BC4274). The genotype of this animal was thought to be *let-56(s2321) unc-22(s7) lev-1(x22) ; sDp11*. However, my results could also be explained if *let-56(s2321)* had recombined away from the *unc-22(s7) lev-1(x22)* marked chromosome during the course of my experiments. If this had occurred, the genotype of BC4274 individuals would be *unc-22(s7) lev-1(x22)*. A recombination event resulting in the loss of the lethal mutation was unlikely for two reasons. First, in the strain that was mutagenized, the mutations *let-56(s2321), unc-22(s7)* and *lev-1(x22)* were balanced by the *nT1(IV;V)* rearrangement, which eliminates crossing over between them. Second, *let-56* is recombinationally close to *unc-22*, separated from it by approximately 0.05 map units (Schein et al. 1993). I confirmed that the lethal mutation had not recombined away from the other markers on chromosome *IV* as follows. BC4274 individuals were mated with wild-type (N2) males, and wild-type L4 hermaphrodites resulting from this cross were picked to separate plates and allowed to produce self progeny. 101 young Unc-22 larvae were selected from among the self progeny, and were set on separate plates in order to observe their development over a period of two weeks. Of these 101 individuals, three

failed to develop into fertile adults, displaying the sterile adult terminal phenotype characteristic of *let-56(s2321)* homozygous individuals. I concluded that *let-56(s2321)* was present in BC4274 individuals, and had not recombined away from the *unc-22(s7)* and *lev-1(x22)* markers. However, because the sterile phenotype segregated infrequently in this experiment, it appeared that the duplication of *let-56* I had isolated was attached to chromosome *IV*, approximately 17 map units from *unc-22* (95% C. I. 8.9 map units to 28.3 map units; see Appendix 1). In addition, my results showed that this duplication that contained wild-type *let-56* activity did not contain a functional copy of the *unc-22* locus, as individuals that carried either one or two copies displayed an Unc-22 phenotype. Therefore, one of the duplication breakpoints must lie between *let-56* and *unc-22*, or possibly within the *unc-22* locus itself.

The other levamisole resistant animal I recovered following the gamma ray mutagenesis screen displayed a much weaker "twitch" than did BC4274 individuals, and exhibited greater motility. The phenotypes of the progeny produced by this "weak twitcher" were examined. In the F_1, three distinct phenotypes were noted. I observed the parental fertile weak twitcher phenotype in addition to sterile severe twitcher animals and fertile non-twitcher animals. The sterile twitcher animals resembled *let-56(s2321)unc-22(s7)lev-1(x22)* homozygous individuals. I selected individual non-twitcher animals for progeny testing, and found only non-twitcher offspring among the progeny of these animals, whereas weak twitcher individuals consistently produced the three phenotypic classes of offspring described above. These observations led me to propose that I had recovered a duplication that contained a wild-type copy of the *let-56* locus, and affected the *unc-22* locus in some way. The weak twitching phenotype I had observed was intriguing. Animals homozygous for the *unc-22(e66)* allele and carrying a single copy of the free *unc-22(+)*-bearing duplication *nDp5* (Beitel,

Clark and Horvitz 1990) did not exhibit a weak twitching phenotype, but rather appeared wild-type (this study and G. Beitel personal communication). As a consequence of these observations, I considered two possibilities. Either the duplication I had recovered contained an altered copy of the *unc-22* gene that was incapable of encoding completely wild-type *unc-22* function, or the duplication had become attached to a chromosome where it was subject to some type of position effect down-regulation so that it produced a less than wild-type amount of *unc-22* gene product. I believed the first possibility to be unlikely for the following reason. If I had recovered a duplication that contained an altered copy of the *unc-22* gene, then I should never have observed non-twitcher progeny segregating from the weak twitchers which were, presumably, homozygous for *unc-22(s7)*. None of the self progeny would carry a wild-type copy of *unc-22*, and would all therefore twitch. All known mutant alleles of *unc-22* display the twitching phenotype; the *unc-22* gene has been shown to be sensitive to even apparently innocuous lesions, all of which produce the characteristic twitch (Benian, L'Hernault and Morris 1993). Support for the second hypothesis was provided by the absence of arrested animals among the progeny of the non-twitcher hermaphrodites. If the duplication I had recovered was free, then it seemed likely that an occasional sterile twitcher would be seen among the progeny of the non-twitcher animals. This was not observed. Therefore, the duplication had probably become attached to a chromosome. If so, it seemed possible that it might have become linked to a chromosomal region that was insufficiently transcriptionally active to produce enough wild-type *unc-22* gene product to overcome the twitching phenotype produced by the homozygous *unc-22(s7)* allele. I selected a single non-twitcher hermaphrodite to establish the strain BC4430 and designated the duplication carried by this strain *sDp10(IV)*.

In order to ascertain that *unc-22(s7)* was homozygous in BC4430 animals, I mated N2 males with BC4430 hermaphrodites, and selected male progeny. None of the male progeny that resulted from the above cross twitched when immersed in a 1% nicotine solution, which causes individuals heterozygous for *unc-22* mutations to twitch in a fashion resembling individuals homozygous for *unc-22* mutations (Moerman and Baillie 1979; 1981). I mated single males from this cross with *dpy-20(e1282) unc-26(e345)* hermaphrodites. Outcross progeny were easily distinguished from self progeny, as self progeny were Dpy-Unc, and outcross progeny were wild-type. 30 single male matings were performed, and the progeny resulting from each cross were examined in a 1% nicotine solution. None of the crosses resulted in the production of hermaphrodite progeny that twitched in nicotine although, in each of the crosses, there were approximately equal proportions of twitching and non-twitching males. These results were consistent with the idea that BC4430 individuals were homozygous for *unc-22(s7)* and carried a duplication of *unc-22* that was attached to the *X* chromosome and present on both homologues.

In order to determine whether the *let-56(s2321)* mutation was still present in BC4430 and to test for *sDp10*'s coverage of *dpy-20* and *unc-26*, individual L4 hermaphrodites resulting from the cross described above were picked to individual plates and allowed to produce self progeny. These hermaphrodites were of two possible genotypes, either + *+/dpy-20(e1282) unc-26(e345) ; sDp10* or *let-56(s2321) unc-22(s7)lev-1(x22)/dpy-20(e1282) unc-26(e345) ; sDp10.* The genotype of the latter animals was recognized by the appearance of twitcher progeny. The phenotypes of the progeny produced by these hermaphrodites were scored. Sterile twitcher animals resembling *let-56(s2321) unc-22(s7) lev-1(x22)* homozygous individuals were recovered, indicating *let-56(s2321)* was retained in the BC4430 stock, and had not recombined away from the *unc-22(s7) lev-1(x22)*

markers. In addition, the frequency of Dpy-Unc animals observed in these experiments indicated that the duplication did not contain functional copies of either the *dpy-20* or *unc-26* genes ($chi^2 = 0.019$, D.f. = 1, $p = 0.85$).

I next determined whether a male strain carrying *sDp10 (IV;X)* could be constructed. If the duplication could be transmitted through the male, there would be no need to construct male strains heterozygous for the mutations being tested for complementation with *sDp10,* simplifying subsequent complementation tests. I attempted construction of the male strain as follows. Hermaphrodites of the genotype *let-56(s2321)unc-22(s7) lev-1(x22) / let-56(s2321) unc-22(s7) lev-1(x22) ; sDp10 /sDp10* were mated with wild-type (N2) males. These hermaphrodites twitched in nicotine. All of the males resulting from this cross (genotype *let-56 unc-22 lev-1 / +++ ; sDp10*) did not twitch in nicotine. These males were then backcrossed to hermaphrodites of the maternal genotype. Males resulting from this cross were either of the genotype *let-56 unc-22 lev-1 / +++ ; sDp10*, in which case they would not twitch in nicotine, or of the genotype *let-56 unc-22 lev-1 / let-56 unc-22 lev-1 ; sDp10.* The phenotype of these males was expected to resemble that of hermaphrodites of this genotype, i.e. they were expected to exhibit a constitutive weak twitch. No consitutive weak twitch was observed. However, in the presence of nicotine, these males twitched. Therefore, a sex-specific difference in the phenotype of genotypically identical males and hermaphrodites was observed. These males appeared slightly uncoordinated, and did not mate well.

In an attempt to improve the mating efficiency of males carrying *sDp10*, I removed *sDp10* from the *let-56 unc-22 lev-1* background, and constructed a male strain carrying *sDp10* in which both chromosome *IV*'s were wild-type. These males (genotype *++ / sDp10*) mated well, and a male strain (BC4475) was

established. Males of this genotype were used in subsequent complementation tests.

To determine whether *sDp10* covered *unc-31* , ++ ; *sDp10* males were mated to *dpy-4(e1166) unc-31(e169)* hermaphrodites. L4 hermaphrodites of the genotype *dpy-4(e1166) unc-31(e169)/++ ; sDp10* resulting from this cross were picked to individual plates and allowed to produce self progeny in four 24 hour broods. The progeny of ten hermaphrodites were scored. The frequency of non-Unc-31 Dpy-4 animals observed among the self progeny indicated that the duplication did not contain a functional copy of the *unc-31* gene ($chi^2 = 515$, D.f. = 1, $p < 0.005$).

To test whether the duplication carried functional copies of any of the vital loci between *dpy-20* and *unc-31*, ++;*sDp10* males were crossed with hermaphrodites of the genotype *let-(sx) unc-22(s7) {unc-31(e169)}/nT1(IV); +/nT1(V)* [where *let-(sx)* was an allele of one of the essential genes indicated in Table 5; some lethal alleles were on chromosomes marked with *unc-31(e169)*]. At least 15 F_1 hermaphrodite progeny that failed to twitch in nicotine were picked to individual plates. These F_1 animals were either *let-(sx) unc-22(s7) {unc-31(e169)}/+++; sDp10* or *+++/nT1(IV);+/nT1(V) ;sDp10.* The progeny of these animals were examined for the presence of twitcher animals. The presence of multiple fertile adult weak twitcher animals [genotype *let-(sx) unc-22(s7) {unc-31(e169)}* and carrying a single copy of *sDp10*] indicated that *sDp10* carried a wild-type copy of the lethal mutation being tested. In cases where lethal mutations had been induced on chromosomes marked with *unc-31(e169)*, the presence of fertile adult Unc-31 animals that twitched weakly indicated that *sDp10* complemented the lethal mutation. When either weak twitchers or weakly twitching Unc-31 animals were seen, several were picked to individual plates for examination and progeny testing. The *sDp10* complementation results are

presented in Table 5. These results have allowed me to position the genetic breakpoints of *sDp10* on chromosome *IV* as shown in Figure 10. Some lethal mutations contained within the duplicated region were not rescued by a single copy of *sDp10*. No weakly twitching individuals indicative of *sDp10* rescue were found in complementation tests involving *let-93(s734)* or *let-656(s1767)*.

In addition to the putative point mutations listed in Table 5, *sDp10* was capable of rescuing a small deficiency. Individuals homozygous for the deficiency *sDf65* and carrying a single copy of *sDp10* were viable and fertile. *sDf65* affects all the genes between and including *let-56* and *let-52* (Figure 10).

TABLE 5.

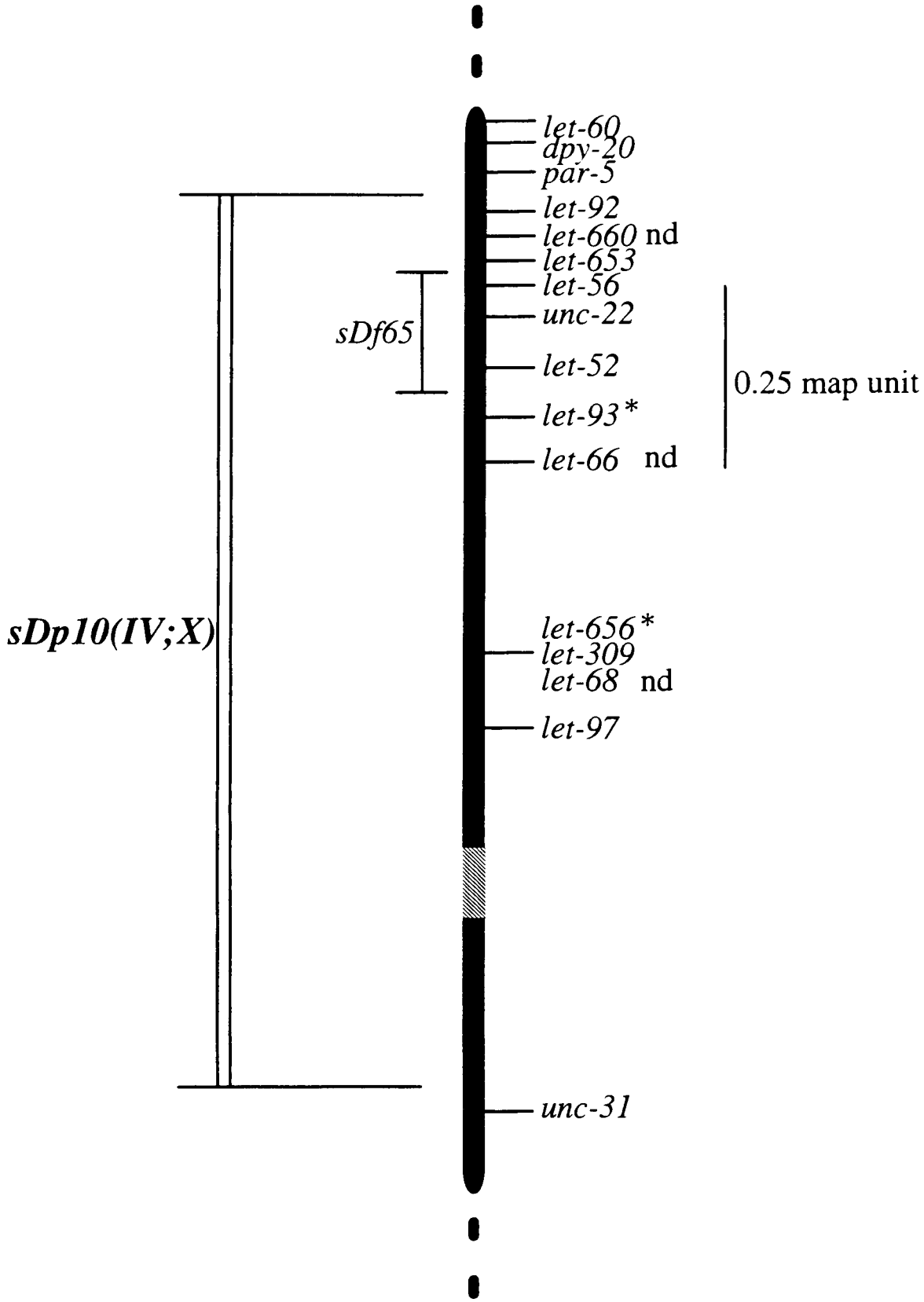
sDp10 complementation results.

Mutation	Result[§]
let-60(s1124)	-
dpy-20(e1362)	-
par-5(it121)	-
let-92(s504)	+
let-653(s1733)	+
let-56(s173)	+
let-56(s2321)	+
let-52(s42)	+
let-93(s734)	-
let-656(s1767)	+
let-309(s1770)	-
let-97(s1121)	+
unc-31(e169)	-
unc-26(e345)	-
sDf65	+

§ A "+" indicates the ability of *sDp10* to rescue individuals homozygous for the mutation. A "-" indicates that *sDp10* could not rescue the mutant phenotype of individuals homozygous for the mutation. The order in which the mutations are listed in the Table corresponds to their order on the genetic map shown in Figure 10.

Figure 10:

A partial genetic map of the *let-60* - *unc-31* interval showing the extent of *sDp10*'s coverage of the fourth chromosome. The duplication's genetic breakpoints are indicated by the long horizontal lines. Markers used in complementation tests with *sDp10* are listed in Table 5, and are shown here. "**nd**" indicates markers that were not examined for complementation with *sDp10*. The asterisks indicate the two genes likely to be covered by *sDp10*, but which a single copy of *sDp10* was incapable of rescuing. The hatched bar between *let-97* and *unc-31* indicates the map in this area has been truncated to fit on the Figure. The extent of *sDf65*, which is also rescued by *sDp10*, is indicated on the Figure.


let-60
dpy-20
par-5
let-92
let-660 nd
let-653
let-56
unc-22
sDf65
let-52
0.25 map unit
let-93*
let-66 nd
sDp10(IV;X)
let-656*
let-309
let-68 nd
let-97
unc-31

DISCUSSION

In this section of the thesis, the recovery and analysis of two chromosomal duplications of *let-56* are discussed. Both of the duplications are attached to chromosomes. *sDp11* appears to be attached to chromosome *IV*; therefore, an intensive analysis of the extent of this duplication's size was not attempted. *sDp10* appears to be linked to the *X* chromosome. Complementation tests with lethal mutations in the vicinity of *unc-22* indicated that *sDp10* contains functional copies of several essential loci in addition to a functional copy of the *unc-22* locus. However, *sDp10* was not capable of rescuing either *let-93(s734)* or *let-656(s1767)* in hermaphrodites. This observation could be explained if the copies of these two genes carried by *sDp10* had been rendered inactive by the initial gamma ray mutagenesis. However, because the dose of gamma radiation chosen for treatment was 1500 r, as recommended by Rosenbluth, Cuddeford and Baillie (1985), this possibility seems unlikely. These authors concluded that with a dose of 1500 r, only 5% of induced rearrangements would be accompanied by an additional chromosomal break anywhere else in the genome. Inactivation of both *let-93* and *let-656* on *sDp10* would involve at least two more breaks.

Another explanation for the inability of *sDp10* to rescue the *let-93(s734)* and *let-656(s1767)* mutations can be provided if the difference in phenotype between *let-56 unc-22 lev-1 /let-56 unc-22 lev-1 ; sDp10* males and *let-56 unc-22 lev-1 / let-56 unc-22 lev-1 ; sDp10* hermaphrodites is considered. Males of this genotype did not exhibit a constitutive weak twitch, whereas hermaphrodites did. The sex specific difference in the *unc-22* phenotype can be explained if the *unc-22* gene carried on *sDp10* is subject to *X* chromosome dosage compensation. In *C. elegans*, *X* chromosome dosage compensation (Meneely and Wood, 1987) is achieved by equalizing the amounts of mRNA produced in the two sexes (Meyer and Casson, 1986). It is presumed that equalization of the amounts of mRNA is

the result of reduction of transcription from the two *X* chromosomes of the hermaphrodite (Delong et al, 1993). Perhaps, in hermaphrodites carrying sDp10, *X* chromosome dosage compensation has extended into *sDp10*. This might have resulted in insufficient amounts of wild-type *unc-22* product to completely mask the Unc-22 phenotype produced by the homozygous *unc-22(s7)* allele. A general down-regulation of transcription from *sDp10* due to "invading" dosage compensation from the *X* chromosome could affect the ability of *sDp10* to rescue mutations in genes that have strict dose requirements. Perhaps *let-93(s734)* and *let-656(s1767)* are examples of mutations of this sort.

Another *X*-linked autosomal duplication has been described by Rosenbluth et al.(1988). This duplication, *sDp30*, carries a number of wild-type chromosome *V* genes. Rosenbluth et al. (1988) obtained results similar to those I obtained with *sDp10*. That is, not all lethal mutations were rescued by the corresponding (presumably) wild-type gene carried on the duplication. Rosenbluth et al. (1988) proposed two alternative scenarios to account for this observation. In the first scenario, the variability of suppression of lethality exhibited by *sDp30* might be due to the variable sensitivity of the mutant phenotypes to reduced dosages of wild type gene product. Alternatively, perhaps the different autosomal genes carried on the duplication were variably affected by the "spreading" of dosage compensation into the duplication from the X chromosome. Either of these scenarios are consistent with the results I obtained with *sDp10*.

The experiments discussed in this section of the thesis show that the scheme I had devised to allow me to screen for chromosomal duplications of vital loci linked to *lev-1(x22)* is capable of recovering such duplications. These duplications are presumably derived from the *nT1(IV;V)* rearrangement. The method is powerful; it was possible to easily identify adult mobile hermaphrodites in a background of approximately 24000 hypercontracted and paralyzed or

developmentally retarded individuals. The scheme should, in theory, be adaptable to accomodate screens for duplications of lethal mutations that map anywhere within the *nT1(IV;V)* balanced region. Due to the pseudo-linkage imposed by *nT1(IV;V)*, this includes not only the right half of chromosome *IV*, but also the left half of chromosome *V* (Ferguson and Horvitz 1985; Clark et al. 1988). Lethal mutations in the latter region would appear pseudo-linked to *lev-1(x22)(IV)*.

Section III.

Identification and sequence analysis of genes in the *let-653* - *unc-22* interval.

PART A.

Identification and DNA sequence analysis of genes on the cosmid C11F2.

INTRODUCTION

In Section I, essential genes in the *let-65* - *let-93* interval were identified and analyzed with respect to complementation behaviour and terminal phenotypes. These experiments provided *in vivo* information on the function of a number of genes in this region of chromosome IV.

In this Section of the thesis, I discuss the identification and characterization of genes in the *let-653* - *unc-22* region that had not been identified by mutation. These genes might present small targets for EMS, which was the mutagen most commonly employed in this region to induce point mutations. Alternatively, these genes might not be required for the normal growth and reproduction of the nematode under laboratory conditions. In the latter case, knowledge of the potential function of this type of gene might indicate the culture conditions required to observe a phenotype when the locus is disrupted by mutation.

One way to obtain information about the potential function of a gene is to determine its DNA sequence. Currently, approximately one in three *C. elegans* genes have informative matches to other sequences present in the public domain databases (Wilson *et al.*, 1994). Therefore, systematic DNA sequencing is a reasonably effective method of identifying and characterizing multiple genes in analyses of chromosomal regions. This is especially true in *C. elegans* due to the high density of genes in the genome. The *C. elegans* genome project has shown that, in the gene cluster on chromosome III, a gene is found, on average, every 5 kb. This number is likely an underestimate of the true density of genes in this region because the computer program "GENEFINDER" (P. Green and L. Hillier, personal communication), currently used by the *C. elegans* genome project to predict genes from genomic DNA sequence, only recognizes potential protein coding regions and will not detect other novel features present in the sequence.

This Section focuses on the identification and characterization of genes located in a 0.2 map unit (m.u.) region between *let-653* and *unc-22*. This region contained the left-hand breakpoints of the deficiencies *sDf9*, *sDf19* and *sDf65*, and contained only one genetically defined locus, the essential gene *let-56* (Figure 10). This locus is deleted by all three of these deficiencies (Clark and Baillie, 1992). The principal method employed in my study was DNA sequence determination and analysis of cDNA clones.

The principal reason for choosing this particular interval for study was the degree of completeness of the genome map. In addition to being intensively analyzed at the genetic level by myself and my predecessors in the Baillie laboratory (D. G. Moerman, T. M. Rogalski and D. V. Clark), a considerable amount of investigation at the molecular level had been conducted, primarily by S. Prasad (Prasad, 1988; Prasad and Baillie, 1989), D. Clark (Clark and Baillie, 1992) and G. Benian and his colleagues (Benian et al., 1989; Benian, L'Hernault and Morris, 1993). In order to place the work described in this Section in context, these investigations are summarized in the remainder of this Introduction.

Nine genes on four different cosmids in the *dpy-20 -unc-22* interval had been identified by Prasad using interspecies hybridization to *C. briggsae*. On certain *Pst*I restriction fragments derived from the cosmid C18D3, Prasad obtained evidence for the existence of four genes, and noted that the rightmost end of this cosmid contained the 3' coding sequences of the *unc-22* gene (Prasad and Baillie, 1989; discussed in detail in Part B of this Section). On each of three *Pst*I restriction fragments obtained from the adjacent cosmid, C11F2, Prasad found evidence for additional genes. He subcloned two of these restriction fragments, generating pCes122 (containing *Pst*I fragment "R") and pCes124 (containing *Pst*I fragment "Q"). Prasad found that fragment R hybridized to three differently sized

bands on Northern blots of total RNA, while fragment Q hybridized to a single 2.4 kb band on similar blots. These data are summarized in Figure 11.

Prasad recovered a partial-length 1.3 kb cDNA (called "C2"), using fragment Q as a probe in hybridization screens of a cDNA library, and subsequently localized the C2 cDNA to the region deleted by the deficiency *sDf9* by Southern blot (Prasad and Baillie, 1989). Prasad (1988) determined the nucleotide sequence of this cDNA, and showed the predicted amino acid sequence exhibited identity to the central third of a human Na^+/H^+ antiporter molecule. However, no clue as to the possible identity of the gene(s) located on fragment R was obtained. Subsequently, during her efforts to correlate the physical and genetic maps in the *dpy-20* - *unc-22* interval, Clark demonstrated that the cosmid C11F2 could rescue *let-56* mutant individuals (Clark and Baillie, 1992). In addition, she demonstrated that the cosmid C29E6 could rescue the *let-653* mutant phenotype (Clark and Baillie, 1992). The data of Clark helped to delimit the region of the physical map pertaining to the *let-653* - *unc-22* region, and are summarized in Figure 12.

DNA sequence determination of the *unc-22* gene (Benian et al., 1989; Benian, L'Hernault and Morris, 1993) and the adjacent regions generated, in addition to the complete sequence of *unc-22*, approximately 13 kb of DNA sequence downstream of the 3' end of *unc-22*. This sequence represented approximately one-third of the worm DNA contained within the cosmid C18D3. An open reading frame had been predicted from this sequence (D. L. Baillie and C. Fields, personal communication), and used to search nucleotide and amino acid databases. These database searches revealed the open reading frame displayed weak similarity to a glucose transporter-like protein. In addition to this gene, for which an informative match was found in the databases, there were other open reading frames for which no informative matches had been obtained. No

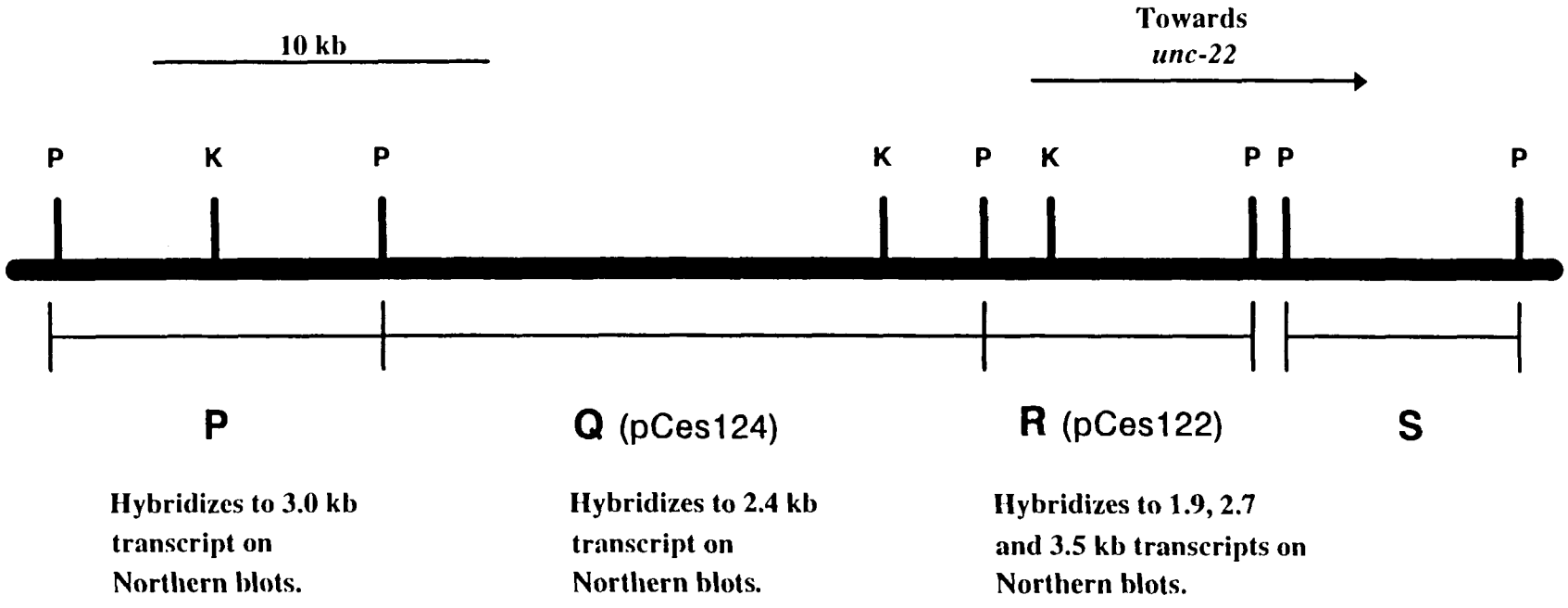


Figure 11:

A restriction map of the cosmid C11F2, adapted from Prasad (1988) and Prasad and Baillie (1989). K = *Kpn*I. P = *Pst*I. The direction in which *unc-22* lies is indicated by the arrow. The terminal *Pst*I sites are located within the cosmid vector. The *Pst*I restriction fragments used by Prasad (1988) to probe Northern blots and the results of these hybridizations are indicated. The gene content of fragment S had not been examined.

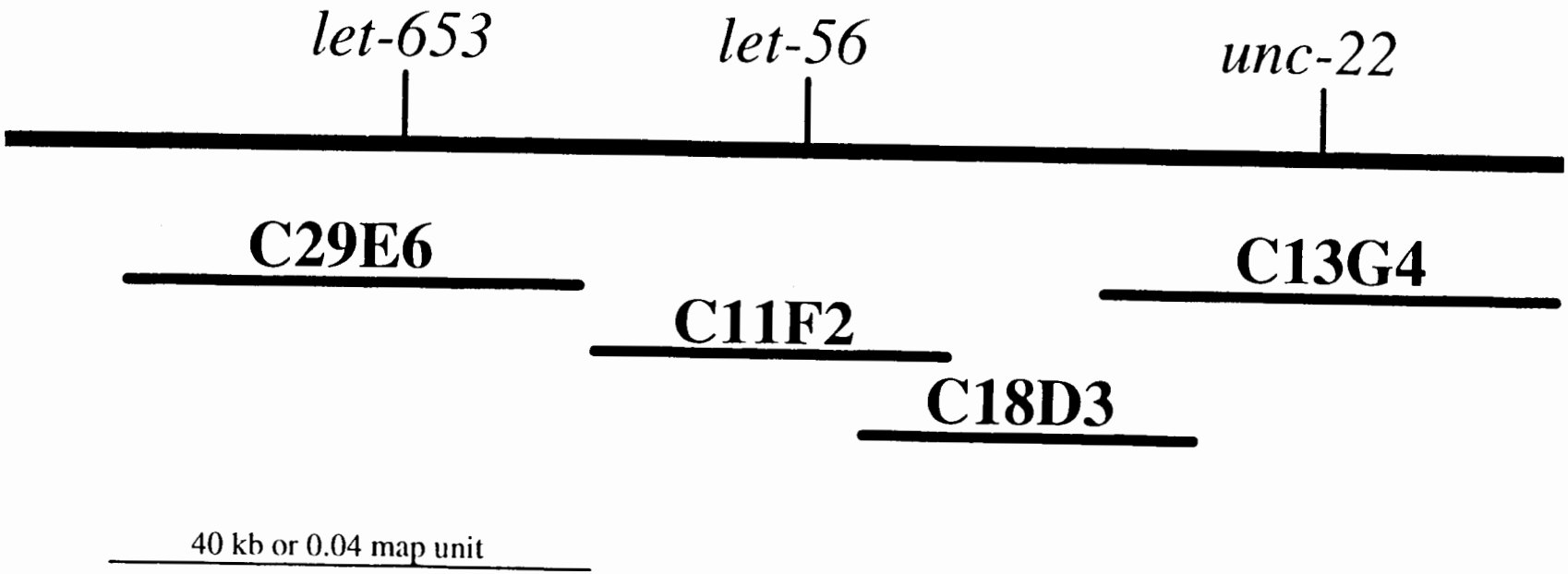


Figure 12:

A summary of a portion of the correlation of the physical and the genetic maps in the *unc-22* vicinity achieved by Clark (1990), adapted from Clark (1990) and Clark and Baillie (1992), and **ACeDB** (Version 2.9). In addition to the cosmids used by Clark, I have indicated the position of the cosmid C18D3, which was analyzed by Prasad (1988), and will be the subject of Part B of this Section.

mutations were known to correspond to any of these predicted open reading frames.

Here, I describe efforts to obtain additional sequence data for the Na^+/H^+ antiporter-like molecule originally identified by Prasad (1988). Second, I describe the DNA sequence determination and analysis of a cDNA that was recovered using fragment R as a probe to screen a cDNA library. Third, I discuss the identification of genes within fragment S. Last, I discuss work conducted in collaboration with J. Schein in which the DNA sequence determined by Benian *et al.*, (1989) is analyzed. This analysis has included prediction of open reading frames and recovery of deletion mutations affecting all known genes between (and including) *let-56* and *unc-22*.

A paper describing the cloning and sequencing of *nhe-1* and *hdl-1* has been published (M. A. Marra, S. S. Prasad and D. L. Baillie, **Mol Gen Genet** (1993) **236**: 289 - 298.

RESULTS

Orienting fragment R with respect to fragment Q on C11F2: An 8kb *Pst*I restriction fragment (fragment R) obtained from C11F2 (Figure 11) had previously been cloned into the *Pst*I site of pUC19 to generate the plasmid pCes122 (Prasad and Baillie, 1989). An 18 kb *Pst*I restriction fragment (fragment Q), also derived from C11F2 and adjacent to fragment R, was similarly cloned into pUC19 (Prasad and Baillie, 1989) generating pCes124. In order to draw inferences concerning the direction of open reading frames contained within these two restriction fragments, it was necessary to relate the orientation of the clones contained within pCes122 and pCes124 to the cosmid from which they were derived. This was accomplished as follows. The ends of fragment Q and fragment R were sequenced using commercially available M13 reverse and M13 -40 primers (Materials and Methods). Oligonucleotide primers were designed from the complement of each sequence using the program OLIGO (Rychlik and Rhoads 1989). A primer predicted from each end of fragment Q was used in conjunction with a primer predicted from each end of fragment R in PCR experiments using approximately 1ng of C11F2 DNA as the amplification template. Only the primers QFOR1 and RF1 (Table 2), designed from the complements of the sequences obtained using the M13 -40 primers, produced an amplification product. This served to unambiguously order the inserts contained within pCes124 and pCes122 with respect to each other as they are arranged on C11F2.

Recovery of Na^+/H^+ antiporter-like cDNAs: Fragment Q had detected a 2.4 kb RNA when used to probe Northern blots of *C. elegans* total RNA, and was subsequently used as a probe in screens of a λgt10 cDNA library (provided by Barbara Meyer) that resulted in the recovery of a 1.3 kb cDNA (Prasad and Baillie, 1989). The sequence of this cDNA (called "C2") was determined (Prasad, 1988), and the longest predicted open reading frame was found to exhibit similarity to a

human Na^+/H^+ antiporter. The extent of the similarity between the two sequences was confined to a 300 amino acid region representing the central one-third of the predicted human protein. Because it was of interest to investigate further the potential identity of this gene, and because analysis of the *C. elegans* sequence had indicated that the 1.3 kb cDNA was of incomplete length (Prasad, 1988), I determined additional sequence for the antiporter-like gene. I obtained additional Na^+/H^+ antiporter-like cDNA clones by screening approximately 8 x 10^5 phage from the Barstead and Waterston λZAP library (Materials and Methods) using the C2 cDNA as a probe. A set of twelve cDNAs was obtained. Restriction analysis of these cDNAs showed the set contained representatives of three different size classes of 2.2 kb, 1.9kb and 1.3 kb. Five cDNAs that included all of the different size classes were selected for DNA sequence analysis. Sequence determination and analysis of the ends of the smaller cDNAs revealed they appeared to contain only sequence that was identical to the C2 cDNA sequence. Sequence analysis of the ends of a 1.9 kb cDNA revealed that it encoded additional, previously undetermined, worm Na^+/H^+ antiporter-similar amino acid sequence as well as containing the previously determined C2 cDNA sequence. This cDNA, designated pCes801, was retained for additional analysis. Sequence determination of the ends of the largest cDNA revealed that it contained, in its 5' region, novel sequence that appeared unrelated to the human antiporter. This cDNA, designated pCes800, was also retained for additional analysis.

pCes800 was found to contain the previously determined C2 cDNA sequence and approximately 1000 bp of novel sequence. pCes801 contained the previously determined C2 sequence, and approximately 600 bp of sequence that was also contained within pCes800. I determined the novel sequence contained within pCes800 and pCes801 using a combination of commercially available and custom designed oligonucleotide primers. Custom oligonucleotide primers that

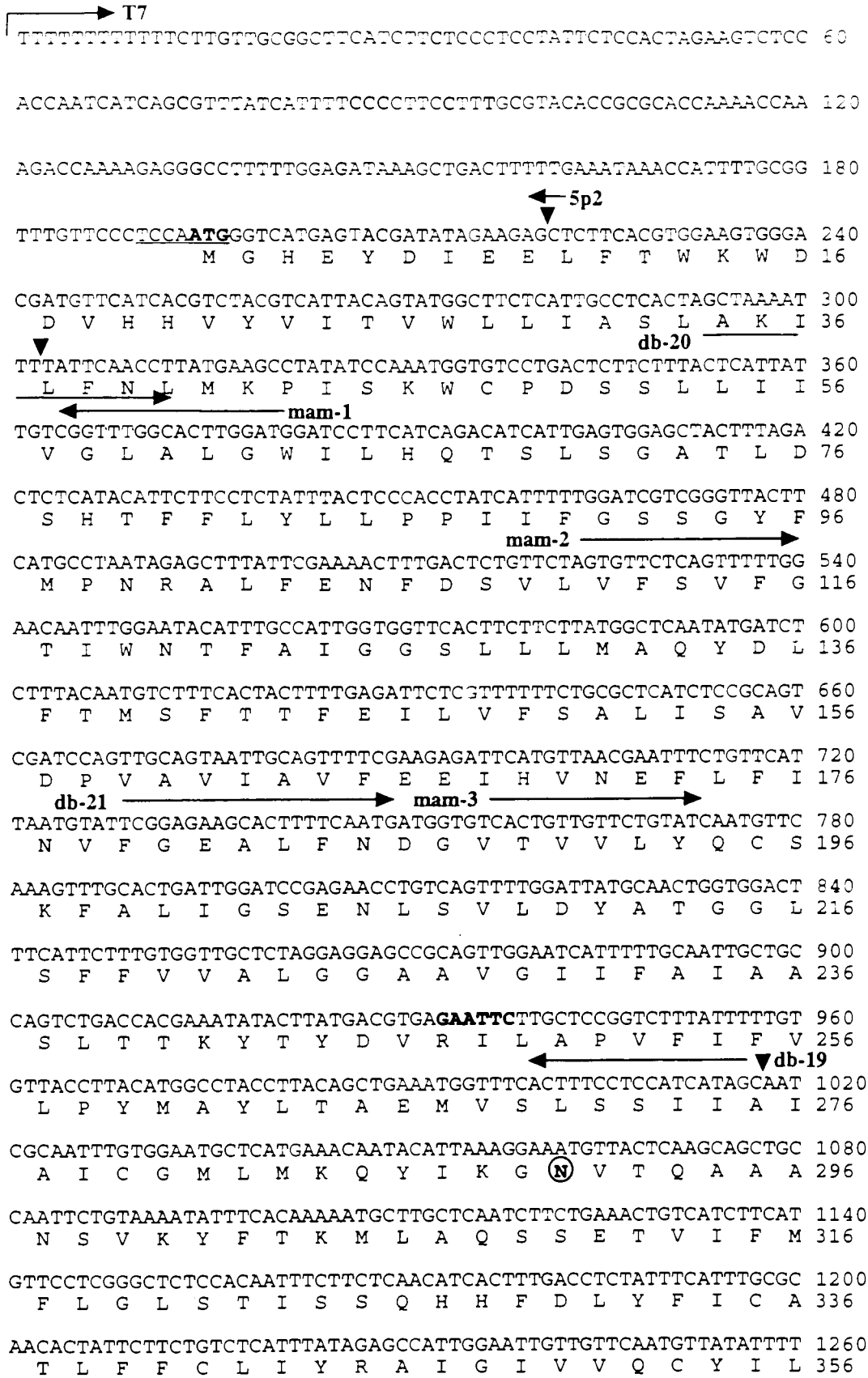
were used are listed in Table 2, and are indicated on the DNA sequence of the antiporter-like gene (Figure 13).

Analysis of the *C. elegans* Na^+/H^+ antiporter-like open reading frame: The sequence obtained from C2, pCes800 and pCes801, when combined, resulted in a "composite" cDNA of 2227 nucleotides. This cDNA possessed a long open reading frame having the potential to encode a protein of 646 amino acids. A TGA stop codon and a single consensus poly-(A) addition site (AATAAA) commence at nucleotide positions 2132 and 2202, respectively (Figure 13). A poly-(A) tail commences at position 2220. A potential initiation site for translation that weakly resembles the consensus for *C. elegans* genes (A/c - A/g - a/c - A/C - ATG; M.D. Perry, G.Z. Hertz and W.B. Wood, personal communication) is found near the 5' terminus of the cDNA sequence immediately upstream of the putative initiating methionine codon commencing at nucleotide position 194. At position 166, in the same frame as the potential initiating methionine codon, is a stop codon. No other uninterrupted open reading frame or potential initiating methionine codon is found upstream of the one at position 194. This observation, coupled with the presence of the consensus site for translation initiation, is compatible with the notion that the composite cDNA contains all of the coding sequence for the worm antiporter-like protein.

No exhaustive attempt to determine the complete genomic sequence corresponding to the worm Na^+/H^+ antiporter-like gene was made. As mentioned, Prasad (1988) had determined the complete genomic sequence corresponding to the C2 cDNA. He found four introns in this portion of the gene. I have located two additional introns by comparing cDNA sequence I obtained from pCes800 to sequence I determined from pCes846, a subclone of pCes124 (Figure 17). This sequence was determined using custom designed oligonucleotide primers. I

Figure 13:

The cDNA sequence of the *C. elegans* Na^+/H^+ antiporter - like gene, *nhe-1*. The predicted amino acid sequence is shown beneath the nucleotide sequence of the *nhe-1* composite cDNA. The potential initiating methionine residue is highlighted in bold-face type. The underline indicates the weak match to the *C. elegans* consensus for translation initiation. The positions and orientations of the various custom designed oligonucleotide primers used in determination of the *nhe-1* sequence are indicated above the nucleotides (see also Table 2). The "T7" notation following the arrowhead at the beginning of the nucleotide sequence indicates the beginning of the cDNA sequence, determined using the T7 primer. The positions of known introns are indicated with inverted arrowheads. The oligonucleotide primer **5P2** was designed from intronic sequence. The sequence determined in this thesis corresponds to the 5' portion of the *nhe-1* cDNA, from nucleotide position 1 to nucleotide position 932, where an *Eco*RI site (indicated in bold-face type) is found. The remainder of the sequence was previously determined by Prasad (1988). A potential site for N-linked glycosylation at amino acid position 290 found to be conserved in all Na^+/H^+ antiporters is circled, and a potential phosphorylation site for protein kinase C, found at amino acid position 627, is indicated by a box and bold-face type. A TGA stop codon and potential polyadenylation signal are also indicated in boxes. The beginning of the poly-(A) tail is indicated in bold-faced type.


T7
5p2
db-20
mam-1
mam-2
db-21
mam-3
db-19

```
GAACCGATTCCGTGCCAAAAAGTTCGAAATGGTTGATCAATTCATCATGTCATATGGAGG 1320
  N  R  F  R  A  K  K  F  E  M  V  D  Q  F  I  M  S  Y  G  G 376

TCTTCGTGGAGCCATTGCCTATGGTCTTGTCGTCTCAATTCCAGCTTCAATTACTCGTAA 1380
  L  R  G  A  I  A  Y  G  L  V  V  S  I  P  A  S  I  T  R  K 396

ACCAATGTTTATCACTGCGACAATTGCATGGATCTACTTCACTGTATTCCTTCAAGGAAT 1440
  P  M  F  I  T  A  T  I  A  W  I  Y  F  T  V  F  L  Q  G  I 416

CACAATTCGACCATTGGTCAACTTTTTGAAGATTAAGAAGAAGGAGGAAAGAGATCCAAC 1500
  T  I  R  P  L  V  N  F  L  K  I  K  K  K  E  E  R  D  P  T 436

GATGGTTGAAAGTGTTTACAATAAATATTTGGATTACATGATGTCTGGAGTGGAAGATAT 1560
  M  V  E  S  V  Y  N  K  Y  L  D  Y  M  M  S  G  V  E  D  I 456

TGCTGGACAGAAGGGGCATTACACTTTCATTGAGAATTTCGAGAGATTCAATGCAAAAGT 1620
  A  G  Q  K  G  H  Y  T  F  I  E  N  F  E  R  F  N  A  K  V 476

AATAAAACCAGTATTGATGAGACACCAGAAAAGAGAAAGTTTCGATGCTTCATCGATTGT 1680
  I  K  P  V  L  M  R  H  Q  K  R  E  S  F  D  A  S  S  I  V 496

TCGTGCTTACGAGAAAATCACATTGGAAGATGCCATCAAACTTGCCAAAGTCAAGAATAA 1740
  R  A  Y  E  K  I  T  L  E  D  A  I  K  L  A  K  V  K  N  N 516

TATTCAAAATAAGCGTCTCGAACGAATTAAGAGCAAAGGTAGAGTTGCACCAATTCTTCC 1800
  I  Q  N  K  R  L  E  R  I  K  S  K  G  R  V  A  P  I  L  P 536

CGATAAAATATCCAATCAAAAGACGATGACACCGAAGGATCTTCAATTGAAGAGGTTTAT 1860
  D  K  I  S  N  Q  K  T  M  T  P  K  D  L  Q  L  K  R  F  M 556

GGAATCTGGTGAAAACATTGATTCTCTGTACACGCTCTTCAGTGATCTGCTTGATAGAAA 1920
  E  S  G  E  N  I  D  S  L  Y  T  L  F  S  D  L  L  D  R  K 576

GTTACACGAAATGAATAGACCATCAGTTCAAATTACGGACGTTGATGGACAGGATGATAT 1980
  L  H  E  M  N  R  P  S  V  Q  I  T  D  V  D  G  Q  D  D  I 596

TCAAGACGATTACATGGCTGAAGTGTCACGATCGAACCTCTCAGCAATGTTCCGAAGTAC 2040
  Q  D  D  Y  M  A  E  V  S  R  S  N  L  S  A  M  F  R  S  T 616

GGAACAACTGCCATCAGAAACGCCATTTCATAGTGGTAGAAGACAATCGACAGGAGATTT 2100
  E  Q  L  P  S  E  T  P  F  H  S  G  R  R  Q  S  T  G  D  L 636

AAATGCAACACGAAGAGCTGATTTCAATGTTTGACATTCAATTGAAACCTCTTCGATATG 2160
  N  A  T  R  R  A  D  F  N  V  *                             646

TTTTTGTAGTAATTCTCGCCATAACGAAAAAAAACTGTAATAATAAATTTATCGCTTTTA 2220
```

determined the complete sequence of one of these introns, and found it to be 297 bp in length. The locations of the introns determined by myself or Prasad (1988) are indicated in Figure 13.

The sequence of the worm antiporter-like cDNA was used in a **blastx** search of the "non-redundant" database as described in Materials and Methods. This search revealed strong similarities to a number of Na^+/H^+ antiporters, including isoforms from rat, an isoform from trout, and an isoform from human. These results are summarized in Table 6.

The observation that cDNAs representing multiple Na^+/H^+ exchanger isoforms had been identified in a number of systems including human, rat, and *E. coli* prompted me to ask whether there were likely to be additional *C. elegans* Na^+/H^+ exchanger isoforms. This was accomplished by low stringency hybridization (see Materials and Methods) of labeled C2 cDNA to a Southern blot containing *Eco*RI digested *C. elegans* genomic DNA. In addition to the hybridization of the C2 cDNA to its cognate genomic sequence, additional hybridization to a larger *Eco*RI restriction fragment was seen (Figure 14), providing evidence that related DNA sequences exist elsewhere in the genome of the worm.

The antiporter amino acid sequences most similar to the worm sequence were obtained from the database at the NCBI using the RETRIEVE e-mail server, and portions of these were aligned with the *C. elegans* amino acid sequence. The initial alignments were performed with CLUSTAL V (Higgins and Sharp, 1989; Higgins, Bleasby and Fuchs, 1992), and inspected manually to ensure the alignments resulted in maximal similarity to the *C. elegans* sequence. Gaps were inserted in the sequence by CLUSTAL V or by myself to maximize sequence similarity. An alignment of portions of the *C. elegans* amino acid sequence with portions of the rat NHE-3 Na^+/H^+ exchanger isoform, an isoform from trout red

TABLE 6.

Summary of *C. elegans* Na^+/H^+ antiporter similarity scores.

	Locus I.D.†	P(N)‡	Similarity (over 520 residues) Number of a.a.	Similarity (over 520 residues) percent similar	Identity (over 520 residues) Number of a.a.	Identity (over 520 residues) percent identical
Rat	sp | P26433	6.5×10^{-104}	393	75.6	230	44.2
Trout	pir | A46188	4.3×10^{-93}	407	78.3	212	40.8
Human	sp | P19634	6.5×10^{-64}	412	79.2	213	41.0

† Indicates the database and the accession number of the particular sequence matching the *C. elegans nhe-1* predicted protein. **sp** is the SWISS-PROT database and **pir** is the NBRF PIR database.

‡ Indicates the Poisson P-values reported by the **blast** algorithm for the group of high scoring segment pairs (HSPs) for each database search. This value is the probability that at least as many HSPs would occur by chance alone, each with a **blast** score at least as high as the lowest scoring member of the group of HSPs. Thus, these probabilities provide a measure of statistical significance.

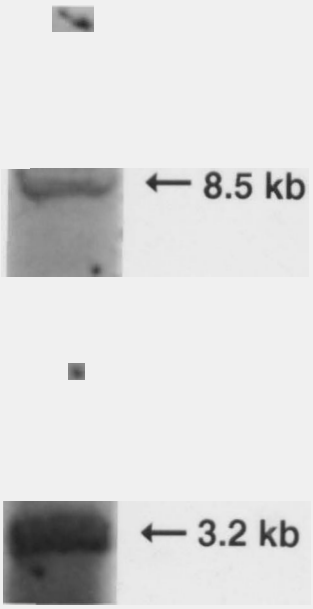


Figure 14:

A chemi-lumigraph of a Southern blot containing *Eco*RI digested genomic DNA, hybridized under low stringency conditions (Materials and Methods) to a probe constructed from the C2 cDNA recovered by Prasad (1988). The intense band at 3.2 kb corresponds to hybridization to the *nhe-1* gene. The less intense signal at 8.5 kb corresponds to related sequences, possibly an additional Na^+/H^+ exchanger isoform, elsewhere in the genome.

cells (βNHE), and the NHE-1 isoform from human is presented in Figure 15. Residues identical to amino acids in the *C. elegans* sequence are indicated, as are conservative amino acid substitutions. I considered an amino acid substitution to be conservative if the substitution received a score greater than zero in the PAM250 amino acid replacement scoring matrix of Dayhoff et al. (1978).

From this analysis, it is apparent that the *C. elegans* predicted protein shares a high degree of similarity to the other antiporters with which it is aligned. The region of similarity extends from *C. elegans* amino acid number 20 to *C. elegans* amino acid number 539. The overall sequence similarity to the antiporters retrieved from the databank is presented in Table 6. Based on the high degree of similarity, I have named the *C. elegans* antiporter-like gene *nhe-1*. The *nhe-* prefix stands for **N**a$^+$/**H**$^+$ **e**xchanger-like.

Previous studies have noted that the hydrophobicity profiles of the antiporter proteins included in Table 6 are consistent with the hypothesis that these molecules possess multiple membrane spanning domains (Sardet *et. al.*, 1989; Borgese *et al.*, 1992; Tse *et al.*, 1992; Counillon and Pouyssegur, 1993). Hydrophobicity plots of the *C. elegans* sequence using the method of Kyte and Doolittle (1982) also predict the existence of regions of high hydrophobicity. A Kyte and Doolittle (1982) hydrophobicity plot of the *C. elegans* amino acid sequence, and one of the human amino acid sequence is shown in Figure 16. It is immediately obvious that the majority of the *C. elegans* and human protein exhibit striking similarity of secondary structure when analyzed in this fashion. The presence and location of putative membrane spanning domains were predicted in both sequences using the method of Klein, Kanehisa and DeLisi (1985).

Figure 15:

A multiple sequence alignment of the *C. elegans* Nhe-1 predicted protein with the amino acid sequences most closely resembling it. The accession numbers of these sequences are noted in Table 6. The sequences have been aligned to provide the best match to the *C. elegans* sequence. Residues identical to a *C. elegans* residue are in bold-face capitals. Residues similar to a *C. elegans* residue are capitalized. Non-conservative substitutions are in lower case. Regions of similarity to the *C. elegans* sequence are indicated by shading. The arrow indicates the conserved potential N-linked glycosylation site. The presence and location of putative membrane spanning domains are indicated in the *C. elegans* and human exchanger amino acid sequences by underlines.

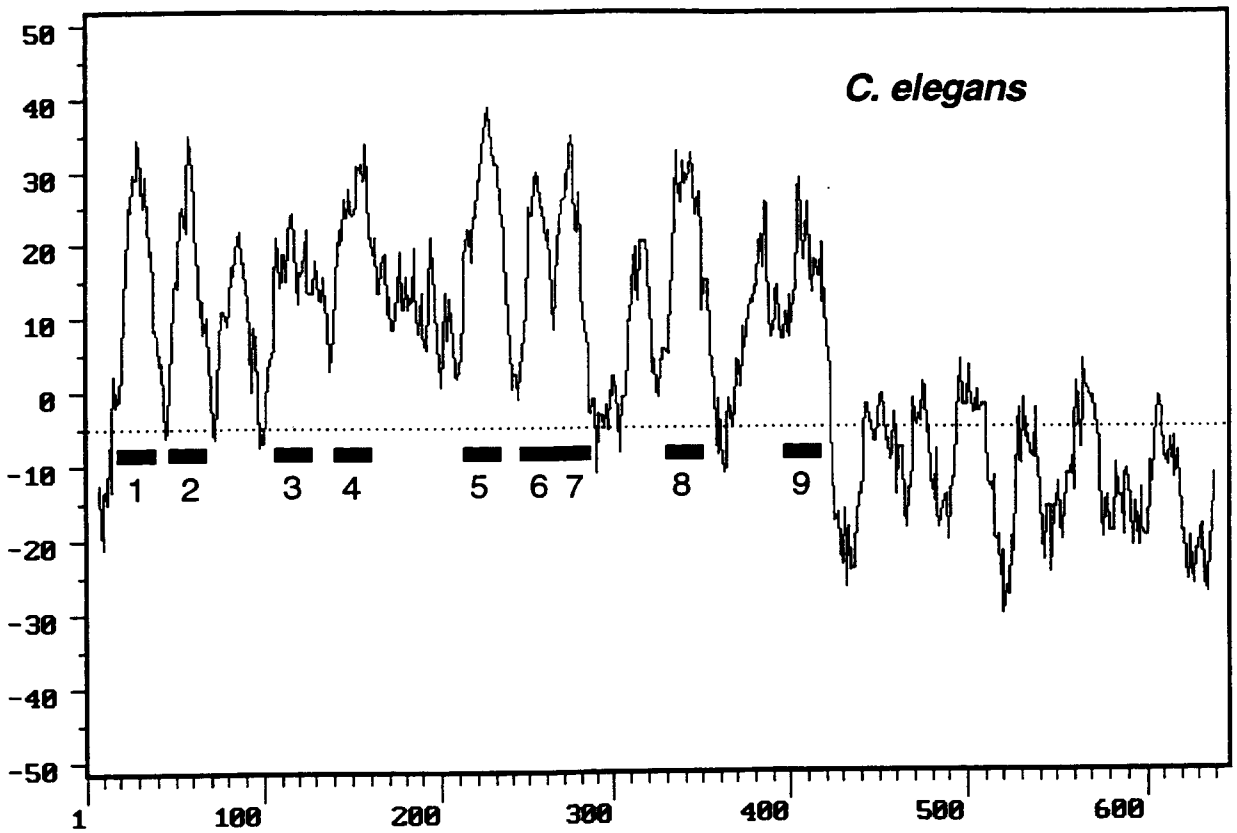
Figure 15:

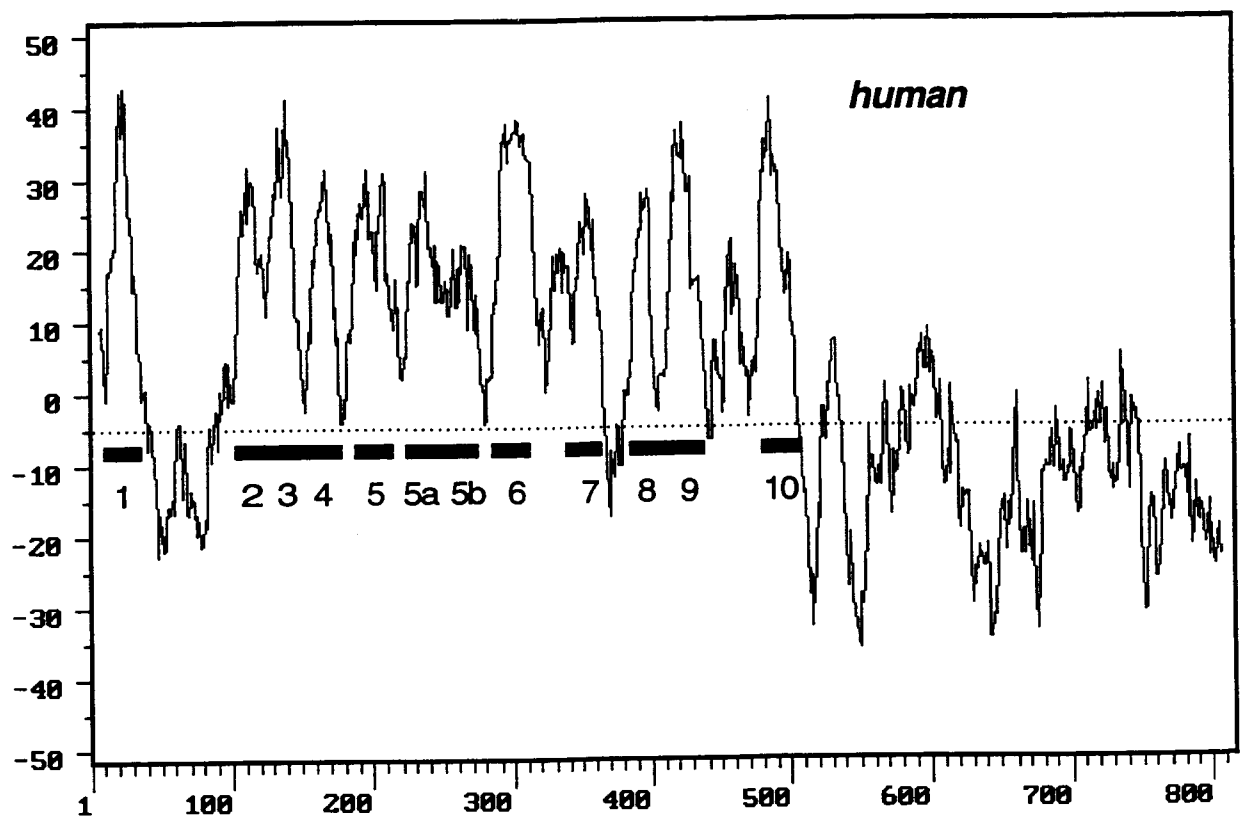
A multiple sequence alignment of the *C. elegans* Nhe-1 predicted protein with the amino acid sequences most closely resembling it. The accession numbers of these sequences are noted in Table 6. The sequences have been aligned to provide the best match to the *C. elegans* sequence. Residues identical to a *C. elegans* residue are in bold-face capitals. Residues similar to a *C. elegans* residue are capitalized. Non-conservative substitutions are in lower case. Regions of similarity to the *C. elegans* sequence are indicated by shading. The arrow indicates the conserved potential N-linked glycosylation site. The presence and location of putative membrane spanning domains are indicated in the *C. elegans* and human exchanger amino acid sequences by underlines.


C. elegans
rat
trout
human

Figure 16:

Kyte - Doolittle hydrophobicity plots of the *C. elegans* Nhe-1 protein and the human NHE-1 protein, calculated with a window of 15 amino acids. Numbers along the X axes refer to length of the protein in amino acids. Values along the Y axes are hydrophobicity scores. The dotted line represents the cutoff between a value that is considered to be hydrophobic and a value that is considered to be hydrophilic. Values above the cutoff are hydrophobic. The heavy lines and the numbers beneath them indicate potential transmembrane domains. In the plot of the human antiporter sequence, the numbering of these domains corresponds to the convention established by Borgese et al. (1992). The numbering of the putative membrane spanning domains in the *C. elegans* protein reflects the order in which the domains are found, starting at the amino terminus, and is not meant to indicate the corresponding domain in the human exchanger. The hydrophilic domains, commencing at amino acid 430 in the *C. elegans* protein and amino acid 510 in the human protein, are predicted to be intracellular, and contain potential phosphorylation sites.


C. elegans
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
100
200
300
400
500
600
1 2 3 4 5 6 7 8 9


human
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
100
200
300
400
500
600
700
800
1 2 3 4 5 5a 5b 6 7 8 9 10

The numbering of the potential transmembrane domains in the human exchanger has been preserved from that previously reported (Borgese et al., 1992). The numbers of the putative transmembrane domains in the *C. elegans* protein do not necessarily indicate an equivalent domain in the human protein.

The locations of the putative membrane spanning domains predicted in both the *C. elegans* and human sequences are indicated in the sequence alignment (Figure 15). In addition, a conserved potential site for asparagine-linked glycosylation has been indicated.

The carboxy-terminal portion of the human Na^+/H^+ antiporter is hydrophilic, and has been predicted to form a large, positively charged cytoplasmic domain containing several potential PKA phosphorylation sites (Sardet et al., 1989). The carboxy-terminal portion of the *C. elegans* protein has a similar hydrophilic domain (Figure 16) and a protein kinase C consensus phosphorylation site (Figure 13).

Determining the direction of the *C. elegans* antiporter-like open reading frame: In order to better understand the organization of the *C. elegans* antiporter-like gene in relation to neighbouring genes in this region of the genome, I determined the direction of the antiporter-like open reading frame. A 600 bp *Eco*RI fragment representing the 5' most portion of pCes801 (called 7M-5') was found to hybridize to an 11 kb *Xba*I restriction fragment on a Southern blot containing *Xba*I digested C11F2 DNA. This 11 kb *Xba*I fragment was gel isolated and subcloned from pCes124 DNA (Materials and Methods), and is contained within the plasmid pCes846. A detailed restriction map of pCes124 is shown in Figure 17. The ends of pCes846 were sequenced with commercially available T3 and T7 primers. The sequence obtained using T3 primer overlapped with sequence obtained from pCes124 using M13 reverse primer. The complement of

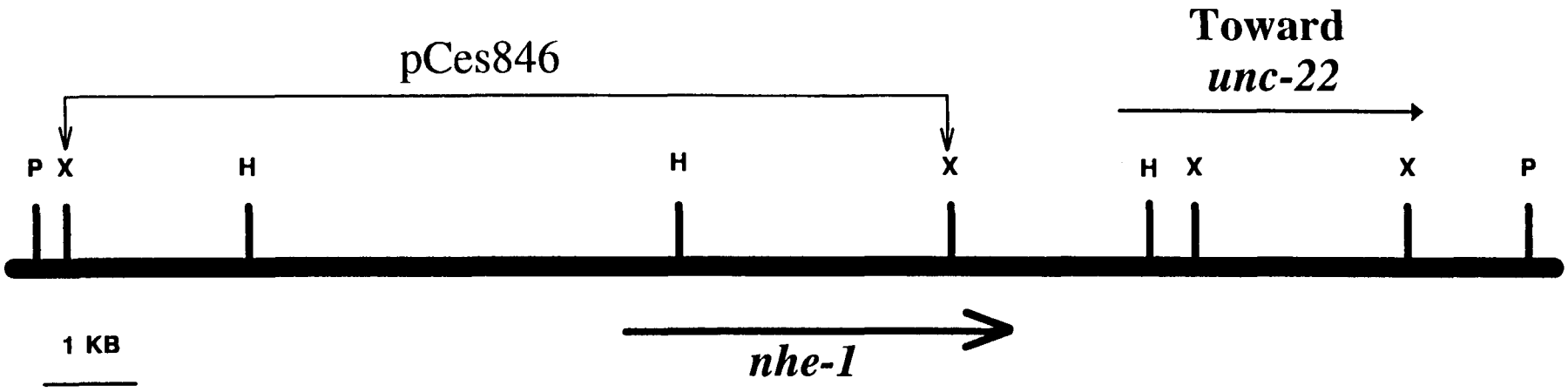


Figure 17:

A restriction map of *Pst*I fragment Q, the insert contained within the plasmid pCes124. P = *Pst*I. H = *Hin*dIII. X = *Xba*I. The direction in which *unc-22* lies is indicated on the map. The extent of the 11 kb *Xba*I fragment contained within pCes846 is also indicated. The direction of the *nhe-1* open reading frame is indicated by the arrow beneath the map.

the pCes846 sequence obtained with T7 primer was found to overlap with *C. elegans* Na^+/H^+ antiporter-like cDNA sequence. This information, along with the orientation of pCes124 with respect to pCes122, shows that the direction of the *C. elegans nhe-1* open reading frame is from left to right on the genetic map (Figure 17). This is opposite to the direction of *unc-22* transcription.

Deficiency mapping the *nhe-1* gene: As mentioned previously, there were a number of chromosomal deficiencies that defined the *let-653 - unc-22* interval. These deficiencies included *sDf9*, *sDf19*, and *sDf65*. Prasad had already demonstrated, by Southern blot hybridization followed by densitometry analysis, that the C2 cDNA hybridized to a region of the chromosome deleted by the deficiency *sDf9* (Prasad and Baillie, 1989). I examined whether the deficiency *sDf19* deleted the DNA containing *nhe-1* using PCR methodology developed by Barstead and Waterston (see Materials and Methods). In this method, the oligonucleotide primers DB-19 and MAM-2 (indicated on the *nhe-1* cDNA sequence in Figure 13) were used in PCR amplifications involving embryos homozygous for *sDf19*. Control primers specific for sequences on chromosome V were included in all PCR amplifications (Table 3). PCR products of the expected size were not seen in reactions using *sDf19* homozygous embryos as the amplification template (Figure 18). A similar analysis yielding the same result was conducted using individuals homozygous for the deficiency *sDf65*. These data are compatible with the notion that both *sDf19* and *sDf65* delete or affect at least the db-19 primer annealing site, and therefore delete or disrupt *nhe-1*.

Analysis of a cDNA recovered using fragment R as a probe: Fragment R had detected three RNAs of 1.9, 2.7, and 3.5 kb in length when used to probe Northern blots of total RNA (Prasad and Baillie 1989; summarized in Figure 11). In order to obtain information on the potential identity of these (or other)

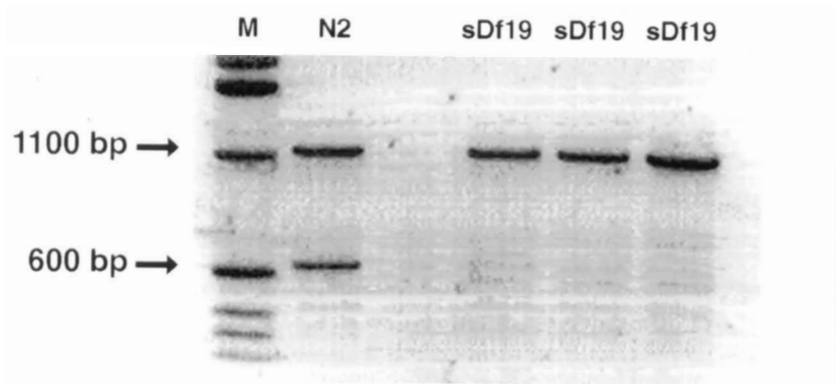


Figure 18:

A photograph of an ethidium bromide-stained agarose gel showing the results of PCR on *sDf19* homozygotes. M indicates the marker lane (1 kb ladder, BRL). The remaining lanes contain amplification products produced in PCRs using the primers DB-19 and MAM-2 (see Figure 13) and control primers specific for chromosome V (265-F and 265-R, Table 2) which produce the 1.1 kb band. None of the *sDf19* lanes contain the 600 bp DB-19 - MAM-2 amplification product.

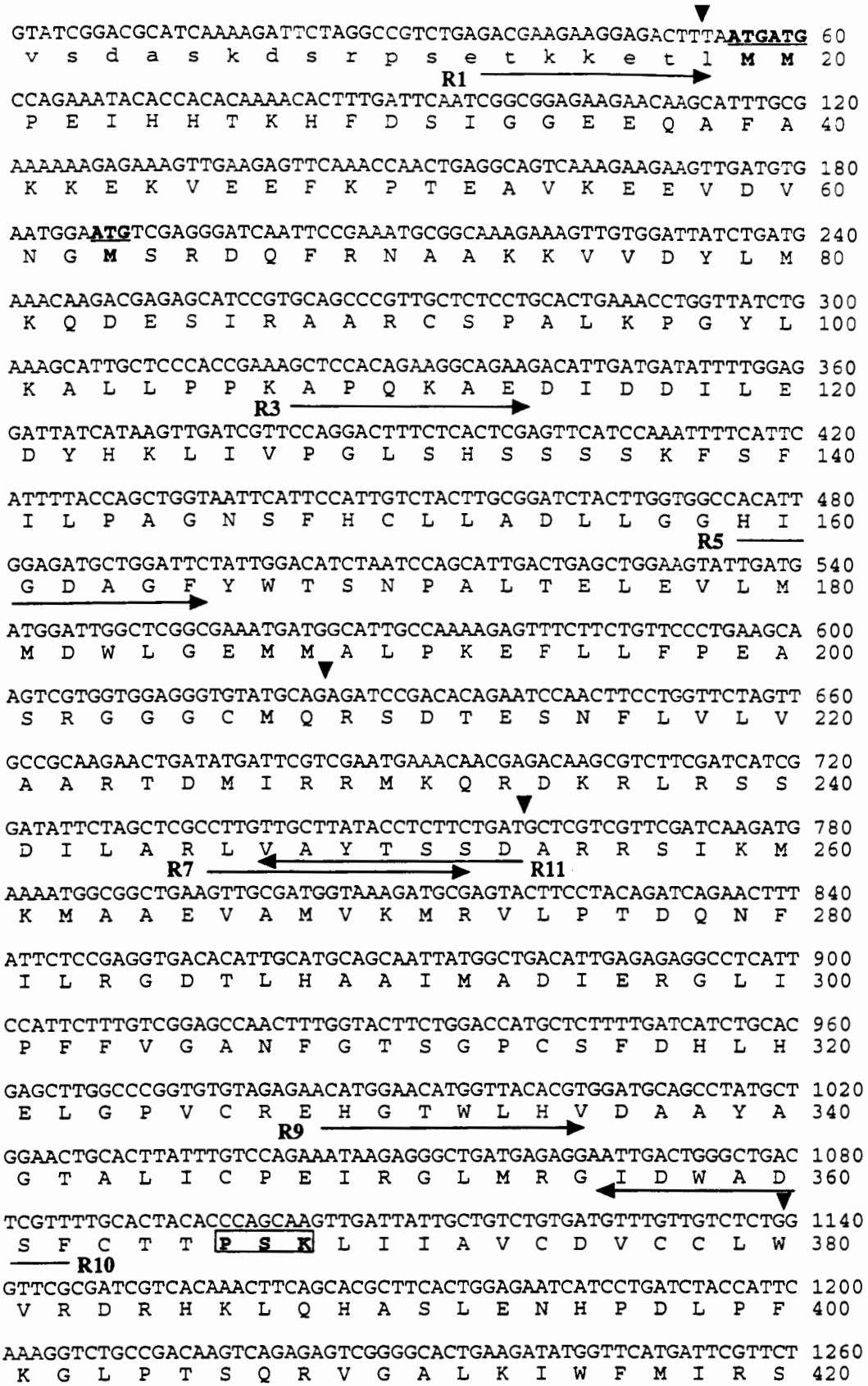
transcripts from genes located on fragment R, I used fragment R to probe the Barstead and Waterston cDNA library (1989) and identified a single positive plaque out of 3.0 X 10^5 screened. This cDNA clone, designated pCes814, was recovered and subjected to restriction enzyme and sequence analysis. The sequence of pCes814 was determined using both commercially available and custom designed oligonucleotide primers (listed in Table 2). The cDNA is 1920 bp in length and contains a single long open reading frame having the potential to encode a protein of 637 amino acids. A TGA stop codon and a consensus poly(A) addition site commence at positions 1876 and 1907, respectively, in the nucleotide sequence. Potential initiating methionines are found at positions 19, 20, and 63 of the amino acid sequence (Figure 19).

No exhaustive attempt to determine the complete genomic sequence corresponding to the decarboxylase-like gene was made. However, the custom designed primers that were used to sequence the cDNA were also used to obtain sequence from pCes122. Comparison of the sequence determined from pCes122 with the sequence of the cDNA has resulted in the identification of at least seven introns in the decarboxylase-like gene. The complete sequence of six of these introns was determined. The sizes of the six introns are (in order from the 5'-most intron to the 3'-most intron) 135 bp, 46 bp, 47 bp, 41 bp, 52 bp and 47 bp. The locations of these introns are indicated in Figure 19.

The cDNA nucleotide sequence was used in **blastx** searches of the "non-redundant" databank using the BLAST server at the NCBI (Materials and Methods). Strong similarities to members of a family of pyridoxal -5' - phosphate dependent amino acid decarboxylases were detected, including histidine decarboxylases, DOPA decarboxylases, tyrosine decarboxylases, and glutamate decarboxylases. The search results obtained with selected decarboxylases are summarized in Table 7.

Figure 19:

Nucleotide sequence of the cDNA contained within pCes814. Potential initiating methionine residues and the corresponding ATG codons are indicated in bold-faced type and are underlined. The positions of the custom-designed oligonucleotide primers used to determine the sequence are indicated above the nucleotides by arrows. The inverted arrowheads indicate the positions of known introns. A consensus sequence for pyridoxal-5'-phosphate co-factor binding is boxed beginning at amino acid residue number 366. A TGA stop codon is found commencing at nucleotide position 1876, and is boxed. A potential poly-(A) addition site is found commencing at nucleotide position 1907, and is also boxed.


GTATCGGACGCATCAAAAGATTCTAGGCCGTCTGAGACGAAGAAGGAGACTTTAATGATG 60
v s d a s k d s r p s e t k k e t l M M 20
R1
CCAGAAATACACCACACAAAACACTTTGATTCAATCGGCGGAGAAGAACAAGCATTTGCG 120
P E I H H T K H F D S I G G E E Q A F A 40
AAAAAAGAGAAAGTTGAAGAGTTCAAACCAACTGAGGCAGTCAAAGAAGAAGTTGATGTG 180
K K E K V E E F K P T E A V K E E V D V 60
AATGGAATGTCGAGGGATCAATTCCGAAATGCGGCAAAGAAAGTTGTGGATTATCTGATG 240
N G M S R D Q F R N A A K K V V D Y L M 80
AAACAAGACGAGAGCATCCGTGCAGCCCGTTGCTCTCCTGCACTGAAACCTGGTTATCTG 300
K Q D E S I R A A R C S P A L K P G Y L 100
AAAGCATTGCTCCCACCGAAAGCTCCACAGAAGGCAGAAGACATTGATGATATTTTGGAG 360
K A L L P P K A P Q K A E D I D D I L E 120
R3
GATTATCATAAGTTGATCGTTCCAGGACTTTCTCACTCGAGTTCATCCAAATTTTCATTC 420
D Y H K L I V P G L S H S S S S K F S F 140
ATTTTACCAGCTGGTAATTCATTCCATTGTCTACTTGCGGATCTACTTGGTGGCCACATT 480
I L P A G N S F H C L L A D L L G G H I 160
R5
GGAGATGCTGGATTCTATTGGACATCTAATCCAGCATTGACTGAGCTGGAAGTATTGATG 540
G D A G F Y W T S N P A L T E L E V L M 180
ATGGATTGGCTCGGCGAAATGATGGCATTGCCAAAAGAGTTTCTTCTGTTCCCTGAAGCA 600
M D W L G E M M A L P K E F L L F P E A 200
AGTCGTGGTGGAGGGTGTATGCAGAGATCCGACACAGAATCCAACTTCCTGGTTCTAGTT 660
S R G G G C M Q R S D T E S N F L V L V 220
GCCGCAAGAACTGATATGATTCGTCGAATGAAACAACGAGACAAGCGTCTTCGATCATCG 720
A A R T D M I R R M K Q R D K R L R S S 240
GATATTCTAGCTCGCCTTGTTGCTTATACCTCTTCTGATGCTCGTCGTTCGATCAAGATG 780
D I L A R L V A Y T S S D A R R S I K M 260
R7
R11
AAAATGGCGGCTGAAGTTGCGATGGTAAAGATGCGAGTACTTCCTACAGATCAGAACTTT 840
K M A A E V A M V K M R V L P T D Q N F 280
ATTCTCCGAGGTGACACATTGCATGCAGCAATTATGGCTGACATTGAGAGAGGCCTCATT 900
I L R G D T L H A A I M A D I E R G L I 300
CCATTCTTTGTCGGAGCCAACTTTGGTACTTCTGGACCATGCTCTTTTGATCATCTGCAC 960
P F F V G A N F G T S G P C S F D H L H 320
GAGCTTGGCCCGGTGTGTAGAGAACATGGAACATGGTTACACGTGGATGCAGCCTATGCT 1020
E L G P V C R E H G T W L H V D A A Y A 340
R9
GGAACTGCACTTATTTGTCCAGAAATAAGAGGGCTGATGAGAGGAATTGACTGGGCTGAC 1080
G T A L I C P E I R G L M R G I D W A D 360
TCGTTTTGCACTACACCCAGCAAGTTGATTATTGCTGTCTGTGATGTTTGTTGTCTCTGG 1140
S F C T T P S K L I I A V C D V C C L W 380
R10
GTTCGCGATCGTCACAAACTTCAGCACGCTTCACTGGAGAATCATCCTGATCTACCATTC 1200
V R D R H K L Q H A S L E N H P D L P F 400
AAAGGTCTGCCGACAAGTCAGAGAGTCGGGGCACTGAAGATATGGTTCATGATTCGTTCT 1260
K G L P T S Q R V G A L K I W F M I R S 420

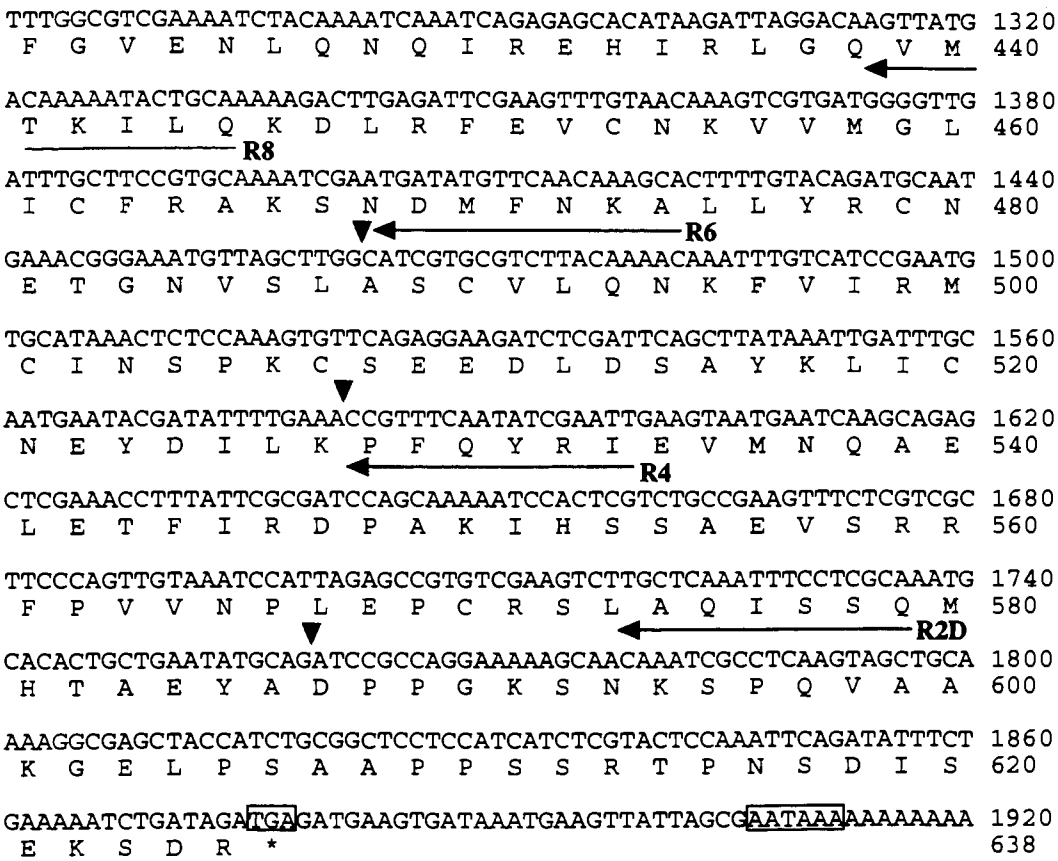

TTTGGCGTCGAAAATCTACAAAATCAAATCAGAGAGCACATAAGATTAGGACAAGTTATG 1320
F G V E N L Q N Q I R E H I R L G Q V M 440
ACAAAAATACTGCAAAAAGACTTGAGATTCGAAGTTTGTAACAAAGTCGTGATGGGGTTG 1380
T K I L Q K D L R F E V C N K V V M G L 460
R8
ATTTGCTTCCGTGCAAAATCGAATGATATGTTCAACAAAGCACTTTTGTACAGATGCAAT 1440
I C F R A K S N D M F N K A L L Y R C N 480
R6
GAAACGGGAAATGTTAGCTTGGCATCGTGCGTCTTACAAAACAAATTTGTCATCCGAATG 1500
E T G N V S L A S C V L Q N K F V I R M 500
TGCATAAACTCTCCAAAGTGTTCAGAGGAAGATCTCGATTCAGCTTATAAATTGATTTGC 1560
C I N S P K C S E E D L D S A Y K L I C 520
AATGAATACGATATTTTGAAACCGTTTCAATATCGAATTGAAGTAATGAATCAAGCAGAG 1620
N E Y D I L K P F Q Y R I E V M N Q A E 540
R4
CTCGAAACCTTTATTCGCGATCCAGCAAAAATCCACTCGTCTGCCGAAGTTTCTCGTCGC 1680
L E T F I R D P A K I H S S A E V S R R 560
TTCCCAGTTGTAAATCCATTAGAGCCGTGTCGAAGTCTTGCTCAAATTTCCTCGCAAATG 1740
F P V V N P L E P C R S L A Q I S S Q M 580
R2D
CACACTGCTGAATATGCAGATCCGCCAGGAAAAAGCAACAAATCGCCTCAAGTAGCTGCA 1800
H T A E Y A D P P G K S N K S P Q V A A 600
AAAGGCGAGCTACCATCTGCGGCTCCTCCATCATCTCGTACTCCAAATTCAGATATTTCT 1860
K G E L P S A A P P S S R T P N S D I S 620
GAAAAATCTGATAGATGAGATGAAGTGATAAATGAAGTTATTAGCGAATAAAAAAAAAAA 1920
E K S D R * 638

TABLE 7.

Summary of *C. elegans* amino acid decarboxylase similarity scores.

			Similarity (over 473 residues)		Identity (over 473 residues)	
	Locus I.D.†	P(N)‡	Number of a.a.	percent similar	Number of a.a.	percent identity
Human HDC§	sp | P19113	7.5×10^{-93}	306	64.7	184	38.9
Rat HDC	sp | P16453	1.9×10^{-89}	312	66.0	183	38.7
Human DDC	sp | P20711	3.6×10^{-84}	308	65.1	179	37.8
Rat DDC	sp | P14173	3.6×10^{-84}	303	64.1	176	37.2

§ HDC denotes histidine decarboxylase. DDC denotes DOPA decarboxylase.

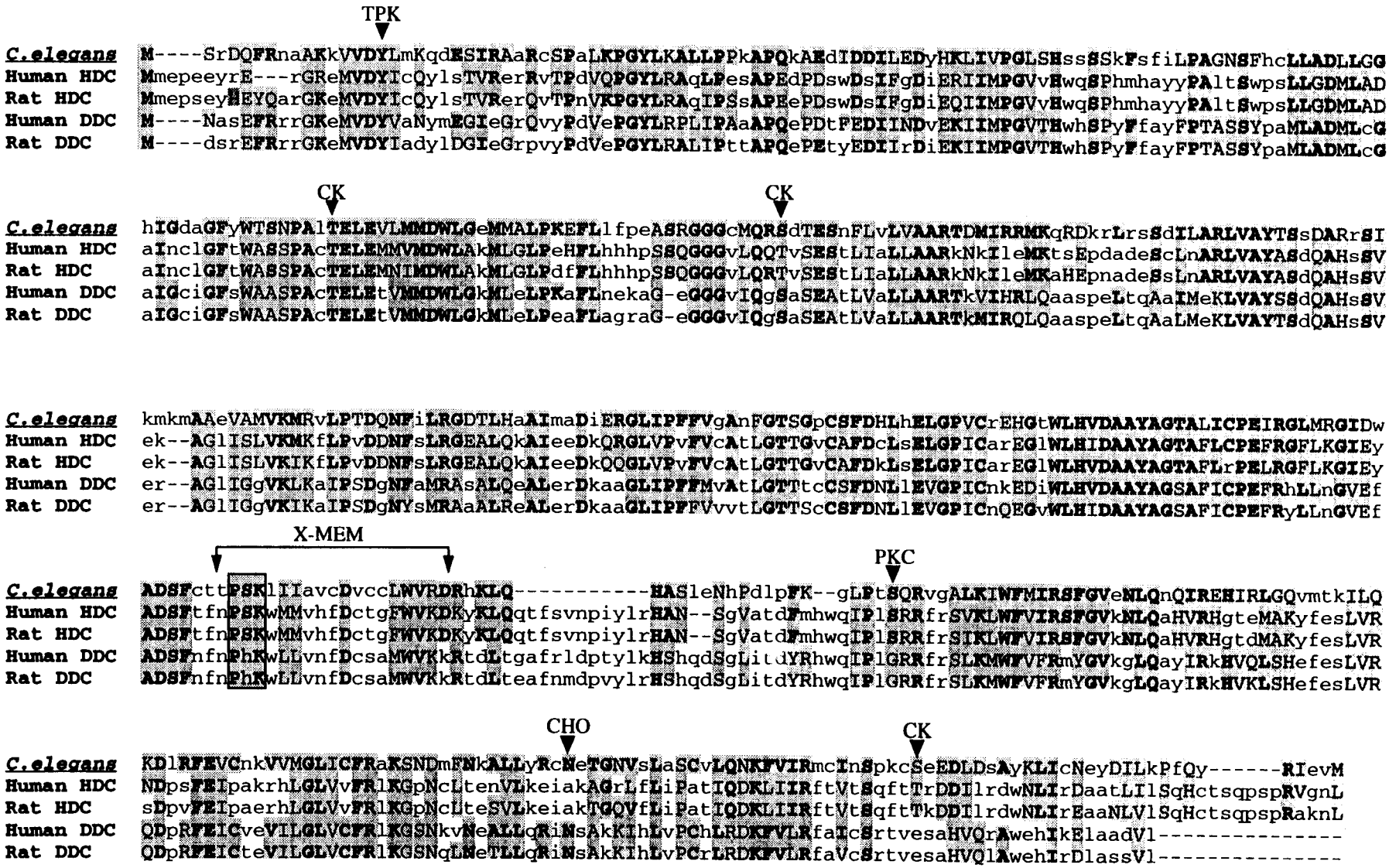
† Indicates the database and the accession number of the particular sequence matching the *C. elegans hdl-1* predicted protein. **sp** is the SWISS-PROT database.

‡ Indicates the Poisson P-values reported by the **blast** algorithm for the group of high scoring segment pairs (HSPs) for each database search. This value is the probability that at least as many HSPs would occur by chance alone, each with a **blast** score at least as high as the lowest scoring member of the group of HSPs. Thus, these probabilities provide a measure of statistical significance.

The decarboxylase amino acid sequences most similar to the *C. elegans* sequence were obtained from the database at the NCBI using the RETRIEVE e-mail server, and portions of these were aligned with the *C. elegans* amino acid sequence (Figure 20). The initial alignments were performed with CLUSTAL V (Higgins and Sharp, 1989; Higgins, Bleasby and Fuchs, 1992), and inspected manually to ensure that the alignments resulted in maximal similarity. Gaps were inserted in the sequence by CLUSTAL V or by myself to maximize similarity to the *C. elegans* sequence. As a result of the **blastx** searches and this alignment, it became evident that the first 62 amino acids of the *C. elegans* sequence shared no informative similarity to any sequence present in any of the databases maintained at the NCBI. Informative similarities commenced at the methionine residue located at amino acid position 63 of the *C.elegans* sequence (Figure 19). This methionine residue could be aligned with the initiating methionine residues of human and rat histidine decarboxylases, and human and rat DOPA decarboxylases (Figure 20). This observation may indicate that the initiating methionine residue for the *C. elegans* sequence is the one located at position 63. However, this methionine does not occur in the context of the consensus sequence for translation initiation in *C. elegans*. An alignment of the *C. elegans* amino acid sequence (commencing at position 63) with human histidine (Human HDC) and DOPA decarboxylases (Human DDC) and rat histidine (Rat HDC) and DOPA decarboxylases (Rat DDC) is presented in Figure 20. Residues identical to amino acids in the *C. elegans* sequence are indicated, as are conservative amino acid substitutions. I considered an amino acid substitution to be conservative if the substitution received a score greater than zero in the PAM250 amino acid replacement scoring matrix of Dayhoff *et al.*, 1978.

Figure 20:

A multiple sequence alignment of the *C. elegans* Hdl-1 predicted protein with histidine and DOPA decarboxylases from rat and human. Gaps have been introduced to maximize the similarity to the *C. elegans* sequence. Amino acids identical to residues in the *C. elegans* sequence are indicated in bold-faced capitals. Amino acids similar to residues in the *C. elegans* sequence are capitalized. Non conservative substitutions are in lower case. Regions of similarity to the *C. elegans* sequence are indicated by shading. The inverted arrowheads indicate *C. elegans* residues that were detected during PROSITE (Bairoch, 1993) analysis of the Hdl-1 protein. These particular residues have been indicated because they are conserved in at least two of the other decarboxylase sequences. TPK = potential tyrosine protein kinase phosphorylation site. CK = potential casein kinase II phosphorylation site. PKC = potential protein kinase C phosphorylation site. CHO = potential glycosylation site. The pyridoxal 5'-phosphate co-factor binding site (**PSK** in the *C. elegans* sequence) has been boxed. **X-MEM** indicates a hydrophobic region of the *C. elegans* protein that is predicted to be membrane-spanning.


TPK
C. elegans
Human HDC
Rat HDC
Human DDC
Rat DDC
CK
CK
X-MEM
PKC
CHO
CK

From this analysis, it is apparent that the amino acid sequence of the *C. elegans* predicted protein exhibits a high degree of similarity to the other decarboxylases with which it is aligned. The *C. elegans* sequence appears to be most similar to the histidine decarboxylases of human and rat, with the region of similarity extending from *C. elegans* amino acid 63 to *C. elegans* amino acid 536. The carboxy terminal 89 amino acids exhibit no informative similarity to other sequences in the databank. A summary of the similarity between the *C. elegans* sequence and other decarboxylases is shown in Table 7. I have named the *C. elegans* gene *hdl-1*. The *hdl-* prefix stands for **h**istidine **d**ecarboxylase-**l**ike.

The *C. elegans* Hdl-1 predicted protein contains a potential binding site for pyridoxal-5'-phosphate, a co-factor required by eukaryotic amino acid decarboxylase enzymes that synthesize neurotransmitters and neuromodulatory compounds. These enzymes include histidine, glutamate, and DOPA decarboxylases. The location of this binding site (Proline-Serine-Lysine) is indicated on the sequences in Figure 20. All of the DOPA decarboxylases recovered from the databank had, as their pyridoxal co-factor binding site, Proline-Histidine-Lysine, while all of the histidine decarboxlase sequences had a serine residue in place of the histidine. The potential pyridoxal co-factor binding site found in the Hdl-1 predicted protein is Proline-Serine-Lysine, as in the histidine decarboxylases.

The Hdl-1 predicted amino acid sequence was analyzed using Version 11 of PROSITE (Bairoch, 1993) running under PCGENE (release 6.8). A number of potential phosphorylation sites for tyrosine kinase, protein kinase C and casein kinase II were detected. While the functional significance of these sites in the *C. elegans* protein is unknown, certain of these sites are conserved in the human and rat histidine and DOPA decarboxylases. These conserved potential

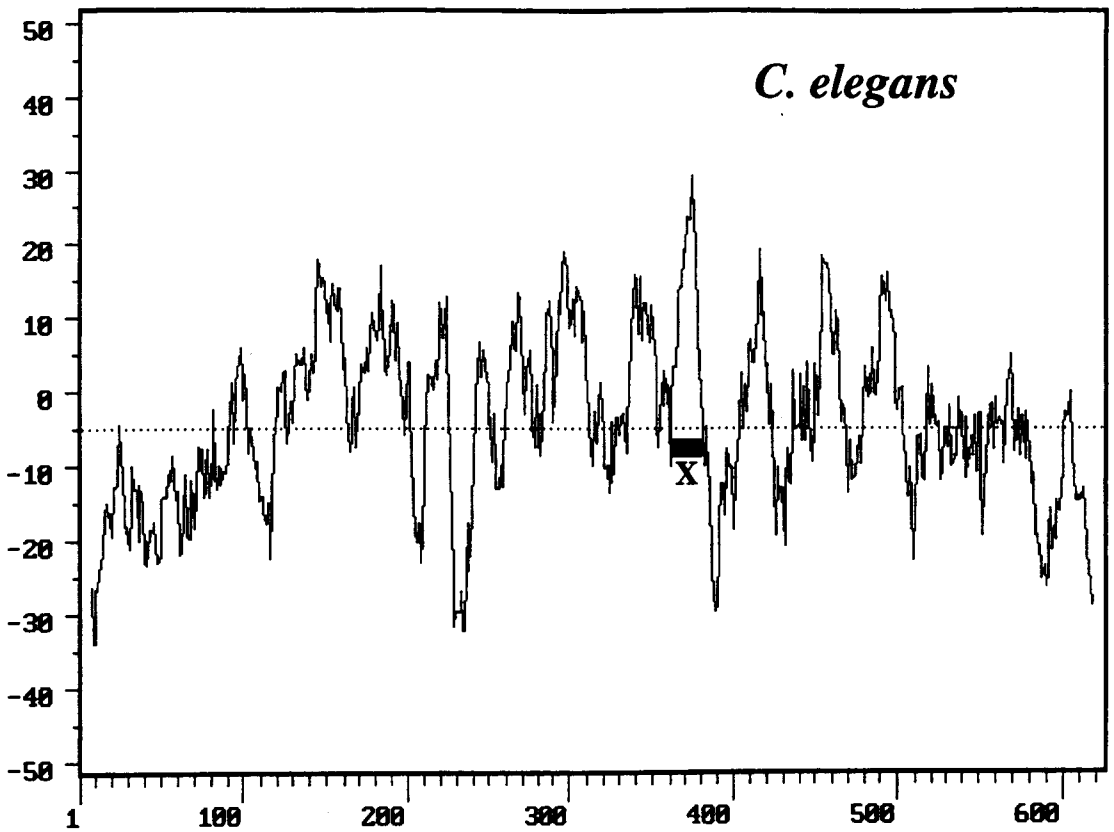
phosphorylation sites are indicated in Figure 20. These sites are also conserved within the *D. melanogaster* HDC sequence (Burg et al., 1993; not shown).

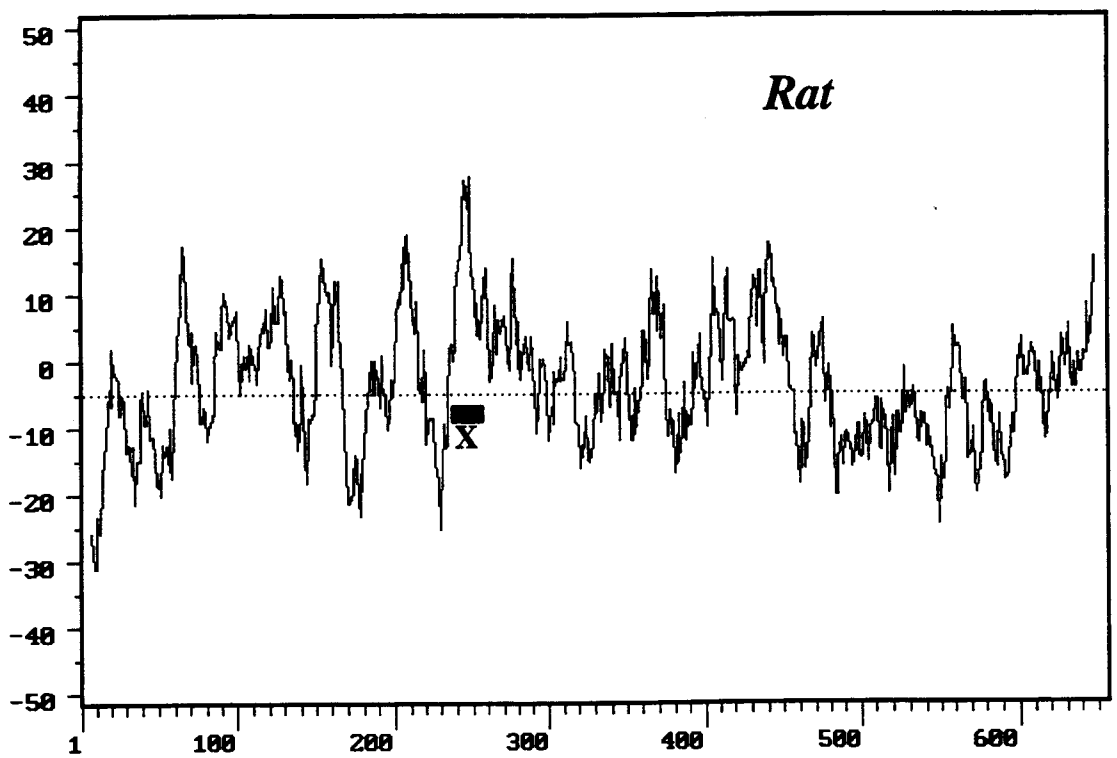
Kyte - Doolittle hydrophobicity analysis was performed on the Hdl-1 amino acid sequence, and is shown in Figure 21. Using the method of Klein, Kanehisa and DeLisi (1985), a transmembrane domain was predicted. The location of this transmembrane domain is indicated in Figure 21. Evidence has been obtained that histidine decarboxylase activity isolated from rat brain is found in a soluble and a membrane - bound form (Toledo et al., 1988). In support of this, the method of Klein, Kanehisa and DeLisi (1985) predicts the existence of a transmembrane domain in the rat histidine decarboxylase protein, as indicated in the hydrophobicity profile in Figure 21.

In order to identify putative regulatory elements in the genomic DNA immediately upstream of the site at which the cDNA was found to commence, I sequenced appropriate deletion derivatives of pCes122 that I had constructed. Twenty nucleotides upstream of the start of the cDNA, a potential 3' splice acceptor that perfectly matched the *C. elegans* consensus (TTTTCAG) was found. This site can function as a splice acceptor in *trans*- splicing reactions that add a splice leader sequence, or may function as a splice acceptor in *cis*- splicing reactions. Examination of the genomic sequence further upstream revealed two additional potential 3' splice acceptors that also perfectly matched the *C. elegans* consensus. These were found 136 and 208 nucleotides ahead of the position at which the cDNA commenced. The presence of these potential 3' splice acceptor sites prompted me to search for potential 5' splice donor sites. Using the weak consensus for splice donors given in Emmons (1988) as a guide, I identified three potential 5' splice donors in the genomic DNA upstream of the cDNA start site. Each of these potential splice donors was

Figure 21:

Kyte - Doolittle hydropathy plots of the *C. elegans* Hdl-1 protein and the rat histidine decarboxylase. The hydropathy profiles were constructed using a window of 15 amino acids. The X axes are scaled in number of amino acids. The Y axes are scaled in hydrophobicity scores. The dotted lines intersecting the Y axes represent the cutoff between hydrophobic values and hydrophilic values. Values above the line are hydrophobic, and values below the line are hydrophilic. The potential transmembrane domain in each profile is indicated by the **X** and the heavy line.


C. elegans
X
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
100
200
300
400
500
600


Rat
X
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
100
200
300
400
500
600

appropriately located to serve as a splice donor for one of the potential 3' splice acceptor sequences I had identified (Figure 22). When these "introns" were spliced out, the resulting DNA sequence was found to encode a long uninterrupted open reading frame specifying a peptide of 85 amino acids that fused, in-frame, to the *hdl-1* open reading frame. The spliced sequence was used in a **blastx** search of the non-redundant database, and a **blastn** search of the **dbest**. The searches failed to yield an informative similarity to any sequence currently in either database.

Determining the direction of the *hdl-1* open reading frame: In order to better understand the orientation of the *hdl-1* gene in relation to neighbouring genes in this region of the genome, I determined the direction of the *hdl-1* open reading frame. This was accomplished as follows. The *hdl-1* cDNA was found to have a single *Sph*I restriction site. The pBluescript SK- vector in which the cDNA was cloned contained no *Sph*I site. Double digestion of the cDNA with *Sph*I and *Eco*RI produced three fragments, one of which (approximately 3 kb) corresponded to vector. A fragment of approximately 1200 bp represented the 3' portion of the cDNA, while a fragment of approximately 700 bp represented the 5' portion of the cDNA. Digestion of pCes122 (the plasmid containing fragment R) with *Sph*I produced two bands on agarose gels, due to the presence of an internal *Sph*I site in fragment R and another in the pUC19 multiple cloning site. One of these bands was 3.5 kb, and the other was 8.0 kb. Southern blotting of the *Sph*I digested pCes122 DNA followed by hybridization with the labeled 5' specific *Sph*I - *Eco*RI restriction fragment and subsequent autoradiography indicated that the 3.5 kb *Sph*I restriction fragment contained the decarboxylase 5' end. This result, in combination with the previously established orientation of fragment R with respect to fragment Q (see above), indicated the direction of the *hdl-1* open reading frame is from right to left with respect to the the genetic map, in the same direction as *unc-22*, and

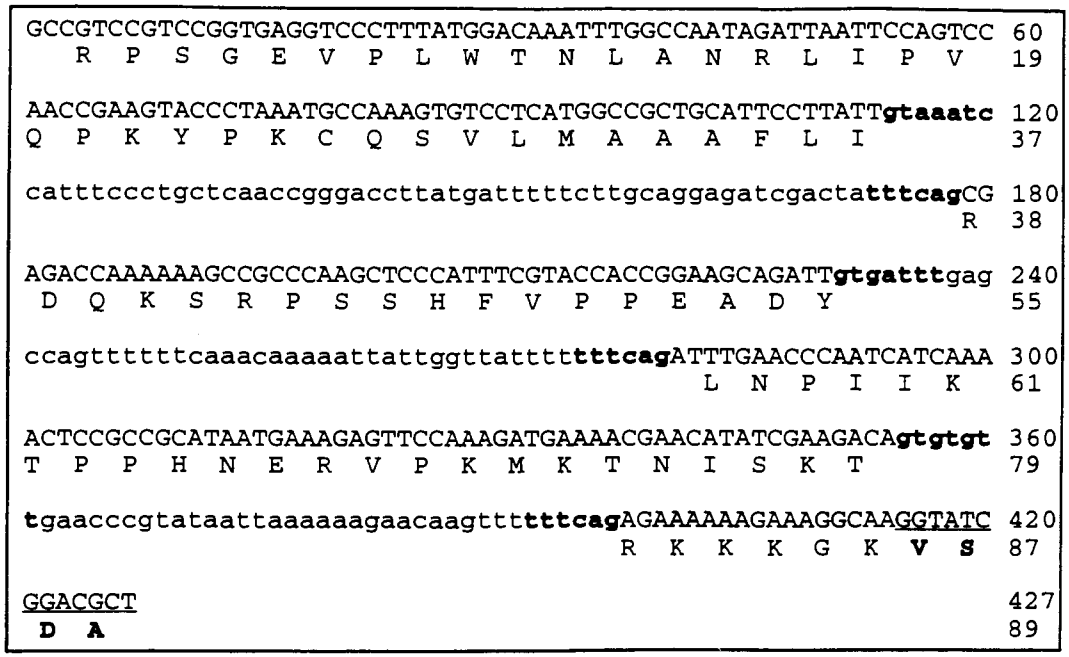


Figure 22:

Genomic DNA sequence upstream of the start of the *hdl-1* cDNA. Putative intron sequence is indicated by lower-case characters. Uppercase characters indicate putative coding sequence. The open reading frame produced by the excision of the putative introns is shown beneath the DNA sequence. Putative intron donor and acceptor sequences are indicated by bold-faced type. DNA sequence found in the *hdl-1* cDNA is underlined. The amino acids in bold-faced type directly beneath these underlined sequences are also encoded by the *hdl-1* cDNA (Figure 19).

opposite to that of the *nhe-1* gene located on fragment Q. These data are summarized in Figure 23.

I had determined that both of the deficiencies *sDf19* and *sDf65* disrupted or deleted *nhe-1*. In addition, Prasad (Prasad and Baillie, 1989) had determined that the deficiency *sDf9* deleted the C2 cDNA, which contained worm antiporter sequence. Because *nhe-1* lies to the left of *hdl-1* on the genetic map, these deficiencies must also delete *hdl-1*. This was confirmed by PCR, as described in Materials and Methods, using the oligonucleotide primers R1 and R8 (Table 2), indicated on the cDNA sequence of the *hdl-1* gene in Figure 19 (data not shown).

Identification and characterization of genes located on Fragment S:

In order to continue identifying and characterizing genes residing on the cosmid C11F2, I used a portion of fragment S as a probe in screens of the Barstead and Waterston cDNA library. Fragment S is the 8.5 kb fragment immediately to the right of fragment R on C11F2 (Prasad 1988; Prasad and Baillie, 1989; see Figure 11). Fragment S represented the remainder of the *C. elegans* DNA contained within the right-most portion of C11F2, and had not been analyzed by Prasad, either by interspecies hybridization to *C. briggsae* or by Northern blot.

Fragment S had not been cloned due to a lack of convenient restriction sites (S. Prasad, personal communication). In order to subclone C11F2 DNA containing fragment S, I digested C11F2 DNA with *Bam*HI, re-ligated the digestion mixture and transformed competent *E. coli* as described (Materials and Methods). Restriction enzyme digestion followed by Southern blot analysis of the re-circularized molecule revealed it contained a portion of fragment R, all of fragment S, DNA representing fragment P, and pJB8 cosmid vector sequences. This plasmid was designated pCes891 (Figure 24). Using pCes891 DNA as my cloning source, I subcloned a 4.2 kb *Hin*dIII restriction fragment into pBluescript SK- to generate pCes1016 (Figure 24). The pCes1016 insert was gel isolated

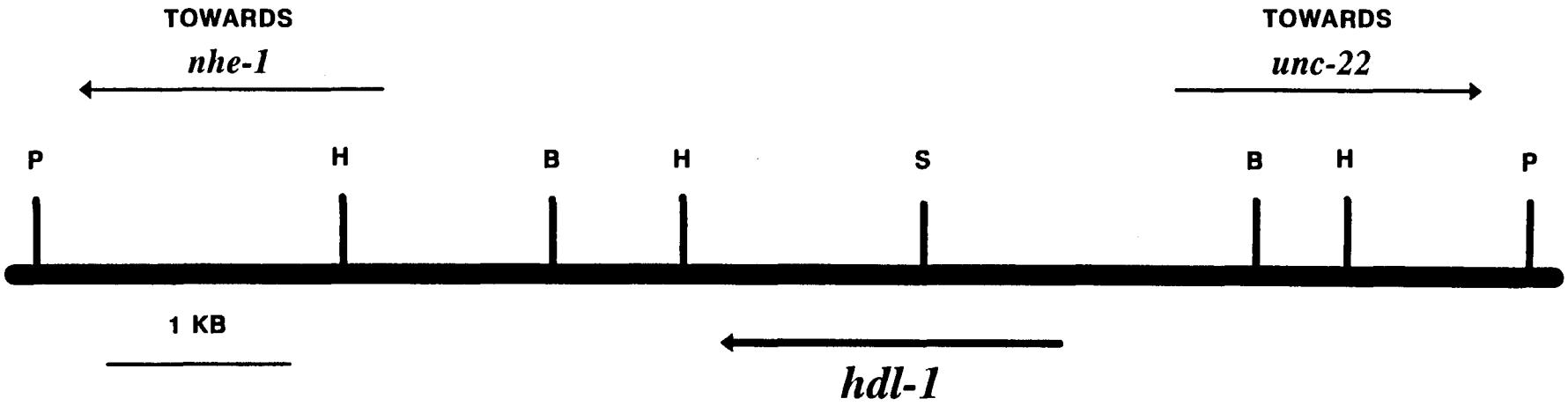


Figure 23:

A restriction map of fragment R indicating the direction of the *hdl-1* open reading frame. P = *Pst*I. H = *Hin*dIII. B = *Bam*HI. S = *Sph*I. The direction in which *unc-22* lies is indicated, as is the direction in which *nhe-1* lies. The direction of the *hdl-1* open reading frame is the same as the direction of *unc-22* transcription, and is opposite to the direction of the *nhe-1* open reading frame.

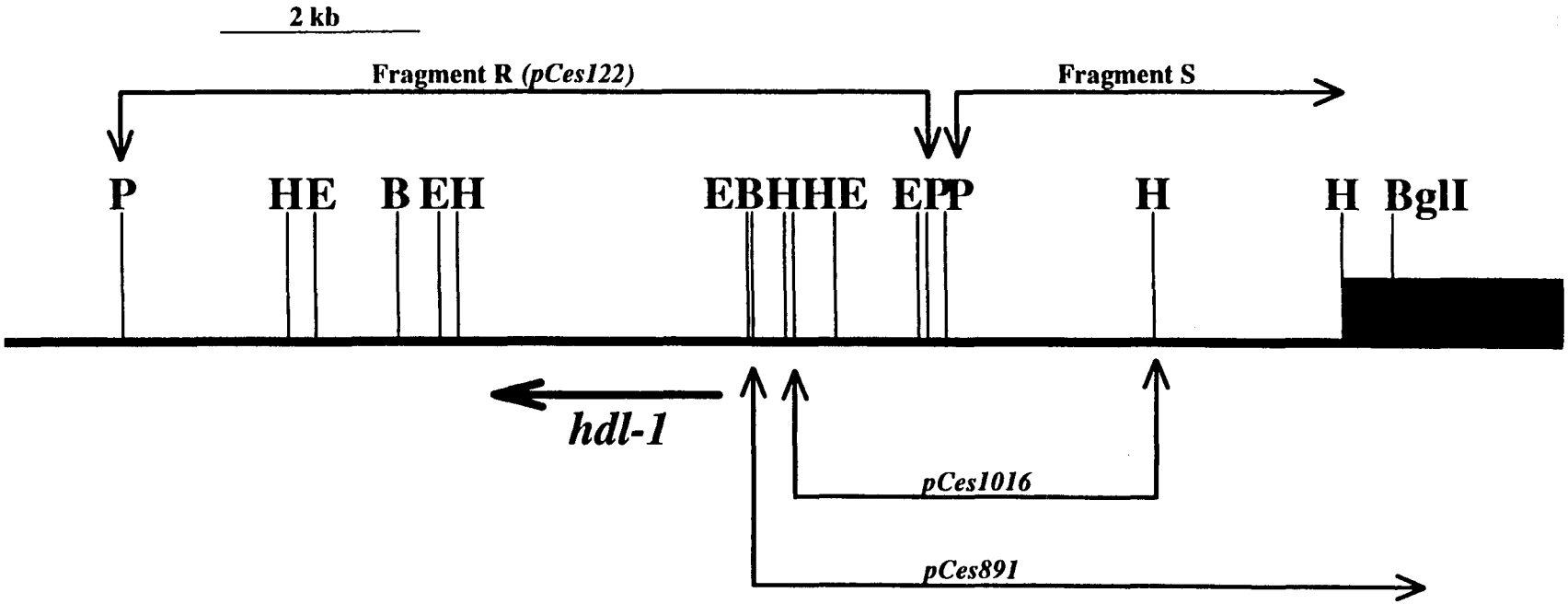


Figure 24:

A restriction map of a portion of the cosmid C11F2 showing fragment R and fragment S. P = *Pst*I. B = *Bam*HI. H = *Hin*dIII. E = *Eco*RI. The location of the *hdl-1* gene is indicated. pCes891 consists of C11F2 DNA from this end of the cosmid, the pJB8 cosmid vector (indicated in this figure by the heavy line on the right), and C11F2 DNA corresponding to fragment P (see Figure 11). pCes891 was the cloning source for the pCes1016 insert, which is also indicated on this map.

(Materials and Methods) and used as a probe in a screen of the Barstead and Waterston (1989) cDNA library (Materials and Methods). Approximately 1.5 X 10^5 plaques were screened, resulting in the identification and subsequent purification of 23 positive plaques. The corresponding cDNA clones were recovered in pBluescript SK- using a protocol supplied by Statagene (Materials and Methods).

The cDNA clones were found to fall into three classes, based upon their *Eco*RI restriction pattern. Twenty of the cDNAs, which I will call the Class A clones, (pCes1018 - pCes1026; pCes1028 - pCes1031; pCes1033 - pCes1036; pCes1038- pCes1040) exhibited virtually identical restriction patterns, liberating fragments of approximately 3.0 kb (corresponding to the vector) 700 bp, 400 bp, 300 bp and 200 bp upon digestion with *Eco*RI. One cDNA (pCes1027), in addition to these fragments, contained another fragment of 1.3 kb. This single clone defined Class B. The two remaining cDNAs, pCes1032 and pCes1037, produced restriction patterns identical to each other, but distinct from any of the other 21 cDNA clones. Digestion of these cDNAs with *Eco*RI released an insert of approximately 1.2 kb. These clones comprised Class C. I selected pCes1024 and pCes1025, (Class A clones) pCes1027 (the sole Class B clone) and pCes1032 and pCes1037 (the two Class C clones) for additional characterization.

Sequence determination and analysis of the Class A and Class B cDNA clones: I commenced characterizing the Class A cDNAs by sequencing the ends of the cDNAs using M13 -21 and M13 reverse primers. The sequences thus obtained were used in **blastx** searches of both the non-redundant database and **dbest** (**d**atabase of **e**xpressed **s**equence **t**ags), using the BLAST server at the NCBI (Materials and Methods). As a result of this preliminary analysis, I determined that the first seven nucleotides of one of the Class A cDNA clones (pCes1025) were identical to the last seven nucleotides of the SL1 splice leader sequence.

pCes1024 was slightly shorter, and did not contain the first 30 nucleotides found in pCes1025. Both cDNAs had poly-(T) tracts at the opposite, presumably 3', ends. These observations indicated that pCes1025 was likely a full-length cDNA. An open reading frame predicted from the 5' ends of these cDNAs exhibited strong similarity to the amino terminii of glutamate dehydrogenase enzymes from different species including eels, rats, cows and mice. A search of the **dbest** with the DNA sequence from the Class A clones resulted in a near identical match with a cDNA (cm11e12) sequenced by the *C. elegans* genome sequencing consortium. This result served to confirm the suspected 5' end of the Class A clones, as the cDNA sequence determined by the consortium was obtained from the 5' end of a directionally cloned cDNA.

The sequence obtained from the single Class B clone (pCes1027) using the M13 -21 primer was identical to the sequence obtained from pCes1025 using the M13 reverse primer, indicating that pCes1027 also contained glutamate dehydrogenase-like sequences. However, the sequence obtained from pCes1027 using the M13 reverse primer did not encode a known glutamate dehydrogense-like protein, nor did it contain an open reading frame. The sequence obtained from both ends of this cDNA contained poly-(T) tracts. These observations were compatible with the notion that pCes1027 was a fusion of two, possibly unrelated, cDNAs in a 5'-5' orientation.

The *Eco*RI restriction patterns of pCes1024, pCes1025 and pCes1027 were identical except that pCes1027 contained an additional *Eco*RI restriction fragment of approximately 1.3 kb. All of these cDNAs shared fragments of 200 bp, 300 bp, 400bp and 700 bp. The observation that an additional *Eco*RI restriction fragment was contained within the potential fusion cDNA indicated that the additional band of 1.3 kb probably represented the unknown cDNA attached to the glutamate dehydrogenase-like cDNA.

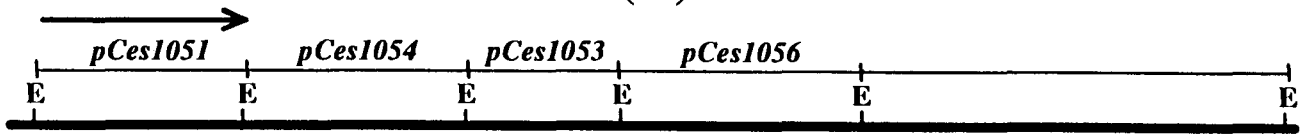
In order to obtain additional sequence data aimed at verifying the identity of the potential glutamate dehydrogenase encoding cDNAs and the suspected fusion cDNA contained within pCes1027, I adopted the following strategy. Each of these cDNAs were first restriction digested with *Eco*RI, re-ligated, and then transformed into competent *E. coli*. This resulted in the capture of fragments, that were approximately 400 bp, 300 bp and 200 bp, in pBluescript SK-. The 700 bp *Eco*RI fragment was not recovered. Because the subcloned *Eco*RI fragments were small, sequencing the ends of these individual fragments would result in the relatively rapid determination of approximately 60% of the cDNA sequence. A subclone containing only a 1.3 kb *Eco*RI fragment (representing the suspected fusion cDNA) was similarly obtained from pCes1027, and the ends of this molecule were likewise sequenced.

The sequences obtained from the ends of four individually subcloned *Eco*RI fragments were used in **blastx** searches of the non-redundant database, and the **dbest**. In the cases of open reading frames predicted from the sequences determined from the 465 bp, 317 bp, 271 bp and 174 bp *Eco*RI fragments of pCes1024 or pCes1025, strong similarities to glutamate dehydrogenase enzymes cloned from eels, rats and cows were observed. While the order of the restriction fragments in the native pCes1024 and pCes1025 cDNAs was not determined directly, I assigned a tentative order to the restriction fragments by comparison of the amino acid sequences predicted from each of the subclones to the Rat glutamate dehydrogenase protein. The inferred order of the restriction fragments and sequences of the subclones are given in Figure 25. Alignments of portions of the *C. elegans* predicted amino acid sequence with corresponding regions of the Rat glutamate dehyrogenase protein are presented in Figure 26. I have named the *C. elegans* gene *gdl-1*. The prefix *gdl-* stands for **g**lutamate **d**ehydrogenase-**l**ike.

Figure 25:

(**A**) A restriction map of pCes1025 inferred from alignments of conceptual translations of these sequences to the Rat glutamate dehydrogenase sequence. E = *Eco*RI. Individual *Eco*RI restriction fragments of pCes1025 were subcloned as described in the text. The plasmids in which these subcloned fragments are contained are indicated above the restriction map. The arrow above the restriction map indicates the direction in which *unc-22* lies, and also indicates the direction of the *gdl-1* open reading frame. The sequence determined from these individually subcloned fragments is shown beneath the restriction map. *Eco*RI sites found within the sequences are indicated in bold-faced type. The lengths of each sequence are also indicated in the right margins. Sequencing these subclones yielded a total of 1228 nucleotides of the *gdl-1* cDNA of an estimated total of approximately 1900 bp. The sequence **AAGTTTGAG** which was found in the pCes1051 sequence, two nucleotides after the 5' *Eco*RI site, is identical to the last nine nucleotides of the 22 nucleotide SL1 splice leader sequence, providing evidence that the pCes1025 cDNA is nearly complete at its 5' end.
(**B**) A poly-(A) tract is found in sequence obtained from the opposite end of pCes1025, indicating the pCes1025 cDNA may be full length. In addition, a *C. elegans* poly-(A) addition consensus sequence (**AATAAA**) is seen 14 nucleotides upstream of the poly-(A) tail. The consensus poly-(A) addition site is indicated in bold-faced type and is underlined. The poly-(A) tail is in bold-faced type.

(A)



pCes1051

GAATTCCGAAGTTTGAGGTAATGTTGAGCACTCTTTCCAGAGGCGCCCGCTCGGTCGCTGTCCGAAGCTATTCGGCCGCTACTCTCGATG 90

CCCACTCCCAGGTGCTCGACGAGCAGAAGCCAATGGAGGAGCAGGTCAACCCATCTTTCTACAAGATGGTCGACTTTTACTTCAACAAGG 180

GAGCCGAGGTCATCGCTCCAAAACTCGCCGAGGAGCTCAAGTCCAACTCTTTGAGCCAAAAGGATAAGAAGAATCTCGGTGTCCG**GAATTC** 271

pCes1054

GAATTCTTGGAGCTATCAAGCCAGTTAACAAGGTCCTTTACATCACCTTCCCAATCCGTCGTGACAATGGAGAGTTCGAGGTTATTGAGG 90

CCTGGCgCGCACAGCATTCTGAGCACAGAACTCCAACCAAGGGAGGTATCCGTTATTCCCtCGATGTGTGCGAAGATGAGGTCAAAGCTT 180

TGTCTGCTCTTATGACTTACAAGTGCGCTGTTGTCGATGTTCCATTCGGAGGAGCTAAGGGAGGAGTCAAGATTGATCCAAAGCAATACA 270

CTGATTATGAAATTGAGAAGATCACCCGCCGTATCGCTATC**GAATTC** 317

pCes1053

GAATTCGCAAAGAAGGGATTCCTCGGCCCCGGAGTTGATGTCCCAGCCCCAGATATGGGAACCGGAGAGCGTGAAATGGGATGGATCGCT 90

GATACCTACGCCCAAACCATTGGACATTTGGACAGAGATGCTTCGGCTTGTATCACAGGAAAGCCAATTGTCAGTGGAG**GAATTC** 175

pCes1056

GAATTCACGGACGTGTTTCCGCCACCGGACGTGGAGTGTGGAAGGGACTCGAGGTGTTCACCAACGATGCTGATTACATGAAGATGGTCG 90

GTCTTGACACCGGACTCGCTGGAAAGACTGCTATCATCCAAGGATTCGGAAACGTCGGACTTCACACCCACAGATACCTCCACCGTGCTG 180

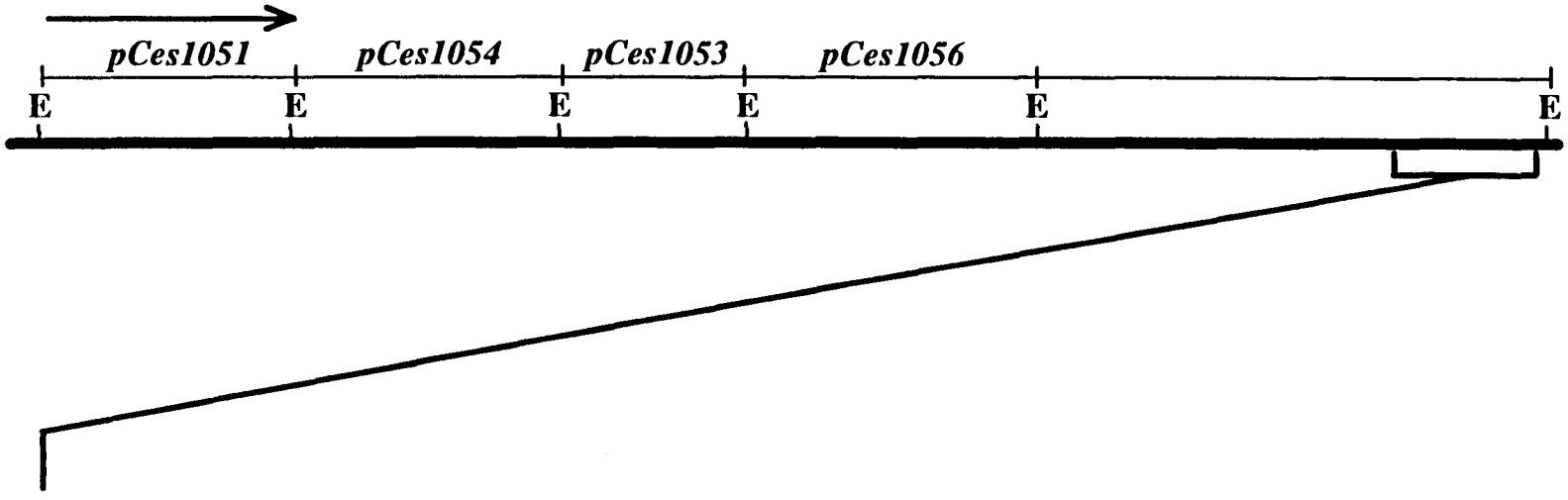
GATCAAAGGTCATCGGAATCCAGGAGTACGATTGCGCCGTCTACAACCCAGATGGTATTCATCCAAAGGAGCTCGAAGATTGGAAGGACG 270

CCAATGGAACTATCAAGAACTTCCCAGGAGCCAAGAACTTCGACCCATTCACTGAGCTTATGTACGAGAAGTGCGATATCTTCGTTCCAG 360

CTGCTTGCGAAAAGTCGATCCACAAGGAAAATGCTAGCCGCATTCAAGCTAAGATCATCGCCGAGGCTGCCAACGGACCAACTACCCCAG 450

CTGCTGACA**GAATTC** 465

(B)



121(c)

AAATAGCCtCGTGTAACCCTTTAGTGCCTTTTCCATAGTTTTCTGTTTCGAAACTATCTAGGTTAACTAAAGTCTAGCTCTGATCTTTTC 90

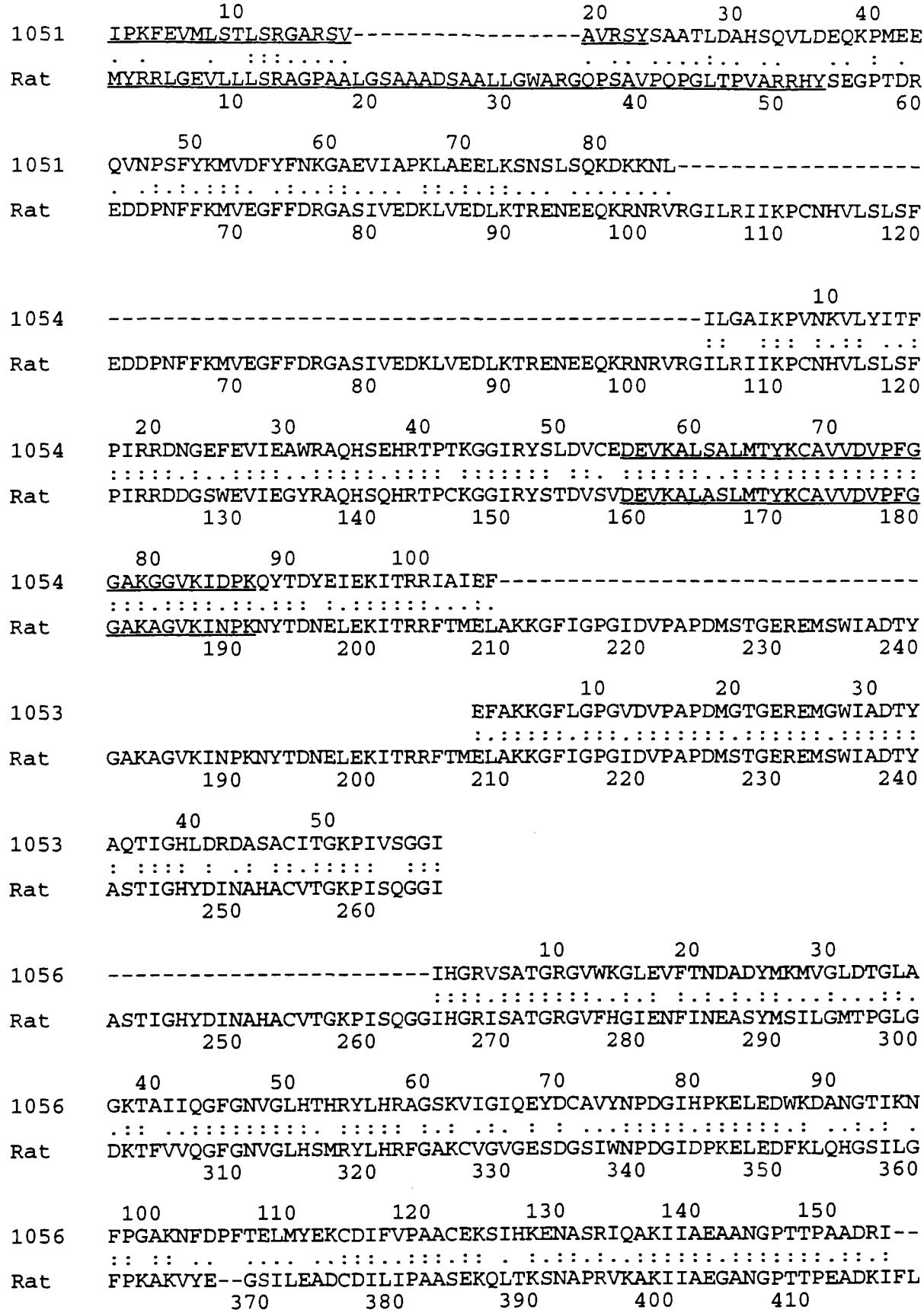
TGCCTGGTTCGTTCTCACTTTAGTATAATTAAATATTTTATTGGTTCCCCACCCCTTTTCATCCCCACACATAATTTTTTCGTTGTTTGT 180

TGGTTCGGGACGGTTTCGAGACGGAGACCATTCTTCAAAAAAAAATCTCAAAAAT**AATAAA**AAGTTTATTTCCTT**AAAAAAAAAAAAAAA** 270

AAAAAA 276

Figure 26:

Sequence alignments of open reading frames predicted from the subclones pCes1051, pCes1054, pCes1053 and pCes1056 to rat glutamate dehydrogenase (SWISS-PROT accession number P10860). The subclone from which each *C. elegans* **o**pen **r**eading **f**rame (ORF) was predicted is indicated to the left of the sequence. The rat amino acid sequence is indicated below the *C. elegans* amino acid sequence. The ORF predicted from pCes1051 matches the rat sequence at the beginning of the rat sequence and extends to rat amino acid 102. The pCes1054 ORF matches the rat sequence commencing at rat amino acid 105. The match extends, ungapped, to rat amino acid 209. The pCes1053 ORF matches the rat sequence commencing at rat amino acid 208, and extends, ungapped, to rat amino acid 265. The pCes1056 ORF matches the rat sequence commencing at rat amino acid 265, and the match extends to rat amino acid 417, with the introduction of a single gap in the rat sequence. A putative mitochondrial transit peptide within the pCes1051 ORF and the corresponding rat sequence is underlined. Cleavage of the transit peptides is predicted to occur between the underlined tyrosine residue and the adjacent serine residue. Putative glutamate dehydrogenase active sites within the pCes1054 ORF and the corresponding rat sequence are likewise underlined.



Of the 402 amino acids resulting from the combination of the pCes ORFs, 57% are identical to residues in the rat sequence, and 31% are considered to be conservative amino acid replacements, according to the PAM250 scoring matrix. Therefore, overall similarity of the *C. elegans* sequence to the rat glutamate dehydrogenase sequence is 87 %. Further support for the notion that the *gdl-1* gene encodes a glutamate dehydrogenase enzyme was obtained from analyzing the pCes ORFs using PROSITE (Version 11; Bairoch, 1993) running under PCGENE (release 6.8). This analysis detected a putative glutamate dehydrogenase active site in the amino acid sequence of the pCes1054 ORF. A similar analysis conducted on the rat protein revealed the existence of the same putative active site. The sequences comprising the active sites and the adjacent areas are highly conserved, and are indicated on the sequence in Figure 26.

Glutamate dehydrogenase enzymes are localized to the mitochondrial matrix (Smith et al., 1975). In order to determine whether the *gdl-1* cDNA encoded a mitochondrial transit peptide, I analyzed the pCes1051 ORF using TRANSPEP, a PCGENE (6.8) utility that detects the existence of these signal sequences using the method of Gavel and Von Heijne (1990). In both the Rat and *C. elegans* amino acid sequences the existence of a mitochondrial transit peptide is predicted. The transit peptides are indicated on the sequence in Figure 26.

The open reading frames predicted from DNA sequence obtained from the ends of the 1.3 kb *Eco*RI fragment derived from pCes1027 did not exhibit similarity to glutamate dehydrogenase enzymes. When these sequences were used in **blastx** searches of the nonredundant database, strong similarities to acyl-CoA dehydrogenase enzymes from *Ascaris suum*, humans and rats were obtained. The sequences obtained from the ends of this fragment verified the initial suspicion that pCes1027 was the result of a fusion of two cDNAs in a 5' - 5' orientation.

Sequence determination and analysis of the Class C cDNA clones: I sequenced both ends of each of the two Class 3 cDNA clones using M13 -21 and M13 reverse primers, and compared the sequences obtained from the two clones. The cDNAs were identical to each other, except that pCes1032 contained an additional 12 nucleotides at the putative 5' end of the clone. The sequence obtained from pCes1032 using the M13 reverse primer contained a poly-(T) tract immediately following the *Eco*RI site of the vector into which the insert was cloned, identifying a putative 3' end of the gene. The sequence obtained from pCes1032 using the M13 -21 primer was used in a **blastx** search of the non-redundant database, and the **dbest**. As a result of the search of the non-redundant database, weak similarities to a portion of the *Brachyury* (T) proteins from zebrafish, *Xenopus*, and mouse were discovered, in addition to a weak similarity to the *Drosophila* lethal(1) optomotor-blind protein. No similarities to sequences in the **dbest** were found.

Because the similarity obtained in the database searches was weak, no inference could be made concerning the putative identity of the protein encoded by the pCes1032 cDNA clone. In order to obtain more information about the potential identity of the gene product, I sequenced the entire pCes1032 cDNA. I constructed unidirectional deletion derivatives of pCes1032 using the method of Henikoff (1984) as described in Materials and Methods, and obtained sequence from these deletion derivatives using the M13 -21 primer.

The cDNA contained within pCes1032 is 1146 bp in length, and contains a single long open reading frame of 996 bp. The cDNA has the potential to encode a peptide of 332 amino acids. A putative TAA stop codon is found commencing at nucleotide position 997. Two consensus poly-(A) addition sites (AATAAA) are found commencing at nucleotide positions 1109 and 1122, respectively. A poly-(A) tract commences at nucleotide position 1131. The functional poly-(A)

addition site may be the one at position 1109, as the poly-(A) tract usually appears 12 - 20 bp downstream of the poly-(A) addition site in *C. elegans* mRNAs (Emmons, 1988). A putative initiating methionine codon (ATG) is found commencing at nucleotide position 4. This ATG occurs in the context of the *C. elegans* consensus sequence for translation initiation (A/c - A/g - a/c - A/C - ATG; M. D. Perry, G. Z. Hertz and W. B. Wood, personal communication), indicating the cDNA may be of complete or near-complete length. The sequence of the cDNA contained within pCes1032 is indicated in Figure 27.

The complete cDNA sequence was used in **blastx** searches of the non-redundant database at the NCBI. No additional information was obtained concerning the possible identity of the peptide encoded by this cDNA. The similarities obtained were weak, and were to the same regions of the same proteins that had been detected previously using only a single read from the M13 -21 primer as the query sequence; i.e., the *Brachyury* (T) proteins from zebrafish, *Xenopus*, and mouse.

In mice, direct evidence has been obtained that the mouse *Brachyury* (T) protein can bind DNA (Kispert and Herrmann, 1993). Furthermore, mouse T protein has been detected in the nucleus (Schulte-Merker et al., 1992), although no consensus nuclear targeting sequences are found within the protein (Herrmann et al., 1990).

The DNA binding activity of the mouse T protein has been shown to reside within the N-terminal 229 amino acids (Kispert and Herrmann 1993). It is within this region, over a span of 35 amino acids, that the *C. elegans* protein exhibits similarity to the various T proteins when the **blast** algorithm is used to search the database (Figure 27). Because the *C. elegans* protein exhibited weak similarity to a DNA binding domain, I used the complete *C. elegans* protein sequence to search the **tfd** (transcription **f**actor **d**atabase) at the NCBI. No strong similarities to

Figure 27:

The sequence of the cDNA contained within pCes1032, and the predicted amino acid sequence. A potential initiating methionine residue is indicated in bold-faced type. The box around amino acids 43 to 77 indicates the region that is conserved with members of the T protein family. Also boxed, and indicated in bold-faced type, are the seven copies of the serine-proline motif and the single copy of the threonine-proline motif. The heavy underline commencing under amino acid 207 and extending to amino acid 224 indicates the location and extent of the bipartite nuclear targeting sequences found in the protein. The targeting sequences are staggered by one amino acid, so that they almost completely overlap over their entire length. A TAA stop codon is boxed. The two potential poly-(A) addition sites are likewise indicated. The poly-(A) tail is indicated in bold-faced type.

```
AAAATGAACGAGCCCACGTACAACTACAATGACTTTTTGATGCCTGCATTGGCAGATGGA 60
 k  M  N  E  P  T  Y  N  Y  N  D  F  L  M  P  A  L  A  D  G  20

CAACCTGATATTTCAATTGAAATAAGTGAAATTACGAAGAAACAGTGGAATCAACTTATT 120
 Q  P  D  I  S  I  E  I  S  E  I  T  K  K  Q  W  N  Q  L  I  40

GGAATTCATGAAGTTGTAGTCTACAAAGACTGCAGAAAACTCTTCCCAGCTGCCAAATTC 180
 G  I  H  E  V  V  V  Y  K  D  C  R  K  L  F  P  A  A  K  F  60

TCAGTACGTGGTCTCGATCCGGCTCTCCTCTACTCGGTGGAGATTTGCTTTGATGTTCTC 240
 S  V  R  G  L  D  P  A  L  L  Y  S  V  E  I  C  F  D  V  L  80

AGCCCGTTCACGTTCTCGTATAATAAGAAACTCGAAAAGTGGATTCCAACGAAAACACCC 300
 S  P  F  T  F  S  Y  N  K  K  L  E  K  W  I  P  T  K  T  P  100

GTCAATTGTCAATCTTCGGCGCTGTCTTCAATGATTTGTGAAACAGGAGATTGGTTTATG 360
 V  N  C  Q  S  S  A  L  S  S  M  I  C  E  T  G  D  W  F  M  120

AATGGAGATATGGATTTTGGTGCATTGAATATTACAAGCTCCCACGAAAAAGCCAAGGAG 420
 N  G  D  M  D  F  G  A  L  N  I  T  S  S  H  E  K  A  K  E  140

AAAATCGAACAAAACAGGATTGCTGTGTGTCTCTCCTCGAGAACAATGTATCAAATCCGA 480
 K  I  E  Q  N  R  I  A  V  C  L  S  S  R  T  M  Y  Q  I  R  160

GTTGTGTTGTATGAGATGCAAAACGACGAACTGATATACCTGAAGGAAACGAAATTCGAC 540
 V  V  L  Y  E  M  Q  N  D  E  L  I  Y  L  K  E  T  K  F  D  180

GAAATGATATTCATCACTTGCACTGAATATCGTGGACAAGTCAATCGTGAATTGAAGAAG 600
 E  M  I  F  I  T  C  T  E  Y  R  G  Q  V  N  R  E  L  K  K  200

AGTATGAATCGTTTCTCAAGAAAAAGAAAGTATGAAATCAGTGGAACAAAGTGCCAGCGT 660
 S  M  N  R  F  S  R  K  R  K  Y  E  I  S  G  T  K  C  Q  R  220

TCGAAACGTCCAGCAAAACTTCCTGTCGAATCTCTAGATGAGAAGGAACTAAAAAATGCT 720
 S  K  R  P  A  K  L  P  V  E  S  L  D  E  K  E  L  K  N  A  240

GAAAAGGGAATCATGAGTTTGCTTGAATCATCACCTTCGTCACCATCGTACCTTTATCAA 780
 E  K  G  I  M  S  L  L  E  S  S  P  S  S  P  S  Y  L  Y  Q  260

TATCAATACAATTCTCCACAATCAATACCATCTTCTATTTATGACTCTCTCTACCCATCG 840
 Y  Q  Y  N  S  P  Q  S  I  P  S  S  I  Y  D  S  L  Y  P  S  280

TCTCCAGAAGAATATCAGATGGTTCATACTTTTCCACAAACAGACCAATTATCTCCTAGT 900
 S  P  E  E  Y  Q  M  V  H  T  F  P  Q  T  D  Q  L  S  P  S  300

TTTCAAACTTCGCCCACAGTCCATTATGATTTTTATAATCAGCAATATTATGACAATTCA 960
 F  Q  T  S  P  T  V  H  Y  D  F  Y  N  Q  Q  Y  Y  D  N  S  320

TTTCATGAAAACTATTTCCAGTCATACAATCGGTTATAACTTTCGATTTATGCTTATGTC 1020
 F  H  E  N  Y  F  Q  S  Y  N  R  L  *                       332

TAGATTTATCCGGTTTCAAATTACTAACCTGTTTTTCTTTTTTATCACAACTAACAAGCT 1080

CACGTTTTATCATTCTTGAAATTTTTTCAATAAATAATGAAAATAAAATCAAAAAAAAAAA 1140

AAAAAA                                                        1146
```

transcription factors in the database were obtained. However, some of the weak similarities that were obtained were to proteins that had multiple copies of the motif "Ser Pro" situated throughout the protein. Examination of the *C. elegans* sequence showed that it too had multiple copies of this motif (Figure 27).

Suzuki (1989) has proposed the existence of the beta turn as a DNA-binding unit of protein. This structure is composed of the amino acids S-P-X-X or T-P-X-X, where the X indicates a residue which is usually basic. Suzuki (1989) has compared the frequency with which these sequences occur in gene regulatory proteins to the frequency with which they occur in "general" (i.e. non-DNA binding) proteins and in proteins that bind DNA in a non-regulatory role. Suzuki notes that these sequences are found more frequently in gene regulatory proteins (e.g. homeotic gene products and steroid hormone receptors) than they are found in DNA binding proteins that are not directly involved in gene regulation (e.g. core histones), or in general proteins.

The protein predicted from the *C. elegans* sequence contains seven copies of the S-P-X-X motif, and a single copy of the T-P-X-X motif (Figure 27). This observation is compatible with the notion that the *C. elegans* protein might function to bind DNA, possibly in a gene regulatory fashion. I have named the *C. elegans* gene ***spm-1***. The ***spm***-prefix stands for **S**erine-**P**roline **M**otif.

Because the Spm-1 peptide contained a number of copies of the **ser-pro** motif, and evidence had been presented that this motif enabled the formation of a beta turn, I examined whether the Spm-1 peptide was predicted to form beta turns using the program BETATURN running under PCGENE (V. 6.8). This program uses the method of Chou and Fasman (1979) to predict the occurrence of beta turns in an amino acid sequence. The program predicted the existence of 42 beta turns over the length of the peptide. Those sequences assigned the highest probability of forming a beta turn were located between amino acids 240 and 300,

which are the regions of the protein containing the majority of the **ser-pro** motifs (Figure 27). Figure 28 illustrates the results of the beta turn predictions.

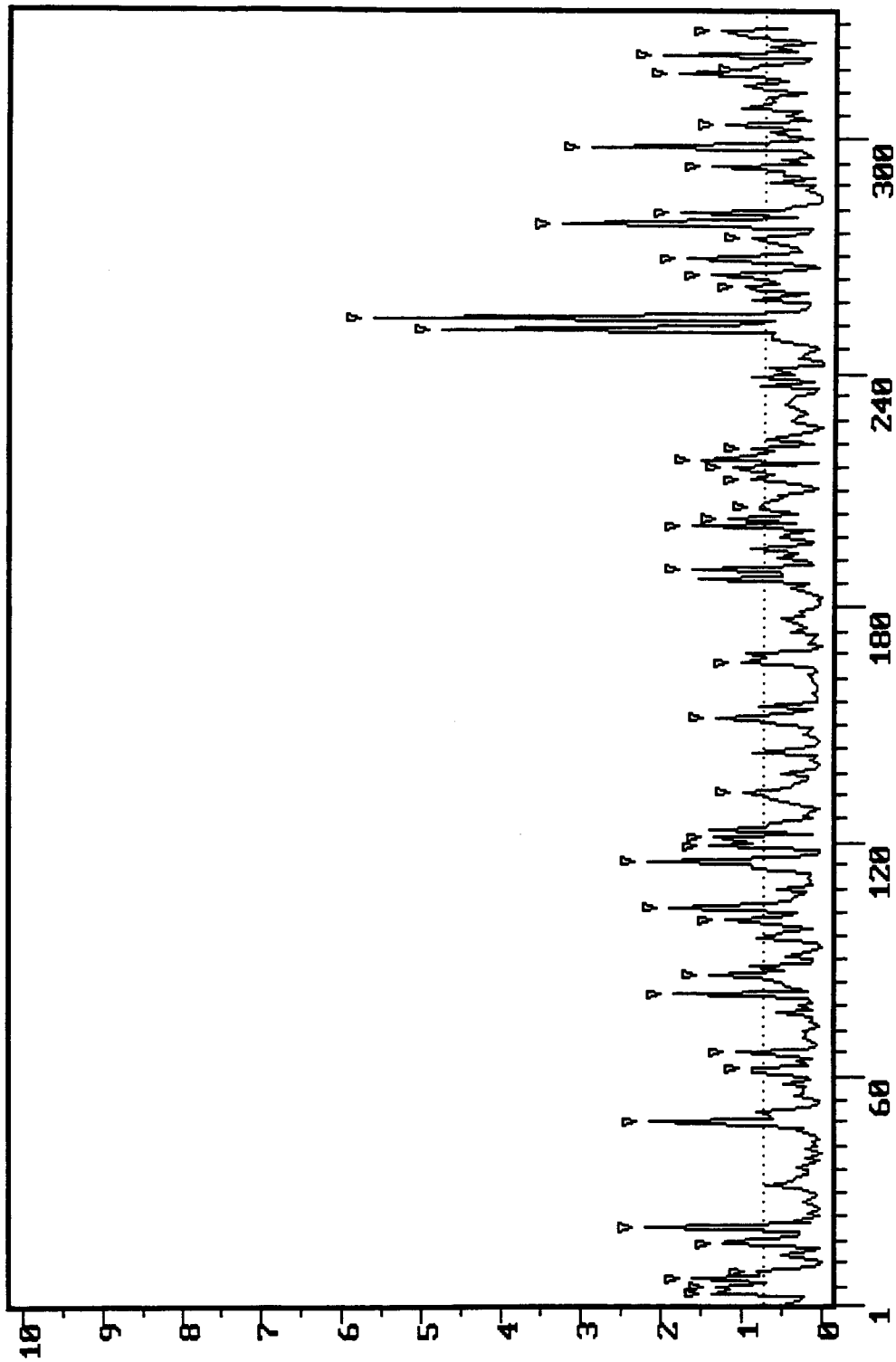
The possibility that the predicted Spm-1 peptide might interact with DNA prompted me to examine the protein for the existence of a nuclear targeting sequence. This was accomplished by analyzing the Spm-1 amino acid sequence with Version 11 of PROSITE (Bairoch, 1993) running under release 6.8 of PCGENE. PROSITE detected a possible "bipartite" nuclear targeting sequence within the *C. elegans* protein. The location and extent of the nuclear targeting sequence is indicated in Figure 27.

I next determined the sequence of the 4.2 kb *Hin*dIII fragment, pCes1016, that I had used as a probe to recover the Class A, Class B and Class C cDNA clones. This was done because cDNAs representing at least two separate genes had been recovered using pCes 1016 as a probe. Therefore, portions of at least two genes must be contained within this restriction fragment, suggesting the density of genes located on pCes1016 was high. Furthermore, pCes1016 sequence would elucidate the orientation of *spm-1* and *gdl-1* with respect to each other and the genome map. Sequence was obtained using M13 -21 primers on unidirectional deletion derivatives constructed from pCes1016.

The sequence of pCes1016 contained the entire *spm-1* coding sequence found in pCes1032. There are two introns in the *spm-1* gene. One of these is 53 bp long, while the other is 424 bp in length. The 5' splice donor and 3' splice acceptor sequences of both introns closely resemble the *C. elegans* consensus (Emmons, 1988). The positions of the introns are indicated in Figure 29. Immediately upstream of the potential initiating ATG is a sequence that perfectly matches the consensus for translation initiation (A/c - A/g - a/c - A/C - ATG). Three bases upstream of the consensus for translation initiation is found the sequence TTCCAG, which is a near perfect 3' splice acceptor sequence.

Figure 28:

Results of the predictions of beta turns within the Spm-1 protein using the method of Chou and Fasman (1979). The X axis represents coordinates along the length of the protein in amino acids. The Y axis values represent the probability of a beta turn [P(turn)] X 10^{-4}. Probable beta turns are those with a P(turn) above 0.75 X 10^{-4} which represents a cutoff value approximately equal to 1.5 times the average probability of any tetrapeptide to be in the beta turn conformation. These probable turns are indicated on the Figure by the inverted arrowheads above some of the peaks. The dotted line indicates the cutoff value for probable beta turns.

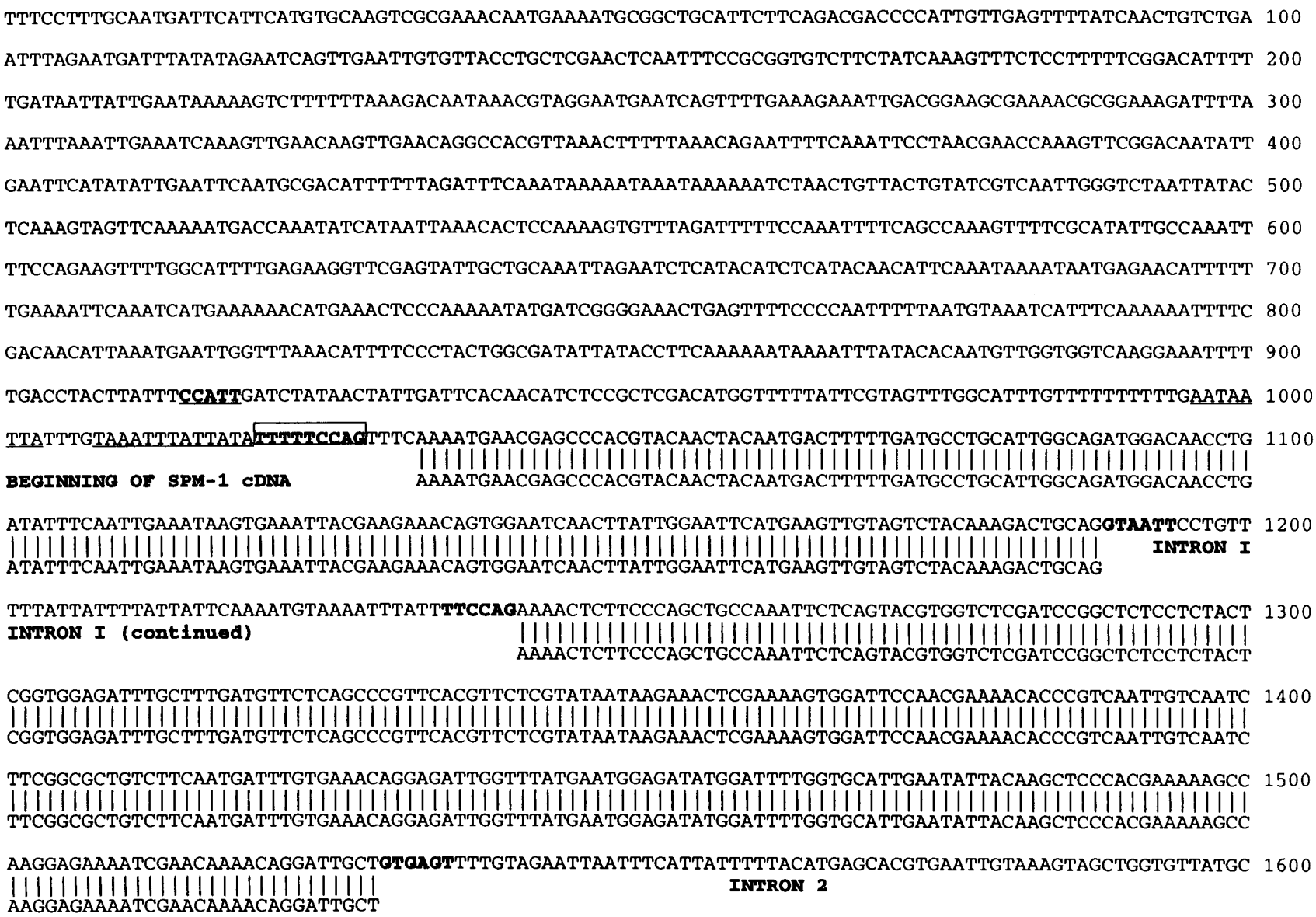

10
9
8
7
6
5
4
3
2
1
0
1
60
120
180
240
300

This may indicate that the *spm-1* cDNA sequence does not contain the 5' end of the *spm-1* gene, or that the *spm-1* mRNA is trans-spliced to a splice leader sequence. The latter possibility was supported by the presence of a putative initiating methionine codon in the context of a perfectly conserved consensus for translation initiation. T and A rich sequences, possibly corresponding to the TATA box, are found between 17 and 41 bases upstream of the putative initiating ATG. A sequence with similarity to the CCAAT motif was found approximately 110 nucleotides upstream of the start of the cDNA sequence.

pCes1016 also contained sequence that aligned perfectly with sequence determined from the *gdl-1* cDNA clones pCes1051 and pCes1054. However, pCes1016 does not contain the entire sequence of the *gdl-1* gene; only 427 bp of the 5' end of the cDNA sequence is contained within pCes1016. Comparison of the pCes1016 sequence and that determined from pCes1051 and pCes1054 reveals that this portion of the gene contains one intron of 47 bp. Both the 5' splice donor and 3' splice acceptor sites of this intron closely resemble the consensus for *C. elegans* (Emmons, 1988). A TTTCAG *trans*-splice acceptor sequence is found in the genomic DNA sequence three nucleotides upstream of the putative initiating methionine. T and A rich sequences, perhaps corresponding to the TATA box, are located immediately upstream of the splice acceptor site, and a sequence that perfectly matches the CCAAT box motif is found 215 bp upstream of the *trans*-splice acceptor sequence. The organization of the *spm*-1 gene and the *gdl-1* gene with respect to pCes1016 is indicated in Figure 29.

Figure 29:

DNA sequence of pCes1016 aligned with cDNA sequence from the *spm-1* and *gdl-1* genes. The orientation of this sequence is from left to right on the genome map. Therefore, both the *spm-1* gene and the *gdl-1* gene have the same transcriptional polarity, which is opposite that of *hdl-1* and *unc-22*, and the same as *nhe-1*. The alignment of pCes1016 sequence with the *spm-1* cDNA commences at pCes1016 nucleotide position 1034 and extends to pCes1016 nucleotide position 2644. The alignment of pCes1016 sequence with the *gdl-1* cDNA commences at pCes1016 nucleotide position 3681, and continues until the end of the pCes1016 sequence, at nucleotide position 4154. The positions of putative CAAT boxes in the suspected 5' regions of both genes are indicated by bold-faced underlined type. The positions of putative TATA boxes are indicated by underlined type in the putative 5' regions of both genes. Putative (in the case of the *spm-1* gene) or demonstrated (in the case of the *gdl-1* gene) *trans*-splice acceptor sites are indicated by boxed bold-faced type. *Cis*-splicing donor and acceptor sequences are indicated in bold-faced type, and the positions and extents of the introns are indicated. The terminal *Hin*dIII site (**AAGCTT**) is indicated in bold-faced type in the *gdl-1* cDNA sequence and in the pCes1016 sequence.


TTTCCTTTGCAATGATTCATTCATGTGCAAGTCGCGAAACAATGAAAATGCGGCTGCATTCTTCAGACGACCCCATTGTTGAGTTTTATCAACTGTCTGA 100
ATTTAGAATGATTTATATAGAATCAGTTGAATTGTGTTACCTGCTCGAACTCAATTTCCGCGGTGTCTTCTATCAAAGTTTCTCCTTTTTCGGACATTTT 200
TGATAATTATTGAATAAAAAGTCTTTTTTAAAGACAATAAACGTAGGAATGAATCAGTTTTGAAAGAAATTGACGGAAGCGAAAACGCGGAAAGATTTTA 300
AATTTAAATTGAAATCAAAGTTGAACAAGTTGAACAGGCCACGTTAAACTTTTTAAACAGAATTTTCAAATTCCTAACGAACCAAAGTTCGGACAATATT 400
GAATTCATATATTGAATTCAATGCGACATTTTTTAGATTTCAAATAAAAATAAATAAAAAATCTAACTGTTACTGTATCGTCAATTGGGTCTAATTATAC 500
TCAAAGTAGTTCAAAAATGACCAAATATCATAATTAAACACTCCAAAAGTGTTTAGATTTTTCCAAATTTTCAGCCAAAGTTTTCGCATATTGCCAAATT 600
TTCCAGAAGTTTTGGCATTTTGAGAAGGTTCGAGTATTGCTGCAAATTAGAATCTCATACATCTCATACAACATTCAAATAAAATAATGAGAACATTTTT 700
TGAAAATTCAAATCATGAAAAAACATGAAACTCCCAAAAATATGATCGGGGAAACTGAGTTTTCCCCAATTTTTAATGTAAATCATTTCAAAAAATTTTC 800
GACAACATTAAATGAATTGGTTTAAACATTTTCCCTACTGGCGATATTATACCTTCAAAAAATAAAATTTATACACAATGTTGGTGGTCAAGGAAATTTT 900
TGACCTACTTATTTCCATTGATCTATAACTATTGATTCACAACATCTCCGCTCGACATGGTTTTTATTCGTAGTTTGGCATTTGTTTTTTTTTTGAATAA 1000
TTATTTGTAAATTTATTATATTTTTTCCAGTTTCAAAATGAACGAGCCCACGTACAACTACAATGACTTTTTGATGCCTGCATTGGCAGATGGACAACCTG 1100
BEGINNING OF SPM-1 cDNA
AAAATGAACGAGCCCACGTACAACTACAATGACTTTTTGATGCCTGCATTGGCAGATGGACAACCTG
ATATTTCAATTGAAATAAGTGAAATTACGAAGAAACAGTGGAATCAACTTATTGGAATTCATGAAGTTGTAGTCTACAAAGACTGCAGGTAATTCCTGTT 1200
INTRON I
ATATTTCAATTGAAATAAGTGAAATTACGAAGAAACAGTGGAATCAACTTATTGGAATTCATGAAGTTGTAGTCTACAAAGACTGCAG
TTTATTATTTTATTATTCAAAATGTAAAATTTATTTTCCAGAAAACTCTTCCCAGCTGCCAAATTCTCAGTACGTGGTCTCGATCCGGCTCTCCTCTACT 1300
INTRON I (continued)
AAAACTCTTCCCAGCTGCCAAATTCTCAGTACGTGGTCTCGATCCGGCTCTCCTCTACT
CGGTGGAGATTTGCTTTGATGTTCTCAGCCCGTTCACGTTCTCGTATAATAAGAAACTCGAAAAGTGGATTCCAACGAAAACACCCGTCAATTGTCAATC 1400
CGGTGGAGATTTGCTTTGATGTTCTCAGCCCGTTCACGTTCTCGTATAATAAGAAACTCGAAAAGTGGATTCCAACGAAAACACCCGTCAATTGTCAATC
TTCGGCGCTGTCTTCAATGATTTGTGAAACAGGAGATTGGTTTATGAATGGAGATATGGATTTTGGTGCATTGAATATTACAAGCTCCCACGAAAAAGCC 1500
TTCGGCGCTGTCTTCAATGATTTGTGAAACAGGAGATTGGTTTATGAATGGAGATATGGATTTTGGTGCATTGAATATTACAAGCTCCCACGAAAAAGCC
AAGGAGAAAATCGAACAAAACAGGATTGCTGTGAGTTTTGTAGAATTAATTTCATTATTTTTACATGAGCACGTGAATTGTAAAGTAGCTGGTGTTATGC 1600
INTRON 2
AAGGAGAAAATCGAACAAAACAGGATTGCT

131(b)

```
ACCATAACTTAAATGCAATCAATCATATTGACGCAAATAAGTGATTTTTTGAAATCATTGTGAAAGCTCGCTCCATGGTCAGGTGAATATGCTCCGCCCT 1700

CTAGCCATGGGTCTCGTTAGGTTTCTGGAGACAAAACCGTAAATTCAAACCTTTTCAATTCGTTCTGCAGACTTCACCAATTTACAGGTTTTCCGCCAAA 1800
                                          INTRON 2 (continued)

ATCAAAAACCTACCGAGACCCAACGTTTTGTGGCGGGGCGTTTTTGCACCTTGATTTCAAAAGGGGCATTTTTTGCGTCCCCAATAATCAGCATGTTTGA 1900

AAACCAAAACGATTAATTGCGCAATGGTCATTTAGACTTTTTTCAAATTTTCAGGTGTGTCTCTCCTCGAGAACAATGTATCAAATCCGAGTTGTGTTGT 2000
                                                    ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                    GTGTGTCTCTCCTCGAGAACAATGTATCAAATCCGAGTTGTGTTGT

ATGAGATGCAAAACGACGAACTGATATACCTGAAGGAAACGAAATTCGACGAAATGATATTCATCACTTGCACTGAATATCGTGGACAAGTCAATCGTGA 2100
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATGAGATGCAAAACGACGAACTGATATACCTGAAGGAAACGAAATTCGACGAAATGATATTCATCACTTGCACTGAATATCGTGGACAAGTCAATCGTGA

ATTGAAGAAGAGTATGAATCGTTTCTCAAGAAAAAGAAAGTATGAAATCAGTGGAACAAAGTGCCAGCGTTCGAAACGTCCAGCAAAACTTCCTGTCGAA 2200
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTGAAGAAGAGTATGAATCGTTTCTCAAGAAAAAGAAAGTATGAAATCAGTGGAACAAAGTGCCAGCGTTCGAAACGTCCAGCAAAACTTCCTGTCGAA

TCTCTAGATGAGAAGGAACTAAAAAATGCTGAAAAGGGAATCATGAGTTTGCTTGAATCATCACCTTCGTCACCATCGTACCTTTATCAATATCAATACA 2300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TCTCTAGATGAGAAGGAACTAAAAAATGCTGAAAAGGGAATCATGAGTTTGCTTGAATCATCACCTTCGTCACCATCGTACCTTTATCAATATCAATACA

ATTCTCCACAATCAATACCATCTTCTATTTATGACTCTCTCTACCCATCGTCTCCAGAAGAATATCAGATGGTTCATACTTTTCCACAAACAGACCAATT 2400
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTCTCCACAATCAATACCATCTTCTATTTATGACTCTCTCTACCCATCGTCTCCAGAAGAATATCAGATGGTTCATACTTTTCCACAAACAGACCAATT

ATCTCCTAGTTTTCAAACTTCGCCCACAGTCCATTATGATTTTTATAATCAGCAATATTATGACAATTCATTTCATGAAAACTATTTCCAGTCATACAAT 2500
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATCTCCTAGTTTTCAAACTTCGCCCACAGTCCATTATGATTTTTATAATCAGCAATATTATGACAATTCATTTCATGAAAACTATTTCCAGTCATACAAT

CGGTTATAACTTTCGATTTATGCTTATGTCTAGATTTATCCGGTTTCAAATTACTAACCTGTTTTTCTTTTTTATCACAACTAACAAGCTCACGTTTTAT 2600
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CGGTTATAACTTTCGATTTATGCTTATGTCTAGATTTATCCGGTTTCAAATTACTAACCTGTTTTTCTTTTTTATCACAACTAACAAGCTCACGTTTTAT

CATTCTTGAAATTTTTTCAATAAATAATGAAAATAAAATCAAAAGTGTAGAATTTGTTCTTAAAGGCGCATGAATTCATTCCAATGGGTCTCGTCGGGAG 2700
||||||||||||||||||||||||||||||||||||||||
CATTCTTGAAATTTTTTCAATAAATAATGAAAATAAAATCAAAAAAAAAAAAAAA          END OF SPM-1 cDNA

AAAAAGTTTATGGTAGCGCGTCTTGGTGCTCGGATTTTTTAATTATTTTTTTTCCATTTTAAAAATATTTTATGCTCTATGTTTGGTTATGGAAGCGACC 2800
```


ACCGAAAAAGCCACATTCTTCAAAAATTTCTGTCAAATCTCTGGGGGAGAAATAAAATTTAAAAATCGGTTGTCTTATTCTGAAGATTATTTACGAAGAA 2900
TCAAAGATGTCGATCAGTAAAACAAATCGAAAGAGAAACAAAAACAAAACTGGTATTCAAAATTCGTTGAATGAGATTTTTGACGCGAATTTAGCGAGAC 3000
ACTGTCTCGATTCTTTCTCAGATGACGTCGATTGACCCTGCTTCTTTTGTTTTGTTGCTATGCAGATCGTATGTCCTATGCAATAGCACCACACAATATG 3100
GGCGAAGATCAATGACACTCGTCCCACTCTTCTTTTTTCCTTCTTCAAATAGGCGGAGATCTTTNCTCTATTTTTCCCCCGAATGTTGCCTCTATTTCTC 3200
TCTTTCTCTAAATATCTATTTGCCAGTCACCTCCTCCTCCTCCTCCATGAAACGCTATCTTTTGCTTGCAGTTTCATCGGTCTCGCCCACCGATGATGGT 3300
TGCGCGTTGTCCTGTAAACAGTTTTGTGTTGGGCTCCTCTAGTTTTGAATAGTTATGGCGCCCGCGTCCCCAGGACGAGCTTTTCGTGGCCTAGAAGCGT 3400
TTGCCTACTCGCGGGCGGTGCCACATTTTTTTTAGGAAAAAAGGTTTTGATGCCTCTGCCAATTTTTTTTGATTTAAATATTTTTAAACGAGCTATAACT 3500
TAAATAATTTCAAACTGCCGCTGAAAATTATTAAAGGCGCATGGATTTCGTGACGAGACCTCCTGGCATTTTCCTTACTTTTCGATTCCGCCCGGATCTT 3600
TTCCTCCTTTTTGCACTTTGGCTTCTTATGTAAAATCTCTTATTTGTTTAAATTTTAATTACAAAATATTTTATTTTTCAGGTAATGTTGAGCACTCTTTC 3700
BEGINNING OF GDL-1 cDNA
GAATTCCGAAGTTTGAGGTAATGTTGAGCACTCTTTC
CAGAGGCGCCCGCTCGGTCGCTGTCCGAAGCTATTCGGCCGCTACTCTCGATGCCCACTCCCAGGTGCTCGACGAGCAGAAGCCAATGGAGGAGCAGGTC 3800
CAGAGGCGCCCGCTCGGTCGCTGTCCGAAGCTATTCGGCCGCTACTCTCGATGCCCACTCCCAGGTGCTCGACGAGCAGAAGCCAATGGAGGAGCAGGTC
AACCCATCTTTCTACAAGATGGTCGACTTTTACTTCAACAAGGGAGCCGAGGTCATCGCTCCAAAACTCGCCGAGGAGCTCAAGTCCAACTCTTTGAGCC 3900
AACCCATCTTTCTACAAGATGGTCGACTTTTACTTCAACAAGGGAGCCGAGGTCATCGCTCCAAAACTCGCCGAGGAGCTCAAGTCCAACTCTTTGAGCC
AAAAGGATAAGAAGAATCTCGTGTCCGGAATTCTTGGAGCTATCAAGCCAGTTAACAAGGTAAGAATTTTGAGTCTCATTAATTTATCATATTTTATTTC 4000
INTRON I
AAAAGGATAAGAAGAATCTCGTGTCCGGAATTCTTGGAGCTATCAAGCCAGTTAACAAG
CTTCAGGTCCTTTACATCACCTTCCCAATCCGTCGTGACAATGGAGAGTTCGAGGTTATTGAGGCCTGGCGCGCACAGCATTCTGAGCACAGAACTCCAA 4100
GTCCTTTACATCACCTTCCCAATCCGTCGTGACAATGGAGAGTTCGAGGTTATTGAGGCCTGGCGCGCACAGCATTCTGAGCACAGAACTCCAA
CCAAGGGAGGTATCCGTTATTCCCTCGATGTGTGCGAAGATGAGGTCAAAGCTT
END OF PCES1016
4180
CCAAGGGAGGTATCCGTTATTCCCtCGATGTGTGCGAAGATGAGGTCAAAGCTT

PART B.

Identification and DNA sequence analysis of genes on the cosmid C18D3, and identification of candidate *let-56* genes.

INTRODUCTION

In spite of the intensive genetic screens for mutants that slowed or arrested normal development (as described in Section I and references therein), the only locus on C11F2 in which mutations of this type had been identified was *let-56* (Clark and Baillie, 1992). In the previous Section, I have described the identification and DNA sequence analysis of four genes located on C11F2. In addition to these four genes, Prasad had obtained evidence that there existed an additional gene (or genes) located on fragment P (Figure 11). Therefore, of a minimum of five genes located on C11F2, only one seemed to contribute a required function to normal development under laboratory conditions.

While sequencing the *unc-22* gene, Benian et al. (1989) sequenced approximately 13 kb of genomic DNA past the 3' end of *unc-22*. Subsequently, during analysis of this DNA sequence (deposited in Genbank under accession number L10351), D. Baillie and C. Fields independently discovered an open reading frame which they found exhibited similarity to a family of mammalian glucose transporter molecules. Subsequently, D. Baillie noted that there were three other open reading frames within this 13 kb of sequence that, when examined by Kyte - Doolittle hydrophobicity plots, appeared to encode potential integral membrane peptides.

A portion of the DNA sequenced by Benian et al. (1989) is contained within the cosmid C18D3, which had been examined by Prasad (1988). Prasad had shown that three *Pst*I restriction fragments from C18D3 hybridized to *C. briggsae* DNA, and two of these hybridized to *C. elegans* RNA on Northern blots (summarized in Figure 30). With this analysis, Prasad had obtained evidence for the existence of at least three genes. No mutations were known to correspond to any of the loci on C18D3. Therefore, of a minimum of eight genes (at least five genes on C11F2 and at least three genes on C18D3) residing in the interval between *let-56* and *unc-22* there existed only a single mutagenically defined locus, *let-56*.

I participated in a collaboration with J. Schein to recover and analyze mutations potentially affecting the genes on C18D3 and C11F2. Because previous genetic screens conducted by myself and others (as described in Section I) had failed to recover EMS induced point mutations in these loci, we chose as our mutagen formaldehyde, the effect of which appears to be the induction of small deficiencies (Johnsen, 1990). We reasoned that a formaldehyde screen for mutations that affected *unc-22* might yield small deficiencies that, in addition to *unc-22*, affected some or all of the genes located on C18D3 and C11F2. The phenotype of individuals homozygous for any such deficiency might then be correlated with the genes that were deleted.

A method was required to determine the extent of any deficiencies that we expected to recover. We desired deficiencies that broke in the region between *unc-22* and *let-56*, and that did not extend further than *unc-22* on the right. There are no genetic markers between *unc-22* and *let-56*, so a mapping strategy based upon complementation tests was not possible. Instead, our strategy was to apply PCR primer pairs specific for some of the genes identified on C18D3 or C11F2 in PCRs that used, as an amplification template, individuals homozygous for an *unc-22* mutation that had been recovered after treatment with formaldehyde (see Materials and Methods). If a PCR product was produced using a specific primer pair, then the *unc-22* mutation could not affect the primer annealing sites used in that particular PCR. The methodology has been diagrammed in Figure 3, and is explained in Materials and Methods.

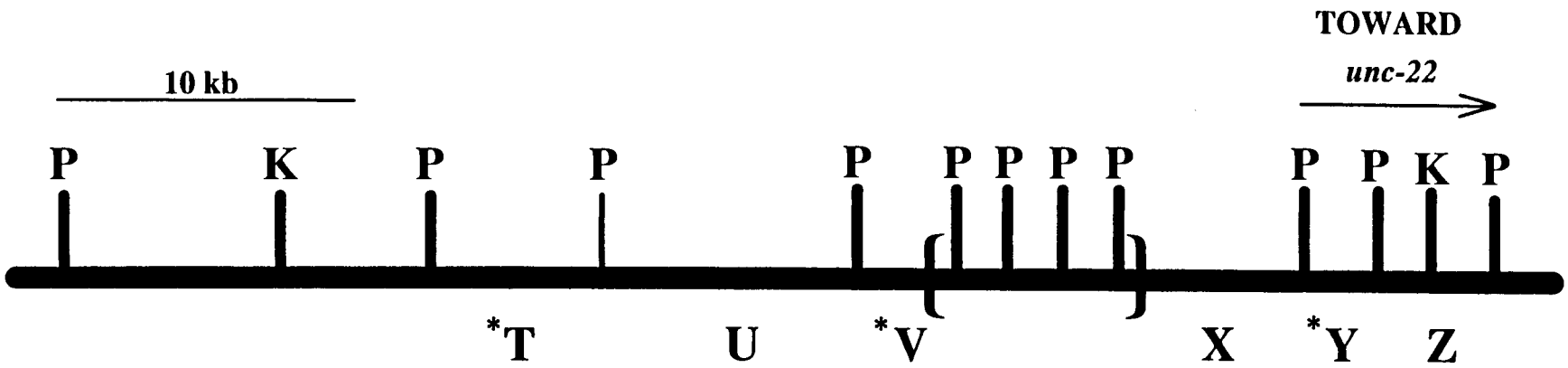


FIGURE 30:

A restriction map of the cosmid C18D3, adapted from Prasad (1988) and Prasad and Baillie (1989). K = *Kpn*I. P = *Pst*I. The direction in which *unc-22* lies is indicated by the arrow. The brackets on the heavy line indicate *Pst*I restriction fragments that were not mapped with respect to each other. The terminal *Pst*I sites are located within the vector. The *Pst*I fragments indicated by the asterisks were used by Prasad to probe Northern blots of total RNA. Fragment T hybridized to a 2.4 kb mRNA. Fragment V hybridized to a 2.6 kb mRNA. Fragment Y had hybridized to *C. briggsae* DNA on Southern blots, but did not hybridize to Northern blots of total RNA.

A paper describing some of the results presented in this section of the thesis has been published (J. E. Schein, M. A. Marra, G. M. Benian, C. Fields and D. L. Baillie, **Genome** (1993) **36**: 1148 - 1156).

RESULTS

Recovery of formaldehyde induced *unc-22* mutations: A screen for formaldehyde induced *unc-22* mutations was conducted by J. Schein. From approximately 190 000 wild-type chromosomes treated with 0.1% formaldehyde, she recovered 27 phenotypically wild-type F_1 individuals that twitched in nicotine. These individuals were presumed to be heterozygous for a recessive *unc-22* mutation. Strains were established for each of the 27 mutations.

Individuals from seventeen of the wild-type twitcher strains gave fertile Unc-22 progeny, while individuals from the remaining ten strains produced no Unc-22 progeny. These latter individuals were suspected of carrying deficiencies that affected *unc-22* and at least one neighbouring essential gene. J. Schein tested individuals from each of the ten strains for complementation with *let-653* (see Figures 6 and 12) and found that all but one failed to complement *let-653*. These deficiencies were therefore too large for our purposes, and were not considered in this study. The single mutation that complemented *let-653* was found to fail to complement *let-56*. The phenotype of individuals homozygous for this mutation appeared indistinguishable from the phenotype of *let-56(s173)* homozygous individuals (see Section I for a complete description of the phenotype). This mutation was designated *sDf83*. J. Schein tested *sDf83* for complementation with *let-52*, the nearest essential gene marker to the right of *unc-22*. *sDf83* was found to complement *let-52*. For this reason, and because *sDf83* homozygous animals exhibited a phenotype indistinguishable from that of *let-56(s173)* homozygous individuals, we retained *sDf83* for further analysis.

Additional putative deficiencies potentially affecting genes on C18D3: In addition to the remaining 18 strains constructed by J. Schein, we had available to us 31 strains that each carried an *unc-22* mutation that had previously been recovered by past members of the Baillie laboratory. Each of the 31 strains segregated fertile

individuals that were homozygous for the mutation carried in that strain. Twenty-four of the thirty-one mutations had been isolated in formaldehyde screens similar to the one conducted by J. Schein (Moerman and Baillie, 1981). Six had been isolated following treatment with gamma radiation (D. L. Baillie, D.V. Clark and D.G. Moerman, personal communication). One mutation was isolated incidentally during a formaldehyde mutagenesis screen in another laboratory and supplied to us (A. Rose, personal communication). A total of 49 mutations were thus available to us to continue our study. The observation that 48 of the 49 available strains segregated fertile individuals homozygous for the mutation prompted us to consider two possibilities. Either the 48 mutations affected *unc-22* and no neighbouring gene(s), or any affected neighbouring genes were not required in a developmental process leading up to and including fertility. However, before we could draw inferences about potential functions of genes in the vicinity of *unc-22*, we needed to demonstrate that the open reading frames that we had predicted from the sequence of Benian et al. (1989) did indeed represent genes.

Prediction of genes and analysis of their products: Examination of the 13 kb of sequence to the left of *unc-22* for open reading frames and for typical nematode introns identified four putative genes. Amino acid sequences were predicted from these putative genes following removal of suspected introns. Database searches were conducted in which these amino acid sequences were used to search the non-redundant database at the NCBI and the **dbest** (Materials and Methods). The protein predicted from the first gene to the left of *unc-22* (coordinates 43 928 to 42 190 in the Benian et al. (1989) sequence), when used in a search of the non-redundant databank, exhibited weak similarity to a family of vertebrate glucose transporter proteins. No similarities were detected in **dbest**. The highest similarity found in the non-redundant databank was with the human GLUT4 protein, a glucose transporter isoform that responds to insulin (reviewed in Gould and Holman, 1993). An alignment of the

predicted *C. elegans* protein with the GLUT4 protein is shown in Figure 31. A partial length 870 bp cDNA corresponding to the *C. elegans* gene was recovered by G. Benian, and Northern blots of total RNA probed with this labeled cDNA have indicated a mRNA size of 2.4 kb (Schein et al., 1993). This demonstrated the existence of a transcript corresponding to the first gene predicted from the sequence of Benian et al. (1989). I have named this gene *gtl-1*, for glucose transporter-like. A Kyte - Doolittle hydrophobicity plot of the predicted Gtl-1 peptide indicating membrane-spanning domains, predicted using the method of Klein, Kanehisa and DeLisi (1985), is shown in Figure 32. A hydrophobicity plot of the GLUT4 protein is included to illustrate the similarity in secondary structure.

A search of the non-redundant database utilizing the protein product of the second gene predicted from the sequence of Benian et al. (1989) (nucleotide positions 46 267 to position 47 476) revealed weak similarity to members of a family of sodium dependent transporter proteins, including L-proline, choline, serotonin and dopamine transporters. The highest similarity discovered in searches of the non-redundant database was to a portion of the rat L-proline transporter. No similarities were found to sequences in the **dbest**. An alignment of the predicted *C. elegans* amino acid sequence with the N-terminal 320 amino acids of the rat L-proline transporter is shown in Figure 33. I demonstrated the existence of a transcript corresponding to this gene by using oligonucleotide primers predicted from the DNA sequence (primers Mys-1 and Mys-3; Table 2) to PCR amplify a 700 bp portion of a cDNA from the cDNA library of Barstead and Waterston (1989) (Materials and Methods). I gel purified this amplification product and supplied it to G. Benian, who subsequently labeled it and used it to probe Northern blots of total RNA. Hybridization to a 4.2 kb mRNA was detected. This gene has been named *ptl-1*, which stands for proline transporter-like. Kyte - Doolittle hydrophobicity plots of the predicted *ptl-1* peptide

Figure 31:

An alignment of the *C. elegans* predicted Gtl-1 peptide with the human insulin responsive glucose transporter GLUT4. Amino acid identities are represented by a double dot. Conservative amino acid replacements are indicated by a single dot. In this alignment the two proteins are 22 % identical and 60 % similar over the length of the *C. elegans* amino acid sequence (437 residues). The regions of the two proteins that contain motifs conserved in glucose transporters are boxed. The first pair of motifs occurs in the N terminal halves of the proteins , and the second pair of motifs occurs in the C terminal halves of the proteins. The repetition of these motifs between the two halves of the protein has led to the suggestion that duplication of an ancestral gene bearing a single pair of the motifs may have resulted in the production of this arrangement of motifs characteristic of the glucose transporter family of proteins (Gould and Holman, 1993). The human glucose transporter contains a domain bearing a resemblance to the leucine zipper motif. This domain is indicated in the GLUT4 protein by a box. The leucine zipper motif consists of four to five iterations of a leucine residue repeated every seventh amino acid (Landschultz et al., 1988), and may play a role in the dimerization of glucose transporter proteins (Landschultz et al., 1988; Buckland and Wild, 1989). Leucine and isoleucine residues in the GLUT4 leucine zipper are indicated in bold-faced type. The *C. elegans* protein exhibits only weak conservation in this region, and does not appear to contain a leucine zipper.

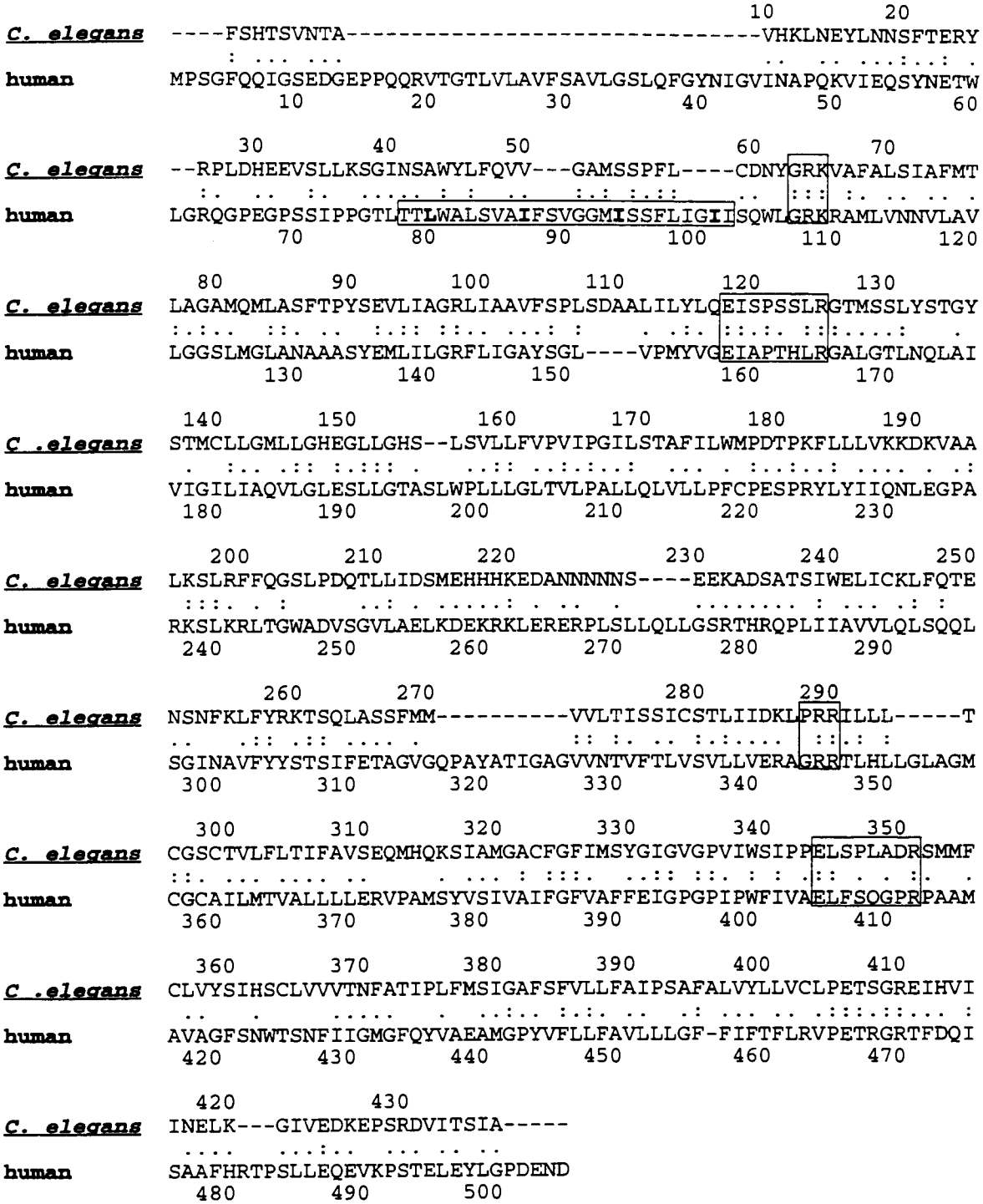
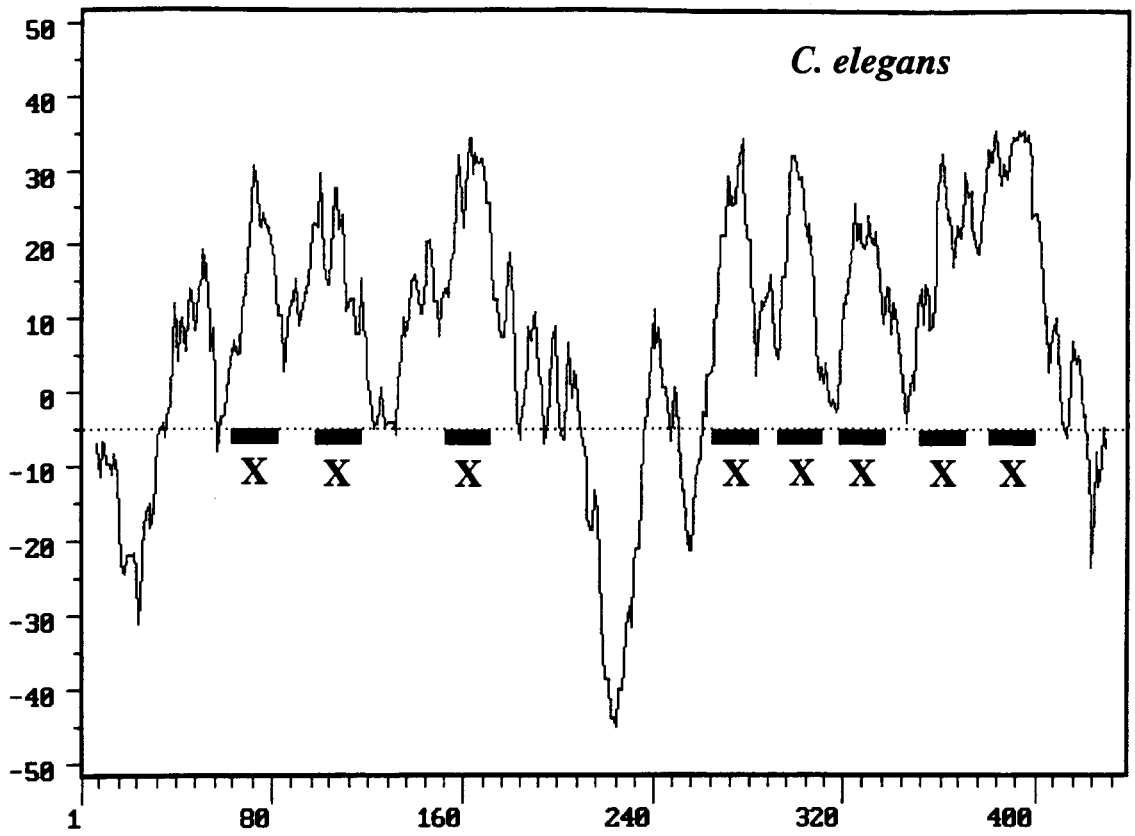

C. elegans
human
C. elegans
human
C .elegans
human
C. elegans
human
C. elegans
human
C .elegans
human
C. elegans
human

Figure 32:

Kyte - Doolittle hydropathy profiles of the predicted *C. elegans* Gtl-1 protein and the human GLUT4 protein. The dotted line indicates the cutoff for hydrophobic values. Values above the cutoff are hydrophobic. The location and extent of potential transmembrane domains, predicted using the method of Klein, Kanehisa and DeLisi (1985), are indicated by the bars with an **X** under them. The hydrophilic domain near the center of the GLUT4 protein is thought to be cytoplasmic, and divides the protein in half, defining the N and C terminal domains. The similarity between the hydrophobicity plots of the N and C terminal domains of the glucose transporters supports the suggestion that this family of transporters arose after a gene duplication event (Gould and Holman, 1993).


C. elegans
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
80
160
240
320
400
X X X X X X X X

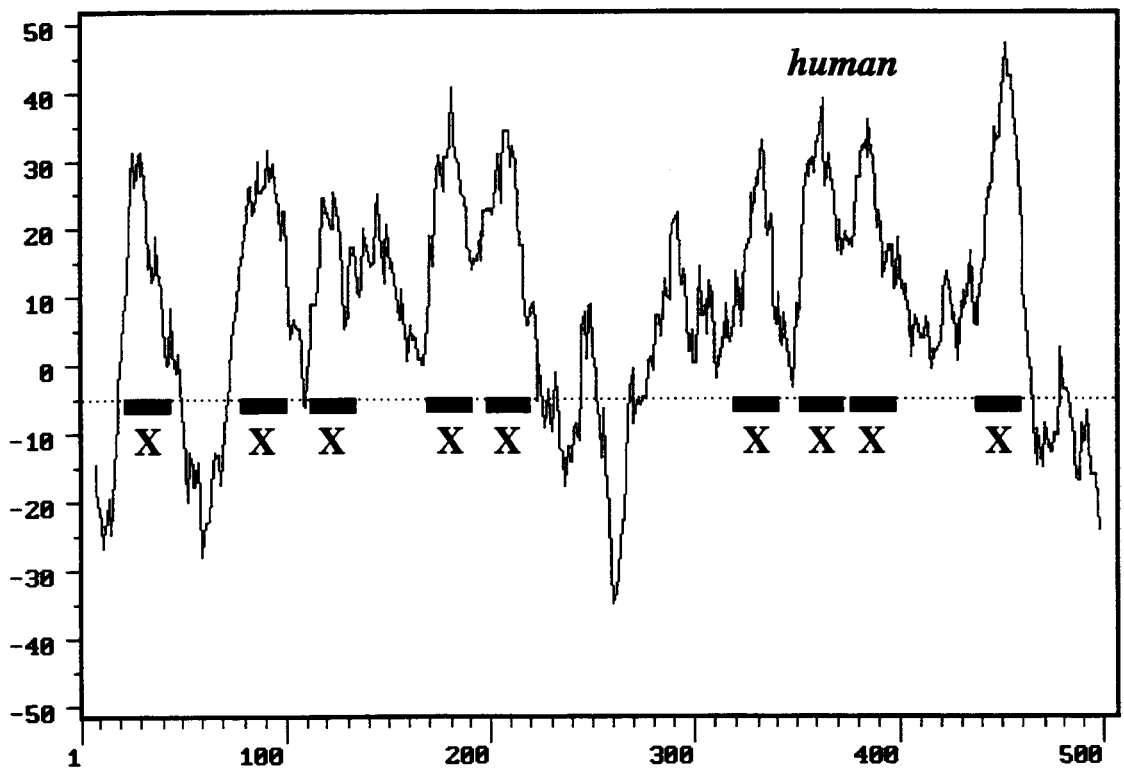
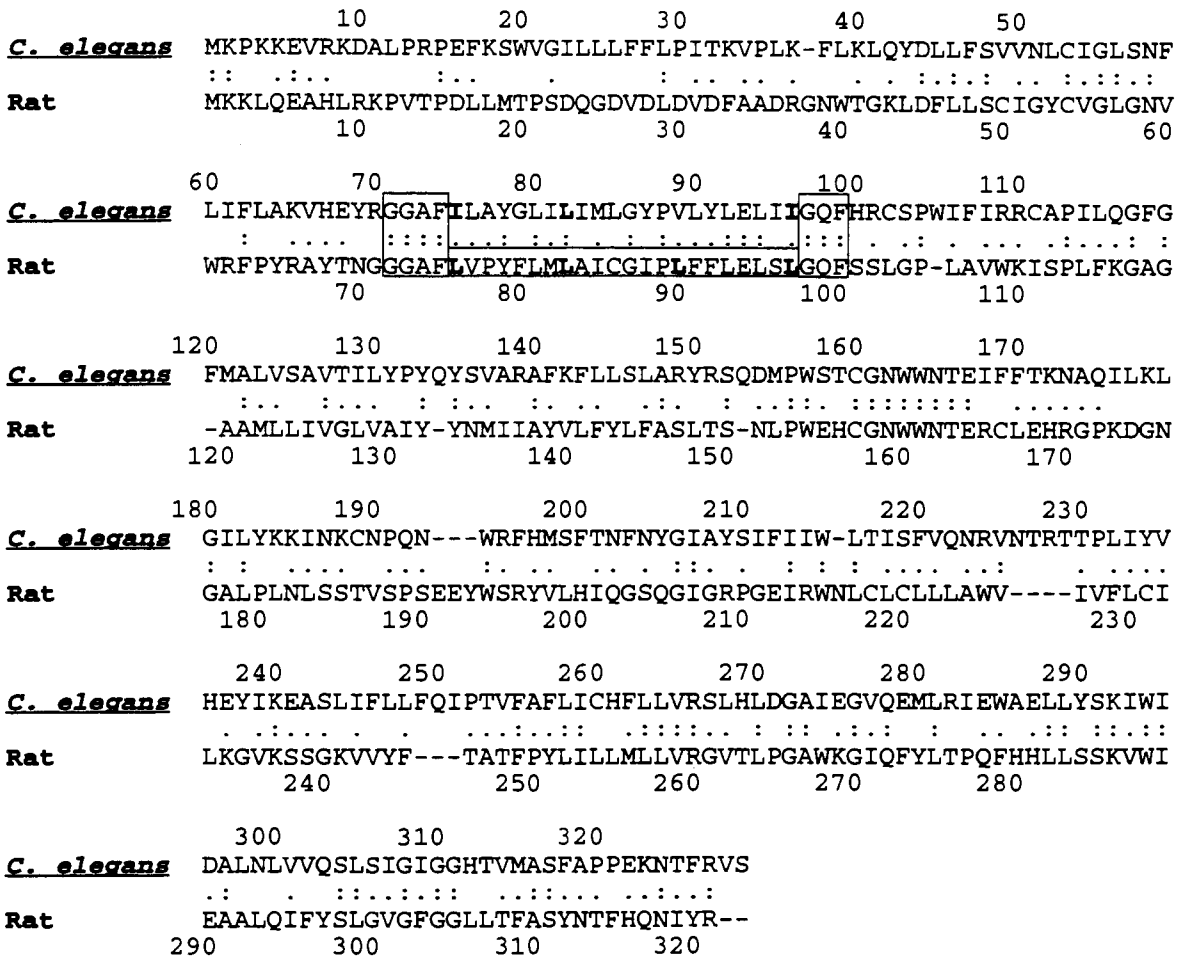

human
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
100
200
300
400
500
X X X X X X X X X

Figure 33:

Alignment of the *C. elegans* predicted Ptl-1 peptide with a portion of the rat L-proline transporter. The carboxyl terminal 317 amino acids of the rat transporter have been omitted from this alignment. The two proteins are 27 % identical over a span of 321 amino acids, which is approximately half the total length of the rat protein. When conservative amino acid replacements are considered, the two proteins are 66 % similar. The boxed region of the rat protein indicates a leucine zipper motif, which is a feature that has been found in many transmembrane proteins, including glucose transporters (McCormack et al., 1989). The human glucose transporter contains a domain resembling the leucine zipper motif (Figure 31). Leucines that form the zipper are indicated by bold-faced type. The leucine zipper region of the rat protein exhibits weak similarity with the *C. elegans* protein, with only one conserved leucine residue found. The regions flanking the zipper, indicated by boxes around *C. elegans* and rat amino acids, are also conserved.


C. elegans
Rat
C. elegans
Rat
C. elegans
Rat
C. elegans
Rat
C. elegans
Rat

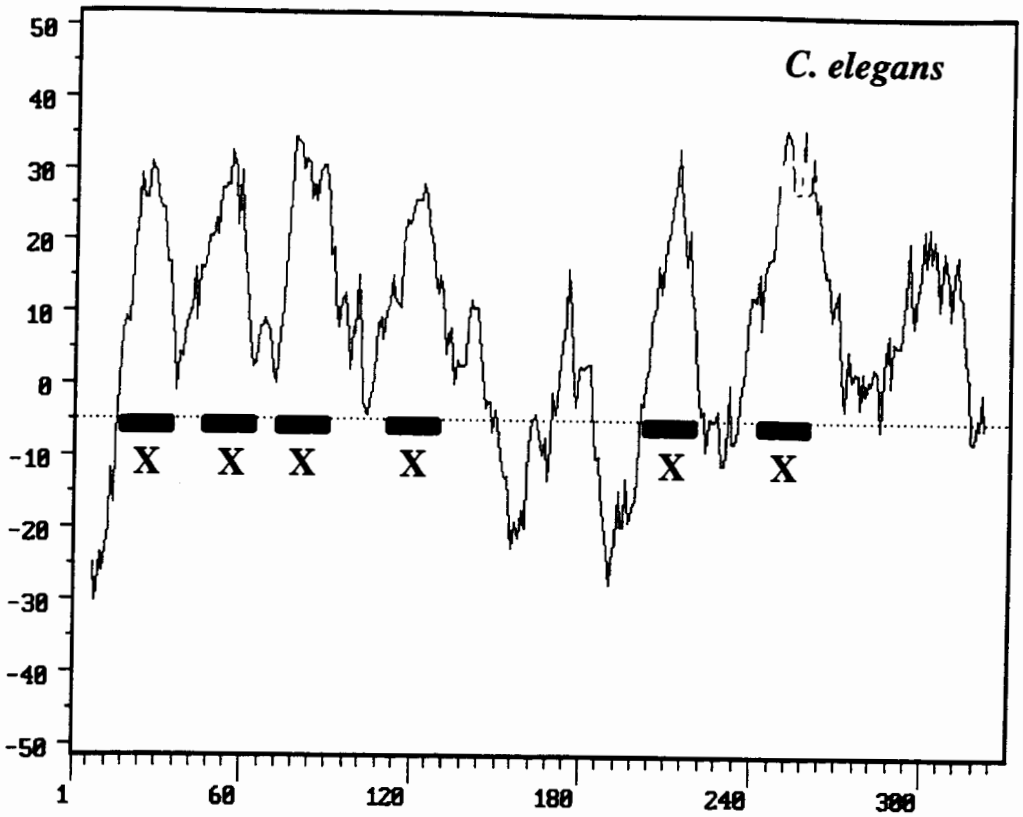
and the rat proline transporter are shown in Figure 34. Potential membrane-spanning domains are indicated.

Database searches utilizing the protein product of the third gene predicted from the sequence of Benian et al. (1989) (position 49 875 to position 52 239) failed to yield similarities to sequences in **dbest** and the non-redundant database at the NCBI. I was unsuccessful in attempts to PCR amplify a portion of a cDNA from the Barstead and Waterston (1989) cDNA library using oligonucleotide primers predicted from the DNA sequence (primers Dup-3 and Dup-4; Table 2). Due to its close proximity to *ptl-1*, it seemed possible that this "gene" was actually additional *ptl-1* gene sequence. I examined this possibility by attempting to PCR amplify a portion of a cDNA from the cDNA library using appropriately oriented oligonucleotide primers, one predicted from *ptl-1* sequence, and another predicted from this gene (primers Mys-1 and Dup-2; Table 2). No amplification product was detected. We named the gene *dup*, for **dup**licated, which reflects an unusual repeated domain of amino acids in the predicted Dup protein (Figure 35).

A search of the non-redundant database utilizing the protein product of the fourth gene predicted from the sequence of Benian et al. (1989) (position 53 438 to the end of the *unc-22* sequence) yielded weak similarities to a number of alcohol and sorbitol dehydrogenase enzymes. The highest similarity was to an alcohol dehydrogenase (ADH) enzyme from the bacterium *A. eutrophus*. An alignment of the predicted *C. elegans* protein to *A. eutrophus* ADH is shown in Figure 36. Identities to two cDNAs sequenced by the *C. elegans* sequencing consortium (cDNAs cm12a11 and cm01b8) were discovered in searches of the **dbest**. In addition, the sequence of Benian et al. (1989) was found to correspond, in part, to the sequence of a partial length cDNA that had been recovered and analyzed by Prasad (Prasad and Baillie, 1989; Prasad, personal communication). Prasad had used this cDNA to probe

Figure 34:

Kyte - Doolittle hydropathy profiles of the *C. elegans* Ptl-1 predicted peptide and the first 320 amino acids of the rat L- proline transporter. Putative transmembrane domains are indicated by the bars with an **X** underneath them.


C. elegans
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
60
120
180
240
300
X
X
X
X
X
X

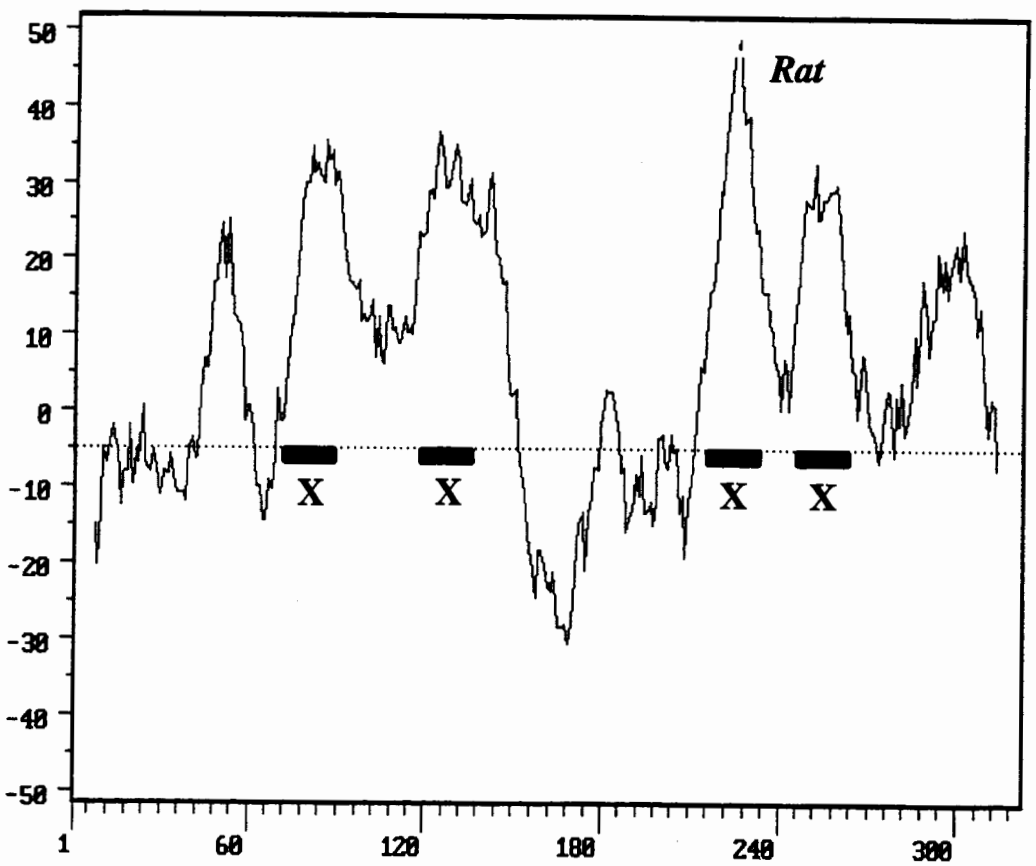

Rat
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
60
120
180
240
300
X
X
X
X

Figure 35:

(**A**). Amino acid sequence predicted from the **dup** open reading frame. The predicted **dup** protein is 388 amino acids in length, and bears no informative similarity to other sequences in the non-redundant database or the **dbest**. The protein consists primarily of a near perfect tandem duplication of amino acid residues. The repeated sequence is shown in (**A**). DUP A indicates the first repeat. It is 167 amino acids long, commencing at amino acid 24 and ending at amino acid 190. Sequences in the DUP A repeat are indicated in bold-faced type and are underlined. The second repeat, DUP B, extends 168 amino acids commencing at amino acid position 202, and ending at amino acid 369. Sequences in the DUP B repeat are also indicated in bold-faced type and are underlined.

(**B**). Alignment of the DUP A and DUP B repeat elements. DUP A and DUP B are 86 % identical and, when conservative amino acid substitutions are considered, are 95 % similar. The unusual repeated structure of **dup** is reflected in its hydrophobicity profile (Figure 37).

(A)

DUP A

MNVCVHICTACCSFFISISCSLQ**HFCNLIDFIKRSRPLAIAQFFRPKTIFLNSLFVMCFGSSWMLIGHIT**

MWPDDHIYDLIDEKFIQFHNTSSRDLAMIVANYEVRIYPVYDWSKSGILGMLIATLITTSIMISYVFFAQ

DUP B

KIHLSLKACTFSGAVKRLHSSLLKSLIAQTIIPLISTIIPCFVIWFLPLGVIIMECLRHYK**HVCNLICLT**

KFSRPLAIAQFFRPKTIFLNSLFVMCFGSSWLLIGHIIMWPDDHIYDLIDEKFKEIHNTNSRDLAMIVAN

YEVRIYPVNDWSKSGILGMLVAAVITNSIMIIYVFFARKIHLSLKACTLSCTTKRLHSSLLKSLIAQTII

PLISTIIPLFVVWFLPLAGDNYGVSQKFCIEDIGIPKR

(B)

```
                10        20        30        40        50        60
DUP A HFCNLIDFIKRSRPLAIAQFFRPKTIFLNSLFVMCFGSSWMLIGHITMWPDDHIYDLIDE
       : :::: ..: :::::::::::::::::::::::::::::::.:::::.::::::::::::
DUP B HVCNLICLTKFSRPLAIAQFFRPKTIFLNSLFVMCFGSSWLLIGHIIMWPDDHIYDLIDE
                10        20        30        40        50        60

                70        80        90       100       110       120
DUP A KFIQFHNTSSRDLAMIVANYEVRIYPVYDWSKSGILGMLIATLITTSIMISYVFFAQKIH
       :: ..:::::::::::::::::::: :::::::::::.:.::.:::: :::::.:::
DUP B KFKEIHNTNSRDLAMIVANYEVRIYPVNDWSKSGILGMLVAAVITNSIMIIYVFFARKIH
                70        80        90       100       110       120

              130       140       150       160
DUP A LSLKACTFSGAVKRLHSSLLKSLIAQTIIPLISTIIPCFVIWFLPL-G
       ::::::::.: ..:::::::::::::::::::::::: ::.:::::: :
DUP B LSLKACTLSCTTKRLHSSLLKSLIAQTIIPLISTIIPLFVVWFLPLAG
                130       140       150       160
```

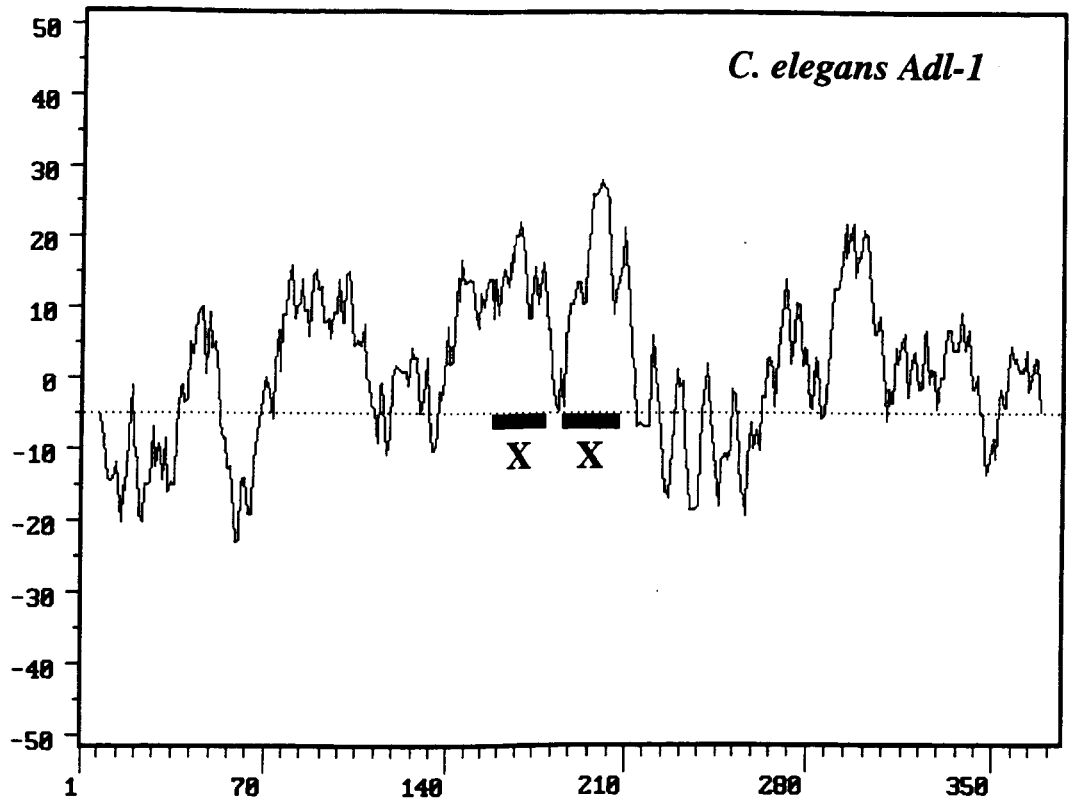
Figure 36:

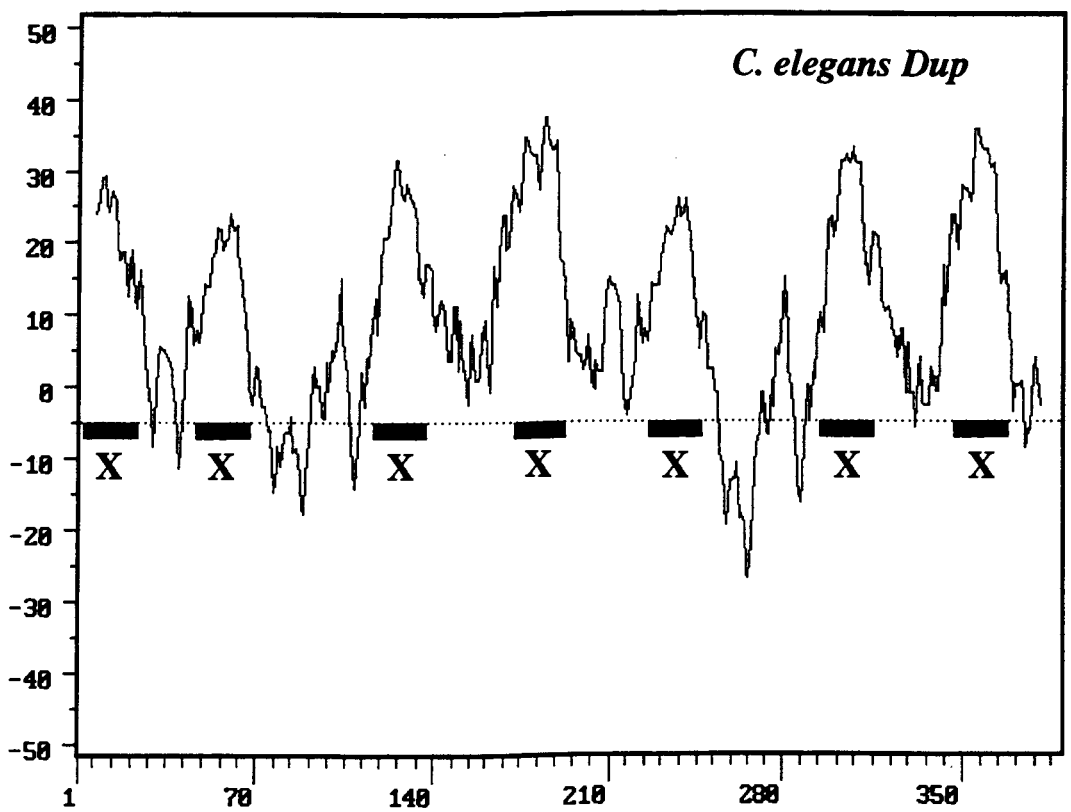
Alignment of the *C. elegans* Adl-1 predicted peptide with the ADH protein from the bacterium *A. eutrophus*. The two proteins are 23 % identical over 337 residues. When conservative amino acid replacements are considered, the similarity between the two proteins rises to 70%.


C. elegans
A. eutrophus

Figure 37:

Kyte - Doolittle hydrophobicity profiles of the *C. elegans* Adl-1 predicted protein and the *C. elegans* Dup predicted protein. The position and extent of putative transmembrane domains are indicated by the bars with an **X** underneath them. The ADH protein from *A. eutrophus*, which was found to exhibit weak similarity to the *C. elegans* Adl-1 peptide, did not contain predicted membrane spanning domains. The Dup protein was originally so named because of the presence of a duplication in the amino acid sequence that is reflected in the hydrophobicity profile of the protein. The region of the protein spanning the second, third and fourth transmembrane domains very closely resembles that region of the protein that contains the fifth, sixth and seventh transmembrane domains.


C. elegans Adl-1
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
70
140
210
280
350
X
X


C. elegans Dup
50
40
30
20
10
0
-10
-20
-30
-40
-50
1
70
140
210
280
350
X
X
X
X
X
X
X

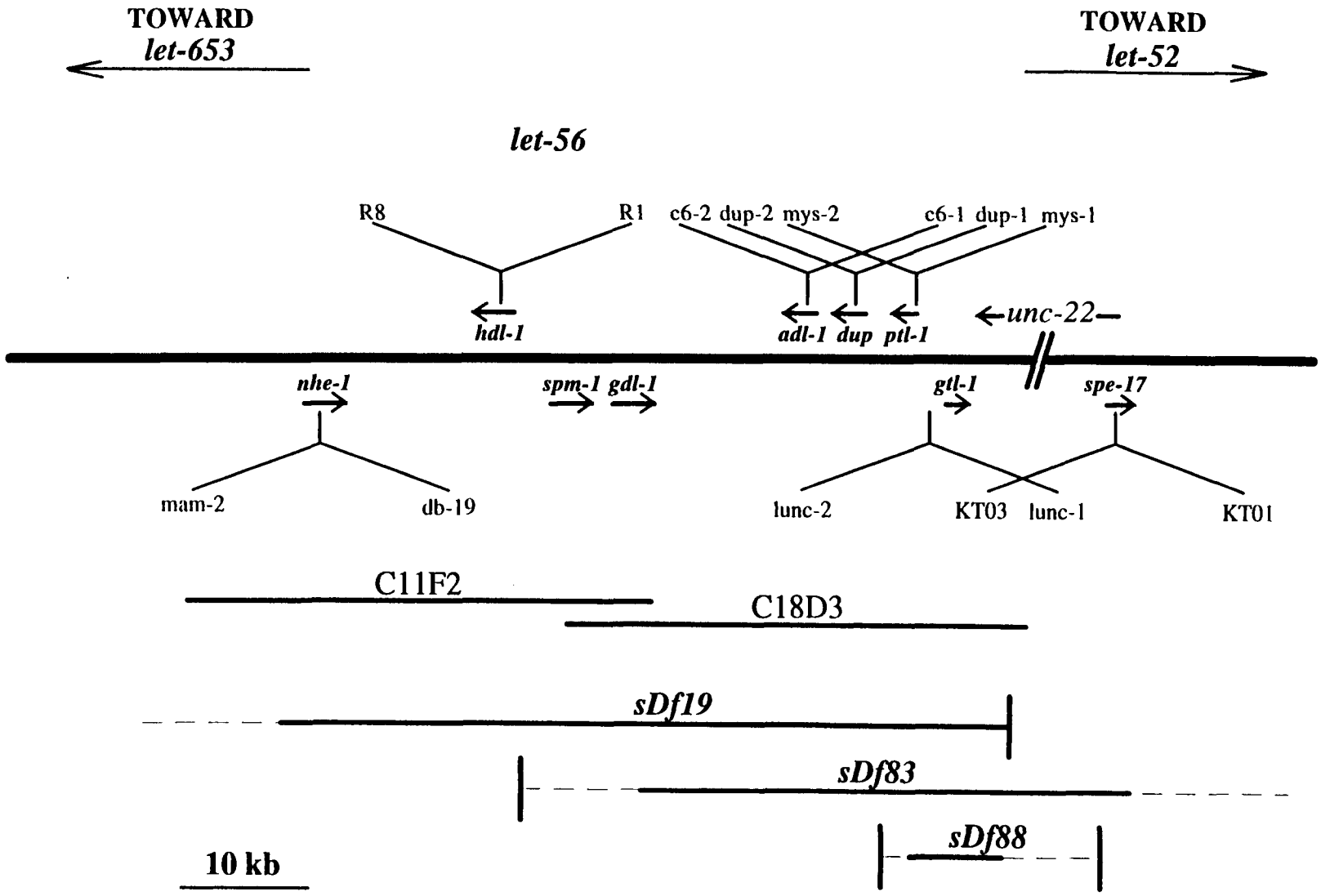
Northern blots of total RNA, and had detected a 2.4 kb message. I have named this gene *adl-1*, for **a**lcohol **d**ehydrogenase-**l**ike. A Kyte - Doolittle hydrophobicity plot of the predicted *adl-1* peptide indicating potential transmembrane domains is shown in Figure 37. ADH from *A. eutrophus* contained no transmembrane domains as predicted by the method of Klein, Kanehisa and DeLisi (1985).

The experiments described above demonstrated that three of the four open reading frames predicted from the sequence of Benian et al. (1989) were transcribed. We next asked whether any of the mutations recovered in the screens conducted by J. Schein and previous members of the laboratory were deletions affecting the genes on C11F2 and C18D3. This was accomplished by employing primers specific for each gene in PCR amplifications that used an animal homozygous for one of the mutations as an amplification template in each reaction. The relative positions of the genes and the PCR primers that were used are indicated in Figure 38. We reasoned that a deficiency that removed one (or more) of the genes we were examining would not have a detrimental effect, when homozygous, if there were no strict developmental requirements for the gene product(s).

PCR analysis of *sDf83* and *sDf88*: Each of the 48 putative deficiencies and *sDf83* were initially tested in PCR amplifications with primers specific for the *gtl-1* gene. As expected, *sDf83* homozygotes failed to produce an amplification product with these primers (data not shown). In addition to *sDf83*, individuals from one other strain derived from a mutation isolated by J. Schein failed to yield an amplification product with these primers (data not shown). The mutation carried in this strain was therefore a deficiency affecting at least *unc-22* and *gtl-1*, and was designated *sDf88*.

Figure 38:

A physical map of the *let-56* - *unc-22* interval. The directions in which *let-653* and *let-52* lie are indicated. *let-56* was rescued by the cosmid C11F2 (Clark and Baillie, 1992) and is shown above the physical map, which is represented by the heavy line. Genes are indicated by arrows above or below the heavy line, depending upon the direction of the open reading frames. The *unc-22* transcription unit is not to scale. This is indicated by the double slashes intersecting the heavy line. The oligonucleotide primers specific for each gene are indicated. The sequences of these primers are given in Table 2. The cosmids C11F2 and C18D3 are shown, as are relevant deficiencies in the area. The dashed lines indicate that the exact molecular breakpoints of these deficiencies have not been determined.


TOWARD
let-653
TOWARD
let-52
let-56
R8
R1
c6-2 dup-2 mys-2
c6-1 dup-1 mys-1
hdl-1
adl-1 dup ptl-1
unc-22
nhe-1
spm-1 gdl-1
gtl-1
spe-17
mam-2
db-19
lunc-2
KT03 lunc-1
KT01
C11F2
C18D3
sDf19
sDf83
sDf88
10 kb

149(b)

In additional PCRs using *sDf88* individuals as the amplification template, we failed to produce the expected PCR product with *ptl-1* primers, but amplification products were observed when *dup* primers (Figure 39) and *adl-1* primers were used. This indicated to us that *sDf88* deleted *gtl-1* and at least the annealing site for the mys-1 primer, thereby disrupting the *ptl-1* gene.

In PCRs using *sDf83* as the amplification template, amplification products were not seen with *ptl-1* primers, Dup primers, or *adl-1* primers (Figure 39; *adl-1* primers shown only). This was expected, as J. Schein had demonstrated that *sDf83* failed to complement *let-56*, and *let-56* had been localized to C11F2 based upon its ability to rescue the lethality associated with *let-56(s173)* in germline transformation experiments (Clark and Baillie, 1992). All of the genes predicted from the sequence of Benian et al. (1989) resided on the portion of C18D3 which does not overlap with C11F2 (Figure 38).

The above data were compatible with the notion that *sDf83* extended into DNA contained within the cosmid C11F2. The phenotype of individuals homozygous for *sDf83* was found to be identical to the phenotype of individuals homozygous for *let-56(s173)*. Animals homozygous for *sDf83* or *let-56(s173)* (see Section I) grew very slowly at 20 ^{0}C, reaching adulthood approximately two weeks after being laid as eggs. We had never observed *sDf83* animals to be fertile, although occasionally some appeared to carry an oocyte. This phenotype suggested two possibilities. Either the genes located on C18D3 were not required in development prior to the requirement for *let-56(+)* function, or there were no strict developmental requirements for any of these genes under the conditions of propagation.

Positioning the breakpoints of *sDf83* on the physical map: We mapped the left breakpoint of *sDf83* using PCR and a primer pair predicted from the *nhe-1* gene, and a primer pair predicted from the *hdl-1* gene. The expected PCR product

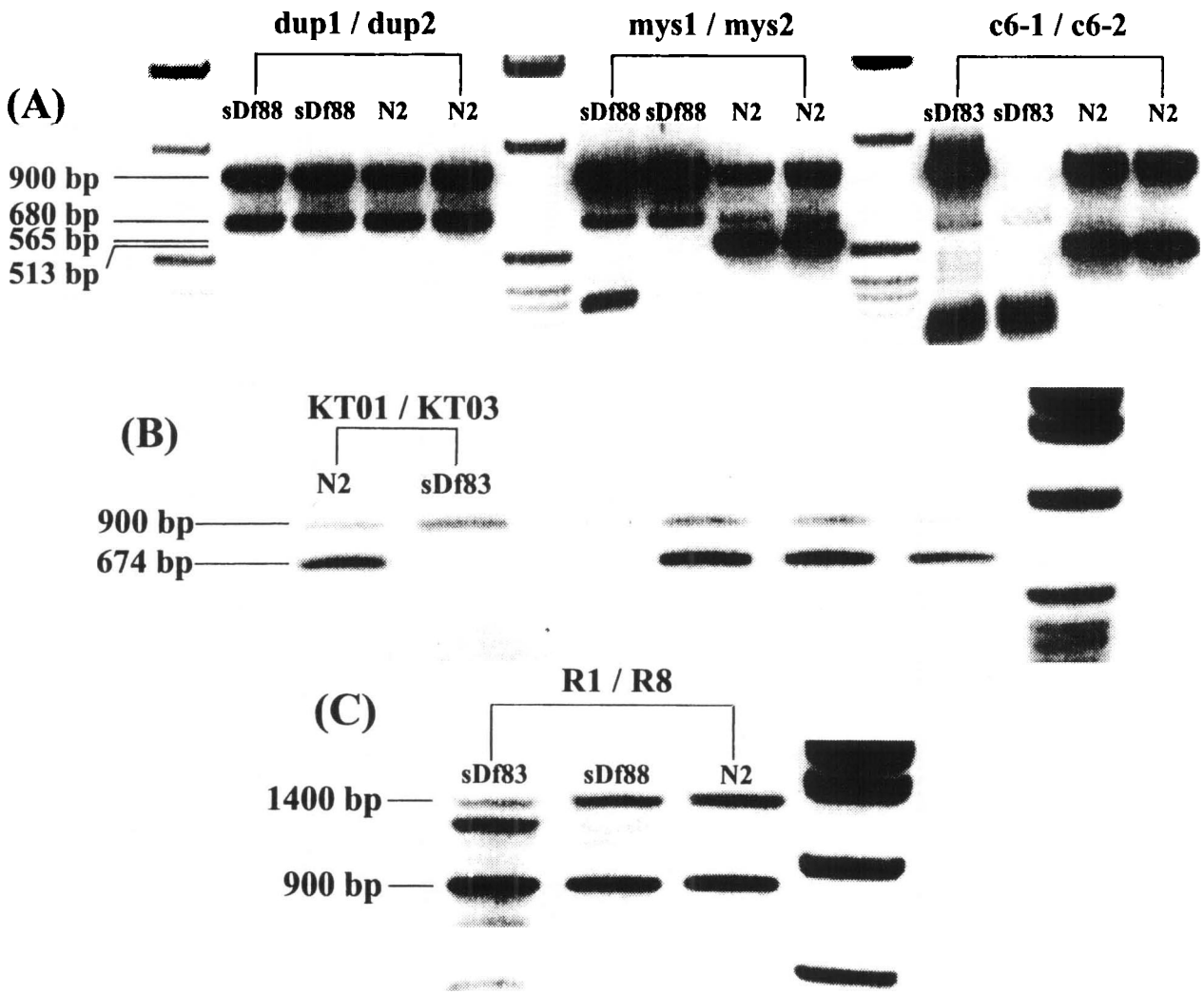
Figure 39:

Agarose gels of PCRs that used *sDf88* or *sDf83* homozygous individuals as the amplification template. The oligonucleotide primers used in the PCR experiments are shown above the lanes. The relative positions of all of these primers are indicated in Figure 38. The expected banding pattern, if the deficiency does not delete at least one of the priming sites, is shown in the lanes that contain amplification products resulting from PCRs that used wild-type (N2) worms as the amplification template. The 900 bp band seen in all lanes corresponds to a control amplification product produced by primers specific for sequences on chromosome V (either 265-F and 265-R or K083 and K087; see Table 2).

(A). PCR mapping the left breakpoint of *sDf88*. The 680 bp band corresponds to the **dup-1 / dup-2** (*dup* primers) amplification product. The 565 bp band corresponds to the **mys-1 / mys-2** (*ptl-1* primers) amplification product. The 513 bp band corresponds to the **c6-1 / c6-2** (*adl-1* primers) amplification product. The second **c6-1 / c6-2** *sDf83* lane represents a failed reaction, as indicated by the absence of the chromosome V specific amplification product. The *sDf83* reactions were included as negative controls.

(B). *sDf83* deletes *unc-22*. The 674 bp band seen in the N2 lane corresponds to the **KT01 / KT03** (*spe-17* primers) amplification product. This product was not seen in reactions that used *sDf83* individuals as the amplification template, indicating that *sDf83* completely deletes *unc-22*, extending at least as far as the KT03 priming site (Figure 38). The unlabeled lanes correspond to reactions performed using other *unc-22* mutations recovered in the mutagenesis screen conducted by J. Schein. A marker lane (1 kb ladder, BRL) is furthest to the right.

(C). *sDf83* does not extend as far as the *hdl-1* R1 primer annealing site, as indicated by the presence of the 1400 bp **R1/R8** amplification product in the *sDf83* lane. A marker lane (1kb ladder, BRL) is furthest to the right.


(A)
dup1 / dup2
mys1 / mys2
c6-1 / c6-2
sDf88 sDf88 N2 N2
sDf88 sDf88 N2 N2
sDf83 sDf83 N2 N2
900 bp
680 bp
565 bp
513 bp
(B)
KT01 / KT03
N2 sDf83
900 bp
674 bp
(C)
R1 / R8
sDf83 sDf88 N2
1400 bp
900 bp

was produced with both primer pairs using *sDf83* homozygous individuals as the amplification template, indicating that *sDf83* did not extend as far as the R1 priming site on C11F2. To map the right breakpoint of *sDf83*, we used PCR and primers specific for a 674 bp region that contains all but the last 174 bp of the *spe-17* gene (L'Hernault et al., 1993). When these primers (KT01 and KT03) were applied to *sDf83* homozygous individuals in PCR experiments, the expected band was not seen (Figure 39), indicating that *sDf83* disrupted at least the KT03 primer annealing site. However, complementation tests conducted by J. Schein had determined that *sDf83* did not delete *let-52*, the nearest marker to the right on the genetic map. This indicated the right breakpoint of *sDf83* was located between the KT03 primer annealing site and *let-52*.

The observation that a primer pair predicted from the *hdl-1* gene produced the expected PCR product from *sDf83* homozygotes (Figure 39) indicated that the left breakpoint of *sDf83* was to the right of the *hdl-1* primer annealing sites. This result, in combination with the following observations, served to define the region of C11F2 that must contain *let-56*. First, *sDf83* had failed to complement mutations in *let-56*, implying that *sDf83* either deleted or disrupted the *let-56* gene. Second, the primer annealing sites within the *hdl-1* gene were located within approximately 10 kb of the rightmost end of C11F2, and C11F2 had been shown to rescue *let-56(s173)* (Clark and Baillie, 1992). Therefore, within this 10 kb region were sequences which provided *let-56(+)* activity, and therefore probably contained the *let-56* gene.

Refinement of the *let-56* region and identification of candidate *let-56* genes: I attempted to refine the *let-56* region by further localization of the left breakpoint of *sDf83* on the physical map. This was performed as follows. Genomic DNA was isolated from both wildtype (N2) animals, and animals of the genotype *sDf83 / [let-?(m435)]nT1(IV) ; + / [let-?(m435)]nT1(V)* as described in

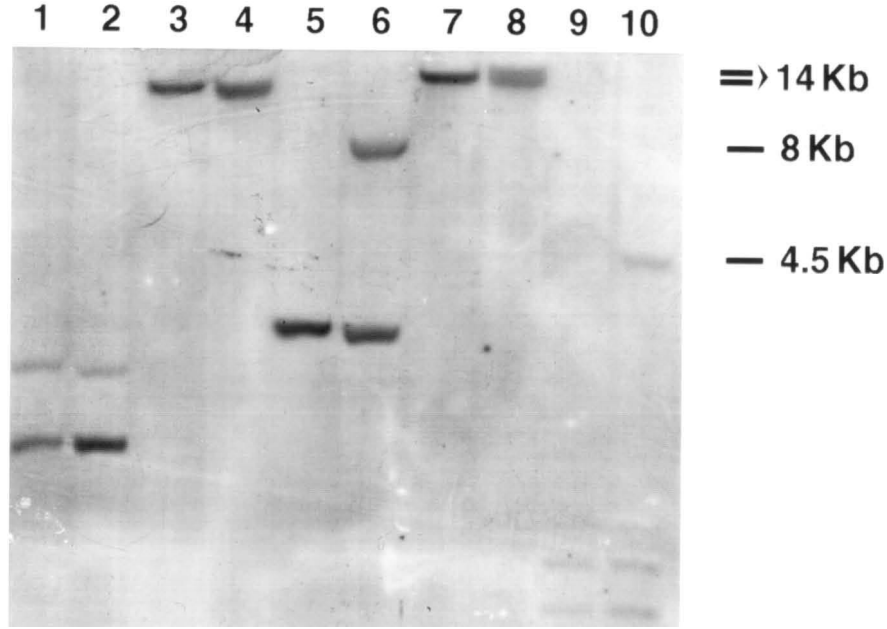


Figure 40:

Chemi-lumigraph of a Southern blot of genomic DNA prepared from wild-type (N2) and *sDf83 / [let-?(m435)]nT1(IV) ; [let-?(m435)]nT1(V)* animals. Genomic DNA was digested to completion with *Acc*I (lanes 1 and 2), *Cla*I (lanes 3 and 4), *Eco*RI (lanes 5 and 6), *Eco*RV (lanes 7 and 8), and *Sca*I (lanes 9 and 10). Digested DNA was electrophoresed on a 0.8 % agarose gel, blotted onto a nylon membrane, and hybridized with a *hdl-1* cDNA probe. Lanes 1, 3, 5, 7, and 9 contain N2 DNA. Lanes 2, 4, 6, 8 and 10 contain DNA prepared from *sDf83* heterozygotes. Polymorphic restriction fragments hybridizing to the *hdl-1* probe are seen in lanes 4 (a doublet), 6, 8 (a doublet) and 10. The approximate sizes of the polymorphic hybridizing restriction fragments are indicated in kb on the right of the chemi-lumigraph.

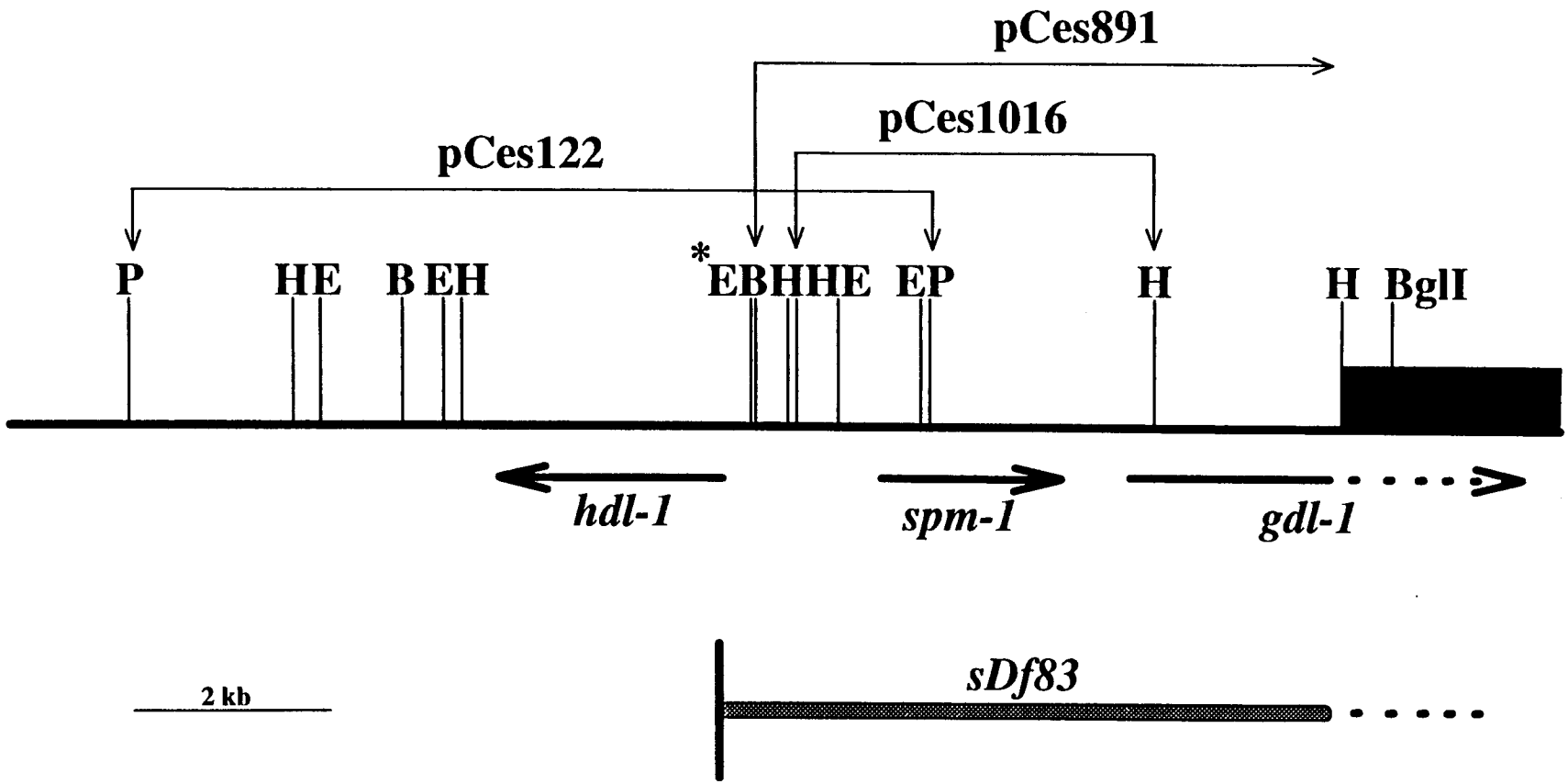
Materials and Methods. N2 and *sDf83* heterozygote DNA was digested to completion with five different restriction enzymes. Digested DNA was electrophoresed on agarose gels and transferred to nylon membranes as described (Materials and Methods). This Southern blot was then probed with the entire gel purified, labeled, *hdl-1* cDNA. Subsequent chemi-luminography of the Southern blot revealed the presence of a polymorphic band in each of the lanes containing DNA prepared from *sDf83* heterozygotes, except for a lane containing *Acc* I digested DNA (Figure 40).

The presence of polymorphic bands on a Southern blot containing DNA prepared from *sDf83* heterozygotes and probed with the *hdl-1* cDNA indicated that the *sDf83* breakpoint was located within the restriction fragments to which the *hdl-1* cDNA hybridized. Furthermore, the observation that a polymorphism was seen in a lane containing *Eco*RI digested DNA allowed a precise localization of the *sDf83* breakpoint, as the *Eco*RI restriction pattern of the *hdl-1* gene is known. An interpretation of the hybridization results shown in Figure 40 is provided in Figure 41. The hybridization results indicate that *sDf83* must affect the *Eco*RI site nearest the 5' end of the *hdl-1* gene. However, *sDf83* cannot extend further than the R1 priming site predicted from the *hdl-1* sequence because this oligonucleotide primer, when used in PCRs of *sDf83* individuals with primer R8, produces an amplification product (Figure 39). Therefore, the left breakpoint of *sDf83* appears to lie within the 200 bp that separate the 5' *Eco*RI site and the R1 primer annealing site. If so, *sDf83* breaks near the 5' region of the *hdl-1* gene, and completely deletes *spm-1* and *gdl-1,* the other genes that reside on this portion of C11F2 (see Part A). These data are summarized in Figure 41.

The precise localization of the left breakpoint of *sDf83* on C11F2 served to define candidate *let-56* genes. As a result of this analysis, it became apparent that *nhe-1* was not *let-56*. Candidate *let-56* genes now included *hdl-1*, *spm-1* and *gdh-1*. In order to test for the presence of functional *let-56* sequences, various subclones that

Figure 41:

An interpretation of the hybridization results presented in Figure 40. The genes that have been identified within this portion of C11F2 and the direction of their open reading frames are indicated by arrows. The restriction fragments contained within the plasmids pCes122, pCes891 and pCes1016 are indicated. The thick line indicates the cosmid vector. E = *Eco*RI. H = *Hin*dIII. B = *Bam*HI. P = *Pst*I. The breakpoint of *sDf83* extends as far to the left as the *Eco*RI site indicated by an asterisk, but does not extend further than the R1 oligonucleotide primer site in the *hdl-1* gene. The hatched box indicates *sDf83*. The dotted line to the right of the hatched box indicates *sDf83* extends further towards *unc-22*. The dotted line extending to the right of *gdl-1* indicates that the *gdl-1* gene is not likely to be fully contained within C11F2, based on the estimated size of full length *gdl-1* cDNAs I have recovered. Full length *gdl-1* cDNAs are approximately 1.9 kb. The portion of C11F2 DNA in which *gdl-1* resides would be sufficient to contain the entire *gdl-1* gene if the *gdl-1* gene lacked introns. One intron in the *gdl-1* gene has been identified so far (Figure 29).


pCes891
pCes1016
pCes122
P
HE
B EH
*EB HHE
EP
H
H BglI
hdl-1
spm-1
gdl-1
2 kb
sDf83

155(b)

had been derived from C11F2 were used by S. Jones or J. Schein to construct transformed strains by germline microinjection. Stable transformed strains resulting from the injection of (in addition to a dominant Roller construct) pCes1016, pCes891, or pCes122 were constructed. Stable transformed strains resulting from the co-injection of pCes891 and pCes122, and from the co-injection of pCes1016 and pCes122 were also constructed. I tested these constructs for their ability to rescue *let-56(s173)* by mating males of the genotype *let-56(s173) unc-22(s7) / + +* to hermaphrodites bearing an extrachromosomal array, and selecting Roller progeny that twitched in nicotine. These animals were set individually, and their progeny were examined for the appearance of fertile constitutive twitcher animals that rolled. The presence of progeny of this phenotype indicated that the extrachromosomal array rescued *let-56(s173)*. Only the extrachromosomal array resulting from the co-injection of pCes122 and pCes891 was found to be capable of rescuing *let-56(s173)* individuals (S. Jones, personal communication). This construct was designated *sEx25*.

***sEx25* restores partial fertility to *sDf83* homozygous individuals:** The observation that the phenotype of *sDf83* homozygous individuals resembled that of *let-56(s173) unc-22(s7)* homozygous individuals indicated that either none of the genes deleted by *sDf83* were required in development prior to the requirement for *let-56* (+) activity, or that there were no indispensable developmental requirements for these genes at all. To test whether any of the genes we had characterized on C18D3 were required for normal development to fertile adulthood, I constructed animals that were homozygous for *sDf83* and carried the *sEx25* construct that had rescued *let-56* (Materials and Methods). *sEx25* contained all of the C11F2 DNA deleted by *sDf83*, and may share approximately 8 kb of overlap with C18D3, but does not contain *adl-1*, *dup*, *ptl-1* or *gtl-1*, nor does it contain a gene identified by Prasad on *Pst*I fragment T (Figure 30). I found that *sDf83* homozygous individuals bearing the *sEx25* extra-

chromosomal array are partially fertile (Table 8). In addition to sterile *sDf83* homozygous progeny and fertile progeny that bear the extra-chromosomal array, these individuals produced varying numbers of unfertilized oocytes (Table 8). Partial fertility and production of unfertilized oocytes are characteristic of null mutations of the *spe-17* gene (L'Hernault et al., 1993). Therefore, consistent with our PCR analysis, this result indicates that the right breakpoint of *sDf83* affects the *spe-17* gene. In addition, these results suggest that the genes identified on C18D3 are not required in development, either singly or in combination with each other.

TABLE 8.

sEx25 ; sDf83 / sDf83 animals are viable and partially fertile.

Hermaphrodite	Number of progeny§	Unfertilized oocytes
1	10	0
2	18	10
3	14	14
4	3	40
5	0	9
6	0	0
7	27	1
8	9	0
9	4	0
10	6	0
11	9	5
12	9	3
13	10	13
14	6	0

§ Young adult *sEx25; sDf83 / sDf83* hemaphrodites were picked to separate plates and allowed to self at room temperature (approximately 20 ^{0}C). The numbers of progeny and unfertilized oocytes present on individual plates were determined over a period of five days.

DISCUSSION

In Section III, I have focused on the 0.1 map unit interval between *let-653* and *unc-22* as a part of my investigation into the organization of genes in the *sDf10* region. I wished to identify and characterize genes between *let-653* and *unc-22*, and their positions with respect to each other. Intensive genetic screens for mutations in this interval that disrupted normal development had been conducted (Section I and references therein), and had resulted in the identification of only a single locus, *let-56*. However, evidence had been provided by the work of Prasad (1988; Prasad and Baillie, 1989) that there existed at least six genes in the interval. Because genetic analysis had failed to identify the majority of the genes residing between *let-653* and *unc-22*, an alternate method of studying the gene content of this interval was required. The method I adopted involved the recovery and sequence analysis of cDNAs that represented genes residing on the cosmid C11F2, and, in collaboration with J. Schein, G. Benian and C. Fields, the analysis of previously determined sequence that corresponded to a portion of the cosmid C18D3 (Benian et al., 1989). Together, C11F2 and C18D3 contained the majority of the DNA between *let-653* and *unc-22*. In addition, C11F2 had previously been shown to rescue the only mutagenically identified gene in the region, the essential gene *let-56* (Clark and Baillie 1992).

This discussion will present brief considerations of the molecular nature of each of the genes that were characterized during the study. Based on similarities to sequences present in nucleic acid and protein sequence databases, I discuss possible functions of certain of these genes. Because no evidence has been obtained concerning the actual function of any of these genes, the proposed functions presented in this discussion are purely speculative.

The *C. elegans nhe-1* gene encodes a protein similar to vertebrate Na^+/H^+ exchangers. DNA sequence analysis of the *nhe-1* gene has revealed that its product exhibits a high degree of similarity to a family of vertebrate proteins that have been demonstrated to function as Na^+/H^+ exchangers. *nhe-1* is the first candidate for an invertebrate metazoan gene encoding a Na^+/H^+ exchanger, and probes derived from this locus are currently being used in attempts to clone related genes from crustaceans (G. Ahearn, personal communication). Na^+/H^+ exchange activity has been observed in every mammalian cell type examined to date (Grinstein et al., 1989). Na^+/H^+ exchangers catalyze the exchange of extracellular Na^+ for intracellular H^+ with a stoichiometry of 1:1. Identified functions for the Na^+/H^+ exchanger include regulation of intracellular pH, in particular the recovery from an acid load (Grinstein et al., 1989), maintenance of cellular volume in response to an osmotic load, and trans-epithelial Na^+ absorption in epithelial cells. In addition to these functions, which have been demonstrated in tissue culture cell lines (reviewed in Alpern et al., 1993), evidence has been presented supporting a role for Na^+/H^+ exchangers in cell proliferation in response to certain mitogens (reviewed in Grinstein et al., 1989).

Pharmacological and kinetic data have provided evidence that multiple isoforms of the Na^+/H^+ antiporter exist in vertebrates (Alpern et al., 1993). Plasma membrane vesicle studies conducted by Knickelbein et al. (1990) have shown that ileal villus epithelial cells have Na^+/H^+ exchangers on both their apical and basolateral membranes which exhibit differential sensitivity to amiloride and its derivatives, drugs which specifically inhibit Na^+/H^+ exchange (Grinstein et al., 1989). The exchanger localized to the basolateral membrane is extremely sensitive to amiloride (exchange is inhibited by 10^{-7} M EIPA, an amiloride derivative) whereas the apical membrane exchanger is 10 to 100 times less sensitive to the effects of EIPA. The exchanger found on the basolateral membrane has been implicated in internal pH regulation and osmotic regulation, while the exchanger found on the

apical membrane appears to mediate secretion of H^+ from cell to lumen in the renal proximal tubule and the small intestine (Alpern et al. 1993).

Using a technique involving functional complementation, a human Na^+/H^+ exchanger was cloned (Sardet et al., 1989). This protein (called NHE-1) was found to contain multiple (10 - 12) putative transmembrane domains in the amino-terminal half of the protein, followed by a carboxy-terminal cytoplasmic domain. Using antibodies directed against the cytoplasmic domain of NHE-1, Tse et al. (1991) demonstrated that NHE-1 was localized to the basolateral membrane of ileal villus epithelial cells. Based on this localization, NHE-1 appears to correspond to the amiloride hyper-sensitive exchanger that is localized to the basolateral membrane and is involved in regulation of cellular pH and cellular volume.

Since the cloning of NHE-1 from human, a number of NHE-1 genes from different vertebrates have been described. The predicted mammalian NHE-1 proteins exhibit high degrees of similarity to each other; for example, the human and rabbit NHE-1 proteins are 95 % identical (Tse et al., 1991). In addition to the rabbit NHE-1 isoform, at least two other distinct rabbit isoforms have been reported. These have been named NHE-2 and NHE-3. All of the isoforms have similar hydrophobicity profiles, and all are predicted to be glycoproteins due to the presence of a conserved potential N-linked glycosylation site.

The *C. elegans nhe-1* predicted protein exhibited strong similarity to Na^+/H^+ exchangers in **blast** searches of the non-redundant databank at the NCBI, most closely resembling the NHE-3 isoform from rat, and a Na^+/H^+ exchanger from trout red cells called beta NHE (Table 6). beta NHE has been shown to be distinct from NHE-1 in that beta NHE is activated by hormones which act through cAMP as a second messenger. NHE-1 is activated by certain hormones, but not in response to cAMP (Borgese et al., 1992).

Both the N and C terminii are very different among members of the Na^+/H^+ exchanger protein family. The C terminus is believed to be the regulatory domain and to contain sites for kinase phosphorylation. NHE-1 has protein kinase C consensus sequences, but no cAMP-dependent protein kinase phosphorylation consensus sequences (Tse et al, 1991). In contrast, NHE-3 and beta NHE contain both protein kinase C and cAMP-dependent protein kinase phosphorylation consensus sites (Tse et al., 1991; Borgese et al., 1992). The C terminus of the *C. elegans* Nhe-1 protein contains protein kinase C phosphorylation consensus sites and cAMP-dependent protein kinase phosphorylation sites, as do beta NHE and NHE-3. Examination of the distribution of rabbit NHE-3 message (Tse et al., 1992) has resulted in the detection of NHE-3 message in intestine and kidney only, and not in the duodenum or descending colon, which are tissues that lack the electroneutral NaCl absorptive process. This localization implies that NHE-3 might be involved in Na^+ absorption in intestinal and renal epithelial cells, which are the functions demonstrated for the apical membrane Na^+/H^+ exchanger protein, as mentioned above. beta NHE might function in a similar fashion in trout red cells.

Potential functions of the *C. elegans nhe-1* gene product: In spite of intensive mutagenesis screens of the region (Section I and references therein), no mutations in the *nhe-1* gene have been recovered. However, these screens were designed to recover mutations resulting in slowed or arrested development, which may not be phenotypes that result from a mutationally altered exchanger. Alternatively, the existence of a related gene or genes elsewhere in the genome (Results, Part A) that could substitute for missing *nhe-1* function may have complicated recovery of mutations in the *nhe-1* gene. Based on the impressive overall similarities to NHE-3 and beta NHE, it is possible that the *C. elegans* protein may be growth factor and kinase regulated. However, the sequence data alone do not indicate whether the *C. elegans* protein functions in a Na^+ absorption process, or has

as its primary function the regulation of cellular pH, or cellular volume, or perhaps both.

It has been noted that the predicted amino acid sequence of an amiloride-sensitive sodium channel gene cloned from rat epithelial cells exhibits similarity to the *mec-4* and *deg-1* gene products from *C. elegans* (Canessa, Horisberger and Rossier, 1993). This sodium channel has not been demonstrated to have Na^+/H^+ exchange activity, and bears little resemblance to known exchanger proteins. Mutations in both *mec-4* and *deg-1* were recovered in genetic screens for touch insensitive animals (Driscoll and Chalfie, 1991; Chalfie and Wolinsky, 1990). The *mec-4* and *deg-1* gene products have been named degenerins, because dominant mutations in either of these genes are sufficient to induce a selective degeneration of the touch cell sensory neurons of the nematode. In these mutants, the cell bodies of degenerating neurons swell to several times their normal diameter before they disappear, presumably due to cell lysis (Driscoll and Chalfie, 1991). This swelling may be due to an inability to osmoregulate. Because one of the functions of the NHE-3 isoform of the Na^+/H^+ exchanger is Na^+ transport and therefore maintenance of cellular volume, certain types of mutation in the *nhe-1* gene might result in a similar cellular phenotype, although not necessarily in neurons.

The *C. elegans hdl-1* gene encodes a protein similar to histidine and DOPA decarboxylases: DNA sequence determination and analysis revealed that the *C. elegans hdl-1* gene encoded a protein that closely resembled members of the family of pyridoxal -5'- phosphate dependent amino acid decarboxylases, including histidine and DOPA decarboxylases. In vertebrates and invertebrates, several pyridoxal phosphate (PLP) cofactor dependent decarboxylases, including DOPA decarboxylase (DDC), glutamate decarboxylase (GAD) and histidine decarboxylase (HDC), catalyze the synthesis of neurotransmitters and neuromodulatory compounds (reviewed in Siegel et al., 1989).

In **blastx** searches of the non-redundant database, an open reading frame predicted from the *hdl-1* cDNA exhibited the strongest similarity to human histidine decarboxylase and rat histidine decarboxylase (Table 7). Histidine decarboxylase is the enzyme that catalyzes the formation of histamine from L-histidine. In mammals, histamine is a chemical mediator involved in allergic reaction, secretion of gastric acid and neurotransmission in the central nervous system (Dimaline and Sandvik, 1991). In rodents, HDC activity is localized primarily to the brain (Watanabe et al., 1984), the glandular regions of the stomach (Savany and Cronenberger, 1982), mast cells (Yamada et al., 1980) and fetal liver (Taguchi et al., 1984). HDC has been purified from rat stomach (Savany and Cronenberger, 1982) and fetal liver (Taguchi et al., 1984) and was shown to be a dimer consisting of two identically sized subunits. In addition, HDC activity has been localized in both the soluble and the membrane fractions of rat brain (Toledo et al., 1988).

A recent report has provided evidence for the function of an invertebrate HDC. In Drosophila, immunocytochemical and biochemical data have provided evidence that histamine is a major transmitter used by all photoreceptors in the compound eye (Sarthy, 1991). A large number of *Drosophila* mutations in different complementation groups that cause defects in the electroretinogram (extracellularly recorded, light evoked, mass response of the eye) had previously been isolated during studies of photoreceptor function (Pak et al., 1969; Pak 1975; 1979). Four mutations defining one of these complementation groups were discovered to be defective in histamine synthesis (Burg et al., 1993). Enzymatic assays conducted on head extracts of animals heterozygous for each of the mutations revealed a reduction of HDC activity from 50 % to 25 % of wildtype activity. The locus responsible for this phenotype was subsequently cloned and sequenced and found to encode a molecule with extensive similarity to mammalian histidine decarboxylases (Burg et al., 1993). The protein predicted from this locus does not contain a candidate membrane

spanning domain, unlike the HDC sequences from human, rat, and *C. elegans*. However, the *Drosophila* protein does possess several potential phosphorylation sites that are conserved in the human, rat, and *C. elegans* sequences (see Results).

The cloning of the *Drosophila* gene encoding HDC, and the demonstration that mutations in the *Drosophila* HDC gene affect photoreception have lent strong genetic support to the suggestion that histamine is a major transmitter used by *Drosophila* photoreceptors. Although the existence of histamine in *C. elegans* has been reported (Pertel and Wilson, 1974), it is unknown as to what role histamine may play in *C. elegans* biology.

The *C. elegans gdl-1* gene encodes a protein similar to glutamate dehydrogenase. Sequence analysis of portions of *gdl-1* cDNAs has shown that the *C. elegans gdl-1* gene encodes a protein which exhibits a very high degree of similarity to glutamate dehydrogenase enzymes from a number of mammalian species, including rats, cows, and humans. In addition, the *C. elegans* protein contains a mitochondrial transit peptide, and a putative glutamate dehydrogenase active site, as predicted by PROSITE analysis (Bairoch, 1993). The presence of these features, in addition to the high degree of similarity to other glutamate dehydrogenase enzymes, is consistent with the notion that the *C. elegans gdl-1* gene encodes a glutamate dehydrogenase.

Glutamate dehydrogenase is a mitochondrial matrix enzyme that catalyzes the reversible oxidative deamination of L-glutamate to α-ketoglutarate and ammonia, using NAD^+ or $NADP^+$ as an oxidizing agent. The ammonia produced in this oxidative deamination is converted to urea, in the case of most terrestrial vertebrates, or uric acid, in the cases of birds and terrestrial reptiles, and subsequently excreted. The equilibrium position for the conversion of glutamate to ammonia and α-ketoglutarate by glutamate dehydrogenase greatly favours glutamate formation over

ammonia formation (Voet and Voet, 1990). This equilibrium position is physiologically important, as high concentrations of ammonia are toxic to the cell.

Glutamate dehydrogenase is thought to be one of a number of enzymes that may modulate the quantity of glutamate available for conversion to GABA (Chase and Kankel, 1987), the most abundant inhibitory neurotransmitter in vertebrates and invertebrates (Cooper, Bloom and Roth, 1991). Glutamate can be converted to GABA by the action of glutamate decarboxylase. Addition of GABA to cut wild-type *C. elegans* (Lewis et al., 1980) or of muscimol, a GABA agonist, to intact animals (McIntire, Jorgensen and Horvitz, 1993; M. Chalfie, personal communication) causes a flaccid paralysis. On the basis of GABA-like immunoreactivity, GABA has been identified as a putative neurotransmitter in 26 *C. elegans* neurons (Chalfie and White, 1988; McIntire, Jorgensen and Horvitz, 1993). Furthermore, several mutants have been recovered which exhibit altered patterns of anti-GABA staining. For example, mutations in *unc-25* lack detectable anti-GABA staining in all GABAergic neurons, while mutations in *unc-47* have enhanced staining (McIntire, Jorgensen and Horvitz, 1993; McIntire et al., 1993). If mutations in *gdl-1* greatly reduced the amount of glutamate available for glutamate decarboxylase-dependent conversion to GABA, these mutations might result in the production of behavioural defects similar to those seen in *unc-25* animals. These include defects in enteric expulsion, a "shrinking" behaviour when prodded, and "loopy" foraging behaviour (McIntire, Jorgensen and Horvitz, 1993). Mutations inactivating glutamate dehydrogenase might also result in an increase in the concentration of intra-cellular ammonia . As high concentrations of ammonia are toxic to the cell, mutations of this sort might be recovered in screens for mutations that slowed or arrested normal development. The deficiency *sDf83*, which completely deletes all genes between *spe-17* and *hdl-1*, including *gdl-1*, exhibits as primary phenotypes slow growth and adult sterility. Other, more subtle, phenotypes

resulting from deletion of the *gdl-1* gene may well be masked by the Unc-22 phenotype associated with *sDf83* homozygous individuals.

The *C. elegans spm-1* gene encodes a novel protein that possesses characteristics which suggest it may bind DNA. Extensive similarities to sequences in the non-redundant and **dbest** databases were not discovered in searches that involved the predicted Spm-1 protein. However, **blast** searches revealed a 35 amino acid long region of similarity between the Spm-1 protein and *Brachyury* (T) proteins from zebrafish, *Xenopus* and mouse, and the *lethal(1) optomotor-blind* protein of *Drosophila melanogaster*. The function, if any, of this 35 amino acid long region is unknown, but in the mouse T protein, these 35 amino acids reside within a 229 amino acid N - terminal region that has been shown to bind DNA, and which is now known as the T domain (Kispert and Herrmann, 1993). Consistent with a role as a DNA binding protein, mouse T protein has been detected in the nucleus (Schulte-Merker et al., 1992). Analysis of the mouse protein's amino acid sequence has not revealed similarities to protein structures known to bind DNA (e.g. a helix - loop - helix motif).

Similarity between the T domain and a 224 amino acid region of the *Drosophila lethal(1) optomotor-blind* (*omb*) gene product (omb) has been noted, and this domain of the *Drosophila* protein has been shown to bind to calf thymus DNA (Pflugfelder et al., 1992). Although a region similar to the T domain exists in omb, omb is not likely to be the homologue of T, based on structural differences and lack of overall similarity. No known DNA binding motifs are present in the *Drosophila* protein. However, Pflugfelder et al. (1992) reported a clustering of S-P-X-X motifs in the C terminal domain of the *Drosophila* protein, outside the DNA binding domain. These motifs have been shown to occur at a high relative frequency in transcription factors in the regions flanking DNA binding domains (Suzuki, 1989), and are also present in the mouse and *Xenopus* T proteins.

The Spm-1 protein, in addition to containing the 35 amino acid region that exhibits similarity to a portion of the DNA binding domain of T proteins and the *Drosophila* omb protein, also contains seven copies of the S-P-X-X motif (Figure 27). Six of the seven copies of the motif are localized to a 43 amino acid C terminal region. This overall arrangement of the S-P-X-X motifs resembles the arrangement found in omb and T. In addition, Spm-1 is predicted to contain a nuclear targeting signal (Figure 27), which supports the notion that it may function in a DNA binding capacity.

If the Spm-1 protein does in fact bind DNA, the discovery of the 35 amino acid region that exhibits similarity to T and omb may define a previously undescribed DNA binding domain. While it is possible that Spm-1 may be distantly related to T and omb, the lack of overall similarity between Spm-1 and either T or omb makes it unlikely that Spm-1 is the *C. elegans* equivalent of either of these proteins.

A cluster of genes adjacent to *unc-22* that are predicted to encode trans-membrane proteins: Three of the four gene products predicted from the sequence of Benian et al. (1989) exhibited weak similarities to sequences in the non-redundant database. These included the products of the glucose transporter-like gene (*gtl-1*), the proline transporter-like gene (*ptl-1*), and the alcohol dehydrogenase-like gene (*adl-1*). In each case, the similarities that were obtained were too weak to be suggestive of conservation of function; however, they may indicate that each of these three proteins is a member of a larger family of proteins that share certain structural characteristics. For example, the *gtl-1* gene product contains two pairs of motifs that are considered to be common to all glucose transporters (Gould and Holman, 1993; Figure 31). However, other structural features present in all known glucose transporters (Gould and Holman, 1993) are not found in Gtl-1. Likewise, Ptl-1 probably belongs to a family which includes the proline transporter proteins, but it is unlikely that *ptl-1*

encodes a molecule that transports proline. A similar statement can be made concerning the identity of the *adl-1* gene product.

The existence of transcripts for each of the above three genes was demonstrated. No transcript was detected for the "*dup*" gene, and no similarity to sequences in the non-redundant database was discovered. The close proximity of *dup* to *ptl-1* is compatible with the notion that both *dup* and *ptl-1* correspond to the same gene. However, attempts to PCR amplify a cDNA using appropriately oriented *ptl-1* and *dup* primers was unsuccessful. This observation is compatible with the idea that *dup* is distinct from *ptl-1,* and possibly produces a rare transcript which was not represented in the cDNA library that was sampled. Alternatively, the possibility exists that *dup* is not a gene at all, but rather an open reading frame generated fortuitously as a result of the splicing out of predicted introns. However, the size and unusual structure of the predicted *dup* gene product (Figure 37) supports the existence of *dup* as a gene (or perhaps pseudogene) and not a fortuitously derived open reading frame.

The most striking feature of the products of the cluster of genes immediately adjacent to *unc-22* is the predicted existence of multiple trans-membrane domains in each of them. Furthermore, there is an interesting pattern seen in the hydrophobicity profiles of Gtl-1, Ptl-1 and Dup. In each of these proteins there are two regions that are predicted to contain multiple transmembrane domains. These are separated by a region of the protein which is predicted to be hydrophilic (Figure 32, Figure 34 and Figure 37). The predicted *dup* gene product, while exhibiting no informative similarity to sequences in the current database set, has characteristics that suggest it has arisen as the result of an intragenic tandem gene duplication event. The repeated domains of Dup exhibit a high degree of similarity to each other, on the order of 95% (Figure 35). In the case of the glucose transporter gene family, which is a subset of the "major facilitator superfamily" (or MFS; Marger and Saier, 1993) and which

apparently includes *gtl-1*, it has been suggested that this repeated pattern has arisen as a result of the tandem intragenic duplication of an ancestral gene, which in turn gave rise to the primordial gene of this family (Gould and Holman, 1993). A similar suggestion has been made to explain the structure of the MIP (major intrinsic protein) family of integral membrane channel proteins. In the case of this protein family, the first member of which was described in bovine lens fiber cell membranes, there are predicted to be six helical trans-membrane domains (Reizer, Reizer and Saier, 1993). It has been proposed that the first and second halves of all MIP family proteins arose as a result of a tandem intragenic duplication event involving an ancestral gene encoding a three-spanner protein (Reizer, Reizer and Saier, 1993).

Of the eight genes that reside in the 50 kb to the left of *unc-22,* for which sequence data are currently available, all but two (*spm-1* and *gdl-1*) are predicted to encode proteins that contain helical trans-membrane domains. Of these, all but one (Hdl-1) are predicted to contain multiple transmembrane domains. In addition, three of the eight gene products resemble proteins that function in neurotransmitter transport (the *ptl-1* gene product) or neurotransmitter synthesis (the products of the *hdl-1* and *gdl-1* genes).

The functional significance, if any, of the organization of genes in the interval I have studied remains to be elucidated. However, the data are compatible with the notion that there appears to exist an organizational theme among genes immediately to the left of *unc-22*. This interpretation of the data provides additional support for a similar conclusion previously reached by Prasad, who showed that each of five transcripts that mapped to either C11F2 or C18D3 exhibited similar temporal patterns of expression (Prasad and Baillie, 1989).

As mentioned, the deficiency *sDf83* has been shown to delete all genes between *hdl-1* and *spe-17*. Furthermore, the left breakpoint of *sDf83* is located either near or in the 5' region of the *hdl-1* gene. The right breakpoint is likely to disrupt the

spe-17 gene, as was indicated by PCR analysis and the phenotype of *sDf83* homozygous individuals bearing the *sEx25* extrachromosomal array. Therefore, *sDf83* affects at least nine genes and possibly ten genes, including *unc-22* and an uncharacterized gene identified by Prasad (1988) on *Pst*I fragment T (Figure 30). Individuals homozygous for *sDf83* were slow growing and sterile, and appeared identical to animals that were *let-56 / let-56*, or *sDf83 / let-56*. These observations were consistent with the idea that there were likely no genes between *spe-17* and *let-56* that provided an indispensable function in development required prior to *let-56*(+) activity. Support for this idea was provided by the finding that *sEx25*, an extrachromosomal array that carries none of the genes characterized on the cosmid C18D3, can rescue *sDf83* homozygotes. This finding explains the inability to recover mutations in these genes in genetic screens designed to recover lesions which slow or arrest development.

The precise localization of the left breakpoint of *sDf83* has defined candidate *let-56* genes. These include *hdl-1*, *spm-1* and *gdl-1*. The observation that the entire *gdl-1* gene is unlikely to reside on C11F2 decreases the possibility that it is *let-56*. The better candidates appear to be *spm-1* and *hdl-1*. The availability of the DNA sequence data for *spm-1* and *hdl-1*, generated in this portion of the thesis, should simplify the identification of the *let-56* gene.

General Discussion.

This thesis has focused on the identification and characterization of genes within the *sDf10(IV)* region of *Caenorhabditis elegans*. In Section I, genes that exhibited indispensable developmental requirements were identified in genetic screens for mutations that slowed or arrested normal development. Of 216 mutations of this type that were recovered in two 0.018 M EMS screens, 17 were found to map to the region of interest. Four of these 17 mutations each defined a new lethal complementation group, increasing the number of essential genes in the *sDf10* region to eleven. Each of the 17 mutations were characterized with respect to their complementation behaviour and their arrest stage phenotypes. This analysis resulted in the identification of complementing alleles of the mucin-like gene *let-653*, which has been recently sequenced by S. J. M. Jones (1994).

Section II described the implementation of a new method for the recovery of duplications of essential loci that are in or near the *sDf10* region. Analysis of a duplication that was recovered using this method revealed that it contained functional copies of several of the essential genes located within *sDf10,* including *let-56* and *let-653*. Interestingly, the duplication could not rescue all of the essential loci it apparently spanned.

In Section III, the *let-653* - *unc-22* region, a sub-interval of *sDf10*, was studied. Although evidence had been provided by Prasad and Baillie (1989) that six genes resided here, repeated genetic screens for mutations that slowed or arrested development had resulted in the identification of only *let-56*. Therefore, an alternate method for the identification and analysis of genes in the interval was employed. This method involved the analysis of sequence data obtained from cDNA clones representing four different genes, and the analysis of portions of four genes predicted from genomic DNA sequence that had previously been determined by Benian et al., (1989). While evidence has been provided for the existence of at least two more genes in the approximately 70 kb *let-653* - *unc-22* region (on *Pst*I fragments P and T),

it is likely that at least some sequence data for the majority of genes in the region now exists.

The *sDf10* region spans approximately 0.5 map units. Based upon her work that achieved correlation between the physical and genetic maps in the *dpy-20 - unc-22* region, Clark (1990) had estimated that the *dpy-20 - unc-22* distance of 0.2 map units corresponded to approximately 200 kb, for an estimate of approximately one megabase per map unit in this region. If this metric is applied to *sDf10*, then *sDf10* is estimated to span approximately 500 kb. This figure assumes that there are no local variations in the estimate of one megabase per map unit over the entire *sDf10* region. This assumption may not be valid.

It is possible to estimate the number of genes contained within *sDf10* given some measure of gene density. The current best estimate of gene density in *C. elegans* has been provided by the *C. elegans* genome sequencing consortium. The consortium has sequenced 2.181 megabases of DNA on chromosome III, and identified 483 putative genes by similarity searches and GENEFINDER analysis (Wilson et al., 1994). 125 of the 4,615 available *C. elegans* cDNA sequence tags match genes predicted from the genomic sequence. Therefore, 17, 832 (483 X 4615 / 125) genes are predicted to exist in the 100,000 kb genome, for an average gene density of one gene every 5.6 kb. This is a slightly higher density of genes than has been found on the cosmids C11F2 and C18D3, which is approximately one gene every 7.0 kb. If these figures are applied to the *sDf10* region, between 70 and 90 genes are predicted to reside within the breakpoints of *sDf10*.

Mutations have identified 14 genes in *sDf10*, or between 16 and 20 % of the predicted total number. Of these 11 (12 to 16 % of the predicted total number) have been identified in EMS screens for mutations that slow or arrest normal development. Therefore, this assay for essential genes shows they occur with a frequency of approximately 1 per 45 kb, or 1 per cosmid length of DNA, in the *sDf10* region.

While there are certain to be additional, as yet unidentified, essential genes in the region, it is likely that this figure represents a practical upper limit to the density of essential genes that can be achieved using this assay in this interval. It remains to be seen whether this density is attainable in other regions of the genome, although results obtained by W. B. Barbazuk (personal communication) and Johnsen and Baillie (1991) on chromosome V suggest it is.

While mapping transcripts in the *dpy-20* - *unc-22* region, Prasad and Baillie (1989) had observed that three genes on the cosmid C11F2 and one gene on the cosmid C18D3 exhibited similar developmental expression profiles, in that maximal hybridization was to RNA isolated from L2 stage animals. An additional gene on C18D3 was maximally expressed in the L3 - L4 stages. The expression profiles of these genes bore a resemblance to that of actin (Prasad and Baillie, 1989). Based on these findings, and the close proximity of these genes to the muscle gene *unc-22*, Prasad and Baillie (1989) speculated that the products of these genes might in some way be inter-related. This speculation prompted me to study a number of genes on C11F2 and C18D3 by DNA sequence analysis. Interestingly, six of the eight genes on these cosmids that were characterized were predicted to encode proteins that contained transmembrane domains. *let-653*, which is the next sequenced gene to the left of *nhe-1*, has also been predicted to encode a protein that contains a transmembrane alpha helix (S. J. M. Jones, 1994). In addition, three of the eight gene products may be involved in the transport or synthesis of neurotransmitter or neuromodulatory compounds. These findings are compatible with the notion that the organization of the genes in this region of the genome exhibit an organizational theme, in that most of them encode products that are predicted to be membrane associated.

Whether there are functional constraints that maintain the association of these genes is unknown. One way to approach this question would be to examine the *unc-*

22 homologue and surrounding region in other nematodes, particularly in the closely related *Caenorhabditis briggsae*, which is estimated to have diverged from a common ancestor at least 55 to 77 million years ago (S. S. Prasad, personal communication). Constraints that maintain the association of these genes may be reflected in the conservation of gene order between *C. briggsae* and *C. elegans*, or between *C. elegans* and a more distantly related nematode.

Proposals for additional research: As stated above, an examination of the gene order in the vicinity of the *Caenorhabditis briggsae unc-22* homologue might reveal whether there are functional constraints to the gene order in this region. It would be valuable to obtain the complete genomic and cDNA sequences for the genes predicted from the sequence of Benian et al. (1989), especially for *ptl-1*. It would be interesting to determine whether this gene plays a role in the neurobiology of *C. elegans*. Mutations inactivating the gene could be recovered using the method of Plasterk and Groenen (1992), and these mutations could be assayed for their effects, which might include uncoordinated movement, enteric expulsion defects, or mating abnormalities. Similar experiments could be done with the *hdl-1* and *gdl-1* genes. Mutations inactivating the *nhe-1* gene and the *spm-1* gene could likewise be recovered.

Further characterization of the *nhe-1* gene might involve creation of a chimeric *nhe-1* mini-gene that contained appropriate regulatory sequences to allow expression in mammalian cell lines, and introduction of this construct into a cell line which lacks Na^+/H^+ exchanger activity. The restoration of the ability of transfected cells to correctly regulate intra-cellular pH would indicate that *nhe-1* encodes a functional exchanger.

An obvious set of experiments is concerned with the continued molecular characterization of *let-56*, which is likely to be either *hdl-1*, *spm-1* or *gdl-1*. It might be informative to examine whether histamine, when added to the growth media, is

capable of rescuing *let-56* homozygous mutant indivduals. The ability of histamine to rescue *let-56* mutants would provide evidence that the *let-56* phenotype results from a lack of endogenous histamine. This would, in turn, support the notion that *let-56* was *hdl-1*. If so, this would be the first demonstrated role for histamine in *C. elegans*. Additional related experiments might include an examination of the distribution of histamine and histidine decarboxylase in *C. elegans* using anti-histamine and anti-histidine decarboxylase antibodies.

APPENDIX 1.

Calculation of map distance for *sDp11*.

Hermaphrodite: *let-56(s2321) unc-22(s7) [lev-1(x22)] sDp11 / + + +*

		1/2 (1-p)	1/2 p	1/2 (1-p)	1/2 p
		let-56 unc-22 sDp11	*let-56 unc-22 +*	*+ + +*	*+ + sDp11*
1/2 (1-p)	*let-56 unc-22 sDp11*	**Viable Unc-22**	**Viable Unc-22**	**Wild-type**	**Wild-type**
1/2 p	*let-56 unc-22 +*	**Viable Unc-22**	**Arrested Unc-22**	**Wild-type**	**Wild-type**
1/2 (1-p)	*+ + +*	**Wild-type**	**Wild-type**	**Wild-type**	**Wild-type**
1/2 p	*+ + sDp11*	**Wild-type**	**Wild-type**	**Wild-type**	**Wild-type**

Let p equal the recombination fraction. Then, the frequency of arrested Unc-22 individuals is $1/4p^2$. Because only Unc-22 individuals were scored, and the Unc-22 class comprises 1/4 of the total progeny, the frequency of arrested Unc-22 individuals must be:

$1/4p^2$ = (number of arrested Unc-22 individuals) / (4 X) ; where X is the total number of Unc-22 individuals. In this case, the number of arrested Unc-22 animals is 3, and the total number of Unc-22 individuals is 101. Solving for p yields 0.172, or 17.2 map units.

References.

Alpern RJ, Yamaji Y, Cano A, Horie S, Miller RT, Moe OW, Preisig PA (1993) Chronic regulation of the Na/H antiporter. J Lab Clin Med 122: 137-140

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

Aronson PS, Boron WF (1986) Na^+/H^+ Exchange, Intracellular pH and Cell Function. Current topics in Membranes and Transport, Volume 26, Academic Press, New York

Ayme A, Tissieres A (1985) Locus 67B of *Drosophila melanogaster* contains seven, not four, closely related heat-shock genes. EMBO J 4:2949-2954

Bairoch A (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. Nucleic Acids Res 21: 3097-3103

Barstead RJ, Waterston RH (1989) The basal component of the nematode dense-body is vinculin. J Biol Chem 264: 10177-10185

Barstead RJ, Waterston RH (1991) Vinculin is essential for muscle function in the nematode. J Cell Biol 114: 715-724

Barstead RJ, Kleiman L, Waterston RH (1991) Cloning, sequencing and mapping of an alpha-actinin gene from the nematode *Caenorhabditis elegans*. Cell Motil Cytoskeleton 20: 69-78

Beitel GJ, Clark SG, Horvitz, HR (1990) *Caenorhabditis elegans ras* gene *let-60* acts as a switch in the pathway of vulval induction. Nature 348: 503-509

Benian GM, Kiff JE, Neckelmann N, Moerman DG, Waterston RH (1989) Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. Nature 342: 45-50

Benian GM, L'Hernault SW, Morris ME (1993) Additional sequence complexity in the muscle gene, *unc-22*, and its encoded protein, twitchin, of *Caenorhabditis elegans*. Genetics 134: 1097-1104

Benton WD, Davis RW (1977) Screening gt10 recombinant clones by hybridization to single plaques *in-situ*. Science 196: 180-182

Blumenthal, T., Thomas, J (1988) Cis and trans mRNA splicing in *C. elegans*. Trends Genet 4: 305-308

Bonifer C, Vidal M, Grosveld F, Sippel AE (1990) Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. EMBO J 9: 2843-2848

Borgese F, Sardet C, Cappadoro M, Pouyssegur J, Motais R (1992) Cloning and expression of a cAMP-activated Na+/H+ exchanger: Evidence that the cytoplasmic domain mediates hormonal regulation. Proc Natl Acad Sci 89: 6765-6769

Boron WF (1983) Transport of H^+ and of ionic weak acids and bases. J Membr Biol 72: 1-16

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

Buckland R, Wild F (1989) Leucine zipper motif extends. Nature 338: 547

Burg MG, Sarthy PV, Koliantz G, Pak WL (1993) Genetic and molecular identification of a *Drosophila* histidine decarboxylase gene required in photoreceptor transmitter synthesis. EMBO J 12: 911-919

Butler PJD (1983) The folding of chromatin. CRC Crit Rev Biochem 15: 57-91

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. CABIOS 5: 233-234

Canessa CM, Horisberger J-D, Rossier BC (1993) Epithelial sodium channel related to proteins involved in neurodegeneration. Nature 361: 467-470

Carlstedt I, Sheehan JK (1989) Structure and macromolecular properties of cervical mucus glycoproteins. In: Mucus and related topics. Chantler E, Ratcliffe NA (eds) University of Cambridge Symp Soc Exp Biol 43: 289

Chalfie M, White J (1988) The nervous system. In: The nematode *Caenorhabditis elegans*. *Edited by* W.B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 337-391

Chalfie M, Wolinsky E (1990) The identification and suppression of inherited neurodegeneration in *Caenorhabditis elegans*. Nature 345: 410-416

Charest DL, Clark DV, Green ME, Baillie DL (1990) Genetic and fine structure analysis of unc-26(IV) and adjacent regions in Caenorhabditis elegans. Mol Gen Genet 221: 459-465

Chase BA, Kankel DR (1987) A genetic analysis of glutamatergic function in *Drosophila.* J Neuro 18: 15-41

Chou PY, Fasman GD (1979) Conservation of chain reversal regions in proteins. Biophys J 26: 385-399

Clark DV, Rogalski TM, Donati LM, Baillie DL (1988) The *unc-22(IV)* region of *Caenorhabditis elegans*: Genetic analysis of lethal mutations. Genetics 119: 345-353

Clark DV (1990) The *unc-22 IV* region of *Caenorhabditis elegans*: genetic analysis and molecular mapping. Ph. D. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Clark DV, Baillie DL (1992) Genetic analysis and complementation by germ-line transformation of lethal mutations in the *unc-22 IV* region of *Caenorhabditis elegans.* Mol Gen Genet 232:97-105

Cooper JR, Bloom FE, Roth RH (1991) The biochemical basis of neuropharmacology 1. 133-166, Oxford University Press, New York

Coulson A, Waterston R, Kiff J, Sulston J, Kohara Y (1988) Genome linking with yeast artificial chromosomes. Nature 335: 184-186

Coulson A, Sulston J, Brenner S, Karn J (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans.* Proc Natl Acad Sci USA 83: 7821-7825

Coulson A, Kozono Y, Lutterbach B, Shownkeen R, Sulston J, Waterston R (1991) YACS and the *C. elegans* genome. Bioessays 13: 413-417

Counillon L, Pouyssegur J (1993) Nucleotide sequence of the Chinese hamster Na+/H+ exchanger NHE1. Biochim et biophys acta 1172: 343-345

Dalley BK, Rogalski TM, Tullis GE, Riddle DL, Golomb M (1993) Post-transcriptional regulation of RNA polymerase II levels in *Caenorhabditis elegans.* Genetics 133: 237-245

Dayhoff MO, Schwartz RM, Orcutt BC (1978) In: Atlas of protein sequence and structure , Vol 5 supplement 3. Dayhoff MO (ed), NBRF, Washington, p 345

DeLong L, Plenefisch JD, Klein RD, Meyer BJ (1993) Feedback control of sex determination by dosage compensation revealed through *Caenorhabditis elegans sdc-3* mutations. Genetics 133: 875-896

Desai C, Garriga G, McIntire S L, Horvitz, HR (1988) A genetic pathway for the development of the *C. elegans* HSN motor neurons. Nature 336: 638-646

Dimaline R, Sandvik AK (1991) Histidine decarboxylase gene expression in rat fundus is regulated by gastrin. FEBS lett 281: 20-22

Donati LAM (1985) A genetic analysis of the right arm of Linkage Group IV of *Caenorhabditis elegans*, with emphasis on the *sDf2* region. M.Sc. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Driscoll M, Chalfie M (1991) The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. Nature 349: 588-593

Edgley ML, Riddle DL (1990) The nematode *Caenorhabditis elegans*. In: Genetic maps: Locus maps of complex genomes. Fifth ed., O'Brien SJ (ed) 3.111-3.133, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Eissenberg JC, Elgin SC (1991) Boundary functions in the control of gene expression. Trends Genet 7: 335-340

Emmons SW, Klass MR, Hirsh D (1979) Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci 76: 1333-1337

Emmons SW (1988) The genome. In: The nematode *Caenorhabditis elegans*. Wood WB (ed) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Eveleth DD, Marsh JL (1986) Evidence for evolutionary duplication of genes in the dopa decarboxylase region of *Drosophila*. Genetics 114:469-483

Eveleth DD, Geitz RD, Spencer CA, Nargang FE, Hodgetts RB, Marsh JL (1986) Sequence and structure of the dopa decarboxylase gene of *Drosophila*: Evidence for novel RNA splicing variants. EMBO J 5: 2663-2672

Feinberg A, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13

Ferguson EL, Horvitz HR (1985) Identification and characterization of 22 genes that affect the vulval cell lineages of the nematode *Caenorhabditis elegans*. Genetics 110: 17-72

Fire A (1986) Integrative transformation of *Caenorhabditis elegans*. EMBO J 5: 2673-2680

Gasser SM, Laemmli UK (1986) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *Drosophila melanogaster*. Cell 46: 521-530

Gavel Y, von Heijne G (1990) Cleavage-site motifs in mitochondrial targeting peptides. Protein Engineering 4: 33-37

Gilchrist EJ, Moerman DM (1992) Mutations in the *sup-38* gene of *Caenorhabditis elegans* suppress muscle attachment defects in Unc-52 mutants. Genetics 132: 431-442

Gilchrist EJ (1993) Genetic and molecular analysis of mutations in the *unc-52* gene of *Caenorhabditis elegans*, and isolation of a suppressor, *sup-38*. Ph.D. Thesis, University of British Columbia, Vancouver, British Columbia, Canada

Gould GW, Holman GD (1993) The glucose transporter family: structure, function and tissue specific expression. Biochem J 295: 329-341

Grinstein S, Rotin D, Mason MJ (1989) Na+/H+ exchange and growth factor induced cytosolic pH changes. Role in cellular proliferation. Biochim et biophys acta 988: 73-97

Hager EJ, Miller OL (Jr) (1991) Ultrastructural analysis of polytene chromatin of *Drosophila melanogaster* reveals clusters of tightly linked co-expressed genes. Chromosoma 100: 173-186

Han M, Sternberg PW (1990) *let-60*, a gene that specifies cell fates during *C. elegans* vulval induction, encodes a *ras* protein. Cell 63: 921-931

Hattori M, Sakaki Y (1986) Dideoxy sequencing method using denatured plasmid templates. Anal Biochem 152: 232-238

Heery DM, Gannon F, Powell R (1990) A simple method for subcloning DNA fragments from gel slices. Trends Genet 6: 173

Heine U, Blumenthal T (1986) Characterization of regions of the *Caenorhabditis elegans* X chromosome containing vitellogenin genes. J Mol Biol 188: 301-312

Henikoff S (1987) Unidirectional digestion with Exonuclease III in DNA sequence analysis. Methods Enzymol 155: 156-165

Henikoff S, Eghtedarzadeh K (1987) Conserved arrangement of nested genes at the *Drosophila Gart* locus. Genetics 117: 711-725

Henikoff S, Henikoff JG (1991) Automated assembly of protein blocks for database searching. Nucleic Acids Res 19: 6565-6572

Herman RK, Madl JE, Kari CK (1979) Duplications in *Caenorhabditis elegans*. Genetics 92: 419-435

Herman RK (1987) Mosaic analysis of two genes that affect nervous system structure in *Caenorhabditis elegans*. Genetics 116: 377-388

Herman RK, Kari CK (1989) Recombination between small X chromosome duplications and the X chromosome in *Caenorhabditis elegans*. Genetics 121: 723-737

Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H (1990) Cloning of the T gene required in mesoderm formation in the mouse. Nature 343: 617-622

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5: 151-153

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. CABIOS 8: 189-191

Hilliker AJ, Clark SH, Chovnick A, Gelbart WM (1980) Cytogenetic analysis of the chromosomal region immediately adjacent to the *rosy* locus in *Drosophila melanogaster*. Genetics 95: 95-110

Hodgkin JA (1983) Male phenotypes and mating efficiency in *C. elegans*. Genetics 103: 43-64

Horvitz HR, Brenner S, Hodgkin J, Herman RK (1979) A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol Gen Genet 175: 129-133

Horvitz HR, Chalfie M, Trent C, Sulston JE, Evans PD (1982) Serotonin and octopamine in the nematode. Science 216: 1012-1014

Howell AM, Gilmour SG, Mancebo RA, Rose AM (1987) Genetic analysis of a large autosomal region in *Caenorhabditis elegans* by the use of a free duplication. Genet Res 49: 207-213

Howell AM, Rose AM (1990) Essential genes in the *hdf6* region of chromosome I in *Caenorhabditis elegans*. Genetics 126: 583-592

Huang X-Y, Barrios LAM, Vonkhorporn P, Honda S, Albertson DG, Hecht RM (1989) Genomic organization of the glyceraldehyde-3-phosphate dehydrogenase gene family of *Caenorhabditis elegans*. J Mol Biol 206: 411-424

Ichinose H, Kurosawa Y, Titanik K, Fujita K, Nagatsu T (1989) Isolation and characterization of a cDNA clone encoding human aromatic L-amino acid decarboxylase. Biochem Biophys Res Comm 164: 1024-1030

Johnsen RC (1990) Genetic analysis of the left half of Linkage Group V in *Caenorhabditis elegans*. Ph.D. Thesis, Simon Fraser University, Burnaby, B.C. Canada

Johnsen RC, Baillie DL (1991) Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. Genetics 129: 735 - 752

Jones SJM (1994) The molecular characterization of the essential gene *let-653* in *Caenorhabditis elegans*. M.Sc. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Joseph DR, Sullivan P, Wang Y-M, Kozak C, Fenstermacher DA, Behrendsen ME, Zahnow CA (1990) Characterization and expression of the complementary DNA encoding rat histidine decarboxylase. Proc Natl Acad Sci USA 87: 733-737

Judd BH, Shen MW, Kaufman TC (1972) The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. Genetics 71: 139-156

Kennison JA (1993) Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. Trends Genet 9: 75-78

Kimble J, Hirsh D (1979) The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Dev Biol 70: 396-417

Kispert A, Herrmann BG (1993) The *Brachyury* gene encodes a novel DNA binding protein. EMBO J 12: 3211-3220

Klein P, Kanehisa M, DeLisi C (1985) The detection and classification of membrane-spanning proteins. Biochim biophys acta 815: 468-476

Knickelbein RG, Aronson PS, Dobbins JW (1990) Characterization of Na/H exchangers on villus cells in rabbit ileum. Am J Physiol 259: 802-806

Kramer JM, French RP, Park EC, Johnson JJ (1990) The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. Mol Cell Biol 10:2081-2089

Krause M, Hirsh D (1987) A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49: 753-761

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-132

Landschulz WH, Johnson PF, McKnight SL (1989) The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 243: 1681-1688

Lewis JA, Wu CH, Berg H, Levine JH (1980) The genetics of levamisole resistance in the nematode *Caenorhabditis elegans*. Genetics 95: 905-928

L'Hernault SW, Benian GM, Emmons RB (1993) Genetic and molecular characterization of the *Caenorhabditis elegans* spermatogenesis-defective gene *spe-17*. Genetics 134: 769-780

Liu Q-R, Mandiyan S, Nelson H, Nelson N (1992) A family of genes encoding neurotransmitter transporters. Proc Natl Acad Sci USA 89: 6639-6643

Livingstone MS (1981) Two mutations in *Drosophila* differentially affect the synthesis of octopamine, dopamine, and serotonin by altering the activities of two different amino acid decarboxylases. Soc Neurosci Abtr 7: 117.1

Lunan KD, Mitchell HK (1969) The metabolism of tyrosine-o-phosphate in *Drosophila*. Arch Biochem Biophys 132: 450-456

Marger MD, Saier MH, Jr (1993) A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem Sci 18: 13-20

Marra MA, Prasad SS, Baillie DL (1993) Molecular analysis of two genes between *let-653* and *let-56* in the *unc-22(IV)* region of *C. elegans*. Mol Gen Genet 236: 289-298

Marra MA, Baillie DL (1994) Recovery of duplications by drug resistance selection in *Caenorhabditis elegans*. Genome, in press

McCormack K, Campanelli JT, Rawaswami M, Mathew MK, Tanouye MA (1989) Leucine zipper update. Nature 340: 103-104

McDonald H, Smailus D, Jenkins H, Adams K, Simpson NE, Goodfellow PJ (1992) Identification and characterization of a gene at D10S94 in the MEN2A region. Genomics 13: 344-348

McIntire SL, Jorgensen E, Horvitz HR (1993) Genes required for GABA function in *Caenorhabditis elegans*. Nature 364: 334-336

McIntire SL, Jorgensen E, Kaplan J, Horvitz HR (1993) The GABAergic nervous system of *Caenorhabditis elegans*. Nature 364: 337-340

McKim KS, Howell AM, Rose AM (1988) The effects of translocations on recombination frequency in *Caenorhabditis elegans*. Genetics 120: 987-1001

McKim KS, Rose AM (1990) Chromosome I duplications in *Caenorhabditis elegans*. Genetics 124: 115-132

McKim KS, Starr T, Rose AM (1992) Genetic and molecular analysis of the *dpy-14* region in *Caenorhabditis elegans*. Mol Gen Genet 233: 241-251

McKim KS, Matheson C, Marra MA, Wakarchuk MF, Baillie DL (1994) The *Caenorhabditis elegans unc-60* gene encodes proteins homologous to a family of actin-binding proteins. Mol Gen Genet 242: 346-357

Mello CC, Kramer JM, Stinchcomb D, Ambros V (1991) Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J 10: 3959-3970

Meneely PM, Herman RK (1979) Lethals, steriles, and deficiencies in a region of the X chromosome of *Caenorhabditis elegans*. Genetics 92: 99-115

Meneely PM, Herman RK (1981) Suppression and function of X linked lethal and sterile mutations in *C. elegans*. Genetics 97: 65-84

Meneely PM, Wood WB (1987) Genetic analysis of X-chromosome dosage compensation in *Caenorhabditis elegans*. Genetics 117: 25-41

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101: 20

Meyer BJ, Casson LP (1986) *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell 47: 871-881

Meyerowitz EM, Hogness DS (1982) Molecular organization of a Drosophila puff site that responds to ecdysone. Cell 28: 165-176

Mirkovitch J, Mirault M-E, Laemmli UK (1984) Organization of the higher order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell 39: 223-232

Mirkovitch J, Spierer P, Laemmli UK (1986) Genes and loops in 320,000 base-pairs of the *Drosophila melanogaster* chromosome. J Mol Biol 190: 255-258

Moerman DG, Baillie DL (1979) Genetic organization in *C. elegans*: Fine structure analysis of the *unc-22 gene*. Genetics 91: 95-103

Moerman DG (1980) A genetic analysis of the *unc-22* region in *Caenorhabditis elegans*. Ph. D. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Moerman DG, Baillie DL (1981) Formaldehyde mutagenesis in the nematode *Caenorhabditis elegans*. Mutat. Res. 80: 273-279

Moerman DG, Plurad S, Waterston RH, Baillie DL (1982) Mutations in the *unc-54* myosin heavy chain gene of *Caenorhabditis elegans* that alter contractility but not muscle structure. Cell 29: 773-781

Moerman DG, Benian GM, Waterston RH (1986) Molecular cloning of the muscle gene *unc-22* in *Caenorhabditis elegans* by Tc1 transposon tagging. Proc Natl Acad Sci 83: 2579-2583

Moerman DG, Benian GM, Barstead RJ, Schreifer L, Waterston RH (1988) Identification and intracellular localization of the *unc-22* gene product of *Caenorhabditis elegans*. Genes Dev 2: 93-105

Nelson WG, Pienta KJ, Barrack ER, Coffey DS (1986) The role of the nuclear matrix in the organization and function of DNA. Ann Rev Biophys Biophys Chem 15: 457-475

Oliver SG, van der Aart QJM, Agostoni-Carbone ML, Aigle M, Alexandraki D,Antoine G, Anwar R, Ballesta JPG, Benit P, Berben G, Bergantino E, Biteau N, Bolle PA, Bolotin-Fukuhara M, Brown A, Brown AJP, Buhler JM, Carcano G, Carignani G, Cederberg G, Chanet R, Contreras R, Crouzet M, Daignan-Fournier B, Defoor E, Delgado M, Demolder J, Doira C, Dubois E, Dujon B, Dusterhoft A, Erdman D, Esteban M, Fabre F, Fairhead C, Faye G, Feldmann H, Fiers W, Francingues-Gaillard MC, Franco L, Frontali L, Fukuhara H, Fuller LJ, Galland P, Gent ME, Gigot D, Gilliquet V, Glansdorff N, Goffeau A, Grenson M, Grisanti P, Grivell LA, de Haan M, Haasemann M, Hatat D, Hoenicke J, Hegemann J, Herbert CJ, Hilger F, Hohmann S, Hollenberg CP, Huse K, Iborra F, Indge KJ, Isono K, Jacq C, Jacquet M, James CM, Jauniaux JC, Jia Y, Jimenez A, Kelly A, Kleinhans U, Kreisl P, Lanfranchi G, Lewis C, van der Linden CG, Lucchini G, Lutzenkirchen K, Maat MJ, Mallet L, Mannhaupt G, Martegani E, Mathieu A, Maurer CTC, McConnell D, McKee RA, Messenguy F, Mewes HW, Molemans F, Montague MA, Muzi Falconi M, Navas L, Newlon CS, Noone D, Pallier C, Panzeri L, Pearson BM, Perea J, Philippsen P, Pierard A, Planta RJ, Plevani P, Poetsch B, Pohl F, Purnelle B, Ramezani Rad M, Rasmussen W, Raynal A, Remacha M, Richterich P, Roberts AB, Rodriguez F, Sanz E, Schaaff-Gerstenschlager I, Scherens B, Schweitzer B, Shu Y, Skala J, Slonimski PP, Sor F, Soustelle C, Spiegelberg R, Stateva LI, Steensma HY, Steiner S, Thierry A, Thireos G, Tzermia M, Urrestarazu LA, Valle G, Vetter I, van Vliet-Reedijk JC, Voet M, Volckaert G, Vreken P, Wang H, Warmington JR, von Wettstein D, Wicksteed BL, Wilson C, Wurst H, Xu G, Yoshikawa A, Zimmermann FK, Sgouros JG (1992) The complete DNA sequence of yeast chromosome III. Nature 357:38-46

Opstelten RJ, Clement JM, Wanka F (1989) Direct repeats at nuclear matrix-associated DNA regions and their putative control function in the replicating eukaryotic genome. Chromosoma 98: 422-427

Ouellette BFF, Clark MW, Keng T, Storms RK, Zhong W, Zeng B, Fortin N, Delaney S, Barton A, Kaback DB, Bussey H (1993) Sequencing of chromosome I from *Saccharomyces cerevisiae*: analysis of a 32 kb region between LTE1 and SPO7 genes. Genome 36: 32-42

Pak WL, Grossfield J, White NV (1969) Nonphototactic mutants in a study of vision of *Drosophila*. Nature 222: 351-354

Pak WL (1975) In King RC (ed) Handbook of genetics, Plenum Press, New York, Vol. 3, pp 703-733

Pak WL (1979) In Breakefield XO (ed) Neurogenetics: Genetic approaches to the nervous system. Elsevier North Holland, New York, pp 67-99

Paulson JR, Laemmli UK (1977) The structure of histone depleted metaphase chromosomes. Cell 12: 817-828

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence analysis. Proc Natl Acad Sci USA 85: 2444-2448

Pertel R, Wilson SH (1974) Histamine content of the nematode, *Caenorhabditis elegans*. Comp Gen Pharmac 5: 83-85

Peters K, McDowall J, Rose AM (1991) Mutations in the *bli-4(I)* locus of *Caenorhabditis elegans* disrupt both adult cuticle and early larval development. Genetics 129: 95-102

Pflugfelder GO, Roth H, Poeck B, Kerscher S, Schwarz H, Jonschker B, Heisenberg M (1992) The lethal(1)optomotor-blind gene of *Drosophila melanogaster* is a major organizer of optic lobe development: Isolation and characterization of the gene. Proc Natl Acad Sci 89: 1199-1203

Plasterk RHA, Groenen JIM (1992) Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J 11: 287-290

Pouyssegur J, Sardet C, Franchi A, L'Allemain G, Paris S (1984) A specific mutation abolishing Na^+/H^+ antiport activity in hamster fibroblasts precludes growth at neutral and acidic pH. Proc Natl Acad Sci USA 81: 4833-4837

Prasad SS (1988) Molecular analysis within the *dpy-20* to *unc-22* region of chromosome IV in *Caenorhabditis elegans*. Ph.D. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Prasad SS, Baillie DL (1989) Evolutionarily conserved coding sequences in the *dpy-20 - unc-22* region of *Caenorhabditis elegans*. Genomics 5: 185-198

Rand JB (1989) Genetic analysis of the *cha-1 - unc-17* gene complex in *Caenorhabditis elegans*. Genetics 122: 73-80

Reizer J, Reizer A, Saier MH, Jr (1993) The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. Critical Reviews in Biochemistry and Molecular Biology, 28: 235-257

Rogalski TM, Moerman DG, and Baillie DL (1982) Essential genes and deficiencies in the *unc-22 IV* region of *Caenorhabditis elegans*. Genetics, 102: 725-736

Rogalski TM (1984) A genetic characterization of the region around the *unc-22(IV)* gene in *Caenorhabditis elegans*. Ph.D. Thesis, Simon Fraser University, Burnaby, British Columbia, Canada

Rogalski TM, Baillie DL (1985) Genetic organization of the *unc-22 IV* gene and the adjacent region in *Caenorhabditis elegans*. Mol Gen Genet 201: 409-414

Rogalski TM, Riddle DL (1988) A *Caenorhabditis elegans* RNA polymerase gene, *ama-1 IV*, and nearby essential genes. Genetics 118: 61-74

Rogalski TM, Williams BD, Mullen GP, Moerman DG (1993) Products of the *unc-52* gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. Genes Dev 7: 1471-1484

Rose AM and Baillie DL (1979) The effect of temperature and parental age on recombination and non-disjunction in *Caenorhabditis elegans*. Genetics 92: 409-418

Rose AM and Baillie DL (1980) Genetic organization of the region around *unc-15(I)*, a gene affecting paramyosin in *Caenorhabditis elegans*. Genetics 96: 639-648

Rose AM, Baillie DL, Candido EPM, Beckenbach KA, Nelson D (1982) The linkage mapping of cloned restriction fragment length differences in *Caenorhabditis elegans*. Mol Gen Genet 188: 286-291

Rose AM, Baillie DL, Curran J (1984) Meiotic pairing behaviour of two free duplications of linkage group I in *Caenorhabditis elegans*. Mol Gen Genet 195: 52-56

Rosenbluth RE, Cuddeford CE, Baillie DL (1983) Mutagenesis in *Caenorhabditis elegans* I. A rapid eukaryotic test system using the reciprocal translocation, *eT1(III;V)* Mutat Res 110: 39-48

Rosenbluth RE, Cuddeford C, Baillie DL (1985) Mutagenesis in *Caenorhabditis elegans*. II. A spectrum of mutational events induced with 1500 R of gamma-radiation. Genetics 109: 493-511

Rosenbluth RE, Rogalski TM, Johnsen RC, Addison LM, Baillie DL (1988) Genomic organization in *Caenorhabditis elegans*: deficiency mapping on linkage group V (left). Genet. Res. 52: 105-118

Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing, and in vitro amplification of DNA. Nucleic Acids Research 17: 8543-8551

Saiki, RK, Gelfand, DH, Stoffel, S, Scharf, SJ, Higuchi, R, Horn, GT, Mullis, KB, Erlich, HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239: 487-491

Sambrook, J, Fritsch, EF, and Maniatis, T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2nd Ed.

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467

Sardet C, Counillon L, Franchi A, Pouyssegur J (1990) Growth factors induce phosphorylation of the Na^+/H^+ antiporter, a glycoprotein of 110 KD. Science 247: 723-726

Sardet C, Franchi A, Pouyssegur J (1989) Molecular cloning, primary structure, and expression of the human growth factor activatable Na^+/H^+ antiporter. Cell 56: 271-280

Sarthy PV (1991) Histamine: a neurotransmitter candidate for *Drosophila* photoreceptors. J Neurochem 57: 1757-1768

Savanay A, Cronenberger L (1982) Properties of histidine decarboxylase from rat gastric mucosa. Eur J Biochem 12: 593-599

Schein JES, Marra MA, Benian GM, Fields C, Baillie DL (1993) The use of deficiencies to determine essential gene content in the *let-56 - unc-22* region of *Caenorhabditis elegans*. Genome 36: 1148-1156

Schulte-Merker S, Ho RK, Herrmann BG, Nusslein-Volhard C (1992) The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. Development 116: 1021-1032

Shenoy S, Choi J, Bagrodia S, Copeland TD, Maller JL, Shalloway D (1989) Purified maturation promoting factor phosphorylates $pp60^{c\text{-}src}$ at the sites phosphorylated during fibroblast mitosis. Cell 57: 763-774

Siegel G, Agranoff B, Albers RW, Molinoff P (1989) Basic neurochemistry. ed 4. Raven Press, New York

Sigurdson DC, Spanier GJ, Herman RK (1984) *C. elegans* deficiency mapping. Genetics 108: 331-345

Sirotkin K, Davidson N (1982) Developmentally regulated transcripts from *Drosophila melanogaster* chromosomal site 67B. Dev Biol 89: 196-210

Smith EL, Austen BM, Blumenthal KM, Nyc JF (1975) Glutamate dehydrogenases. In: The enzymes. Boyer PD (ed) 3rd edition vol. 11 293-367, Academic Press, New York

Smith GE, Summers MO (1980) The bidirectional transfer of DNA and RNA to nitrocellulose or DBM paper. Anal Biochem 109: 123-129

Smith HC, Puvion E, Buchholtz L, Berezney R (1984) Spatial distribution of DNA loop attachment and replicational sites in the nuclear matrix. J Cell Biol 99: 1794-1802

Snutch TP (1980) A molecular and genetic analysis of the heat shock response of *Caenorhabditis elegans*. Ph.D. Thesis, Simon Fraser University, Burnaby, B.C., Canada

Snyder M, Hirsch J, Davidson N (1981) The cuticle genes of Drosophila: A developmentally regulated gene cluster. Cell 25: 165-177

Snyder M, Hunkapiller M, Yuen D, Silvert D, Fristom J, Davidson N (1982) Cuticle protein genes of *Drosophila*: Structure, organization and evolution of four clustered genes. Cell 29: 1027-1040

Southern EM (1975) Detection of specific sequences among DNA fragments separated by by gel electrophoresis. J Mol Biol 98: 503-517

Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T (1993) Operons in C. elegans: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell 73: 521-532

Starr TVB (1989) Molecular analysis of the *dpy-14* region of chromosome I in *Caenorhabditis elegans*. Ph.D. Thesis, University of British Columbia, Vancouver, B.C., Canada

Steward R, Nusslein-Volhard C (1986) The genetics of the *dorsal-bicaudal-D* region of *Drosophila melanogaster*. Genetics 113: 665-678

Stief A, Winter DM, Stratling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position independent gene activity. Nature 341: 343-345

Sulston J, Brenner S (1974) The DNA of *Caenorhabditis elegans*. Genetics 77: 95-104

Sulston J, Dew M, Brenner S (1975) Dopaminergic neurons in the nematode *Caenorhabditis elegans*. J Comp Neurol 163: 215-226

Sulston J, Horvitz HR (1977) Post-embryonic cell lineages of the nematode *Caenorhabditis elegans.* Dev Biol 56: 110-156

Sulston J, Horvitz HR (1981) Abnormal cell lineages in mutants of the nematode *Caenorhabditis elegans.* Dev Biol 82: 41-55

Sulston J, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans.* Dev Biol 100: 64-119

Sulston J, Hodgkin J (1988) Methods. In: The nematode *Caenorhabditis elegans.* Wood WB (ed) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sulston J, Du Z, Thomas K, Wilson R, Hillier L, Staden R, Halloran N, Green P, Thierry-Mieg J, Qui L, Dear S, Coulson A, Craxton M, Durbin R, Berks M, Metzstein M, Hawkins T, Ainscough R, Waterston R (1992) The *C. elegans* sequencing project: a beginning. Nature 356: 37-41

Suzuki M (1989) SPXX, a frequent sequence motif in gene regulatory proteins. J Mol Biol 207: 61-84

Taguchi Y, Watanabe T, Kubota H, Hayashi H, Wada H (1984) Purification of histidine decarboxylase from the liver of fetal rats and its immunochemical and immunohistochemical characterization. J Biol Chem 259: 5214-5221

Tanaka T, Horio Y, Teketoshi M, Imamura I, Ando-Yamamoto M, Kangawa K, Matsuo H, Kurodo M, Wada H (1989) Molecular cloning and sequencing of a cDNA of rat dopa decarboxylase: Partial amino acid homologies with other enzymes synthesizing catelcholamines. Proc Natl Acad Sci USA 86: 8142-8146

Toledo A, Rodriguez R, Sabria J, Rodriguez J, Blanco I (1991) Calcium effects on the solubilization of membrane-bound histidine decarboxylase in the rat brain. J Neurochem 56: 380-384

Tse C-M, Ma AI, Yang VW, Watson AJM, Levine S, Montrose MH, Potter J, Sardet C, Pouyssegur J, Donowitz M (1991) Molecular cloning and expression of a cDNA encoding the rabbit ileal villus cell basolateral membrane Na+/H+ exchanger. EMBO J 10: 1957-1967

Tse C-M, Brant SR, Walker MS, Pouyssegur J, Donowitz M (1992) Cloning and sequencing of a rabbit cDNA encoding an intestinal and kidney-specific Na+/H+ exchanger isoform (NHE-3). J Biol Chem 267: 9340-9346

Thornton DJ, Davies JR, Kraayenbrink M, Richardson PS, Sheehan JK, Carlstedt I (1990) Mucous glycoproteins from normal human tracheobronchial secretion. Biochem J 265: 179

Van der Velden HM, Willigen G, Watzeles RHW, Wanka F (1984) Attachment of origins of replication to the nuclear matrix and the chromosomal scaffold. FEBS lett 171: 13-16

Vaughn JP, Dijkwel PA, Mullenders LHF, Hamlin JL (1990) Replication forks are associated with the nuclear matrix. Nucleic Acids Res 18: 1965-1969

Voet D, Voet JG (1990) Biochemistry. John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore

Ward S, Burke J, Sulston JE, Coulson AR, Albertson DG, Ammons D, Klass M, Hogan E (1988) The genomic organization of transcribed major sperm protein genes and pseudogenes in the nematode *Caenorhabditis elegans*. J Mol Biol 199: 1-13

Watanabe T, Yoshitaka T, Shiosaka S, Tanaka J, Kubota H, Terano Y, Tohyama M, Wada H (1984) Distribution of the histaminergic neuron system in the central nervous system of rats; a fluorescent immunohistochemical analysis with histidine decarboxylase as a marker. Brain res 295: 13-25

Waterston RH, Thomson JN, Brenner S (1980) Mutants with altered muscle structure in *Caenorhabditis elegans*. Dev Biol 77: 271-302

Waterston RH, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N, Metzstein M, Hawkins T, Wilson R, Berks M, Du Z, Thomas K, Thierry-Mieg J, Sulston J (1992) A survey of expressed genes in *Caenorhabditis elegans*. Nature Genet 1: 114-123

Williams BD, Schrank B, Huynh C, Shownkeen R, Waterston RH (1992) A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence -tagged sites. Genetics 131: 609-624

Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Bonfield J, Burton J, Connell M, Copsey T, Cooper J, Coulson A, Craxton M, Dear S, Du Z,

Durbin R, Favello A, Fraser A, Fulton L, Gardner A, Green P, Hawkins T, Hillier L, Jier M, Johnston L, Jones M, Kershaw J, Kirsten J, Laisster N, Latreille P, Lightning J, Lloyd C, Mortimore B, O'Callaghan M, , Parsons J, Percy C, Rifken L, Roopra A, Saunders D, Shownkeen R, Sims M, Smaldon N, Smith A, Smith M, Sonnhammer E, Staden R, Sulston J, Thierry-Mieg J, Thomas K, Vaudin M, Vaughan K, Waterston R, Watson A, Weinstock L, Wilkinson-Sproat J, Wohldman P (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. Nature 368: 32-38

Wright TRF, Black BC, Bishop CP, Marsh JL, Pentz ES, Steward R, Wright EY (1982) The genetics of dopa decarboxylase in *Drosophila melanogaster* V. *Ddc* and *l(2)amd* alleles: Isolation, characterization and intragenic complementation. Mol Gen Genet 188: 18-26

Wright TRF (1987) The genetic and molecular organization of the dense cluster of functionally related, vital genes in the dopa decarboxylase region of the *Drosophila melanogaster* genome. In: Henning W (ed) Results and problems in cell differentiation, vol 14: Eukaryotic chromosomes: Structure and function. Springer, Berlin, Heidelberg, New York, pp 95-120

Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SCR (1979) The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell 16: 797-806

Yamada M, Watanabe T, Harino S, Fukui H, Wada H (1980) The effects of protease inhibitors on histidine decarboxylase activities and assay of enzyme in various rat tissues. Biochim biophys acta 615: 458-464

Yamauchi K, Ruriko S, Ohkawara Y, Tanno Y, Maeyama K, Watanabe T, Satoh K, Yoshizawa M, Shibuhara S, Takishima T (1990) Nucleotide sequence of the cDNA encoding L-histidine decarboxylase derived from human basophilic leukemia cell line, KU-812-F. Nucleic Acids Research 18: 5891