A GENETIC AND MOLECULAR ANALYSIS OF THE *bli-4* LOCUS OF *Caenorhabditis elegans*, AN ESSENTIAL GENE ENCODING KEX2-LIKE SUBTILISIN-TYPE SERINE ENDOPROTEASES

by

KENNETH WILLIAM PETERS

BSc. (Honours) Simon Fraser University, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Genetics Program

Faculty of Medicine

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

JULY, 1992

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature) _____

Department of Medical Genetics

The University of British Columbia
Vancouver, Canada

Date January 7, 1993

ABSTRACT

This thesis reports a characterization of genetic mutations and molecular structure of the *bli-4 (I)* gene of *Caenorhabditis elegans*. The *bli-4* locus had been previously defined by a single recessive mutation, *e937*, which disrupts the structure of adult-stage cuticle causing the formation of fluid-filled separations of the cuticle layers, or blisters. A characterization of *e937* and eleven additional mutations is reported. The mutations were grouped into three classes based on phenotype and complementation analysis: Class I, represented by the single allele causing blistering, *e937*; Class II, nine lethal mutations that arrest development at the end of embryogenesis and that fail to complement all other mutations; and Class III, two larval lethal mutations, *s90* and *h754*, that complement *e937*. The complementation pattern provides evidence that all of these mutations are allelic, and that *bli-4* is a complex locus with an essential function late in embryogenesis.

A region of chromosome I that includes the *bli-4* locus was identified by aligning the *C. elegans* genetic and physical maps. The *bli-4* coding region was identified by using cosmids as hybridization probes to detect chromosomal alterations in the DNA of *bli-4* mutant strains. Two mutations of *bli-4* are small rearrangements of the gene; *e937*, is a 3.5 kilo base (kb) deletion, and *h1010* is an insertion of the 1.6 kb transposable element, Tc1. Three protein products that differ at the carboxyl end, are predicted from alternately spliced *bli-4* cDNA clones. The predicted proteins were designated blisterin A, blisterin B and blisterin C according to the order that the variantly spliced 3′ ends occur on the chromosome. Of the three *bli-4* cDNA clones characterized, only blisterin B includes an open reading frame beginning with an ATG start codon. The blisterin A and blisterin C open

reading frames begin within blisterin B, and are likely to be incomplete. The predicted blisterin B gene product has a potential secretion signal peptide at its amino terminal. Blisterin C has a potential transmembrane domain near its carboxyl terminal while blisterin A and B lack this domain. The blisterins share significant sequence identity with kex2-like serine endoproteases, which are responsible for the cleavage of secreted proteins in yeast and mammals. This characterization of *bli-4* provides the first evidence for an essential role of a *KEX2*-like gene in the development of a multi-cellular organism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEGMENTS

The work presented in this thesis represents the result of six years of my life as a graduate student in the lab of Dr. Ann Rose. During that time I learned from and was supported by many individuals. In particular, I would like to acknowledge the instruction of Ann Marie Howell, who helped me with genetics, and Terry Starr and Shiv Prasad, who taught me molecular techniques. I would also like to thank Kim McKim, Monique Zetka and Joe Babity. My work was greatly facilitated by my collaboration with Jennifer McDowall, who identified many of the mutations of the *bli-4* gene and who shared her germline transformation results with me. I am also grateful to Betty Leung for her assistance with DNA sequencing. The direction and support of my supervisory committee, Dr. David Baillie, Dr. Don Moerman, Dr. Tom Grigliatti and Dr. Fred Dill, is gratefully acknowledged. I would particularly like to thank my supervisor, Dr. Ann Rose, for her patience and support. Finally, I would like to thank my parents, Ron and Cynthia and my wife, Helen for their love and support. My work was made possible by Post Graduate Research Fellowships from the Medical Research Council of Canada, and the Natural Sciences and Engineering Research Council.

This thesis is dedicated to Helen and Jeffrey.

CHAPTER ONE: INTRODUCTION

1.1 GENERAL

Much of our present understanding of how organisms develop has come from analysis of heritable alterations in the genetic material. In many cases, genetic analysis has begun by the identification of a mutation with an unusual phenotype, and the curiosity of an investigator to know the basis of that phenotype. My thesis presents an analysis of the *bli-4* locus. *bli-4* is a member of a group of *C. elegans* genes that can be mutated to cause blisters. The blistered phenotype is fluid-filled separations of cuticle layers (Brenner, 1974). Blistering exhibits both variable penetrance and expressivity. When I began this work, six genes with mutations resulting in the blistered phenotype had been identified, named *bli-1* through *bli-6*. None of the blister genes had been cloned, and the physiological basis for blistering was unknown. The purpose of my thesis work was to determine the nature of the *bli-4* gene product with the goal of elucidating the biological function of the gene.

1.2 BIOLOGY OF *C. elegans*

Caenorhabditis elegans is a free-living soil nematode that exists as self-fertilizing hermaphrodites and males. *C. elegans* is widely used as model organism for genetic studies, following the work of Brenner (1974). The biology of *C. elegans* has been studied extensively, leading to insights into the genetic control of its life cycle, genomic organisation, development, sexual differentiation, anatomy, muscle and nervous system, and cell lineage. This work is reviewed in the book, The nematode *Caenorhabditis elegans* (Wood, 1988). The hermaphrodite gonad produces approximately 300 hundred sperm during the last larval stage, then switches to the production of ova through its adult

life. Ova are fertilized as they pass through a spermatheca, where the sperm are stored. The fertilized egg, or zygote, begins to develop immediately, and is passed through the vulva into the surrounding environment 2 to 4 hours later. The embryo continues development until the embryo hatches out of its egg, about 14 hours post-fertilization at 25C. The worm passes through four larval stages lasting about 45 hours, termed L1, L2, L3 and L4. Each developmental stage is punctuated by a cuticle moult and the synthesis of a new cuticle (Singh and Sulston, 1978). Worms reach adulthood about 2.5 days after fertilization at 25C. An additional developmental stage, the dauer larvae, occurs as a substitute for the L2 stage under conditions of extreme stress such as starvation (reviewed by Riddle, 1988). At the time of hatching, hermaphrodites have 558 somatic cells, males have 560 (Sulston *et al.*, 1983). Adult worms are about one mm in length, and have about 1000 somatic nuclei (Sulston and Horvitz, 1981).

Sex is determined by X to autosome ratio; hermaphrodites have two X chromosomes (XX), males have one (XO) (Reviewed by Hodgkin, 1980). Male worms are produced periodically due to nondisjunction of the X chromosome (Hodgkin, Horvitz, and Brenner, 1979), at a frequency of about 0.1% at 20C (Rose and Baillie, 1979). When males are crossed to hermaphrodites, the male sperm is used in preference to the hermaphrodite sperm, and about 50% of the cross-progeny are male. In the absence of males, *C. elegans* reproduces as a hermaphrodite, producing many progeny from a single individual. A single hermaphrodite worm produces about 300 progeny. This large number of progeny combined with the short life-cycle (3.5 days at 20C) and ease of manipulation makes the worm an ideal organism for genetic studies.

The *C. elegans* genome is small, comprised of 1 X 10^8 nucleotides, of which 17% is repetitive DNA (Sulston, J. and S. Brenner. 1974; J. Sulston, personal communication). *C. elegans* has six chromosomes, or linkage groups (LG); five autosomes, LGI, LGII, LGIII, LGIV and LGV, and a sex-determining chromosome, LGX. Minimum estimates of the total number of essential genes based on lethal saturation screens over large regions of the genome range from 2,000 (Brenner, 1974) to 6,000 (Moerman and Baillie, 1979; Rogalski and Baillie, 1985; Clark *et al.*, 1988; Howell, 1989; McDowall, 1990; Johnsen and Baillie, 1991). Based on the estimated 8.3 X 10^7 bp of unique sequence in the genome, the genetic estimates of gene numbers predict a gene density of 10 kilo bases (kb) per gene to 40 kb per gene. Molecular estimates based on cross species hybridization (Heine and Blumenthal 1986; Prasad and Baillie, 1989) estimate one conserved region per 10 to 15 kb. Computer analysis of DNA sequence data for a 120 kb region of LGIII identified one potential coding region per 3 to 4 kb. Based on this estimate of gene density, the total number of genes could be as high as 15,000 (Sulston *et al.*, 1992). Thus, *C. elegans* is estimated to have one gene for every 2,000 to 15,000 nucleotides.

A variety of mutations altering the development, behavior and appearance of the worm have been identified. Morphological mutations include dumpy (*dpy*), roller (*rol*), squat (*sqt*), long (*lon*), small (*sma*), and blister (*bli*) (Brenner, 1974; Higgins and Hirsh, 1977; Cox *et al.*, 1980; Kusch and Edgar, 1986). Dpy worms are short and fat (Brenner, 1974); Rol worms roll to the right or left due to the helical twisting of their cuticle (Higgins and Hirsh, 1977); Sqt mutants are dominant Rol and recessive Dpy (Cox *et al.*, 1980; Kusch and Edgar, 1986); Lon worms are longer than wild type (Brenner, 1974); Sma are smaller; and Bli have fluid-filled cuticular swellings

(Brenner, 1974). More than 900 *C. elegans* genes have been identified by mutational analysis (Edgley and Riddle, 1990).

1.3 CUTICLE STRUCTURE

C. elegans has a complex, developmentally regulated extracellular cuticle that functions as a hydrostatic exoskeleton (Figure 1). For this reason, the cuticle of *C. elegans* has been proposed as a model system for the study of the assembly, architecture and function of extracellular matrices (Higgins and Hirsh, 1977; Cox *et al.*, 1980). Biochemical and ultrastuctural analyses have revealed that the cuticle is arranged in two layers, a basal layer and a cortical layer, and is composed of collagenous proteins that are extensively cross-linked by disulfide bonds, as well as other proteins that are resistant to collagenase (Cox, Kusch and Edgar, 1981). The structures of the layers vary with developmental stage (Cox, Staprans and Edgar, 1981), illustrated in Figure 1. The adult cuticle has an additional layer consisting of a fluid-filled space spanned by columnar structures termed struts connecting the basal and cortical layers (Cox, Kusch and Edgar, 1981). The *C. elegans* collagen gene family consists of 50 to 150 members, encoding small collagens of 30 to 40 kilodaltons (kd) that are covalently cross-linked in the cuticle (Kramer Cox and Hirsh, 1982; Cox, Kramer and Hirsh, 1984). Collagen gene expression varies with developmental stage (Cox and Hirsh, 1985; Kramer, Cox and Hirsh, 1985). Recently three genes that affect cuticle morphology, *dpy-13* (von Mende *et al.*, 1988), *sqt-1* (Kramer *et al.*, 1988), and *rol-6* (Kramer *et al.*, 1990) have been cloned and shown to be collagen genes.

Figure 1. Developmental changes in cuticle structure

Diagrammatic representation of cuticle structure from adults, L4, dauer, and L1 worms. The basal and cortical layers in adult cuticle are separated by a fluid-filled layer and connected by columnar structures called struts. C = Cortical layer; B = Basal layer; St = Strut; Fl = Fluid layer; Fb = Fibrillar layer; SL = Striated layer. After Cox, Staprans and Edgar, 1981.

1.4 ORIGIN OF *bli-4* ALLELES

bli-4(e937) was induced with ^{32}P (Babu and Brenner, unpublished results, cited in Brenner, 1974). *e937* was the only allele of *bli-4* identified prior to this study. *bli-4* is located in the cluster of genes on LGI, and was positioned between *dpy-5* and *dpy-14* in a region that is covered by the free duplication *sDp2* (Figure 2).

sDp2 has been used as a balancer in lethal screens by the Rose lab (Howell *et al.*, 1987; Howell, 1989; McDowall, 1990). More than 500 lethal mutations in the *sDp2* balanced region have been isolated following EMS mutagenesis. A number of the LGI lethal mutations that failed to complement the blistered phenotype of *e937* were identified by complementation analysis (Peters, McDowall and Rose, 1991). The first such lethal mutation was *h42* (Howell *et al.*, 1987). Subsequently, seven other *sDp2* balanced lethal mutations that failed to complement the blistered phenotype of *e937* were identified by complementation analysis. The complementation tests that identified these mutations were performed by the individuals listed in Table 1. All of the lethal mutations that failed to complement *e937* also failed to complement the *sDp2* balanced mutation *h754*, and the mutation *s90*, which was identified by Rose and Baillie (1980), and assigned the gene name *let-77*. *s90* and *h754* complement *e937*. One goal of my thesis work was to explain this complementation pattern.

Table 1. Alleles of *bli-4*

Allele	Phenotype	Origin
e937	Adult blisters	Brenner, 1974
h42	late embryonic arrest	Howell *et al*, 1987
h199	"	D. Pilgrim and A.M. Rose
h254	"	A.M. Howell and A.M. Rose
h384	"	McDowall, 1990
h427	"	"
h520	"	"
h699	"	"
h754	"	"
h791	L1 arrest	"
s90[a]	"	Rose and Baillie, 1980

[a] *s90* was assigned to *let-77* in Rose and Baillie, 1980.

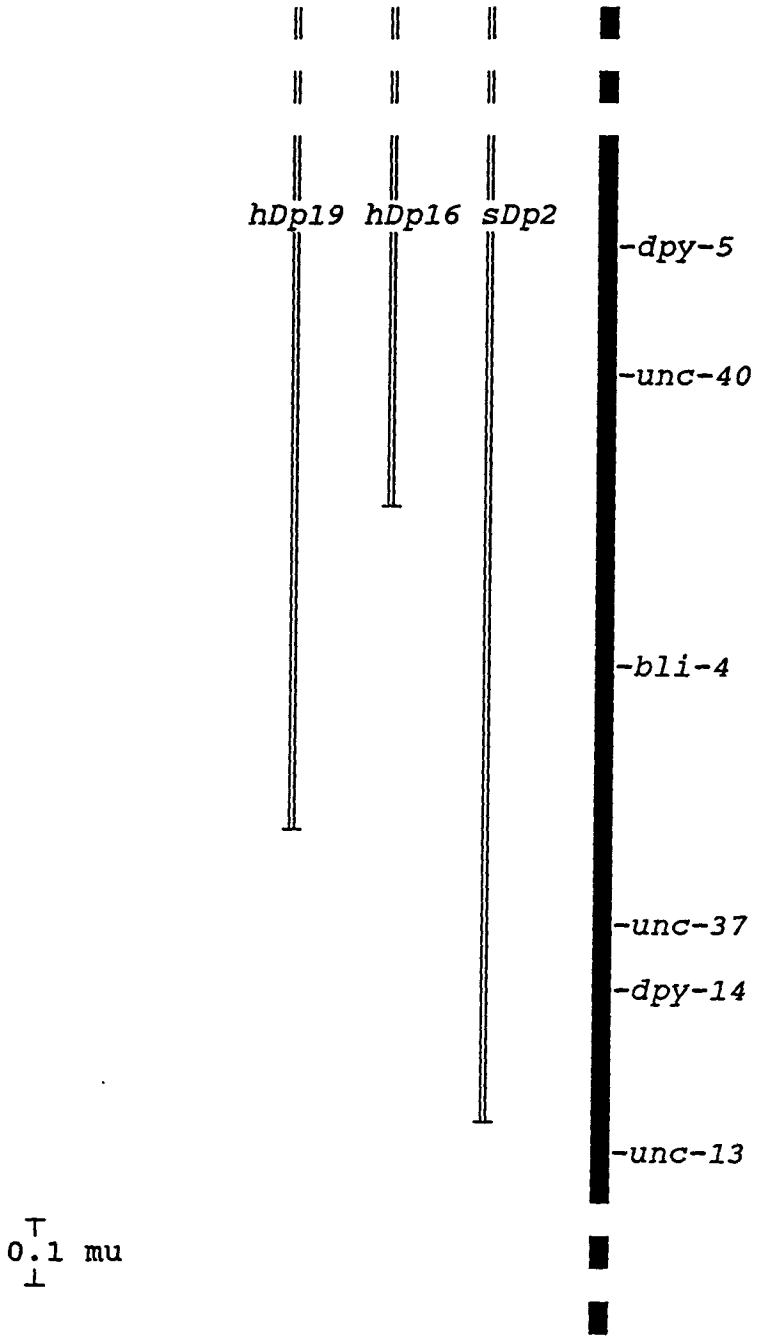


Figure 2. Genetic map of the *bli-4* region

A partial genetic map of the region of LGI around *bli-4* illustrating map positions for closely linked visible markers and duplications. *sDp2* was mapped by Rose, Baillie and Curran, (1984). *hDp16* and *hDp19* were generated and mapped by McKim and Rose, 1990, and further mapped by McDowall (1990). Some map data is from Edgley and Riddle (1990).

1.5 RESOURCES USED TO CLONE *bli-4*

An important resource exploited in this thesis was the *C. elegans* physical map. The physical map consists of contiguous sets (contigs) of overlapping cosmid and yeast artificial chromosome clones (YACs) covering most of the *C. elegans* genome (Coulson *et al.*, 1986; Coulson *et al.*, 1988). Cosmid contigs have been aligned with the genetic map through the cloning and mapping of genetic markers by numerous labs.

Strain-specific fragment length difference (RFLD) mapping has been widely used in *C. elegans* to position genetic markers with respect to molecular markers (examples are found in Rose *et al.*, 1982; Baillie *et al.*, 1985; Williams *et al.*, 1992). A contig spanning an interval that includes *bli-4* was aligned with the genetic map through the mapping of a strain specific RFLD, *hP5* (RFLD) (Starr *et al.*, 1989). Mapping of *bli-4* on the physical map was accomplished by mapping *hP5* and a breakpoint of the deletion *hDf8*. This mapping contributed to the cloning of *bli-4* by restricting the range of cosmids that could include the locus.

The technique of transposon tagging was used to identify the *bli-4* coding region. Insertion of a transposable element into a locus of interest creates a mutation that can be identified by Southern hybridization analysis. *C. elegans* genes that have been cloned by this method include *lin-12* (Greenwald, 1985) and *unc-22* (Moerman, Benian and Waterston, 1986). Tc1 is a 1610 bp transposable element in *C. elegans* (Emmons *et al.*, 1983; Rosenzweig, 1983). Tc1 is not normally mobile in *C. elegans* strain N2 (Eidie and Anderson, 1985a), but transposes at a high rate in *C. elegans* strain BO (Moerman and Waterston, 1984; Eidie and Anderson, 1985b). Loci responsible for the mutator activity have been crossed into an N2 background (Mori, Moerman and Waterston, 1988). In this study, I used the

mutator locus *mut-6(st702)* to isolate a spontaneous allele of *bli-4*, *h1010*. *h1010* provided a detectable rearrangement that contributed to the identification of the cosmid containing the *bli-4* locus.

1.6 REVIEW OF DIBASIC ENDOPROTEASES

As I will describe in chapter three, *bli-4* encodes a KEX2-like endoprotease. Therefore, I introduce this class of proteins here. The existence of prohormone processing enzymes was inferred when it was discovered that some proteins, such as pituitary hormones (Chretien and Li, 1967) and insulin (Steiner *et al.*, 1967; Chance, Ellis and Brommer, 1968), are synthesized as inactive precursors. Many secreted proteins that are excised by cleavage at pairs of basic residues have been described since that initial discovery (reviewed in Docherty and Steiner, 1982; Thomas, Thorne and Hurby, 1988). The enzymes that catalyze the proteolytic processing of secreted proteins are subtilisin-type serine endoproteases related to the yeast KEX2 gene product.

kex2: The *Saccharomyces cerevisiae* KEX2 gene product, kex2, was the first eukaryotic pro-protein processing enzyme to be isolated, and remains the best characterized. As such, the biochemical properties of kex2 form a paradigm for all eukaryotic serine endoprotease processing enzymes. KEX2 was first identified in screens for yeast cells unable to secrete the K1 killer toxin, an M1 dsRNA virus encoded toxin that kills cells lacking the killer plasmid (Liebowitz and Wickner, 1976; Wickner and Liebowitz, 1976). The name KEX derives from killer expression. *kex2* mutants were also found to be defective in the ability to secrete the mature alpha-factor (Liebowitz and Wickner, 1976), a mating pheromone that acts on **a** mating-type cells causing them to arrest at the G1 phase of the cell division cycle (reviewed in Fuller, Sterne, and Thorner, 1988). Both the K1 killer toxin and the alpha-factor are

synthesized as inactive precursor proteins, and are activated by series of processing steps beginning with an endoproteolytic cleavage at pairs of basic residues (reviewed in Fuller, Sterne, and Thorner, 1988). In *kex2* mutants, this cleavage does not take place. Instead, the prepro-alpha-factor is secreted in a heavily glycosolated form. Cleavage of the alpha-factor is restored by replacing the defective *kex2* mutant with the normal gene on a plasmid (Julius *et al.*, 1984).

KEX2 was cloned by complementation of the mutant phenotypes with plasmid DNA containing genomic fragments (Julius *et al.*, 1984). The KEX2 gene product, kex2, is structurally related to the bacterial subtilisin-like serine proteases (Fuller, Brake and Thorner, 1986; Mizuno *et al.*, 1988). Studies assaying the protease activity of kex2 using short synthetic substrates have revealed that the enzyme is membrane bound and dependent on calcium, and cleaves its substrates on the carboxyl side of Arg-Arg and Lys-Arg pairs of basic residues (Julius, Sheckman and Thorner, 1984; Mizuno *et al.*, 1987; Fuller, Brake and Thorner, 1989a).

Considerable evidence demonstrates that kex2 is localized in the Golgi body. First, a short hydrophobic domain with the hallmarks of a signal peptide is found at the amino-terminal of kex2 (Fuller, Brake and Thorner, 1988). Second, the effect of conditional mutations that affect the movement of proteins through the Golgi body on pro-alpha-factor processing indicate that kex2 cleaves the pro-alpha-factor in a late compartment of the Golgi body (Novick and Scheckman, 1979; Julius, Sheckman and Thorner, 1984; Franzuoff and Scheckman, 1989). Finally, kex2 has been localized by direct immunolocalization to the Golgi compartment (Franzuoff *et al.*, 1991; Redding, Holcomb and Fuller, 1991).

Retention of kex2 in the Golgi body requires a hydrophobic domain, thought to be a transmembrane domain (Fuller, Sterne and Thorner, 1988), near the carboxyl-terminal of the protein. Deletion of the hydrophobic domain results in the secretion of a soluble form of the protein, confirming that the hydrophobic domain is required for retention of kex2 in the Golgi body (Fuller, Brake and Thorner, 1989b; Brenner and Fuller, 1992). Retention of kex2 in the Golgi body also requires clathrin, suggesting that kex2 may be cycled through secretory vesicles or the cell surface (Payne and Scheckman, 1989).

kex2 can function in mammalian cells. The properties of kex2 resemble closely those predicted for mammalian prohormone processing enzymes. To determine if kex2 could function as a prohormone processing enzyme, Thomas *et al.* (1988) tested the ability of kex2 to correctly process the mammalian prohormone pro-opiomelanocortin (POMC). When POMC and kex2 were coexpressed in cells lacking an endogenous processing activity, POMC was processed to produce mature gamma-lipotropin (gamma-LPH), beta-endorphin (beta-End), and beta-lipotropin (beta-LPH). Similar experiments demonstrated that kex2 can convert pro-beta-nerve growth factor (pro-beta-NGF) to mature beta-NGF (Bresnahan *et al.*, 1990), and the single-chain zymogen form of human protein C to the mature two-chain active form (Foster *et al.*, 1991). These experiments demonstrate that the functional characteristics of the mammalian prohormone processing enzymes are sufficiently conserved that the yeast kex2 can replace their function in mammalian cell culture lines.

Furin: A human homologue of KEX2, the human fur gene (hfur) was first partially sequenced because of its proximity to the fes oncogene (fur stands for fes upstream region) (Roebroek *et al.*, 1986) and was subsequently

identified as a homologue of KEX2 by sequence similarity (Fuller, Brake and Thorner, 1989b; van den Ouweland *et al.*, 1990). The structure of the fur gene product, hfurin, is similar to that of kex2. Like kex2, hfurin includes an amino terminal signal peptide and a carboxyl-terminal transmembrane domain (Roebroek *et al.*, 1986; van den Ouweland *et al.*, 1990). hfurin was localized by immunoflorescence to the Golgi body (Bresnahan *et al.*, 1990). Hfurin has been shown in transfection studies to correctly process pro vonWillebrand factor (van de Ven *et al.*, 1990; Wise *et al.*, 1990) and pro-beta-NGF (Bresnahan *et al.*, 1990). The fur mRNA has been found in cell lines derived from a wide variety of tissues including non endocrine cell types (Bresnahan *et al.*, 1990). This broad spectrum of expression has led to the speculation that hfurin is a component of the constitutive (tissue general) secretory pathway. Genes with more than 90% sequence identity to human furin have been cloned from mouse (Hatsuzawa *et al.*, 1990) and rat (Misumi, Sohda and Ikehara, 1990).

PC1 and PC2: Recently, candidates for the mammalian prohormone converting (PC) enzymes regulated (tissue specific) secretory pathway have been cloned using the polymerase chain reaction. These include a partial clone, mPC1, and the complete mPC2, cloned from a mouse pituitary cell line (Seidah *et al.*, 1990). In addition, hPC2, a homologue of mPC2, was cloned from human insulinoma total RNA (Smeekens and Steiner, 1990). A complete mPC1 cDNA was subsequently cloned from the mouse AtT20 pituitary cell line and the sequence published as mPC3 (Smeekens *et al.*, 1991). These proteins are similar to kex2 and hfurin in that they have a secretion signal peptide and a subtilisin-type protease domain that has significant sequence identity with hfurin and kex2. However, in contrast to kex2 and hfurin, these proteins lack a carboxyl-terminal transmembrane

domain. It is not yet known if these proteins are retained in the Golgi body by some alternative mechanism, or if they are secreted.

PC1(PC3) and PC2 are transcribed in endocrine and neuroendocrine tissues only (Seidah *et al.*, 1990; Smeekens *et al.*, 1991). The tissue distribution of expression of these genes has lead to the suggestion that they may be involved in providing the tissue specificity of prohormone processing. This suggestion is supported by the finding that tissue specific processing of mPOMC can be reconstituted by coexpression of PC1(PC3) and PC2 either alone or in combination in tissue culture cells (Thomas *et al.*, 1991; Benjannet *et al.*, 1991).

bli-4. This study presents evidence that the *bli-4* locus of *C. elegans* is a kex2-like protease. With the exception of the yeast protein kex2, no mutations have been identified in genes encoding the kex2-like proteins. kex2 loss of function mutants result in the inability to process the alpha-mating factor and the K1 killer toxin. Neither of these functions is essential. In contrast, the most severe lethal phenotype of *bli-4* is developmental arrest as late embryos. This observation indicates an essential role for the *bli-4* gene product prior to or at the end of embryogenesis. This is the first identification of mutations in a KEX2-like gene in a multicellular organism, and the first direct evidence that a KEX2-like gene is essential to development.

CHAPTER TWO: MATERIALS AND METHODS

2.1 NOMENCLATURE

Nomenclature used conforms to the uniform system for *Caenorhabditis elegans* (Horvitz *et al.*, 1979). Nomenclature for translocations conforms to that used by McKim, Howell and Rose (1988). Genes are represented by a three letter name followed by a number indicating the order in which they were discovered. For example, *bli-4* is the fourth locus defined by blister mutations. Allele names are given as one or two lower case letters followed by a number. The letters are a laboratory designation, while the numbers indicate the particular mutation. For example, for the allele *h42*, the designation h indicates that the mutation was isolated in the Rose laboratory. The number 42 indicates that the mutation is the 42nd mutation isolated in the Rose laboratory. Strain names are given as one or two upper case letters followed by a number. As is the case for alleles, the letters are a laboratory designation, while the numbers indicate the particular strain. The Rose laboratory strain designation is KR. To indicate rearrangements, the laboratory designation letter is followed by a uppercase and lower case letters indicating the nature of the rearrangement; T indicates a translocation; Dp indicates a duplication; Df indicates a deficiency (deletion); P indicates a restriction fragment length difference (polymorphism). For example, *hDf8* is the eighth deficiency generated in the Rose laboratory. Phenotypes for a given mutation are presented as the gene name with no italics and starting with a capital letter. For example, *bli-4* is a gene name, while Bli-4 is a phenotype. Names that refer to DNA, such as gene names, allele names, translocation names, polymorphism names, and deletion names, are italicized. Names that refer to whole

animals, such as strain names and phenotypes, are not italicized. In this study, plasmids are pCeh followed by a number, where p stands for plasmid, Ce stands for *C. elegans*, and h stands for the Rose laboratory. For example, pCeh205 is Rose laboratory *C. elegans* DNA plasmid 205. A list of all strains and mutations with laboratory designations used in this study is presented in Appendix B.

2.2 NEMATODE CULTURE CONDITIONS

Worms were maintained and mated on 10 X 35 mm petri plates containing strain OP50 *E. coli* streaked on Nematode Growth Medium (NGM) (Brenner, 1974) at 20C except where noted otherwise. In experiments where counting or examining progeny was necessary, hermaphrodites were transferred to fresh plates every twelve to sixteen hours to prevent overcrowding. Nematode Growth Media is constituted as follows:

NGM medium:

300g NaCl	After autoclaving (20 minutes)
17g Agar	1 ml Cholesterol (5mg/ml in Ethanol)
5g Bactotryptone	1 ml 1M $CaCl_2$
Distilled water to 1 L	1 ml 1M $MgSO_4$
	25 ml 1M KH_2PO_4 (pH 6.0)

2.3 DETERMINATION OF PENETRANCE

Penetrance was determined by scoring adult hermaphrodites for blisters. In populations of worms resulting from hermaphrodite self-fertilization, penetrance was:

$$\text{Penetrance} = \frac{100\% \times [\#\ \text{Blistered worms}]}{[\text{total}\ \#\ \text{of worms}]}$$

In populations of worms resulting from cross-fertilization of *bli-4(e937)* homozygous hermaphrodites with *bli-4(e937)/ +* heterozygous males, half the progeny was expected to be heterozygous for *bli-4(e937)*. Because the males were *bli-4* heterozygotes, a maximum of 50% of the worms were expected to blister if penetrance was 100%. Therefore penetrance in the cross experiment was defined as the percentage of blistered worms out of one half the total number of progeny:

$$\text{Penetrance} = \frac{100\% \times 2[\#\ \text{Blistered worms}]}{[\text{total}\ \#\ \text{of worms}]}$$

2.4 DETERMINATION OF LARVAL STAGE

To follow larval moults, the method of Cassada and Russell (1975) was used. Synchronous populations of worms were established as follows. A population of hermaphrodites was permitted to lay eggs on a 10 cm petri-dish for a period of one to two days. All of the worms were then washed from the plate with M9 buffer, leaving unhatched eggs adhering to the agar surface. Newly hatched L1 larva were harvested at one-hour intervals and placed on a fresh plate. Each population of worms established this way was synchronous with respect to development to within one hour. *C. elegans* undergoes a period of reduced activity prior to each moult, termed the lethargus period. During this period, movement is reduced and pharyngeal pumping ceases. Lethargus periods were monitored at 25^{o} by plotting the percentage of worms that were pumping in synchronous populations with

respect to time. Moults were indicated by rapid and marked reductions in the percentage of worms exhibiting pumping.

2.5 COMPLEMENTATION TESTING

sDp2 lethal alleles. Lethal alleles rescued by *sDp2* were complementation tested *inter se* as described by Howell *et al.* (1987). Heterozygous males of the genotype *dpy-5 let-X unc-13/ + + +* were mated to hermaphrodites of the genotype *sDp2/ dpy-5 let-Y unc-13/ dpy-5 let-Y unc-13*. *sDp2* complements *dpy-5* but not *unc-13* (Figure 2). The absence of fertile Dpy-5 Unc-13 in the cross progeny indicated failure to complement.

sDp2 lethal alleles and *e937*. Complementation tests were done in both of the following ways: a) Heterozygous males of the genotype *dpy-5 let-X unc-13/ + + +* were mated to hermaphrodites of the genotype *bli-4(e937) unc-13/ bli-4(e937) unc-13*. The presence of Bli-4 Unc-13 males indicated failure to complement. b) Heterozygous males of the genotype *dpy-5 let-X unc-13/ + + +* were mated to hermaphrodites of the genotype *bli-4(e937)/ bli-4(e937)*. The presence of Bli-4 males in the cross progeny indicated failure to complement. In reciprocal crosses, both *bli-4 unc-13/ + +* males and *bli-4/ +* males were mated to *sDp2/ dpy-5 let-X unc-13/ dpy-5 let-X unc-13* hermaphrodites. The presence of Bli-4 Unc-13 or Bli-4 hermaphrodites and males indicated failure to complement.

let-77(s90) and *bli-4(h1010)*. *let-77(s90)* and *bli-4(h1010)* were not linked to *dpy-5*, making it necessary to use a different complementation testing protocol from that used for the *sDp2* balanced lethal alleles. *s90* and *h1010* were each balanced by the translocation *szT1(I;X)*, and complementation tests performed as follows: a) *dpy-5 let-X unc-13/ + + +*

males were crossed to *let-77 unc-13; +/ szTI(I;X)[lon-2]* or *unc-63 bli-4(h1010) unc-13; +/ szTI(I;X)[lon-2]* hermaphrodites. The absence of Unc-13 progeny indicated failure to complement. Successful mating was indicated by the presence of wild type males. b) Spontaneous Lon-2 males of the genotype *let-77 unc-13; 0/ szTI(I;X)[lon-2]* or *unc-63 bli-4(h1010) unc-13; 0/ szTI(I;X)[lon-2]* were crossed to *bli-4(e937) unc-13* hermaphrodites. The presence of non bli-4 Unc-13 male progeny and the absence of Bli-4 Unc-13 male progeny indicated complementation (Bli Unc hermaphrodites in this experiment could have resulted from self-fertilization). In reciprocal crosses, the presence of non bli-4 Unc-13 and the absence of Bli-4 Unc-13 hermaphrodite and male progeny indicated complementation.

hDf8. The deletion *hDf8* was isolated using formaldehyde in a screen for mutations that failed to compelement *dpy-14*. (McKim, Starr and Rose, 1992). *hDf8* is associated with a suppressor of chromosome I recombination, and does not carry any flanking markers, making complementation tests difficult. To test *hDf8* for complementation of the *bli-4(e937)* allele, the translocation *szT1* was used (Fodor and Deak, 1985). Males of the genotype *hDf8 (I); O/ szTI(I;X)[lon-2]* were mated to hermaphrodites of the genotype *dpy-5 bli-4(e937)*. In this cross, the only males produced have the genotype *dpy-5 bli-4(e937)/ hDf8*. If *hDf8* deleted *bli-4*, the non Lon-2 males would be blistered; if *hDf8* did not delete *bli-4*, the male progeny would be wild type. Only wild type male progeny were observed, indicating that *hDf8* does not delete *bli-4*.

2.6 DETERMINATION OF DEVELOPMENTAL ARREST STAGE

The stage at which lethal homozygotes arrested development was determined. Several heterozygous hermaphrodites of the genotype

dpy-5 let-X unc-13/ + + + were permitted to lay eggs on an NGM plate for a short period (not more than two hours) and then the homozygous lethal progeny were examined by Normarski differential interference microscopy for terminal phenotype.

2.7 ALLELE MAPPING

h42, h199, h254 and *e937* were three-factor mapped by scoring segregation from strains bearing *cis*-linked flanking markers *dpy-5(e61)* and *unc-13(e450)* in *trans* to an unmarked chromosome. In the case of *e937*, recombinants were picked as Dpy-5 or Unc-13 worms and their progeny screened for the presence of the Bli-4 phenotype. Because the blistered phenotype is not expressed in Dpy-5 worms, the Dpy recombinants were tested for the presence of *e937* by complementation testing. *s90*, which was not induced on a *dpy-5* chromosome, was two-factor mapped with respect to *unc-13(e51)*, consistent with a position between *dpy-5 unc-13*.
Recombination was determined under the conditions recommended by Rose and Baillie, 1979a. Hermaphrodites were kept at 20C and were transferred to fresh plates every 12 to 16 hours to prevent overcrowding.
Recombination frequency was calculated using the mapping function $p=1-(1-2R)^{1/2}$ where R is the fraction of recombinant progeny over total progeny, and total progeny is calculated as 4/3(the number of wild type plus one recombinant class) (Brenner, 1974).
The alleles *h384, h427, h520, h699, h754* and *h791* were mapped to an interval around *bli-4* defined by *hDp16* and *hDp19* (Figure 2) by McDowall (1990).

2.8 SCREEN FOR *bli-4* ALLELES IN A MUTATOR STRAIN

mut-6 causes high levels of transposition of the transposable genetic element Tc1 (Mori, Moerman and Waterston, 1988). A mutator strain (KR1822) of the genotype *unc-63(e384) unc-13(e450); mut-6(st702)* was constructed (*mut-6* was from RW7097, a strain obtained from D. G. Moerman and R. H. Waterston). Mutator activity in KR1822 was confirmed by screening in 1% nicotine for twitcher worms resulting from the insertion of Tc1 into the *unc-22* gene as described by Mori, Moerman and Waterston (1988). KR1822 segregates spontaneous twitchers at a rate of 3×10^{-4}. KR1822 was screened for spontaneous *bli-4* alleles by mating KR1822 hermaphrodites to *dpy-5(e61) bli-4(e937)/ + +* heterozygous males, and screening the progeny for blisters. Three Bli worms, two hermaphrodites and one male, were identified after screening 82,300 chromosomes, an induction frequency of 3.6×10^{-5}. Of the three spontaneous blistered animals recovered, one survived. The surviving hermaphrodite carried a *bli-4* lethal allele designated *h1010* and was maintained using the translocation *szT1(I;X)* in the strain KR1858.

2.9 ISOLATION OF INTACT CUTICLE (Cox, Kusch and Edgar, 1981)

Intact cutcles were isolated by sonicating worms for three minutes in 10 mM Tris (pH 7.4), 1mM EDTA, 1mM phenylmethanesulfonyl floride and incubating at 100C for 2 minutes in 1% SDS, 0.125 M Tris (pH 6.8).

2.10 COSMIDS AND PLASMIDS GROWTH CONDITIONS

Plasmids were constructed using the Stratagene Bluescript SK vector, which carries an ampicillin resistance gene and a multiple enzyme recognition site polylinker for subcloning. Cosmids used in this study were obtained from A. Coulson and J. Sulston at the MRC, Cambridge, England.

Cosmid names beginning with F, T, or K used kanamycin-resistant lorist 2 vectors (Cross and Little, 1986; Gibson *et al.*, 1987). Cosmid names beginning with C, B, or Z used ampicillin resistant pJB8 vectors (Ish-Horowicz and Burke, 1981).

2.11 DNA PREPARATION

2.11.1 COSMIDS AND PLASMIDS

Mini-preps (Maniatis, Fritsch and Sambrook, 1982). 5 mL of L-broth containing either 50 ug/ul ampicillin or 25 ug/ul kanamycin was inoculated with a single bacterial colony picked from an L-broth plate. The culture was incubated in a shaking incubator at 300 rpm at 37C for 16 to 18 hours. Cells were transferred to a 1.5 ml microfuge tube and harvested by pelleting for 20 to 30 seconds in an Eppendorf 5415 microcentrifuge. The pellet was resuspended in 300 uL of solution I and incubated for five minutes at room temperature. 400 uL of solution II was then added, mixed by vigorous inversion, and incubated on ice for five minutes. To this mixture 300 uL of ice-cold solution III was added, mixed by vigorous inversion, and incubated on ice for five minutes. Cellular debris were removed by pelleting at 12,000 rpm in an Eppendorf 5415 microfuge for five minutes. The aqueous layer was transferred to a fresh microfuge tube and extracted once or twice with 850 uL water-saturated phenol and once with Sevag's reagent (24:1 Chloroform:isoamyl alcohol). The DNA was precipitated by adding 850 uL isopropanol and pelleted at 12,000 rpm for 10 minutes. The DNA pellet was washed with 70% ethanol, dried in a vacuum chamber, and resuspended in 100 uL 1X TE containing 20 ug RNase A (BRL). Samples were stored at -20C.

Solution I:

50 mM sucrose

10 mM EDTA

25 mM Tris (pH 8.0)

5 mg/mL lysozyme (Boehringer-Mannheim) freshly added before each use

Solution II

0.2 M NaOH

1% SDS

Solution III

3 M potassium acetate

11.6% glacial acetic acid

TE

10 mM Tris base (pH 7.6)

1 mM EDTA

Sevag's reagent 24:1 Chloroform:isoamyl alcohol.

L-broth

5 g NaCl

5 g Difco Yeast Extract

10 g Bacto-tryptone

Distilled water to 1 L

L-broth Plates

5 g NaCl

5 g Difco Yeast Extract

10 g Bacto-tryptone

Distilled water to 1 L

Large scale plasmid and cosmid DNA isolation (Davis, Botstein and Roth, 1980). 10 mL of L-broth containing either 50 ug/ul ampicillin or 25 ug/ul kanamycin was inoculated with a single colony picked from an L-broth plate. The culture was incubated in a shaking incubator at 300 rpm at 37C for 16 to 18 hours, then transferred to 500 mL of fresh L-broth containing 50 ug/mL ampicillin or 25 ug/ml kanamycin and incubated overnight at 37C with shaking. Cells were harvested in 250 mL Beckman tubes by centrifugation at 5,000 rpm for 5 minutes in a Beckman J21 Centrifuge using a JA-14 rotor.

The supernatant was discarded and the pellet resuspended in 5 mL of lysis buffer. The cell suspension was transferred to a 45 mL Beckman tube and incubated at room temperature for five minutes. 20 mL of solution II was added and mixed by vigorous inversion, followed by incubation on ice 5 minutes. 10 mL of ice-cold solution III was then added, mixed by vigorous inversion and incubated on ice for a further 5 to 10 minutes. The cellular debris were removed by centrifugation at 16,000 rpm for 20 minutes at 4C using a JA-20 rotor. 36 mL of the supernatant was transferred to a 50 ml falcon tube. 24 mL of isopropanol was added an mixed by gentle inversion and the mixture stored at -20C for 30 minutes. The DNA was collected by centrifugation at 2,000 rpm for 10 minutes using an IEC HN-SII Desk top Centrifuge. The precipitate was rinsed twice with 70% ethanol and dissolved in 5 mL of 1X TE; 7.5 mL CsCl saturated solution; 0.65 mL ethidium bromide (10 mg/mL). The mixture was transferred to a 12 mL Beckman heat sealable tube, and centrifuged at 60,000 rpm for 18 hours at 15C in a Beckman ultracentrifuge. The DNA band was removed from the tube using a needle and transferred to a disposable plastic culture tube. Ethidium bromide was removed by repeated extraction with water-saturated butanol until the aqueous layer was colourless. DNA was precipitated from the aqueous layer by adding two volumes of water and six volumes of 95% ethanol, followed by incubation at -20C for one hour. The DNA was collected by centrifugation at 2,000 rpm for 10 minutes, and the pellet washed with 70% ethanol and air-dried. The DNA was resuspended in 1 mL of 1X TE, and stored at 4C.

Lysis Buffer	Solution I
25 mM Tris (pH 8.0)	3 M potassium acetate
10 mM EDTA (pH 8.0)	11.6 % glacial acetic acid
50 mM sucrose	

5 mg/mL lysozyme (Boehringer-Mannheim) added just prior to use.

2.11.2 *C. ELEGANS* GENOMIC DNA

C. elegans genomic DNA was prepared using the method of Emmons, Klass and Hirsh (1979) as modified by J. Curran and D.L. Baillie (personal communication). Worm cultures were grown at 20C on a lawn of wild type *E. coli* in 100 mm petri plates containing NGM (Section 2.2) made with 0.7% agarose instead of agar. Cultures were incubated for five to ten days until the plates were crowded with adult worms, but not starved. Worms were harvested from the plates by washing with 0.5% NaCl and transferred to 50 mL Falcon tubes. The worms were collected by centrifugation at 2,000 rpm for 10 minutes and the supernatant discarded. To eliminate bacteria, worms were rinsed one or more times with 10 mL of 0.5% NaCl until the supernatant was clear, and recentrifuged. Worms were resuspended in 5 mL of proteinase K buffer and incubated with 5 mg proteinase K at 65C for 10 minutes or until the worms were dissolved. A further 10 mL of proteinase K buffer was added to increase the volume to 15 ml. Proteins and cellular debris were removed by extraction of the aqueous solution with water-saturated phenol. The aqueous and organic phases were separated by centrifugation at 2,000 rpm for 10 minutes. The extraction was repeated until the interface was clear of cellular debris. The aqueous layer was then extracted twice with 15 ml of Sevag's reagent and collected by centrifugation at 2,000 rpm for 5 minutes. The solution was transferred to a fresh 50 ml

falcon tube, and the DNA was precipitated by adding 1/10 volume of 4M NH_4OAc and one volume of isopropanol and mixed by gentle inversion. Precipitated DNA was harvested from the solution with a sealed 1 mL Pasture pipette and washed with 70% ethanol. DNA was resuspended in 1 mL of TE containing 20 ug RNase A by gentle inversion overnight at room temperature and stored at 4C.

Proteinase K buffer

0.1 M Tris (pH 8.0)

0.05 M EDTA (pH 8.0)

0.2 M NaCl

1% SDS

2.12 ESTIMATION OF DNA CONCENTRATION

The concentration of DNA in solution was determined by measuring the absorbance of light at 260 nM (A_{260}) of 6 ul of the DNA solution diluted to 600 ul with distilled water. Absorption readings were taken in 1 ml cuvettes using a Perkin-Elmer Lambda-3 UV spectrophotomer, and the results read as optical density units (OD). For double stranded DNA, 1 OD = 50 ug/ml. For single stranded oligonucleotides, 1 OD = 20 ug/ml.

2.13 RESTRICTION DIGESTS

Restriction endonucleases were obtained from Pharmacia, Bio-Rad, and Boehringer-Mannhiem. Digestion of DNA with restriction enzymes was carried out under the conditions recommended by the manufacturer.

2.14 GEL ELECTROPHORESIS

Agarose gels were prepared by melting 0.7% agarose in TBE buffer. 1 ug/ml Ethidium Bromide was added to permit visualization of DNA. Gels were cast in horizontal trays of either 15 cm X 20 cm (LKB 20212 Maxiphor Electrophoresis unit) or 6 cm X 10 cm (Bio-Rad Mini-sub DNA Cell). DNA samples were mixed with 1/10th volume of loading buffer containing 30% glycerol and 0.25% bromophenol blue in water. Electrophoresis was conducted with the gel submerged in TBE buffer at 30 volts for 16 to 18 hours for large gels or 95 volts for 1 to 2 hours for small gels. DNA bands were visualized by illumination with a 300 nm UV transilluminator.

TBE Buffer

0.89 M Tris

0.89 M boric acid

1 mM EDTA (pH 8.0)

2.15 ELECTROELUTION (Maniatis, Fritsch and Sambrook, 1982)

Electroelution was used to recover DNA fragments from agarose gels following electrophoresis. The DNA band to be recovered was visualized using the 300 nm UV transilluminator. The band was excised from the gel using a razor blade and placed into 10 mm diameter Spectra/por standard dialysis tube cut to a length of about 5 cm. The tube was sealed at one end with a plastic clip and filled with approximately 400 ul of 0.5 X TBE (Section 2.14) All air bubbles were removed and the tube was sealed at the other end. The dialysis tube containing the gel slice was then placed in a minigel apparatus submerged in 1 X TBE and 90 volts was applied for about one

hour to elute the DNA from the gel slice. The eluted DNA was recovered from the dialysis bag using a micropipette.

Preparation of dialysis tubing: Prior to use, dialysis tubing was boiled for 10 minutes in 2% sodium bicarbonate solution, rinsed thoroughly in distilled water, and then placed in distilled water and autoclaved for 10 minutes. Prepared tubing was stored in water at 4C.

2.16 SUBCLONING

The vector Bluescript by Stratagene was used for all subcloning. The vector was prepared by restriction enzyme digestion followed by either heat inactivation of the enzyme (incubation at 65C for 15 minutes) or extraction with TE-saturated phenol followed by extraction with Sevag's reagent. Fragments to be subcloned (insert fragment) were digested with the appropriate restriction enzyme and purified by agarose gel electrophoresis through Bio-Rad low melting temperature agarose. Insert fragments were recovered from the gel by electroelution (Section 2.14). 1 ug of insert fragment was mixed with 100 ng of restriction enzyme digested vector and precipitated by adding 1/10 volume of 8 M NH_4OAc and 2 volumes of ethanol. The mixed vector and insert fragments were then resuspended in 40 ul BRL ligation mix and 1 unit of BRL ligase. The ligation mixture was incubated overnight at 16C. 10 ul of the ligation mixture was used to transform BRL competent DH5-alpha *E. coli* cells. The ligation mixture was added to 100 ul of competent cells, incubated on ice for 30 minutes and then heated to 42C for two minutes. The transformed cells were plated on L-broth plates containing 50 ug/ml ampicillin, 40 ug/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactoside (Xgal) in dimethyl formamide and 160 ug/ml Isopropopyl-beta-D-thiogalactoside (IPTG), and incubated overnight at 37C.

Colonies containing inserts were white; colonies without inserts were blue. Single white colonies were picked and screened for the correct insert by restriction analysis of miniprep DNA. Note: Bluescript does not express beta-galactosidase as well as some other cloning vectors, making blue colonies somewhat difficult to see. The blue colour is enhanced by further incubation at 4C for several hours.

1 X BRL ligation mix

50 mM Tris

10 mM $MgCl_2$

0.1 mg/ml BSA

1 mM ATP

10 mM dithiothreitol

5% polyethylene glycol

2.17 SOUTHERN TRANSFERS (Southern, 1975)

Agarose gels containing DNA fragments to be transferred to a membrane were prepared as described (Section 2.14). 100 ng to 500 ng of restriction digested plasmid or cosmid DNA, or three to five ug of restriction digested *C. elegans* DNA was loaded in each lane. The gel was treated by soaking in 0.25 M HCl at room temperature for 10 minutes, 1.5 M NaCl; 0.5 M NaOH for 30 minutes, and 4 M NH_4OAc for 15 minutes. The gel was rinsed in distilled water between each treatment. The gel was then placed on a Whatman 3MM filter paper wick soaked in 10 X SSC and supported by a glass plate. A Schleicher and Scheull Nytran membrane cut to the size of the gel was soaked in 4 M NH_4OAc laid onto the gel. All air bubbles between the gel and the Nytran were removed. Two pieces of Whatman

3MM filter paper cut to the size of the gel were soaked in 4 M NH_4OAc and laid onto the membrane, and a 6 to 10 cm stack of paper towels were laid onto the filter paper. The DNA was transferred to the membrane by capillary action for about 2 hours. To prevent the 10 X SSC from bypassing the gel, the wick was prevented from touching the paper towels by placing strips of plastic around the gel. Following the transfer, the DNA was fixed to the membrane by baking at 80 to 90C for 1 to 2 hours.

10 X SSC

350 g NaCl

176 g sodium citrate

Distilled water to 4 L

2.18 NORTHERN TRANSFERS

Total RNA used was a gift from S. Prasad and D.L. Baillie, Simon Fraser University, Burnaby, BC. 15 ug of *C. elegans* RNA was denatured in 2.2 M formaldehyde; 50% deionized formamide; 1X MOPS for 15 minutes at 60C. The RNA fractionated by electrophoresis through a 1.1% agarose gel (15 cm X 20 cm) containing 1 X MOPS; 2.2 M formaldehyde (pH 7.0) at 80 volts for two to three hours. 1 X MOPS buffer was used as an electrophoresis buffer. After electrophoresis the gel was soaked twice for 15 minutes in distilled water. The RNA was then transferred to a GeneScreen membrane following the GeneScreen protocol. Two pieces of Whatman 3MM paper were wetted with phosphate buffer and placed over an elevated glass plate so that its ends formed wicks in phosphate buffer. The gel was placed on the filter paper and plastic strips placed on each side of the gel. A GeneScreen membrane was cut to the exact size of the gel and soaked for 20 minutes in phosphate buffer. The membrane was then placed gently onto the gel and

all air bubbles removed. Five pieces of Whatman 3MM filter paper cut to the size of the gel were placed on top of the membrane, and covered with a 10 cm stack of paper towels. After 12 hours, membrane was removed and baked at 90C for two hours to fix the RNA to the membrane.

1X MOPS solution

0.2 M morpholinopropanesulfonic acid (pH 7.0)

50 mM sodium acetate

1 mM EDTA (pH 8.0)

Phosphate Buffer

0.025M Na_2HPO_4/NaH_2PO_4

mixed to pH 6.5

2.19 PREPARATION OF HYBRIDIZATION PROBES

All hybridization probes were separated from their cloning vectors. This was accomplished by restriction enzyme digestion followed by purification by agarose electrophoresis and electroelution to recover the probe from the gel. Probes were made radioactive (labelled) with ^{32}P using the oligo labelling technique (Feinberg and Vogelstein, 1984). 30 ul of water containing 1 ng/ul DNA was boiled in a sealed microfuge tube for 10 minutes, then placed immediately on ice. To the DNA was added 10 ul of OLB-A, 5 ul of 100 mg/ml BSA, one unit of DNA Klenow (polymerase II large subunit) and 5 ul ^{32}P-ATP (3,000 Ci/mmol; 10.0 mCi/mL; New England Nuclear). The labelling mixture was incubated at 16C overnight. The mixture was then diluted to 100 ul and passed through a spun column prepared from Pharmacia G-25 fine Sephadex to remove unincorporated

^{32}P-ATP. The probe was boiled in a sealed microfuge tube for 10 minutes and then immediately placed on ice for five minutes prior to use.

OLB-A

solution A:B:C (1:2.5:1.5)

18 uL 2-mercaptoethanol

5 uL 0.1 M dTTP (Pharmacia)

5 uL 0.1 M dCTP (Pharmacia)

5 uL 0.1 M dGTP (Pharmacia)

Solution A

1.0 mL 1.25 M Tris (pH 8.0)

0.125 M $MgCl_2$

Solution B

2 M Hepes (pH 6.6)

solution C

Random hexanucleotides (Pharmacia) suspended in 1X TE at 90 OD/mL.

2.20 HYBRIDIZATION

2.20.1 Southern blots. Southern blot filters were placed in heat sealable bags with 12 to 15 ml of hybridization solution (5 X SSPE; 0.3% SDS) and the hybridization probe (Section 2.19). Air bubbles were removed and the bag sealed. Hybridization was at 68C for 18 hours with constant agitation. The filter washed in 200 ml of 0.2 X SSC; 0.2 % SDS twice at minutes at room temperature for five minutes per wash, and twice at 65C for 30 minutes per wash. The wash solution was changed after each wash. The filter was then air-dried prior to autoradiography.

20 X SSPE

174 g NaCl

27.6 g Na_2HPO_4

7.4 g EDTA

Distilled water to 1 L

2.20.2 Northern blots. Northern blots were hybridized using the GeneScreen protocol. A radioactive probe was prepared as described (Section 2.19). The hybridization membrane was prehybridized by incubation with hybridization buffer in a heat sealed bag for six hours at 42C with constant agitation. The probe was denatured by boiling in 1.5 ml of the hybridization buffer and then added to the hybridization bag. The bag was resealed, and incubated overnight at 42C with constant agitation. The hybridization buffer was then removed, and the membrane washed twice for 5 minutes in 2 X SSC (Section 2.17) at room temperature, twice in 2 X SSC and 1% SDS for 30 minutes at 65C, and twice in 0.1 X SSC for 30 minutes at room temperature. The membrane was then air-dried and autoradiographed as described (Section 2.21).

Hybridization Buffer	P Buffer
Made fresh just prior to use	1% BSA
5 ml deionized formamide	1% ficoll (M.W. 400,000)
2 ml P buffer	250 mM Tris-HCL (pH 7.5)
2 ml 50% dextran sulfate	0.5% sodium pyrophosphate
0.58 g NaCl*	5% SDS
1 ml 1 mg/ml denatured salmon sperm DNA	

*The solution is heated to 42C 10 minutes prior to the addition of NaCl.

2.21 AUTORADIOGRAPHY

Probes bound to hybridization membranes were visualized using Kodak XAR5 or XRP X-ray film in cassettes using Dupont Cronex Lightning Plus or Dupont Par Speed enhancement screens. Autoradiographs of Southern blots of *C. elegans* genomic DNA and northern blots were exposed for one to two days at -70C. Autoradiographs of Southern blots of plasmid and cosmid DNA were exposed for 6 to 18 hours at room temperature.

2.22 CONSTRUCTION OF CB937 LAMBDA-ZAP LIBRARY

A library of *Eco*RI digested CB937 genomic DNA in the Stratagene lambda-zap vector using a kit and protocol supplied by Stratagene. The vector was prepared as follows. First, the lamda cohesive ends (COS sites) were ligated. 10 ug lambda-Zap DNA was ligated in a 20 ul ligation mix overnight at 16C. The ligase reaction was stopped by incubating the mixture for 15 minutes at 68C. Next, the concatamerized vector was digested with *Eco*RI. 2.4 ul of EcoRI digest buffer and 2.0 ul of *Eco*RI (20 units) was added and the mixture incubated at 37C for one hour. The vector was then dephosphorylated with calf intestinal phosphatase (CIP). 5 ul CIP buffer and 1 ul CIP (0.00625 units) was added and the volume brought to 50 ul with distilled water. The dephosphorylation mixture was incubated for 30 minutes at 37C, and a second ul of CIP was added followed by a further 30 minute incubation. After the CIP treatment, 40 ul of water, 10 ul 10 X STE and 5 ul 10 % SDS was added. 105 ul phenol/Sevag's reagent was added, mixed by inversion, and separated by centrifugation. 90 ul of the aqueous phase was removed and the organic phase back extracted with 60 ul of TE.

The resulting aqueous phases were pooled bringing the total aqueous volume to 150 ul. The aqueous solution was then re-extracted with 150 ul of phenol/Sevag's reagent and 135 ul of the aqueous phase removed. The organic phase was back extracted with 65 ul of phenol Sevag's reagent and the two aqueous phases combined to bring the volume to 200 ul. The aqueous phase was then extracted with 200 ul of Sevag's reagent and 185 ul of the aqueous phase removed. The vector was then precipitated by the addition of 18.5 ul of 8 M NH_4OAc and 400 ul 95% ethanol, and harvested by centrifugation for 30 minutes at 4C. The vector DNA was resuspended in 5 ul 2 X ligation buffer. 1 ug of the insert DNA (KR1858 genomic DNA) was prepared by digestion with *Eco*RI in a volume of 20 ul followed by heat inactivation of the *Eco*RI. 2.5 ul of the insert mixture was mixed with 2.5 ul of the vector to bring the volume to 5 ul in 1 X ligation buffer. The ligation mixture was then incubated overnight at 16C. The library was then packaged using a Gigapack Gold lambda packaging extract. The library was amplified and titered prior to use.

Ligation Mix

2 ul BSA (1 mg/ml)

2 ul 10 mM ATP

2 ul 100 mM dithiothreitol

2 ul 50% polyethylene glycol

2 ul ligase buffer

1 ul T4 DNA ligase

Distilled water to 20 ul.

10 X Ligase Buffer

500 mM Tris-HCl (pH 7.5)

70 mM $MgCl_2$

10 X CIP Buffer

500 mM Tris-HCl (pH 8.5)

1 mM EDTA

10 X STE	Phenol/Sevag's reagent
1.0 M NaCl	phenol:Sevag's reagent
100 mM Tris (pH 8.0)	1:1
10 mM EDTA	

2.23 cDNA ISOLATION

cDNA clones were isolated from a *C. elegans* cDNA library constructed by Barstead and Waterston (1989) in the Stratagene lambda-zap vector. 40,000 phage from the cDNA library were screened using the *h1010* probe (1.3 *Eco*RI fragment of pCeh181) as described (section 2.24).

2.24 LAMBDA-ZAP PHAGE LIBRARY SCREENING

Host bacteria *E. coli* strain BB4 was grown in a 4 ml culture in TB broth at 37C overnight in a shaking incubator. The host bacteria were harvested by centrifugation at 2000 rpm for 10 minutes and prepared for use by resuspension in 2 ml 10 mM $MgCl_2$. For each phage plate to be prepared, 5 X 10^4 phage diluted in SM buffer were incubated with 200 ul of BB4 cells for 15 minutes at 37C. The phage were then mixed with 3 ml of molten top agarose at 50C and poured onto 100 mm petri plates containing L-broth agar prewarmed to 37C. The phage plates were incubated at 37C of 16 hours. The plates were then cooled to 4C for two hours. The phage were transferred to 85 mm S&S NC nitrocellulose filters by laying the dry filters onto the bacterial lawn containing the phage plaques. The filters were then removed and soaked in 0.5 M NaOH; 1.5 M NaCl for 5 minutes, then 1 M NH_4Ac for 5 minutes. The treated filters were air-dried and then baked for one hour at 80C. Filters were hybridized and washed as described (Section 2.20.1), and positive plaques identified by autoradiography (Section

2.21). Each hybridizing (positive) plaque was isolated from the phage plate by picking with a sterile toothpick and placed in SM buffer. The screen was then repeated twice more for each positive phage as described above, except that only 200 phage were plated on each plate. After three screens, all of the phage in a given stock were positive. One positive phage from each stock was then used to isolate the phage insert (section 2.25)

Top Agarose	SM Buffer
5 g NaCl	0.1 M NaCl
5 g yeast extract	0.01 M $MgCl_2$
10 g bacto-tyrptone	0.05 M Tris (pH 7.5)
7.5 g Agarose	0.01% gelatin
Distilled water to 1 L	

2.25 ISOLATION OF PLASMIDS FROM LAMDA-ZAP PHAGE

Plasmids were isolated from lambda-zap phage stocks using the StrataGene protocol. *E. coli* strain BB4 cells were prepared as described (Section 2.24). 1 X 10^5 phage containing the desired insert were mixed with 1 X 10^7 R408 helper phage and added to 200 ul of BB4 cells. The mixture was incubated at 37C for 15 minutes and then 5 ml of 2 X YT was added followed by a further incubation of four hours at 37C in a shaking incubator. The cells were then killed by incubation at 70C for 20 minutes and removed from the solution by centrifugation at 2000 rpm for five minutes. The phage insert fragment was then contained in the supernatant in the Bluescript vector packaged in F1 phage particles. To recover the plasmid from the phage particles, 200 ul of the supernatant solution was mixed with 200 ul of BB4 cells and incubated at 37C for 15 minutes. 10 ul of cells were then

plated on L-broth plates containing 50 ug/ml ampicillin and incubated at 37C overnight. The resulting bacterial colonies contained the insert in the Bluescript vector.

2 X YT

5 g NaCl

5 g yeast extract

8 g bacto-tryptone

2.26 DELETIONS

Nested sets of plasmid insert deletions were constructed using the method of Henikoff (1984). First, the plasmid was digested with two restriction enzymes, *Bst*XI and *Xba*I. The five base 3' overhang generated by *Bst*XI digestion is not a substrate for Exonuclease III (Exo III), and therefore protects the vector from deletion. The four base 5' overhang generated by *Xba*I digestion is a substrate for ExoIII digestion, and therefore permits Exo III to delete the insert. Both enzymes digest in Boehringer-Mannheim H buffer. However, *Bst*XI requires 45C while *Xba*I requires 37C. 5 ug of plasmid DNA was added to 5 ul of 10 X H buffer, 5 ul of 1 mg/ml BSA, and 5 units of *Bst*XI, brought to 50 ul with distilled water, and incubated at 45C for one hour. 5 units of *Xba*I was then added and the mixture incubated at 37C for one hour. The digestion mixture was then extracted with 50 ul of phenol/Sevag's reagent (1:1), the aqueous phase removed to a fresh microfuge tube and extracted with Sevag's reagent only. The aqueous phase was removed to a fresh microfuge tube and the DNA precipitated by the addition of 1/10th volume of 2 M NaCl and two volumes of 95% ethanol

followed by incubation at -20C for 20 minutes. The DNA was pelleted in an Eppendorf 1514 microfuge at 12,000 rpm for 10 minutes, washed with 70% ethanol and dried under vacuum. The DNA pellet was resuspended in 60 ul of Exo III buffer and placed in a 37C heating block. 300 units of Exo III was added. At 30 second intervals, 2.5 ul aliquots of the Exo III digest mixture were removed and placed in individual tubes containing 7.5 ul of S1 mix on ice. After all the aliquots had been removed, the samples were incubated at room temperature for 30 minutes. S1 stop buffer was then added and the tubes heated to 70C for 10 minutes to inactivate the S1 nuclease. The tubes were then transferred to a 37C heating block, 1 ul of Klenow mix was added to each tube and the samples incubated for three minutes. 1 ul of dNTP mix was added and the samples incubated for a further five minutes at 37C. The samples were removed from the heating block, 40 ul ligation mix was added and the tubes incubated at 16C overnight. Deletion samples pooled in groups representing four adjacent time points and used to transform BRL DH5-alpha cells as described (Section 2.16). DNA prepared from bacterial cultures derived from individual bacterial colonies was extracted by miniprep (Section 2.11.1) and screened for deletion size by restriction digestion and agarose gel electrophoresis (Section 2.14).

10 X Exo III Buffer

60 mM Tris (pH 8.0)

6.6 mM $MgCl_2$

7.4 X S1 Buffer

0.3 M potassium acetate (pH 4.6)

2.5 M NaCl

10 mM $ZnSO_4$

50% glycerol

S1 Mix

172 ul deionized water

27 ul 7.4 X S1 buffer

60 units S1 nuclease

S1 Stop Buffer

0.3 M Tris base

0.05 M EDTA

1 X Klenow Buffer

20 mM Tris-HCl (pH 8.0)

100 mM $MgCl_2$

Klenow Mix

30 ul 1 X Klenow buffer

5 units Klenow

10 X Ligase Buffer

500 mM Tris-HCl (pH 7.6)

100 mM $MgCl_2$

10 mM ATP

Ligase Mix

790 ul deionized water

100 ul 50% PEG

10 ul 100 mM dithiothreitol

5 units T4 DNA ligase

2.27 DNA SEQUENCING

All sequence reactions were performed using sequencing kits and protocols from Applied Biosystems Inc (ABI) and an ABI model 373A sequencing machine. The sequence methods used are based on the linear amplification technique of Craxton (1991) and the dideoxy-terminator method (Sanger and Coulson, 1975; Sanger, Nicklen and Coulson, 1977). ABI kits use florescent dyes covalently linked to either sequencing primers or dideoxynucleotide terminators. Reactions terminating in ddT, ddA, ddG and ddC were loaded into a single lane on a poly-acrylamide gel in the 373A sequencer, and the base order determined by fluorescence emitted from the dye as the band passed a scanning laser.

2.27.1 Template preparation. DNA templates were prepared using a QIAGEN miniprep kit. A bacterial culture was prepared as described (Section 2.11.1). 1.5 ml of the bacterial culture was harvested by centrifugation, resuspended in 300 ul of buffer P1, and incubated at room temperature for five minutes. 300 ul of buffer P2 was added and mixed by gentle inversion, and incubated at room temperature for a further five minutes. 300 ul of buffer P3 was then added, mixed by gentle inversion and the cellular debris removed by centrifugation in an Eppendorf 1514 microfuge for 15 minutes at 12,000 rpm at 4C. The supernatant was applied to a QIAGEN column, and the eluate precipitated with 0.7 volumes of isopropanol. The precipitate was washed with 70% ethanol, dried briefly under vacuum, and resupended in 20 ul TE (Section 2.11.1). The concentration of the DNA was estimated as described (Section 2.12).

QIAGEN columns are anion exchange columns that were used as follows. The column was washed with 1 ml buffer QBT. The sample was then applied and allowed to enter the resin by gravity flow. The bound sample was washed twice with buffer QC. Finally, the DNA was eluted with 0.8 ml of buffer QF.

Buffer P1	Buffer P2	Buffer P3
50 mM Tris-HCl	200 mM NaOH	2.55 M KOAC
10 mM EDTA	1% SDS	
100 ug/ml RNase A		
pH 8.0		

Buffer QBT	Buffer CF	Buffer QF
750 mM NaCl	1.0 M NaCl	1.5 M NaCl
50 mM MOPS	50 mM MOPS	50 mM MOPS
15% Ethanol	15% Ethanol	15% Ethanol
0.15% Triton X-100	pH 7.0	pH 8.2
pH 7.0		

2.27.2 Dye-labelled primer sequencing.

1 - 1.5 ug of double stranded template DNA in 6.0 ul of TE was used for each set of reactions. Four 0.5 ml microfuge tubes were labelled A, C, G and T. The fluorescent dyes used with the A and C mixes were more fluorescent than those used with the G and T mixes. To compensate, the G and T reactions were doubled. One ul of each of the following reagents was added to A and C tubes, and two ul to G and T tubes: d/ddNTP mix; dye primer (0.4 pM/ul); 5 X cycle sequence buffer; DNA template; diluted Taq polymerase. The reactions were overlaid with 20 ul of mineral oil to prevent evaporation, and the tubes placed in a Cetus-Perkin-Elmer thermocyler preheated to 95C. Cycling was as follows: 15 cycles with 95C for 30 seconds, 55C for 30 seconds, and 70C for 60 seconds. This was followed by 15 cycles with 95C for 30 seconds and 70C for one minute. The samples were then removed from the tubes. Mineral oil was removed by rolling the samples on parafilm. The samples were then pooled in a microfuge tube containing 100 ul of 95% ethanol and 2 ul of 3 M sodium acetate and incubated for 15 minutes at room temperature. The reaction products were pelleted by centrifugation for 30 minutes at 12,000 rpm in an eppendorf 1514 microfuge at 4C, washed with 70% ethanol and dried briefly under vacuum. One hour prior to loading, the samples were resuspended in 4 ul

of deionized formamide/ 50 mM EDTA 1:1 (v/v). Immediately before loading, the samples were heated to 90C for two minutes and quick-chilled on ice.

d/ddNTP Mixes

d/ddA Mix	d/ddC Mix
1.5 mM ddATP	0.75 mM ddCTP
62.5 uM dATP	250 uM dATP
250 uM dCTP	62.5 uM dCTP
375 uM c^7dGTP	375 uM c^7dGTP
250 uM dTTP	250 uM dTTP

d/ddG Mix	d/ddT Mix
0.125 mM ddGTP	1.25 mM ddTTP
250 uM dATP	250 uM dATP
250 uM dCTP	250 uM dCTP
94 uM c^7dGTP	375 uM c^7dGTP
250 uM dTTP	62.5 uM dTTP

Dye-labelled primers

Forward primer: 5′ GTAAAACGACGGCCAGT 3′

Reverse primer: 5′ AACAGCTATGACCATG 3′

5 X Cycle Sequencing Buffer	Diluted Taq Polymerase
400 mM Tris-HCl (pH 8.9)	0.5 ul AmpliTaq^R (8 units/ul)
100 mM $(NH_4)_2$ SO_4	1.0 ul 5 X Cycle sequencing buffer
25 mM $MgCl_2$	5.5 ul distilled Water

2.27.3 Dye-labelled terminator sequencing

One ug of double-stranded DNA template was mixed with 7.25 ul reaction premix and 3.2 pmol of the sequencing primer in a 0.5 ml microfuge tube and brought to a volume of 20 ul with distilled water. The reaction mixture was overlaid with 40 ul of mineral oil and placed in a Cetus-Perkin-Elmer thermocycler preheated to 90C. The incubation temperatures were cycled as follows: 96C for 30 seconds, 50C for 15 seconds, and 60C for 4 minutes for 25 cycles. The samples were then removed from the tubes. Mineral oil was removed by rolling the reaction mix on parafilm. The sample was then loaded onto a Select D-50 spun column (5 Prime -> 3 Prime Inc) and eluted by centrifugation at 1000 X g for 2 minutes (2400 rpm, IEC HN-SII Desk top Centrifuge). The eluted sample was then precipitated by the addition of 40 ul of 95% ethanol and 0.3 M sodium acetate followed by incubation for 15 minutes at room temperature. The reaction products were pelleted by centrifugation for 30 minutes at 12,000 rpm in an eppendorf 1514 microfuge at 4C, washed with 70% ethanol and dried briefly under vacuum. One hour prior to loading, the samples were resuspended in four ul of deionized formamide/ 50 mM EDTA 1:1 (v/v). Immediately before loading, the samples were heated to 90C for two minutes and quick-chilled on ice.

Reaction Premix (4 Reactions)

16 ul 5 X TACS Buffer

4 ul dNTP mix

2 ul DyeDeoxy^{TM} A Terminator

2 ul DyeDeoxy^{TM} C Terminator

2 ul DyeDeoxy^{TM} G Terminator

2 ul DyeDeoxy^{TM} T Terminator

1 ul AmpliTaq^{R} DNA Polymerase

5 X TACS Buffer

400 mM Tris-HCl

10 mM $MgCl_2$

100 mM $(NH_4)_2SO_4$

pH 9.0

dNTP Mix	Abbreviations
750 uM dITP	dNTP Deoxynucleoside triphosphate
150 uM dATP	ddNTP Dideoxynucleoside triphosphate
150 uM dTTP	c^7dGTP 7-deazaguanosine triphosphate
150 uM dGTP	ATP Adenosine triphosphate
	CTP Cytidine triphosphate
	TTP Thymidine triphosphate
	ITP Inosine triphospate

Custom Sequencing Primers

Numbers in brackets indicate the priming position in the blisterin B cDNA (Section 3.2.2.2). Sense indicates that the primer is derived from the sense strand sequence; antisense indicates that the primer is derived from the antisense strand.

KRp5 5' TGCTGGTTGACGGAAATC 3' (1349, antisense)

KRp6 5' CTACTCGGCTACTCCTGC 3' (2103, sense)

KRp7 5' TCCTTTTCCACCTCTGCC 3' (752, antisense)

KRp8 5' GAAGGCAACACCGACACC 3' (950, antisense)

KRp9 5' TCCACCAACTGCTCCACC 3' (692, antisense)

KRp10 5' ACTCTCTTCTTCGGTCGC 3' (499, antisense)

KRp11 5' GTGTCCTTGTTGTTTCCG 3' (414, antisense)

KRp13 5' TGGTGGAGCAGTTGGTGG 3' (708, sense)

2.28 POLYMERASE CHAIN REACTION (Saki *et al.*, 1988)

PCR was performed using Perkin-Elmer-Cetus thermocyler and reagent kit. Cycling was performed as follows: 92C 60 sec, 58C 1 60 sec, 72C 60 sec, and 25 cycles. The reagents were mixed as listed below, and overlayed with 40 ul of mineral oil. Because KR1858 was heterozygous for the amplification target DNA, the actual quantity of template was 50 ng. The primers used were KRp13 (Section 2.27) and P618, a Tc1 specific primer

(Williams *et al.*, 1992). P618 was a gift from D. Moerman, University of British Columbia, Vancouver, BC. The sequence of P618 is:

P618 5'- GAA CAC TGT GGT GAA GTT TC - 3' (Tc1 specific).

PCR Reaction mix (100 ul)

100 ng Template DNA

10 umoles P618

10 umoles KRp13

4 uM dATP

4 uM dCTP

4 uM dTTP

4 uM dGTP

1 X Reaction Buffer

1 unit AmpliTaq^R DNA Polymerase

10 X Reaction buffer

100 mM Tris HCl (pH 8.3)

500 mM KCl

15 mM $MgCl_2$

0.1% (W/V) Gelatin

2.29 SEQUENCE ANALYSIS

Sequences were assembled using the Delany Sequence program to identify overlaps and the Eye-ball Sequence Editor ESEE by Eric Cabbot to edit and assemble sequences. Kyte-Doolittle hydropathy analysis (Kyte and Doolittle, 1982) was done using the GREASE program (Pearson and Lipman, 1988). Potential glycosolation sites were identified using the prosite program of PCGene. Searches of the Swissprot protein database was done using the FASTA algorithm with a ktup value of 2 (Pearson and Lipman, 1988).

CHAPTER THREE: RESULTS

The results section of this thesis is divided into two major sections: the first section will report the results of a genetic analysis of *bli-4*; the second section will report results of a molecular analysis of *bli-4*.

3.1 GENETIC ANALYSIS OF THE *bli-4* LOCUS

This section begins with a characterization of the blistered phenotype resulting from the *e937* allele of *bli-4*. This characterization includes a determination of where blisters form, determination of penetrance, when blisters form, and how blisters are affected by other mutations that affect cuticle morphology. Next, the complementation patterns of *bli-4* alleles is examined, the arrest point of *bli-4* lethal alleles is determined, the map position of *bli-4* alleles is determined, and finally, the isolation of a mutator-induced allele is described.

3.1.1 Characterization of *e937*

3.1.1.1 Blisters form between cuticle layers

bli-4(*e937*) is a recessive mutation that results in fluid-filled blisters in the cuticle of adult-stage worms (Figure 3). The cuticle of larval-stage worms is not affected in a way that is observable under the dissecting microscope or by Normarski optics. No attempt at ultrastructural or biochemical characterization of larval-stage cuticles of *bli-4* homozygote worms was made in this study. Therefore, the possibility of subtle alterations in larval cuticle by *bli-4*(*e937*) is not excluded.

Under Normarski optics, blisters appeared to be a separation between the adult cuticle layers, rather than between the hypodermal tissue and the cuticle. The hypothesis that blisters result from a separation of cuticle layers predicts that blisters would remain intact in isolated cuticles. To test this

prediction the cuticle of blistered adult hermaphrodites was isolated as described in Section 2.9. Figure 4 shows the appearance of a blister in an adult hermaphrodite (A) compared to one in isolated cuticle (B). Visible in the isolated cuticle are some internal cuticle structures, including the pharynx and the gut, eggs, which are resistant to the treatment used to isolate the cuticle, and an intact blister. This experiment supports the conclusion that the blisters occur between the cuticle layers.

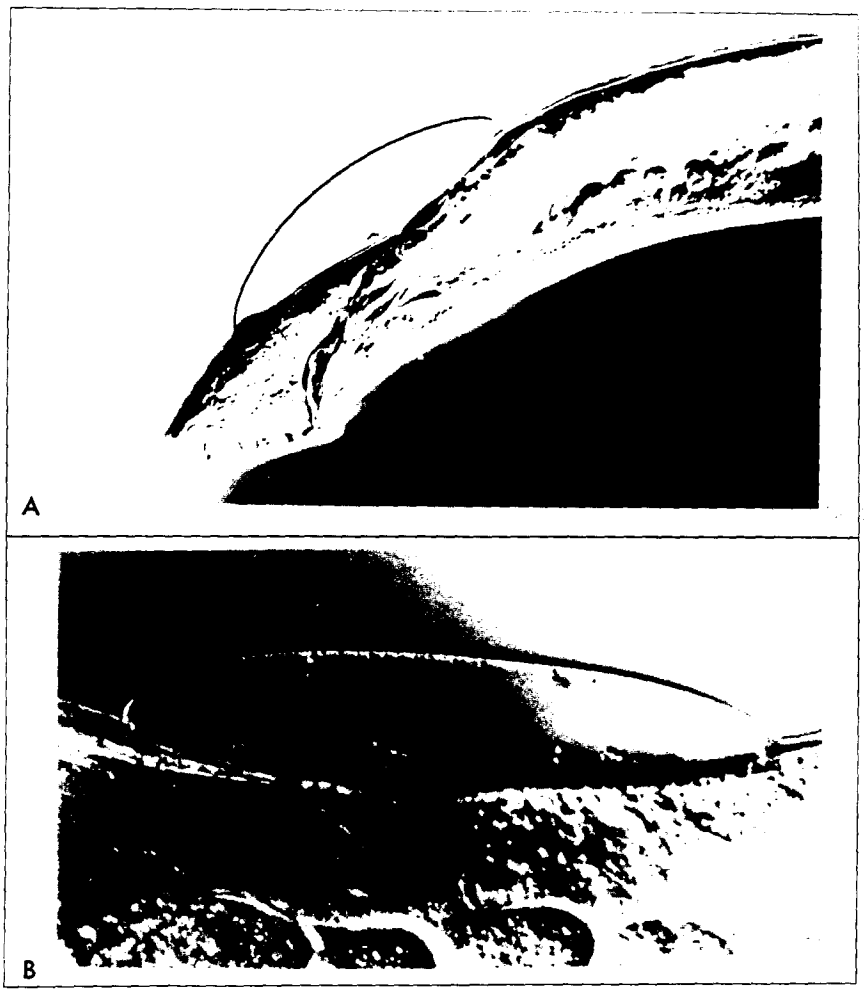


Figure 3. Blistered adult CB937 *bli-4(e937)* worms

A. Blister on adult male tail.

B. Blister on adult hermaphrodite.

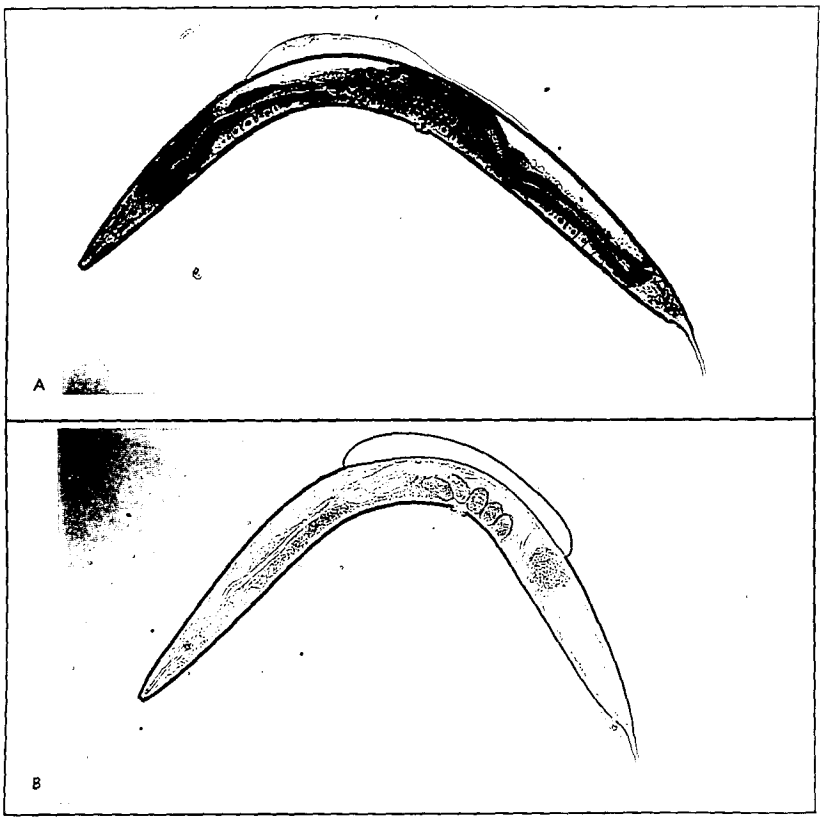


Figure 4. Blistering in a whole worm and isolated cuticle
A. Intact blistered CB937 adult hermaphrodite.
B. Isolated blistered cuticle from a different CB937 hermaphrodite. Eggs and collagenous pharynx and gut lining remain intact in this preparation.

3.1.1.2 Blistering is incompletely penetrant and is not temperature sensitive

Blistering in CB937 worms is incompletely penetrant, and expressivity is variable. Penetrance is defined as the portion of mutant homozygotes in a population that are blistered out of the total number of mutant homozygotes in the population. Expressivity is defined as the extent to which an affected individual expresses the phenotype. To determine the penetrance of blistering in CB937 worms, adult hermaphrodites were scored for blisters. Of these, 1043 of 1092 (95.5%) were blistered at 20C. To determine if blistering is temperature sensitive, this experiment was repeated at 15C. 1706 out of 1811 (94.3%) adult hermaphrodites were blistered at 15C, indicating that blistering is not temperature sensitive within the 15 - 20C range.

No experiments were done to quantify the expressivity of the blistered phenotype. However, anecdotally, in every population of blistered worms that were examined, expressivity varied from very small localized blisters to very severe blisters covering most of the animal. In general, in genetic backgrounds that enhance penetrance, blister expressivity in *e937* worms is increased. In genetic backgrounds that reduce penetrance, blister expressivity is decreased.

3.1.1.3 Blistering is adult specific

To determine exactly when during development *e937* worms first blister, the method of Cassada and Russell (1975) was used to follow moulting as an indicator of developmental stage. Wild type worms go through four larval stages prior to maturing to adulthood and express the adult cuticle after the fourth moult. The larval to adult moult occurs at 35.5 hours after hatching at 25C (Wood *et al.*, 1980). *C. elegans* undergoes a period of reduced activity prior to each moult, termed the lethargus period.

During this period, movement is reduced and pharyngeal pumping ceases. Lethargus periods in *bli-4*(*e937*) hermaphrodites were monitored at 25C by plotting the percentage of worms that were pumping in synchronous populations with respect to time (Figure 5A). The blistered phenotype is adult specific: larval stage worms did not blister. *e937* hermaphrodites first expressed blisters about 2 hours after the adult moult at 25C. *e937* worms reached the adult moult at 46 hours after hatching, 11 hours later than wild type worms. Thus, although *e937* lacks visible effects on larval-stage worms, it slows growth by about 30%.

Experiments were performed to determine if the adult specificity of blistering in *e937* worms requires expression of adult cuticle or other adult-specific structures. For example, would blistering be expressed early if the adult cuticle is expressed early? Would blistering not be expressed if adult-stage worms expressed a larval cuticle or failed to express an adult cuticle? To address these questions, mutants of two heterochronic genes, *lin-29* (III) and *lin-14* (X), that affect the timing of the adult cuticle were used. *lin-29(n1440)* causes reiteration of larval-stage cuticle in adults (that is, adult worms do not express the adult cuticle). *lin-14(n179ts)* causes precocious expression of the adult cuticle (ie adult-stage cuticle is expressed at larval stage four). The interaction of *e937* with mutations of the heterochronic genes *lin-29(n1440)* and *lin-14(n179ts)* was determined.

lin-29 loss-of-function alleles fail to make the L4 to adult cuticle switch, and reiterate the L4 stage cuticle, causing the animals to undergo extra moults (Ambros and Horvitz, 1984). This is the only known effect of *lin-29* mutations. I predicted that mutations in *lin-29* would suppress blistering if expression of the *e937* phenotype requires an adult cuticle. A *bli-4*(*e937*); *lin-29(n1440)* double mutant was constructed and screened for

the expression of blisters. *bli-4*(e937); *lin-29(n1440)* worms did not express blisters: 0/1160 F1 progeny of *bli-4*; *lin-29* hermaphrodites expressed blisters at any age. While the possibility that *lin-29* has effects other than the simple reiteration of L4 cuticle cannot be ruled out, it is most likely that blisters could not form in Lin-29 hermaphrodites because they lacked an adult cuticle. This observation supports the conclusion that blisters cannot form in adult worms not expressing the adult cuticle.

The *lin-14* loss-of-function allele *n179ts* results in the precocious expression of the adult cuticle after the third moult at the restrictive temperature of 25C (Ambros and Horvitz, 1984; Ambros and Horvitz, 1987). Hermaphrodites undergo a fourth moult producing a second adult cuticle. If the expression of the blistered phenotype requires an adult cuticle, then the precocious expression of the adult cuticle in *n179* worms was predicted to permit blistering one moult earlier than in wild type. A *bli-4*(e937); *lin-14(n179ts)* double mutant was constructed. Lethargus periods in *bli-4*(e937); *lin-14(n179ts)* hermaphrodites at 25C were determined by plotting the percentage of worms that were pumping in synchronous populations with respect to time (Figure 5B). The interaction of *bli-4* and *lin-14* was unexpectedly complex. At the restrictive temperature of 25C, most *bli-4*; *lin-14* animals were sterile. 20% of these worms arrested development prior to reaching the fourth moult. In addition, the rate of growth of the double mutant was variable. Consequently, synchronous populations quickly ceased to be synchronous. This may be seen by comparing the graph in Figure 5A with that in Figure 5B. Sterility, variable growth rates and variable larval arrests are not characteristics of either *lin-14(n179)* or *bli-4*(e937) alone. Therefore, *bli-4*(e937) and *lin-14(n179)* produce an incompletely penetrant synthetic lethality in the double mutant.

From the data presented in Figure 5B, it appears that blistering in the *bli-4*; *lin-14* double mutant did not occur until after the fourth moult. All of the blistered worms appeared to be adults based on size. Moreover, no blistered worms were observed undergoing a lethargus period or moulting. Thus, *lin-14(n179)* does not seem to alter the expression of blistering with respect to the number of moults. The lack of blistering in *lin-14(n179ts) bli-4(e937)* L4 worms may indicate that the presence of the adult cuticle in L4 *e937* animals is not sufficient for the expression of blisters. However, it is also possible that the cuticle expressed by *lin-14(n179ts)* worms is not a wild type adult cuticle.

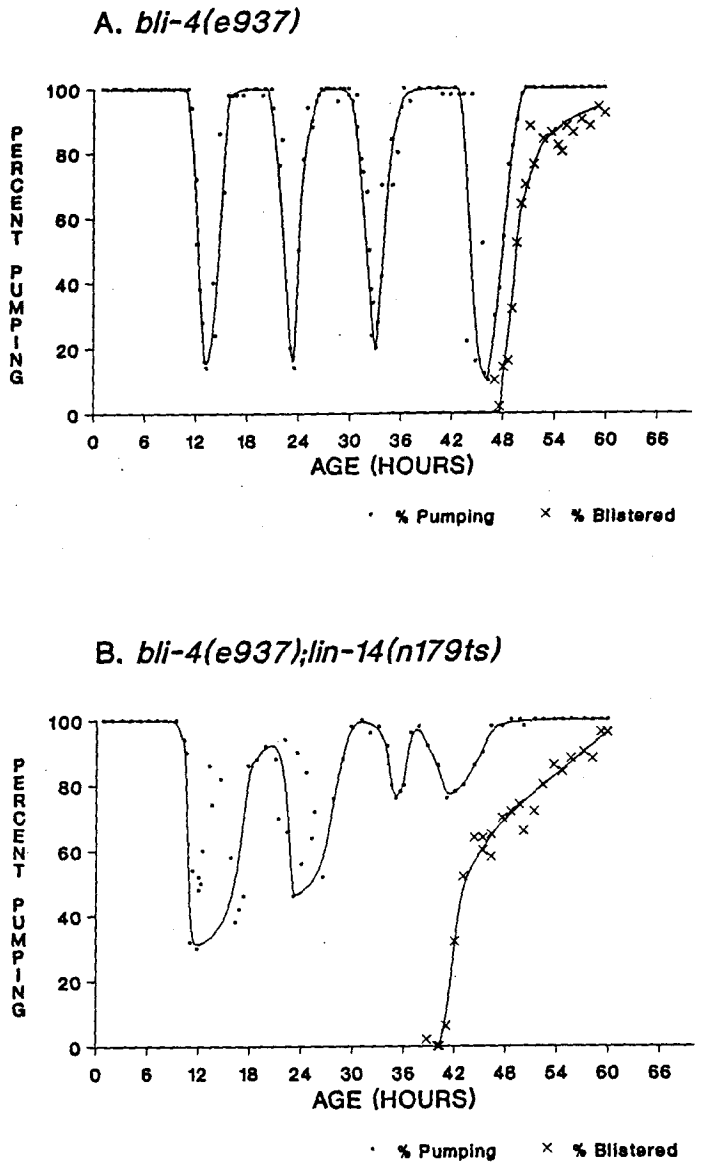


Figure 5. Developmental stage of blistering

Lethargus periods and expression of blistering in A. Bli-4, and B. Bli-4; Lin-14 hermaphrodites. Synchronous populations of several hundred animals were hatched at time zero (+/- 1 hr) and grown on NGM plates at 25C using the method of Cassada and Russell (1975). At frequent intervals, 50 animals were observed for 5 seconds and the percentage with pharyngeal pumping recorded.

3.1.1.4 Interaction of *e937* with cuticle genes

Roller, squat and some dumpy phenotypes are generally epistatic to blistering in Bli-1 and Bli-2 mutants (Higgins and Hirsh, 1977; Cox *et al.*, 1980). It was observed that Dpy-5 also has a dominant effect on blistering. To determine the effect of Dpy mutations, double mutants of Bli-4 and 11 Dpy mutations were constructed. The penetrance of blistering in *e937* homozygotes was determined in *dpy-X* homozygous and heterozygous backgrounds (Table 2). The average body length for each Dpy mutant is presented as a measure of Dpy severity. Non-Dpy Bli-4 length is about 1.2 mm.

In non-Dpy Bli-4 animals, blister penetrance is 95%. In Dpy homozygotes, blistering was completely or almost completely suppressed by 9 of the 11 Dpy mutations. The exceptions were Dpy-14 and Dpy-17, in which blister penetrance was reduced to 41% and 54% respectively. *dpy-14*(*e188*) and *dpy-17*(*e164*) are two of the least severe Dpy mutations (Table 2). This raises the possibility that Bli-4 penetrance in Dpy worms is related to Dpy severity. However, two other mild *dpy* mutations tested, *dpy-18*(*e364*) and the *e24* allele of *sqt-3*, did not permit the expression of blisters. In contrast, the *e1* allele of *dpy-1*, one of the most severe Dpy mutations tested, produced three Dpy Bli worms. Thus, Bli-4 penetrance does not correlate with severity of the Dpy phenotype.

To determine the magnitude of blister suppression by Dpy heterozygotes, *bli-4*; *dpy-X* double mutants were crossed to *bli-4*/ + males. Because the males were heterozygous, only half of the cross progeny were homozygous for *bli-4*. non Bli-4 worms could either be *bli-4* heterozygotes or *bli-4* homozygotes that did not express blistering. To compensate for *bli-4* heterozygotes, the penetrance was determined by multiplying the

percentage of blistered worms by two. Thus the penetrance is only an estimate in this experiment, and the penetrance can be greater than 100% if more that half the progeny are homozygous for *bli-4*, as was the case for *dpy-14*, which gave a penetrance of 118%

Blistering was completely or almost completely suppressed by 3 of the 11 heterozygous Dpy mutations. Dpy-5 and Dpy-13 both dominantly suppressed blistering completely in hermaphrodites and nearly completely in males; Dpy-6, which is X-linked, nearly completely suppressed blistering in hermaphrodites (Table 2). *dpy-13*(e*184*) is semi-dominant, but *dpy-5*(e*61*) and *dpy-6*(e*14*) are recessive. Dpy-9 and Dpy-10 dominantly reduce Bli-4 penetrance to about 20%. *dpy-9*(e*424*) and *dpy-10*(e*128*) are both recessive. Dpy-3, Dpy-14, Dpy-17, Dpy-18 and Sqt-3 all had moderate or no dominant effects on Bli-4 penetrance. Of these, only *sqt-3(e24)* is semi-dominant. The dominant suppression of blistering by recessive Dpy mutations indicates that the cuticle of the Dpy heterozygotes is not fully wild type.

Table 2. *bli-4(e937)* penetrance in *dpy* homozygote and heterozygote backgrounds

dpy-X	Average length[a] (mm)	Penetrance[b] in *bli-4*; *dpy-X* / *bli-4 dpy-X*	Penetrance[c] in *bli-4*; *dpy-X* / *bli-4 +*	
		hermaphrodites	hermaphrodites	males
dpy-1	0.50 (0.036)	<1% (n=761)	58% (n=214)	51% (n=203)
dpy-3	0.54 (0.040)	0% (n=909)	40% (n=421)	0% (n=405)[d]
dpy-5	0.49 (0.051)	0% (n=735)	0% (n=459)	6% (n=442)
dpy-6	0.50 (0.051)	0% (n=245)	2% (n=346)	0% (n=337)
dpy-9	0.55 (0.036)	0% (n=1435)	19% (n=448)	22% (n=461)
dpy-10	0.53 (0.050)	0% (n=885)	19% (n=446)	12% (n=427)
dpy-13	0.49 (0.046)	0% (n=1143)	0% (n=437)	<1% (n=427)
dpy-14	0.57 (0.055)	41% (n=343)	118% (n=159)	88% (n=164)
dpy-17	0.72 (0.065)	54% (n=1564)	68% (n=361)	81% (n=370)
dpy-18	0.68 (0.066)	0% (n=687)	66% (n=373)	98% (n=361)
sqt-3	0.81 (0.054)	0% (n=345)	58% (n=442)	51% (n=424)
dpy +	1.2[e]		95.5% (n=1092)	

[a]Average length of 20 *dpy-X unc-13* adult hermaphrodites. Standard errors are presented in brackets.

[b]Penetrance is defined as the percentage of blistered worms out of the total number observed.

[c]Because the *bli-4;dpy-X* hermaphrodites were mated to *bli-4* heterozygous male in this experiment (Section 2.3), a maximum of 50% of the worms are expected to blister if penetrance is 100%. Therefore, penetrance is defined as the percentage of blistered worms out of one half the total number of observed.

[d]*dpy-3* and *dpy-6* are X-linked and all male cross progeny are Dpy.

[e]From Wood, 1988.

3.1.2 Mapping of *bli-4* alleles

h42, h199, h254 and *e937* were three factor mapped by scoring segregation from strains bearing *cis*-linked flanking markers *dpy-5* and *unc-13 in trans* to an unmarked chromosome (Table 3). Each allele mapped near the center of the interval, with the exception of *h199* (see below). Recombinants were picked as Dpy-5 or Unc-13 worms and progeny tested for the presence of *bli-4*. *s90*, which was not induced on a *dpy-5* chromosome, was two factor mapped to 0.6 m.u. from *unc-13*(*e51*), consistent with a position between *dpy-5 unc-13*.

h199 mapped 0.8 m.u. from *unc-13*, but failed to recombine with *dpy-5*. This apparent crossover suppression could indicate that *h199* is a deficiency spanning *bli-4*, or that *h199* is linked to a second mutation in an essential gene closely linked or to the left of *dpy-5*. *h199* is unlikely to be a deficiency, because it complements alleles of *unc-40*, which is between *bli-4* and *dpy-5*. However, the chromosome carrying *h199* fails to complement *sDf4*, a deficiency of *dpy-5* (Howell, 1989). *h199* could be an inversion with breakpoints in *bli-4* and an essential gene in *sDf4*. More likely, however, KR513, the strain carrying *h199*, carries a second site mutation, in a gene to the left of *unc-40*. This conclusion is supported by the observation that *Eco*RI restriction endonuclease digested genomic DNA prepared from *h199* heterozygous worms showed wild type band sizes on Southern blots when probed with *bli-4* genomic DNA clones (Section 3.2.1.2.2). The putative second site mutation in KR513 was designated *h1360*.

The position of the *e937* allele of *bli-4* on LGI between the markers *dpy-5* and *unc-13* was confirmed by three factor mapping. Recombinant F1 progeny of hermaphrodites of the genotype *dpy-5 bli-4 unc-13*(*e450*)/ + + + were picked. Eleven Bli Unc, six Dpy, ten

Unc, and seven Bli Dpy recombinants were recovered. This gives the map position *dpy-5* (17/34) *bli-4* (17/34) *unc-13*, placing *bli-4* at the center of the *dpy-5 unc-13* interval, which is 1.6 m.u. (Howell *et al.*, 1987). Having an accurate map position for *bli-4* was important to the cloning of this gene.

3.1.3 Screen for a mutator allele

To facilitate the isolation of the *bli-4* coding region, a screen was performed to identify potential transposable element insertion mutations of *bli-4* (Figure 6). The strain KR1822 was constructed for use in this screen and assayed for mutator activity as described in Section 2.8. KR1822 has the genotype *unc-63*(e*384*) *unc-13*(*e450*); *mut-6*(*st702*). *mut-6* causes high levels of transposition of the transposable genetic element Tc1 (Mori, Moerman and Waterston, 1988). Spontaneous alleles of *bli-4* would occur in KR1822 whenever a transposable element inserted into the *bli-4* locus. In the screen, blistered worms in the F1 generation could only arise if a spontaneous *bli-4* mutation occurred in the gonad of the hermaphrodite. Because the males were heterozygous for *bli-4*, one-half of the F1 progeny were not informative in this cross (ie, one-half of the spontaneous *bli-4* alleles would not be detected because they would receive a wild type *bli-4* allele from the male). Three Bli worms, two hermaphrodites and one male, were identified after screening 164,600 worms (82,300 informative chromosomes), an induction frequency of $3.6X10^{-5}$. Of the three spontaneous blistered animals recovered, one survived. The surviving hermaphrodite carried a class II *bli-4* lethal allele designated *h1010* and maintained using the translocation *szT1(I;X)* in the strain KR1858.

Table 3. *bli-4* genetic map data

Allele	Maternal genotype	Wild type	Dpy	Unc	Recombination frequency (m.u.)	
					dpy-5 to *bli-4*	*bli-4* to *unc-13*
h42	*dpy-5 h42 unc-13* + + +	1376	10	8	0.5 m.u. (0.3-0.9)[a]	0.4 m.u. (0.2-0.8)
h199	*dpy-5 h199 unc-13* + + +	1781	0	10	0.0m.u. (0.0-0.1)	0.4 m.u. (0.2-0.9)
h254	*dpy-5 h254 unc-13* + + +	1033	9	3	0.7 m.u. (0.3-1.2)	0.2 m.u. (0.1-0.5)
s90	*s90 unc-13* + +	1896	N/A	14	N/A	0.6 m.u. (0.3-0.9)

[a]95% confidence intervals are given in parentheses. Confidence intervals were calculated using the table of Crow and Gardener (1959).

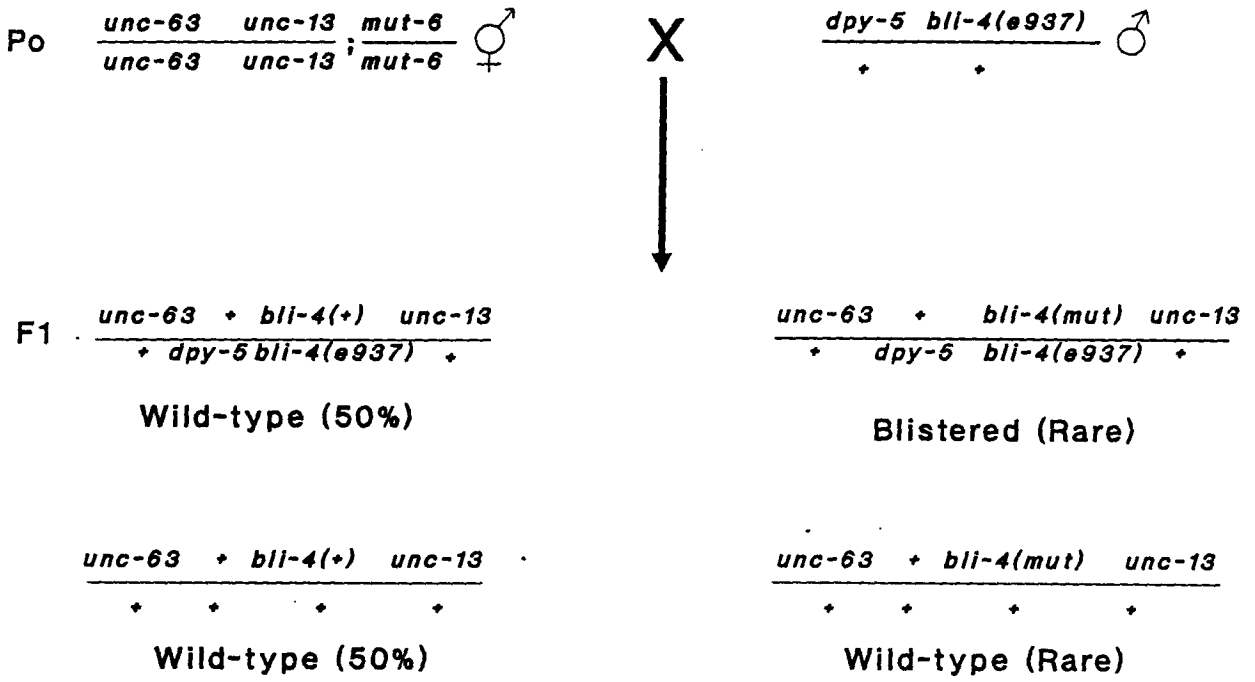


Figure 6. Screen for mutator-induced alleles of *bli-4*

Screen for spontaneous *bli-4* lethal alleles in a *mut-6*(*st702*) background. KR1822 hermaphrodites were crossed to *dpy-5 bli-4*(*e937*)/ + + males and the F1 progeny screened for blisters. Blisters would result when an ovum carrying a novel *bli-4* mutation in the germline of the KR1822 hermaphrodite was fertilized by a sperm carrying *bli-4* (*e937*). Because *dpy-5*(*e61*) suppresses blistering in *e937* homozygotes, only severe *bli-4* mutations, that is, lethal class II alleles, would be detected in this screen.

3.1.4 Determination of developmental arrest stage

The stage at which lethal homozygotes arrested development was determined. Several heterozygous hermaphrodites of the genotype *dpy-5 let-X unc-13/ + + +* were permitted to lay eggs on an NGM plate for a short period (not more than two hours) and then the homozygous lethal progeny were examined by Normarski differential interference microscopy for time of arrest. All of the class II lethal alleles of *bli-4* (including *h1010*) arrest development at or just before hatching, which is the end point of embryogenesis. Thus, the essential role of the *bli-4* affected by the class II alleles occurs at or before the end of embryogenesis. The class III complementing alleles (*s90* and *h754*) arrest development during the L1 stage, later than the non-complementing alleles.

3.1.5 Evidence that *e937* is hypomorhic

Muller (1937) defined a hypomorphic mutation as one that resulted in a reduction, but not a total loss, of function. Genetically, hypomorphic mutations have a phenotype that is more severe *in trans* to a deletion, or a null (loss of function) allele. No deletions of the *bli-4* locus were available. However, the blistered phenotype of *e937* exhibited higher penetrance, and greater expressivity *in trans* to class II alleles in a genetic background that did not normally permit blistering. The blistered phenotype of *e937* hermaphrodites is suppressed in *dpy-5(e61)* heterozygotes (Section 3.1.1.4). This suppression was reversed when *e937* was heterozygous to a class II lethal allele. That is, the blistered phenotype of *e937* is more severe *in trans* to class II alleles. When *bli-4(e937)/ bli-4(e937)* hermaphrodites were crossed to *dpy-5 bli-4(h42) unc-13/ + + +* males, 46% of hermaphrodites and 50% of males were blistered. The maximum percentage of blistered progeny expected in this cross was 50%. Similar results were obtained with

other lethal alleles (data not shown). This result was interpreted as follows: that the function required to prevent blistering in adult cuticle is reduced in *e937* hermaphrodites, and is more severely reduced or absent in the lethal alleles tested. This result is consistent with Muller's definition (Muller, 1937) of a hypomorphic mutation.

3.1.6 Complementation analysis of *bli-4* alleles

The results of *inter se* complementation tests of *bli-4* alleles are presented in Table 4. Mutations fell into three complementation classes that correlated with the phenotypes. Class I was the viable allele *e937*. Class II was the lethal alleles, *h42, h199, h254, h384, h427, h520, h699, h791* and *h1010*. Class II alleles failed to complement all other alleles. Class III was the lethal alleles *s90* and *h754*. These alleles complemented *e937*. Two explanations for this complementation pattern are as follows. In the first hypothesis, *e937* and the class III complementing alleles are mutations of one gene that affect complementary domains or alternatively spliced exons of the gene. The second hypothesis is that *e937* and the class III alleles are mutations of two separate genes, and all of the class II lethal mutations are deletions affecting both genes. In this hypothesis, the nine class II mutations are all deletions that delete the gene affected by *e937* and the gene affected by the class III mutations. The two gene hypothesis can be proven incorrect if just one of the class II alleles is shown not to be a deletion. At least one class II allele, *h1010* is not a deletion, it is an insertion (Section 3.2.1.2.1). Furthermore, four other class II alleles, *h42, h199, h384* and *h487*, do not contain chromosomal alterations that are detectable by Southern analysis (Section 3.2.1.2.2). On this basis, I infer that *e937* and the class III mutations *s90* and *h754* are complementing alleles of the same gene.

In summary, the class I allele *e937* is hypomorphic, and is the only allele that results in blisters. Blisters are adult specific and occur between cuticle layers. Class II alleles fail to complement *e937* and have the most severe phenotype, developmental arrest at the point of hatching. That class II alleles are the most severe is consistent with the hypothesis that they are null alleles. Class III alleles complement the blistered phenotype of *e937*, and arrest development at the L1 stage.

Table 4. *Inter se* complementation data[a] for *bli-4* alleles

	Class I	Class II									Class III	
Allele	*e937*	*h42*	*h199*	*h254*	*h384*	*h427*	*h520*	*h699*	*h791*	*h1010*	*h754*	*s90*
e937											+	+
h42		-	-	-	-	-	-	-	-	-	-	-
h199			-	-	-	-	-	-	-	-	-	-
h254				-	-	-	-	-	-	-	-	-
h384					-	-	-	-	-	-	-	-
h427						-	-	-	-	-	-	-
h520							-	-	-	-	-	-
h699								-	-	-	-	-
h791									-	-	-	-
h1010										-	-	-
h754											-	-
s90												-

[a]Complementation tests were conducted as decribed (Section 2.5). Failure to complement is indicated by "-". Allelic combinations resulting in a blistered phenotype are shaded grey; other complementation failures are lethal. complementation is indicated by "+".

3.2 MOLECULAR ANALYSIS OF THE *bli-4* LOCUS

A physical map of overlapping cosmid and YAC clones has been generated for the *C. elegans* genome (Coulson *et al.*, 1986; Coulson *et al.*, 1988). The *bli-4* locus was cloned by aligning *bli-4* with the physical map. Cosmids spanning *bli-4* were used to as hybridization experiments to probe Southern blots of restriction enzyme digested DNA from *bli-4* mutant strains. Clones of genomic DNA that detected restriction fragment length differences (RFLD) in *bli-4* mutant strains were then used as probes to isolate cDNA clones. The DNA sequence of the cDNA clones was determined, and the sequence used to search DNA and protein computer data bases. This approach revealed that *bli-4* encodes protein products that structurally resemble the kex2-like proteinases, enzymes that process secreted proteins by proteolytic cleavage.

3.2.1 Identification of the *bli-4* coding region

3.2.1.1 Alignment of the genetic and physical maps

A partial physical map of cosmids from the *bli-4* region is presented in Figure 7. *bli-4* had been mapped between *dpy-5* and *dpy-14* (Rose and Baillie, 1980). Approximately 35 cosmid clones are needed to cover the *dpy-5* to *dpy-14* interval. The average size of the cosmid clone inserts is 34 kb (Coulson *et al.*, 1986). Therefore, the *dpy-5* to *dpy-14* interval includes about 1,200 Kb of DNA. To define more precisely the position of *bli-4* within the physical map, I used two strain-specific RFLDs markers flanking the locus. The RFLD *hP5* and an RFLD associated with the left breakpoint of deletion *hDf8* were used to place *bli-4* within a 200 Kb interval.

3.2.1.1.1 *hP5* defines the left-most position of *bli-4* in the physical map

A first step in placing *bli-4* within the physical map was to position the gene with respect to the molecular marker *hP5*. *hP5* is an N2 (Bristol) BO (Bergerac) strain restriction fragment length polymorphism (RFLP) *hP5*.

The *hP5* RFLD was detected by the hybridization probe pCeh51. pCeh51 detected a 2.4 kb band in *Eco*RI digested N2 DNA and a 4.0 kb band in *Eco*RI digested BO DNA. To three factor map *bli-4* with respect to *hP5*, I constructed N2 BO recombinant mapping strains as described in Figure 8. Southern blots of *Eco*RI digested genomic DNA prepared from recombinant strains segregated from *hP5* mapping strains were hybridized to the pCeh51 probe. An example of such an experiment is presented in Figure 9. The strains used and the data obtained are presented in Table 5. 29 strains derived from independent recombinants were analyzed. The map order inferred from these experiments is *dpy-5* (7/29) *hP5* (3/29) *bli-4* (18/29) *unc-13*. This result places *hP5* between *dpy-5* and *bli-4*, an interval of 0.9 cM. 3/10 of 0.9 cM is about 0.3 cM. The DNA density in the *bli-4* region is about 500 kb per cM (Starr *et al.*, 1989), predicting that *bli-4* would be located about 150 kb to the right of *hP5*. The *hP5* mapping data lead to the conclusion that *hP5* defines the left-most position of *bli-4* in the physical map (Figure 7).

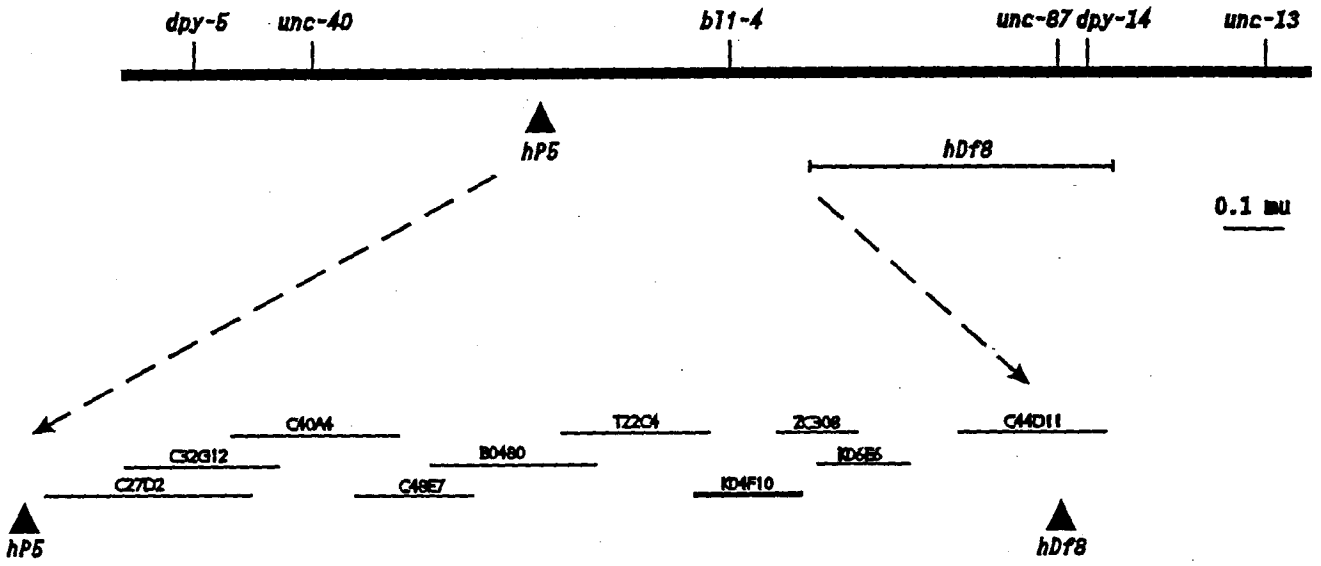
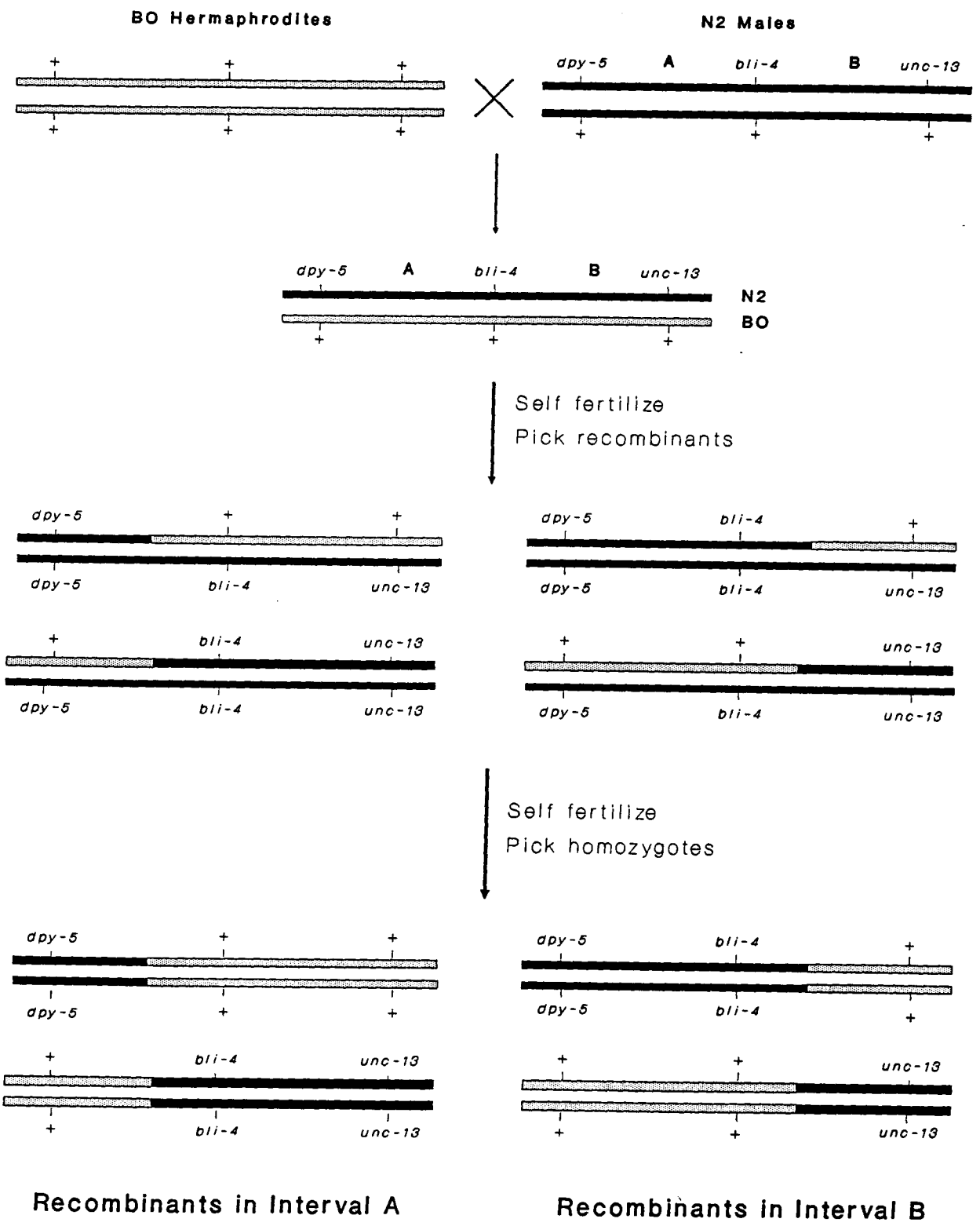


Figure 7. Genetic and physical maps of the *bli-4* region

bli-4 was mapped genetically between the N2/BO strain RFLD *hP5* (0.3 mu to the left of *bli-4*), and the left breakpoint of the deletion *hDf8*. This interval is spanned by 200 kb of contiguous cosmid clones, except for a small gap between K06E6 and C44D11, which is spanned by YAC clones (Coulson *et al.*, 1986; Coulson *et al.*, 1988). The positions of *hP5* and the left break point of *hDf8* in the physical map are indicated by arrows. All of the cosmids shown were used to probe *bli-4* mutations.

Figure 8. Construction of N2/BO RFLP mapping strains

Wild type BO hermaphrodites were crossed to *dpy-5*(e61) *bli-4*(e937) *unc-13*(e420)/ + + + N2 males. L4 hermaphrodites were picked from the F1 progeny and were placed individually on fresh plates. F2 progeny recombinant in interval A were picked as Dpy-5 or Bli-4 Unc-13 hermaphrodites. F2 progeny recombinant in interval B were picked as Unc-13 or Dpy-5 Bli-4 hermaphrodites. Because *dpy-5*(e61) suppresses blistering in e937 homozygotes, all Dpy-5 progeny were crossed to a *bli-4* lethal bearing strain to determine presence of the *bli-4*(e937) allele.


Construction of N2/BO mapping strains
BO Hermaphrodites
N2 Males
dpy-5
A
bli-4
B
unc-13
N2
BO
Self fertilize
Pick recombinants
Self fertilize
Pick homozygotes
Recombinants in Interval A
Recombinants in Interval B

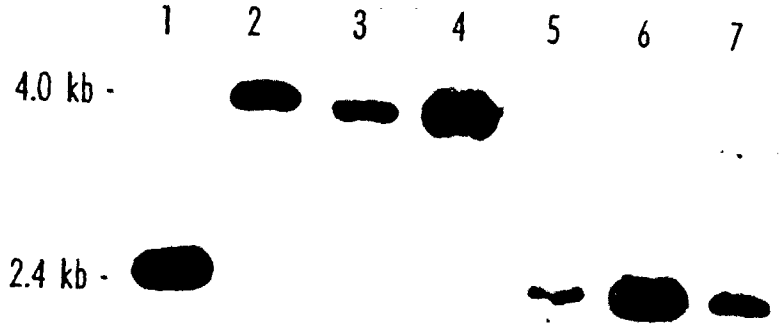


Figure 9 Southern analysis of *hP5* mapping strains

Genomic DNA digested with *Eco*RI and probed with pCeh51, which detects the RFLD *hP5*. Lane 1, N2; lane 2, BO; lane 3, KR1210; lane 4, KR1211; lane 5, KR1212; lane 6, KR1220; lane 7, KR1221. The genotypes of the recombinant strains are presented in Table 5.

Table 5. *hP5* Three factor mapping data.

Strain	Recombinant Genotype	Recombinant Interval[a]	*hP5* Pattern[b]	*hP5* Recombined with *bli-4?*
KR1187	+ (BO) *bli-4 unc-13* (N2)	A		
KR1189	+ (BO) *bli-4 unc-13* (N2)	A		
KR1190	+ (BO) *bli-4 unc-13* (N2)	A	BO	Yes
KR1191	+ (BO) *bli-4 unc-13* (N2)	A		
KR1192	+ (BO) *bli-4 unc-13* (N2)	A	N2	No
KR1193	+ (BO) *bli-4 unc-13* (N2)	A		
KR1194	+ (BO) *bli-4 unc-13* (N2)	A		
KR1195	+ (BO) *bli-4 unc-13* (N2)	A	BO	Yes
KR1210	*dpy-5* (N2) + + (BO)	A	BO	No
KR1211	*dpy-5* (N2) + + (BO)	A	BO	No
KR1212	*dpy-5* (N2) + + (BO)	A	N2	Yes
KR1213	*dpy-5* (N2) + + (BO)	A		
KR1214	*dpy-5* (N2) + + (BO)	A	BO	No
KR1215	*dpy-5* (N2) + + (BO)	A	BO	No
KR1216	*dpy-5* (N2) + + (BO)	A	BO	No
KR1217	*dpy-5* (N2) + + (BO)	A	N2	Yes
KR1218	*dpy-5* (N2) + + (BO)	A	BO	No
KR1196	+ + (BO) *unc-13* (N2)	B	BO	No
KR1197	+ + (BO) *unc-13* (N2)	B	BO	No
KR1198	+ + (BO) *unc-13* (N2)	B	BO	No
KR1199	+ + (BO) *unc-13* (N2)	B	BO	No
KR1200	+ + (BO) *unc-13* (N2)	B	BO	No
KR1201	+ + (BO) *unc-13* (N2)	B	BO	No
KR1202	+ + (BO) *unc-13* (N2)	B	BO	No
KR1203	+ + (BO) *unc-13* (N2)	B	BO	No
KR1204	+ + (BO) *unc-13* (N2)	B	BO	No
KR1205	+ + (BO) *unc-13* (N2)	B	BO	No
KR1206	+ + (BO) *unc-13* (N2)	B	BO	No
KR1207	+ + (BO) *unc-13* (N2)	B	BO	No
KR1208	+ + (BO) *unc-13* (N2)	B	BO	No
KR1209	+ + (BO) *unc-13* (N2)	B	BO	No
KR1219	*dpy-5 bli-4* (N2) + (BO)	B	N2	No
KR1220	*dpy-5 bli-4* (N2) + (BO)	B	N2	No
KR1221	*dpy-5 bli-4* (N2) + (BO)	B	N2	No
KR1222	*dpy-5 bli-4* (N2) + (BO)	B	N2	No

[a]Recombinant interval A indicates recombination between *dpy-5* and *bli-4*; interval B indicates recombination between *bli-4* and *unc-13*.

[b]*hP5* pattern N2 indicates that pCeh51 detects a 2.4 kb *Eco*RI fragment in the recombinant strain; pattern BO indicates that pCeh51 detects a 4.0 kb band.

3.2.1.1.2 *hDf8* defines the right-most position of *bli-4* on the physical map

A deletion of *dpy-14*, *hDf8* (McKim, Starr and Rose, 1992), provided a convenient tool to determine the right-most position of *bli-4* within the cosmid map. To determine the position of *hDf8* relative to *bli-4*, the *hDf8* chromosome was tested for complementation of *bli-4(e937)* allele. Because *hDf8* did not carry any genetic markers, the following protocol was used. Males of the genotype *hDf8 (I); O/ szTI(I;X)[lon-2]* were mated to hermaphrodites of the genotype *dpy-5 bli-4(e937)*. In this cross, the only males produced have the genotype *dpy-5 bli-4(e937)/ + + hDf8*. If *hDf8* deleted *bli-4*, the non Lon-2 males would be blistered; if *hDf8* did not delete *bli-4*, the male progeny would be wild type. Only wild type male progeny were observed, indicating that *hDf8* does not delete *bli-4* and demonstrating that the gene lies outside of the deletion, and to the left of the deletion breakpoint.

To localize *hDf8* on the cosmid map, I determined the dosage of restriction fragments of cosmids on either side of the deletion breakpoint. These fragments were used to probe Southern blots of restriction digested wild type and deletion heterozygote genomic DNA. Using an LKB scanning densitometer, the intensity of autoradiograph bands was compared to an internal control probe, pCes233, which hybridizes to the molecular marker sP4 on LGIV (Baillie, Beckenbach, and Rose, 1985). *Eco*RI fragments of cosmids K06E6 (K06E6-E6) and C44D11 (C44D11-E5) were used (Figure 10). K06E6-E6 was present at equal intensity in both N2 and KR1816 (*hDf8*) strains, while C44D11-E5 was present at 1/2 the intensity in the deletion heterozygote strain as in the wild type strain (Table 6). Thus, the left break point of *hDf8* is located between K06E6 and C44D11. The *hDf8* mapping

data indicates that the *bli-4* coding region is located to the left of C44D11 on the physical map (Figure 7). This region includes approximately 200 kb of DNA, consistent with the *hP5* mapping data.

In Southern analysis of *Eco*RI digested *hDf8* heterozygote DNA, C44D11 detected the loss of an 8.5 kb band and a novel band at 5.5 kb (Figure 11). The RFLD detected by C44D11 is a candidate for the *hDf8* left breakpoint. The right breakpoint of *hDf8* occurs between *dpy-14* and *unc-13*, and is detected by the cosmid C14A12 in Southern analysis (McKim, Starr and Rose, 1992).

Figure 10. Dosage analysis of *hDf8*

A. N2 and KR1000 DNA digested with *Eco*RI and probed with K06E6-E6 (the 6th largest *Eco*RI fragment of K06E6) and pCes233, which detects the RFLD sP4.
B. LKB laser densitometer scan of the N2 lane from panel (A).
C. LKB laser densitometer scan of the KR1000 lane from panel (A).
D. N2 and KR1000 DNA digested with *Eco*RI and probed with C44D11-E5 (the 5th largest *Eco*RI fragment of C44D11) and pCes233, which detects the RFLD sP4.
E. LKB laser densitometer scan of the N2 lane from panel (D).
F. LKB laser densitometer scan of the KR1000 lane from panel (D).

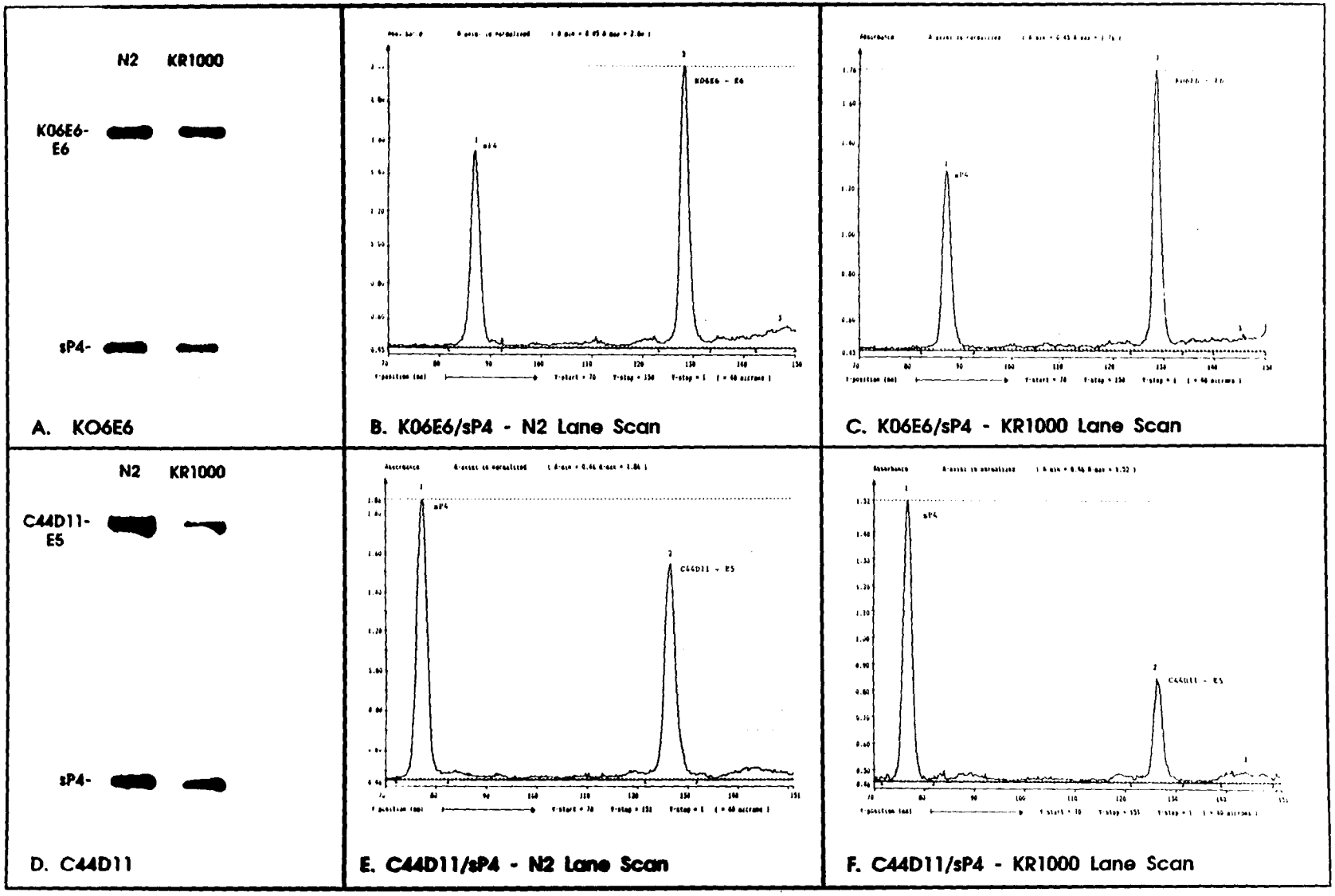

N2
KR1000
K06E6-
E6
sP4-
A. KO6E6
B. K06E6/sP4 - N2 Lane Scan
C. K06E6/sP4 - KR1000 Lane Scan
N2
KR1000
C44D11-
E5
sP4-
D. C44D11
E. C44D11/sP4 - N2 Lane Scan
F. C44D11/sP4 - KR1000 Lane Scan

Table 6. *hDf8* dosage data

Cosmid[a] fragment	Genomic[b] DNA	Relative Area[c] Cosmid Probe	Relative Area[c] *sP4* probe	Cosmid probe[d] *sP4* probe	KR1816[e] N2	Deleted in *hDf8*?
K06E6 - E6	N2	52.31	37.12	1.41	1.05	No
	KR1816	52.88	35.46	1.49		
C44D11 - E5	N2	45.81	54.19	0.42	0.49	Yes
	KR1816	26.93	64.16	0.85		

[a]Fragment to be tested for dosage relative to the LGIV control probe *sP4*. Probes were isolated by electroelution. K06E6 - E6 is the sixth largest *Eco*RI fragment of the cosmid K06E6. C44D11 - E5 is the fifth *Eco*RI fragment of the cosmid C44D11.

[b]Source of genomic DNA to be probed with cosmid and test fragments.

[c]Relative area of hybridization autoradiograph bands determined by scanning densitometer (see Figure 10).

[d]Ratio of cosmid probe band relative area to control probe band relative area.

[e]Ratio of cosmid probe band area to control probe band area determined in the N2 lane to that of the KR1816 lane. A ratio of 1 indicates that the cosmid test fragment is not deleted by *hDf8*; a ratio of 0.5 indicates that the test band is deleted by *hDf8*.

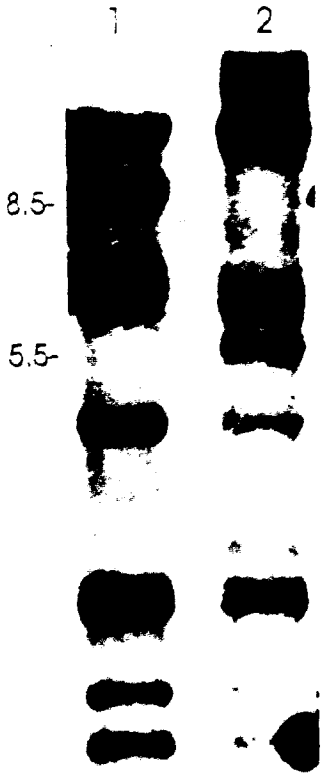


Figure 11. Detection of RFLDs by C44D11 in *hDf8*

N2 (Lane 1) and KR1000 (Lane 2) DNA digested with *Eco*RI and probed with C44D11. An 8.5 kb band detected in N2 DNA is absent in KR1000 DNA, and a novel 5.5 kb band is detected by C44D11 in KR1000 DNA.

3.2.1.2 Identification of rearrangements in the DNA of *bli-4* mutant strains

3.2.1.2.1 A Tc1 insertion mutation in KR1858(*h1010*) DNA

To determine the exact position of *bli-4* within the interval defined by *hP5* and *hDf8*, I used each of the cosmids shown in Figure 7 as a probe to compare the genomic DNA restriction patterns of two strains, wild type (N2) and KR1858, which carried the putative Tc1 mutation *h1010* (Section 3.1.3). Each cosmid was used to probe *Eco*RI and *Sal*I restriction digested genomic DNA prepared from wild type and KR1858 DNA (Figure 12). I chose to use *Eco*RI and *Sal*I to digest the genomic DNA because, with some exceptions (Eidie and Anderson, 1985b), *Eco*RI does not cut inside Tc1, and would produce a novel band 1.6 kb larger than in the parental strain. *Sal*I cuts twice within Tc1, and would produce three novel bands smaller than that of the parental strain. One of these novel bands, the *Sal*I fragment internal to Tc1, would be only 247 bp in length, a size that is difficult to see on a Southern blot. Therefore, only two new bands would be observed. Because KR1858 is heterozygous for the *h1010* mutation, the N2 bands would also be observed.

Only one of the cosmids, K04F10, detected band shifts. A novel 2.9 kb band was observed in *Eco*RI digested KR1858 DNA, and two novel bands of 16.6 kb and 6 kb were observed in *Sal*I digested DNA (Figure 12a). If the 2.9 kb band in KR1858 DNA detected by K04F10 was a result of a 1.6 kb Tc1 insertion, then the 1.3 kb *Eco*RI would contain the insertion site. To test this prediction, *Eco*RI and *Sal*I digested KR1858 DNA was probed with using the 1.3 kb *Eco*RI fragment of K04F10. The same novel bands were detected (Figure 12b).

To confirm my interpretation of the *h1010* restriction data, I used the polymerase chain reaction to amplify across the *h1010* Tc1 insertion site

(Figure 13). The primers used were p618 and KRp13. p618 is specific to a Tc1 sequence starting 72 bp from one end of the transposable element. KRp13 is specific to a sequence 80 bp to the 5′ side of the 1.3 kb *Eco*R1 fragment containing the insertion (Section 3.2.3 reports the genomic sequence used to design the KRp13 primer). An amplification band of approximately 770 bp was obtained (Figure 13b, lane 2). The presence of an amplification band confirms that the insertion is a Tc1 element. The size of the amplification band is consistent with an insertion site approximately 620 bp into the 1.3 kb *Eco*RI fragment.

As an additional control, DNA from KR1822, the mutator strain that produced KR1858, was probed. K04F10 detects wild type band patterns in *Eco*RI and *Sal*I restriction digested KR1822 DNA, the mutator strain that produced *h1010*. The Tc1 insertion in KR1858 is a candidate for the *h1010* mutation for two reasons. First, the occurrence of the Tc1 insertion was concurrent with the occurrence of the *h1010* mutation. Second, the position of the insertion is consistent with the mapping data of *bli-4* with respect to *hP5* (Section 3.2.1.1.1; Table 5).

3.2.1.2.2 A deletion mutation in CB937 *(e937)* DNA

To identify additional rearrangements of the *bli-4* locus, I used a 11 kb *Xho*I subclone of K04F10 in the plasmid pCeh181 to probe Southern blots of genomic DNA prepared from other *bli-4* mutant strains. I probed *Eco*RI digested DNA from strains carrying *bli-4* alleles *e937*, *h42*, *h199*, *h384*, and *h427*. A rearrangement was detected in CB937 DNA, homozygous for *e937*, but not in the other strains tested (Figure 14). The 11 kb XhoI fragment of K04F10 detected the disappearance of two adjacent *Eco*RI fragments of 2.2 and 3.0 kb and the appearance of a novel 1.7 kb band in CB937 DNA. The simplest interpretation of this pattern is that the rearrangement is a 3.5 kb

deletion spanning one *Eco*RI site, and fusing the 2.2 and 3.0 kb bands into one smaller 1.7 kb band. To confirm this interpretation and to further characterize the deletion, I constructed a lambda-zap genomic library using DNA from CB937. Using a mixture of the 2.2 and 3.0 kb *Eco*RI fragments of pCeh181 as probes, I isolated a 1.7 kb *Eco*RI fusion fragment from the CB937 library in the plasmid pCeh206. The fusion fragment was used to probe CB937 genomic DNA (Figure 14). The pattern shown in Figure 14 is consistent with a deletion of 3.5 kb. My interpretation of the of restriction data was confirmed by sequencing through the deletion breakpoints (See Section 3.2.5.2).

The 11 kb *Xho*I fragment of K04F10 detects rearrangements in the DNA of strains carrying two independent *bli-4* mutations; KR1858, carrying *h1010*, and CB937, carrying *e937*. Based on my positioning of *bli-4* relative to *hP5* and *hDf8*, and my detection of two chromosomal alterations affecting the same 11 kb region, I conclude that the *bli-4* coding region is at least partially contained within K04F10. This conclusion is supported by the finding that an artificial duplication constructed by the germline injection of K04F10 and the wholly overlapping cosmid C29F10 rescues the lethal phenotypes of the class II allele *h42*, and the class III allele *h754* (Jennifer McDowall, personal communication). The 1.3 kb *Eco*RI fragment of K04F10 will be referred to as the 1.3 kb *h1010* probe. The 1.7 kb fusion fragment in pCeh206 will be referred to as the 1.7 kb *e937* probe.

3.2.1.3 Detection of RNA bands with K04F10 DNA fragments

The 1.3 kb *h1010* probe, and the 1.0 and 1.9 kb *Eco*RI fragments flanking the 1.3 kb *h1010* probe each detected four bands on a northern blot of *C. elegans* mixed stage total RNA (Figure 15). These bands were 3.5,

3.1, 2.6. and 1.6 kb in size. The detection of RNA bands by the DNA fragments flanking the Tc1 insertion site in *h1010* indicates that this region includes a gene.

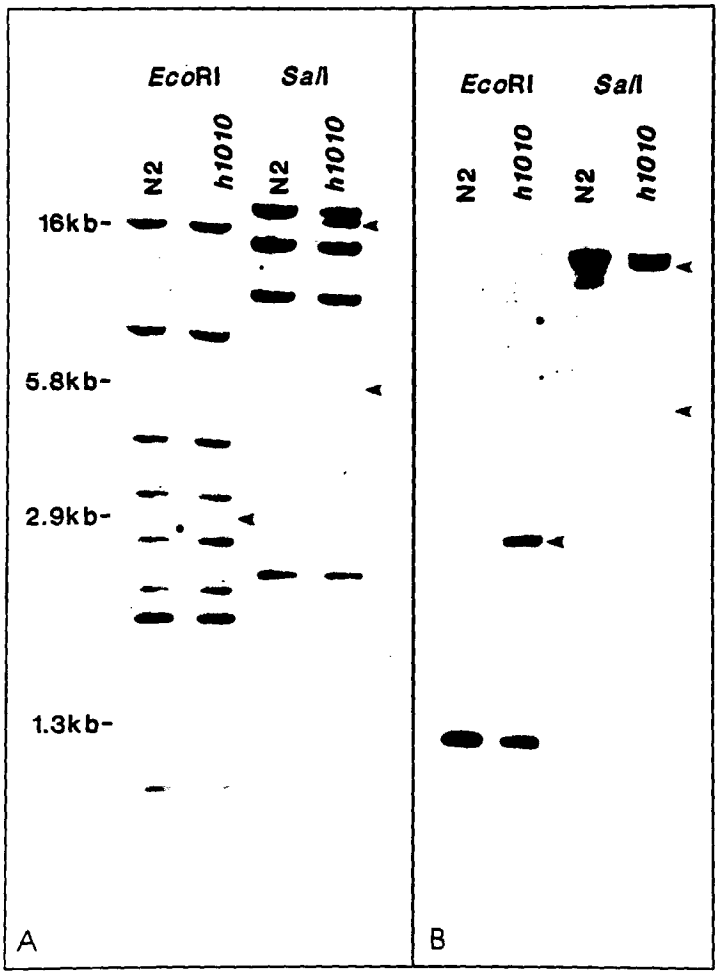
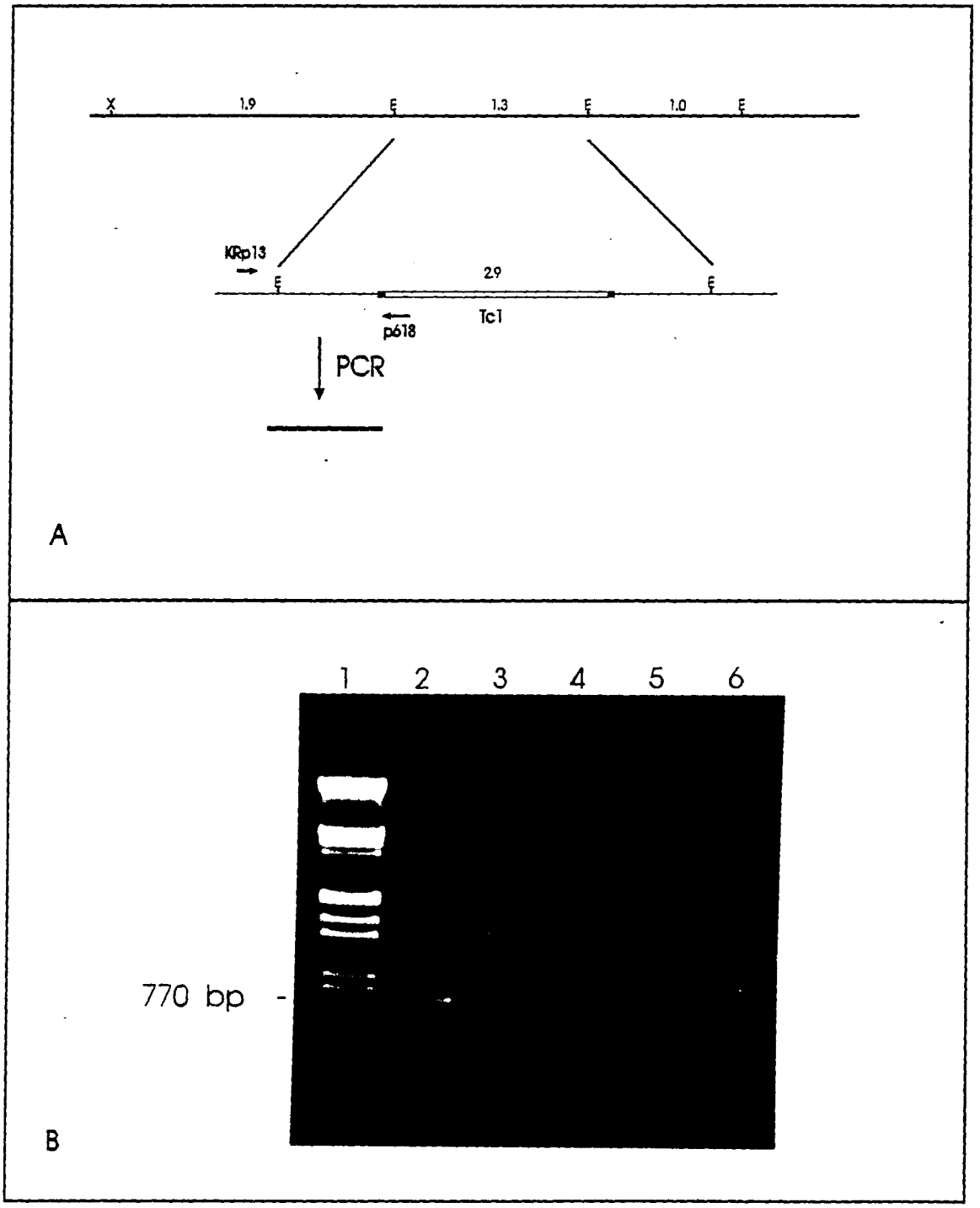


Figure 12. Southern analysis of KR1858 *bli-4(h1010)*

A. Genomic DNA from N2 (wild type) and KR1858, carrying the *h1010* allele *in trans* to a wild type allele, was digested with *Eco*RI or *Sal*I, probed with whole cosmid K04F10. Bands that are altered in size in KR1858 are marked by arrows.
B. Genomic DNA from the same strains digested by the same enzymes as in (A) probed with the 1.3 kb *Eco*RI fragment of K04F10, designated the *1.3 kb h1010 probe*.

Figure 13. Amplification of the *h1010* Tc1 insertion site

Polymerase chain reaction amplification of the Tc1 insertion site in *h1010*. A. Restriction map of the genomic region cloned in the cosmid K04F10 spanning the insertion site, and the primers used to amplify the insertion point. B. Agarose gel analysis of amplification products. Lane (1) lamda *Hind*III/*Eco*RI marker; lane (2) Amplification using primers KRp13 and p618 with 100 ng of KR1858 DNA as template; lanes (3) to (6), Controls, same as lane (2) except as follows: lane (3) No template; lane (4) No primers. lane (5) Primer p618 only. lane (6) Primer KRp13 only.


X
1.9
E
1.3
E
1.0
E
KRp13
E
2.9
E
p618
Tc1
PCR
A
1 2 3 4 5 6
770 bp
B

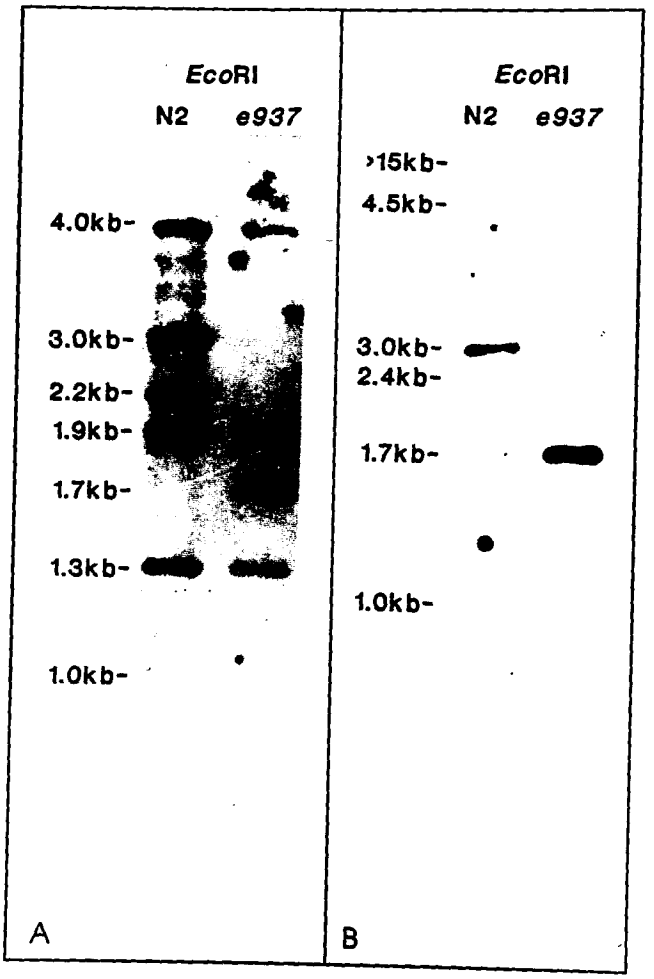


Figure 14. Southern analysis of CB937 *bli-4*(*e937*)
A. Genomic DNA from N2 (wild type) and CB937, carrying the *e937* allele, was digested with *Eco*RI and probed with pCeh181.
B. Genomic DNA from the same strain as in (A) digested with *Eco*RI and probed with a 1.7 kb fusion band in the plasmid pCeh206 isolated from a CB937 genomic library.

Figure 15. Northern analysis of *bli-4* transcripts
Mixed stage total RNA was probed with: A. The 1.9 kb *Eco*RI of pCeh181. B. The 1.3 kb *Eco*RI fragment of pCeh181 flanking the Tc1 insertion site of *h1010*. C. The 1.0 kb *Eco*RI of pCeh181. D. The 0.8 kb *Eco*RI/*Sal*I fragment of pCeh181 containing genomic DNA encoding the unique 3′ end of blisterin A. E. 211 bp *Eco*RI fragment from the unique 3′ end of blisterin B. F. 204 bp *Bam*HI/*Bgl*II fragment from the unique 3′ end of blisterin C.

3.2.2 Analysis of *bli-4* cDNA clones

3.2.2.1 Isolation and hybridization analysis

Six *bli-4* cDNA clones were isolated from the cDNA library of Barstead and Waterston (1989) using the *1.3 kb h1010 probe* (Section 3.2.1.2.1) to screen approximately 40,000 phage plaques. To determine the approximate extent of the *bli-4* locus, the cDNA clones were used as probes in a Southern analysis of the genomic region around *bli-4* contained in the cosmid K04F10 (Figure 16a). cDNA clones hybridized to restriction fragments of K04F10 corresponding to about 15 kb of genomic DNA. As shown in Figure 16a the hybridization pattern was discontinuous. Three of the cDNA clones were chosen for further characterization. These cDNA clones and their predicted products will be referred to as blisterin A (pCeh200), blisterin B (pCeh197) and blisterin C (pCeh196).

3.2.2.2 Sequence analysis of cDNA clones

The following sections present sequence data derived from the cDNA clones blisterin A, blisterin B, and blisterin C. The three cDNA clones share overlapping sequence at the 5′ end, but differ at the 3′ end. The overlapping sequence will be referred to as the "common" sequence. The different 3′ ends of the cDNA clones will be referred to as "unique" sequences. The point at which the 3′ sequence of the three cDNA cones diverge from one another will be referred to as the "divergence point" (Figure 16b).

Blisterin B is the only one of the three cDNA clones that appears to include a complete open reading frame (section 3.2.2.2.1). No evidence exists to support the conclusion that any of the cDNA clones are complete cDNAs. Blisterin C and blisterin A begin abruptly within the blisterin B open reading frame at points indicated in Figure 18. Blisterin A and blisterin C do

not contain open reading frames begining with an ATG start codon. Instead, they begin abruptly within the blisterin B open reading frame at points indicated in Figure 18. Therefore, a working hypothesis is that the blisterin A and blisterin C open reading frames are not complete at the 5′ end. Blisterin A and blisterin C are identical to blisterin B from their repective 5′ ends up to the point where the three cDNA clones diverge from each other. All three cDNA clones terminate in a poly-A track, and are therefore likely to be complete cDNA clones at the 3′ end. Because blisterin A and blisterin C begin within the blisterin B open reading frame, these cDNA clones will be described in relationship to blisterin B. For this reason, I will discuss blisterin B first. Sequencing strategies are presented in Figure 17. A summary of the sequence data is presented in Table 7.

Table 7. Summary of *bli-4* cDNA sequence data

cDNA	Plasmid	Total Length bp	ORF[a]	Start Point in Blisterin B[b]	Stop codon	Protein length
Blisterin A	pCeh200	1459	1281	905	TGA (1284)	427
Blisterin B	pCeh197	2421	2193	0	TGA (2370)	731
Blsiterin C	pCeh196	1956	1713	1295	TAA (1755)	571

cDNA	Divergence Point	Unique Sequence Length	Unique ORF Length	Unique Protein length	5' UTR[c]	3' UTR
Blisterin A	1244	195	39	13	n/a	150
Blisterin B	2149	245	219	73	175	26
Blisterin C	855	1081	858	13	n/a	223

[a]ORF: Open Reading Frame.

[b]Blisterin A and blisterin C contain incomplete open reading frames that begin within the blisterin B ORF.

[c]UTR: Untranslated Region

3.2.2.2.1 Blisterin B

The sequence of the blisterin B cDNA is presented in Figure 18. blisterin B contains an *Eco*RI insert of 2,448 nucleotides. This number includes 27 adenosine residues at the 3′ end of the sequence, indicating that the cDNA is complete at the 3′ end. Blisterin B has two potential ATG methionine initiation codons at the 5′ end. The first ATG, at nucleotide 36, is followed by an in-frame stop codon, TAA, at the fourth codon position, and eight additional stop codons within the next 300 nucleotides. In addition, the first ATG is preceded by a pyrimidine at the -3 position, in violation of the Kozak initiation sequence (Kozak, 1986). The second, ATG, at nucleotide 176, has a purine at the -3 position, consistent with the Kozak sequence. The second ATG begins an open reading frame of 2,193 nucleotides encoding a potential protein of 731 amino acid residues. This reading frame terminates with the nonsense codon TGA beginning at position 2,370. Two in-frame stop codons are present upstream from the second ATG. The second ATG, therefore, is more likely to be a *bona fide* translation initiation site. The open reading frame begining with the second ATG is preceded by a 5′ untranslated region of 175 nucleotides, and is followed by a 3′ untranslated region of 26 nucleotides, excluding the poly-A track. On the basis of this long open reading frame and presence of a poly-A track, I conclude that blisterin B contains a complete open reading frame.

3.2.2.2.2 Blisterin C

The sequence of the blisterin C cDNA is presented in Figure 19. blisterin C contains an *Eco*RI insert of 1,946 nucleotides. This number includes 10 adenosine residues at the 3′ end. The 5′ end of the blisterin C insert begins at nucleotide position 1,295 of blisterin B. The first 835 nucleotides of blisterin C is identical to blisterin B from nucleotide 1,295 to

nucleotide 2,149, the point at which the three cDNA clones diverge from each other. The 1091 nucleotides (excluding the poly-A track) following the divergence point are different from the 3′ end of blisterin B. The open reading frame of blisterin C begins at the third nucleotide at its 5′ end, and is 1,713 nucleotides in length encoding a potential protein fragment of 571 amino acid residues (Figure 19). The blisterin C open reading frame terminates with a TAA stop codon at nucleotide 1,715. The blisterin C open reading frame is followed by an untranslated region of 223 nucleotides, excluding the poly-A track.

3.2.2.2.3 Blisterin A

The sequence of the blisterin A cDNA is presented in Figure 20. The 3′ end of blisterin A terminates in 20 adenosine residues. The 5′ end of the blisterin A insert begins at nucleotide position 905 of blisterin B, and is identical to blisterin B as far as it was sequenced, to blisterin B nucleotide 1,140. The 3′ end sequence is identical to blisterin B starting at blisterin B nucleotide 2,118, and then diverges from the blisterin B sequence at blisterin B nucleotide 2,149. The remaining 215 nucleotides are unique to blisterin A. I will assume that the internal sequence of blisterin A is identical to the portion of blisterin B that it overlaps. This assumption is supported by the estimated size of the blisterin A insert and restriction fragment pattern as determined by agarose gel electrophoresis. The sequence of blisterin A is identical to blisterin B from the blisterin A 5′ end to nucleotide 1244. blisterin A diverges from the blisterin B sequence at blisterin B nucleotide 2,149. The 195 nucleotides (excluding the poly-A track) following the divergence point are different from the 3′ end of both blisterin B and blisterin C. The open reading frame of blisterin A begins at the fourth nucleotide at its 5′ end, and is 1,1281 nucleotides in length, encoding a

potential protein fragment of 427 amino acid residues (Figure 20). The blisterin A open reading frame terminates with a TGA stop codon at nucleotide 1,284. The blisterin A open reading frame is followed by a 3′ untranslated region of 150 nucleotides, excluding the poly-A track.

I conclude from the cDNA sequence data that *bli-4* encodes at least three different gene products generated by alternative splicing. This conclusion is supported by two observations. First, the sequences of the three cDNA clones differ starting at the same point. Second, the point at which the three cDNA sequences diverge is identical in all three clones and corresponds to an Exon/Intron junction (Section 3.2.3).

3.2.2.3 Northern Analysis

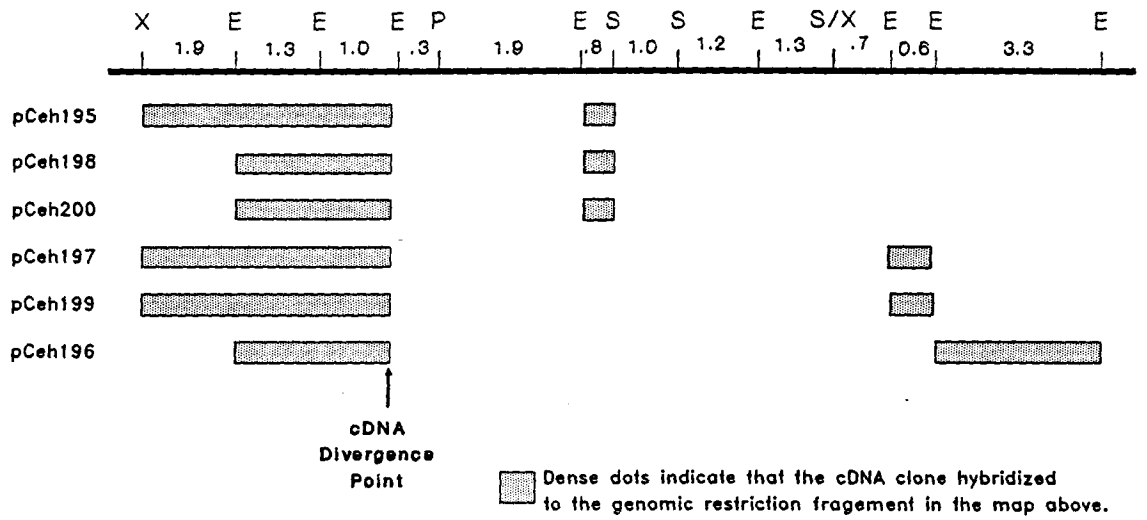
The 1.9 kb, 1.3 kb, and 1.0 kb *Eco*RI fragments of the genomic subclone pCeh181 (Figure 16), encoding the 5′ common sequence of the blisterin cDNA clones, each detected four bands on a northern blot of *C. elegans* mixed stage total RNA (Section 3.2.1.3, Figure 15). These bands were 3.5, 3.1, 2.6. and 1.6 kb in size. To correlate these bands with the cDNA clones, northern blots of N2 RNA were probed with DNA fragments predicted to be unique to each of blisterin A, blisterin B and blisterin C (Figure 15). The probes used were as follows. To obtain a probe to detect blisterin A, the 0.8 kb *Eco*RI genomic DNA fragment encoding the unique 3′ end of blisterin A (Section 3.2.3) was subcloned into plasmid pCeh205. To obtain a probe to detect blisterin B, a 211 bp *Eco*RI fragment from the 3′ unique end of blisterin B was isolated by electroelution from an *Eco*RI digest of the blisterin B plasmid. To obtain a probe to detect blisterin C, a 204 bp *Eco*RI fragment from the 3′ unique end of blisterin C was isolated by electroelution from an *Eco*RI digest of the blisterin C plasmid.

Of the three probes used, only the blisterin A probe detected a unique band on the northern blot, at 3.1 kb. The blisterin A cDNA clone insert is 1.4 kb in length. The fact that the RNA band detected by the blisterin A probe indicates that blisterin A is not a complete cDNA. The blisterin B and blisterin C probes failed to detect unique bands. Both probes detected bands at 3.5, 3.1 and 1.6 kb. The blisterin B probe detected the 3.5 most intensely, while the blisterin C probe detected the 3.1 band most intensely. The failure of these probes to detect unique RNA bands may indicate that they are contaminated by other fragments due to the electroelution procedure used to isolate the probes. Alternatively, the prediction that the blisterin B and blisterin C probes would detect unique RNA bands could be incorrect.

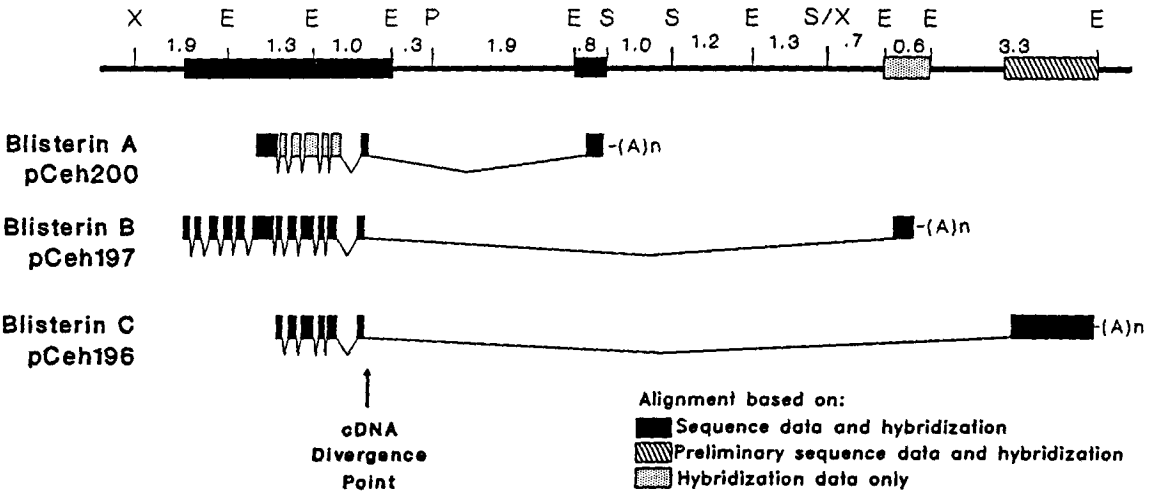
None of the three cDNA probes detected the major band of 2.6 kb detected by the genomic fragments encoding the 5′ common sequences. The 2.6 kb band may, therefore, represent a major transcript class not included in the cDNA clones that have been characterized.

Figure 16. Alignment of cDNA clones with the *bli-4* region
A restriction map of the portion of K04F10 containing the *bli-4* locus is shown at the top of each panel. *bli-4* cDNA clones are shown in the bottom portion of each panel.
A. An alignment of *bli-4* cDNA clones based on hybridization data is presented. Stippling indicates that the cDNA clone hybridized to the corresponding genomic restriction fragment in the restriction map immediately above. No conclusions about the positions of the coding regions within the restriction fragments is implied.
B. An alignment of *bli-4* cDNA clones based on both sequence (Section 3.2.3) and hybridization data. Exons are represented by boxes in the cDNA clones; introns are represented by lines. Exons and introns are indicated in the cDNA clones only in the regions of the genomic DNA that have been sequenced completely. The number and size of exons encoding the unique 3′ ends of cDNA clones encoded by cDNA regions that have not been completely sequenced (blisterin B and blisterin C) is not implied.

The genomic restriction map was constructed as follows. First, a *Xho*I *Sal*I restriction map of the region was made directly from the cosmid K04F10. Second, the 11 and 11.6 kb *Xho*I fragments were subcloned and further restriction mapped. The 11 kb *Xho*I fragment in pCeh181 was further subcloned as two *Xho*I *Pst*I fragments, and these fragments were independently restriction mapped. Ambiguities in the map were resolved by Southern analysis. X=*Xho*I; P=*Pst*I; S=*Sal*I; E=*Eco*RI.



A. Alignment of cDNA clones based on hybridization data.



B. Alignment of cDNA clones based on sequence data.

Figure 17. Sequencing strategies

Arrows indicate individual sequence reactions. # indicates that the sequencing reaction used a custom primer with dye-labeled di-deoxy terminators. All other reactions used dye-labeled forward or reverse M13 primers on templates generated by Exonuclease III deletions (Henikoff, 1984).

A. Genomic DNA clones. The upper portion is a restriction map of the *bli-4* region. Heavy black lines indicated coding regions determined by hybridization and sequencing. The DNA encoding the 5′ ends of the blisterins is labeled. A, B, and C indicate the DNA encoding the unique 3′ ends of blisterin A, blisterin B and blisterin C.

B. cDNA clones. The cDNA divergence point is indicated.

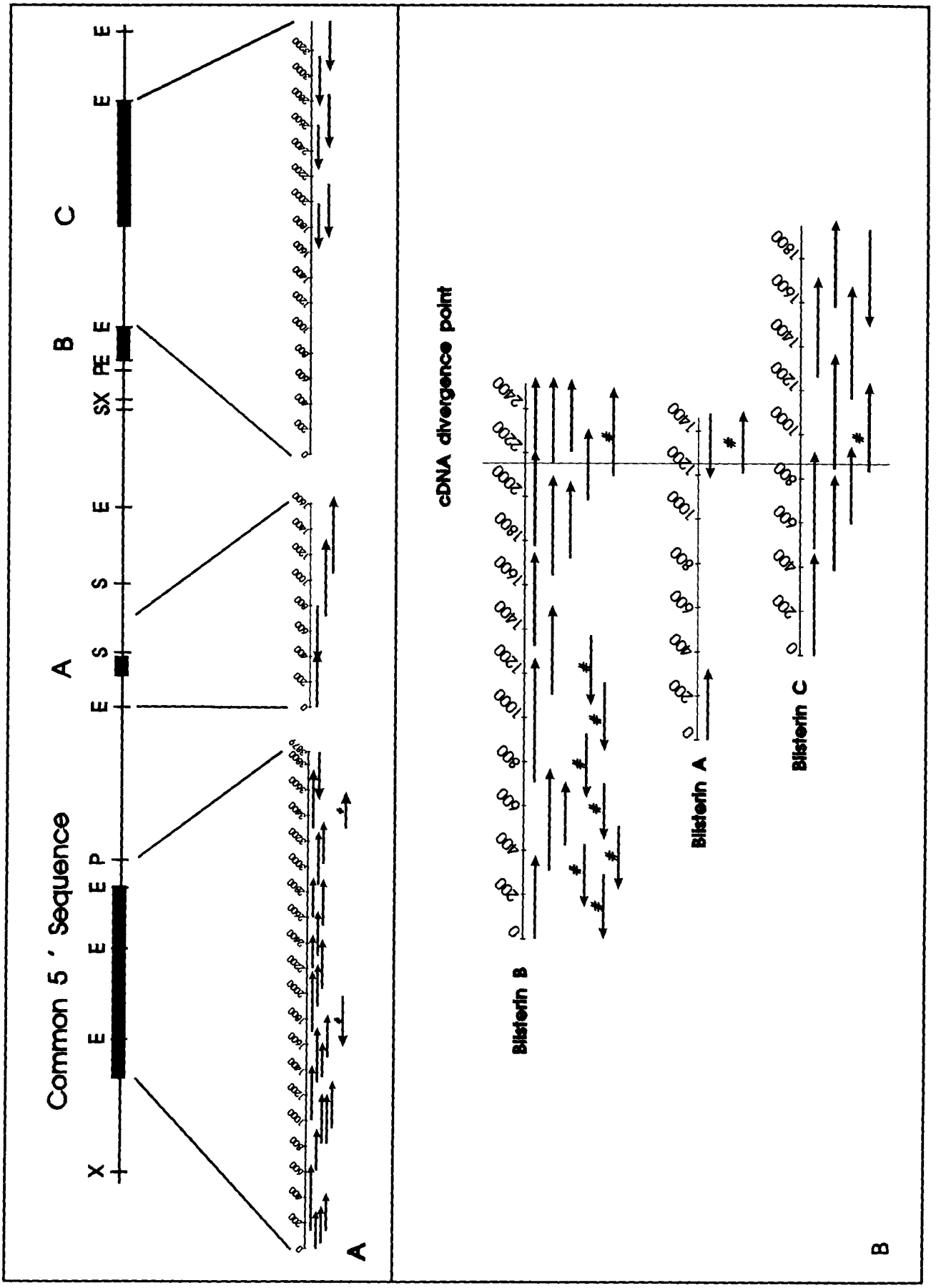

Common 5 ' Sequence
X
E
E
E
P
E
S
S
E
A
B
SX
PE
E
C
E
E
A
cDNA divergence point
Blisterin B
Blisterin A
Blisterin C
B

Figure 18. Blisterin B.

Sequence of blisterin B and predicted protein product. Points at which blisterin A and blisterin C begin and the cDNA divergence points are indicated. The predicted secretory signal peptide is underlined. The protease domain is double underlined. The catalytic amino acids, Asp, His and Ser are indicated by an up arrow (↑). Potential autocatalytic sites are indicated by an upward arrowhead (▲).

```
gcttgccagtaaaaaagtacgttcgcttcgggtcgatgtcagaataaacggaaaagaataacccgttgcaccagcgaatcgtcgaacattttcaatactc 100

acccatatcagtcacccaaaacggatttattattattattagagcatatcatccacatttattctcaaacgcgtgATGCGTATATCGATAGGCCGGATAG 200
                                                                          M  R  I  S  I  G  R  I    8

CATGGCAAATTCTGGCAGTTTTAATCGCAGTTGCATTCACTATTGAACATGATTCCATTTGCGATGAAAGTATAGGTGCCTGTGGGGAACCAATACATAC 300
A  W  Q  I  L  A  V  L  I  A  V  A  F  T  I  E  H  D  S  I  C  D  E  S  I  G  A  C  G  E  P  I  H  T   42

CGTAATACGTTTAGCAAAAAGAGATGATGAGCTTGCACGGCGAATAGCTGCTGATCATGACATGCATGTAAAAGGTGATCCGTTTTTGGATACTCACTAC 400
 V  I  R  L  A  K  R  D  D  E  L  A  R  R  I  A  A  D  H  D  M  H  V  K  G  D  P  F  L  D  T  H  Y   75
                  ▲                    ▲
TTCCTTTATCACTCGGAAACAACAAGGACACGGCGACATAAAAGAGCGATTGTTGAACGATTGGATTCACATCCAGCCGTCGAATGGGTTGAAGAACAGC 500
 F  L  Y  H  S  E  T  T  R  T  R  R  H  K  R  A  I  V  E  R  L  D  S  H  P  A  V  E  W  V  E  E  Q   108
                        ▲
GACCGAAGAAGAGAGTCAAAAGAGATTATATTCTCCTGGATAATGATGTTCATCATTCTAACCCCTTCCGCCGTTCGGTTTTGAACCGTGATGGTACTCG 600
R  P  K  K  R  V  K  R  D  Y  I  L  L  D  N  D  V  H  H  S  N  P  F  R  R  S  V  L  N  R  D  G  T  R  142
         ▲  ▲        ▲                                          ▲
TAGAGCTCAACGACAGCAGCCACAGTCTCCAGCAGAAATTCCATCACTTCCATTTCCTGATCCACTTTATAAAGACCAGTGGTATTTGCATGGTGGAGCA 700
 R  A  Q  R  Q  Q  P  Q  S  P  A  E  I  P  S  L  P  F  P  D  P  L  Y  K  D  Q  W  Y  L  H  G  G  A   175

GTTGGTGGATATGATATGAATGTTCGTCAAGCGTGGCTTCAAGGATATGCAGGCAGAAATGTTTCAGTTTCGATTCTTGACGATGGAATTCAAAGAGATC 800
 V  G  G  Y  D  M  N  V  R  Q  A  W  L  Q  G  Y  A  G  R  N  V  S  V  S  I  L  D  D  G  I  Q  R  D   208
                                                                          ↑
ATCCTGATTTGGCAGCGAACTACGATCCACTCGCGTCAACAGATATCAATGATCACGATGATGATCCAACACCACAAAACAATGGAGATAACAAACATGG 900
H  P  D  L  A  A  N  Y  D  P  L  A  S  T  D  I  N  D  H  D  D  D  P  T  P  Q  N  N  G  D  N  K  H  G   242
      ┌▸ Blisterin A  5' end                                                                        ↑
TACAAGATGTGCTGGAGAAGTAGCTGCACTCGCCGGAAACAATCAATGTGGTGTCGGTGTTGCCTTCAAAGCAAAAATAGGAGGAGTTCGTATGCTTGAT 1000
 T  R  C  A  G  E  V  A  A  L  A  G  N  N  Q  C  G  V  G  V  A  F  K  A  K  I  G  G  V  R  M  L  D   275

GGAGCTGTTTCGGATTCTGTCGAAGCTGCGTCATTGTCTTTGAATCAAGATCATATTGATATATACTCAGCATCATGGGGACCTGAAGACGACGGAAAGA 1100
G  A  V  S  D  S  V  E  A  A  S  L  S  L  N  Q  D  H  I  D  I  Y  S  A  S  W  G  P  E  D  D  G  K   308

CTTTTGATGGTCCAGGTCCACTTGCCCGAGAAGCTTTTTATCGTGGAATCAAGAATGGCAGAGGTGGAAAAGGAAACATTTTTGTATGGGCCAGTGGAAA 1200
T  F  D  G  P  G  P  L  A  R  E  A  F  Y  R  G  I  K  N  G  R  G  G  K  G  N  I  F  V  W  A  S  G  N   342
                                                                          ┌▸ Blisterin C
CGGTGGATCAAGCCAAGACTCATGTTCAGCTGATGGTTACACAACTTCAGTCTACACGCTTTCCATATCTTCGGCTACTTATGATAATCACAGACCATGG 1300
 G  G  S  S  Q  D  S  C  S  A  D  G  Y  T  T  S  V  Y  T  L  S  I  S  S  A  T  Y  D  N  H  R  P  W   375
5' end
TATTTGGAAGAATGCCCATCATCAATTGCAACAACATACAGTTCTGCGGATTTCCGTCAACCAGCAATTGTGACCGTTGATGTTCCTGGAGGATGTACTG 1400
Y  L  E  E  C  P  S  S  I  A  T  T  Y  S  S  A  D  F  R  Q  P  A  I  V  T  V  D  V  P  G  G  C  T   408
```

Figure 18. Blisterin B. Continued

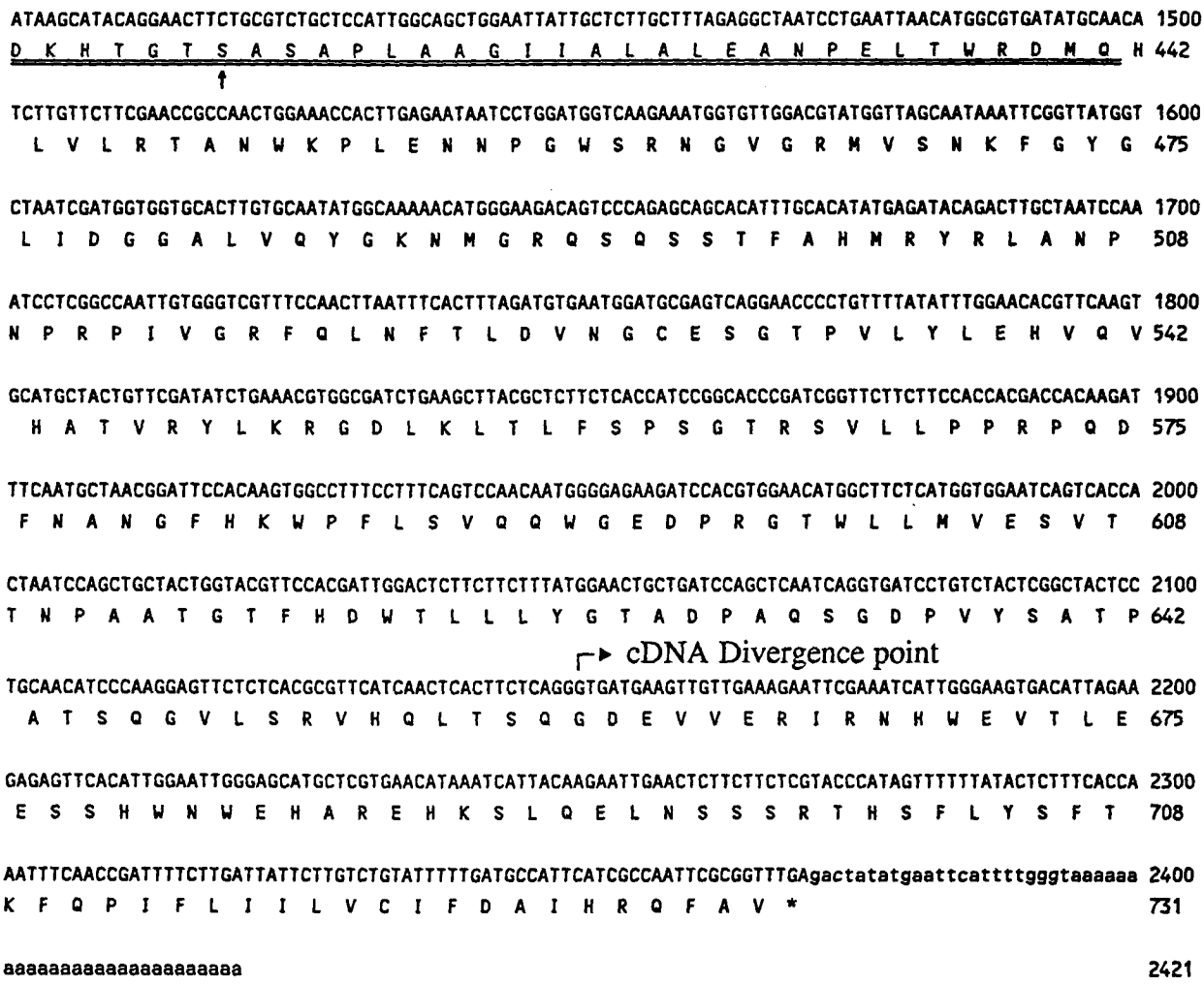


Figure 19. Blisterin C.
Sequence of blisterin C and predicted protein product. The cDNA divergence point is indicated. The predicted transmembrane domain is underlined in bold. The protease domain is double underlined. The catalytic amino acid, Ser, is indicated by an up arrow (↑).

```
CCATGGTATTTGGAAGAATGCCCATCATCAATTGCAACAACATACAGTTCTGCGGATTTCCGTCAACCAGCAATTGTGACCGTTGATGTTCCTGGAGGAT 100
 P  W  Y  L  E  E  C  P  S  S  I  A  T  T  Y  S  S  A  D  F  R  Q  P  A  I  V  T  V  D  V  P  G  G  33

GTACTGATAAGCATACAGGAACTTCTGCGTCTGCTCCATTGGCAGCTGGAATTATTGCTCTTGCTTTAGAGGCTAATCCTGAATTAACATGGCGTGATAT 200
 C  T  D  K  H  T  G  T  S  A  S  A  P  L  A  A  G  I  I  A  L  A  L  E  A  N  P  E  L  T  W  R  D  M  67
                          ↑

GCAACATCTTGTTCTTCGAACCGCCAACTGGAAACCACTTGAGAATAATCCTGGATGGTCAAGAAATGGTGTTGGACGTATGGTTAGCAATAAATTCGGT 300
 Q  H  L  V  L  R  T  A  N  W  K  P  L  E  N  N  P  G  W  S  R  N  G  V  G  R  M  V  S  N  K  F  G  100

TATGGTCTAATCGATGGTGGTGCACTTGTGCAATATGGCAAAAACATGGGAAGACAGTCCCAGAGCAGCACATTTGCACATATGAGATACAGACTTGCTA 400
 Y  G  L  I  D  G  G  A  L  V  Q  Y  G  K  N  M  G  R  Q  S  Q  S  S  T  F  A  H  M  R  Y  R  L  A  133

ATCCAAATCCTCGGCCAATTGTGGGTCGTTTCCAACTTAATTTCACTTTAGATGTGAATGGATGCGAGTCAGGAACCCCTGTTTTATATTTGGAACACGT 500
 N  P  N  P  R  P  I  V  G  R  F  Q  L  N  F  T  L  D  V  N  G  C  E  S  G  T  P  V  L  Y  L  E  H  V 167

TCAAGTGCATGCTACTGTTCGATATCTGAAACGTGGCGATCTGAAGCTTACGCTCTTCTCACCATCCGGCACCCGATCGGTTCTTCTTCCACCACGACCA 600
  Q  V  H  A  T  V  R  Y  L  K  R  G  D  L  K  L  T  L  F  S  P  S  G  T  R  S  V  L  L  P  P  R  P  200

CAAGATTTCAATGCTAACGGATTCCACAAGTGGCCTTTCCTTTCAGTCCAACAATGGGGAGAAGATCCACGTGGAACATGGCTTCTCATGGTGGAATCAG 700
 Q  D  F  N  A  N  G  F  H  K  W  P  F  L  S  V  Q  Q  W  G  E  D  P  R  G  T  W  L  L  M  V  E  S  233

TCACCACTAATCCAGCTGCTACTGGTACGTTCCACGATTGGACTCTTCTTCTTTATGGAACTGCTGATCCAGCTCAATCAGGTGATCCTGTCTACTCGGC 800
 V  T  T  N  P  A  A  T  G  T  F  H  D  W  T  L  L  L  Y  G  T  A  D  P  A  Q  S  G  D  P  V  Y  S  A 267
                                                                  ⌐► cDNA Dvergence point
TACTCCTGCAACATCCCAAGGAGTTCTCTCACGCGTTCATCAACTCACTTCTCAGGTTGAAGAGTCTGCTCGATCTTCATTTCCAGATTTGACGTCCGGC 900
  T  P  A  T  S  Q  G  V  L  S  R  V  H  Q  L  T  S  Q  V  E  E  S  A  R  S  S  F  P  D  L  T  S  G  300

TGGAAATTGTCATGTGATGAATGTAATGGAGGTTGCACAGAATCGAGCTCAGCAACATCATGTTTCGCTTGCAAACATTTAACTCAAACTCTTCGAAATA 1000
 W  K  L  S  C  D  E  C  N  G  G  C  T  E  S  S  S  A  T  S  C  F  A  C  K  H  L  T  Q  T  L  R  N  333

AAGGAGGAAGTGGATTCAAGCGTGTTCAAAAATGCGATGATACTTACTACTTGGATGGTGACAAGTGTAAAATGTGCTCATCCCATTGTCATACTTGTAC 1100
 K  G  G  S  G  F  K  R  V  Q  K  C  D  D  T  Y  Y  L  D  G  D  K  C  K  M  C  S  S  H  C  H  T  C  T 367

GAAAGCTGAGGCGTGTGAAACGTGTCCTGGAAGTCTTTTGCTCATTGATGTGGATAATATGCCACACTATGATCATGGAAAGTGTGTTGAGTCATGTCCT 1200
  K  A  E  A  C  E  T  C  P  G  S  L  L  L  I  D  V  D  N  M  P  H  Y  D  H  G  K  C  V  E  S  C  P  400

CCAGGATTGGTTGCTGATTATGAATCGAATCTTGTTCAAGCTAAATGTATCTGGAGAAAAGATCTTTGTGGTGACGGATATTACATCAACGCTGTTGGAA 1300
 P  G  L  V  A  D  Y  E  S  N  L  V  Q  A  K  C  I  W  R  K  D  L  C  G  D  G  Y  Y  I  N  A  V  G  433

AATGTGATCTTTGCGACTCCTCATGCGAAACTTGCACCGCTCCTGGTCCAATGAGCTGTGAGAAATGTTCAAAAGGTTATGGTAAAGGATCGATTGGATA 1400
 K  C  D  L  C  D  S  S  C  E  T  C  T  A  P  G  P  M  S  C  E  K  C  S  K  G  Y  G  K  G  S  I  G  Y 467
```

Figure 19 Blisterin C continued.

```
CTGTAGACCGTGTTGTCCTGAAGGATCAACAAAGAGTTGGCAATGTGAGGACTGCTCCAAGCCGGATCCTACACTTTTGATTGATTCTAATAAATCATCT 1500
  C  R  P  C  C  P  E  G  S  T  K  S  W  Q  C  E  D  C  S  K  P  D  P  T  L  L  I  D  S  N  K  S  S  500

GGATTTGGATTGATGTTCTGGATTGTAGTTAGTTTGATTGCGGCTTGTGGAATCTGTGCCTGTAAAAAGTGTGCAAGTGAGACGAAAAGCTCAAACGTAG 1600
  G  F  G  L  M  F  W  I  V  V  S  L  I  A  A  C  G  I  C  A  C  K  K  C  A  S  E  T  K  S  S  N  V  533

AATATGCGCCGCTTGCTCAATATAATGCCACAAATGGTGCTATCAATTTAGGAGCACACACTGACGATGAAGACGATGATGAGGATGAAGTATTTGTGAA 1700
E  Y  A  P  L  A  Q  Y  N  A  T  N  G  A  I  N  L  G  A  H  T  D  D  E  D  D  D  E  D  E  V  F  V  N 567

CCCTCAAATTGTTTAAaccaaaacctttcaaaattcatgtttttaattgtaatttttctgccaacttctttgtgtatgcttctgaaatccgggtaattgtt 1800
  P  Q  I  V  *                                                                                       571

tctttttcaaaatttaaaacctgaaacgttttcaagtggccttttatcatgtgattgtattgtttcttttgtcttataccgggtttatttttattatac 1900

cgtactgtttccatgttataaatagattgttgaattaaaaaaaaaa                                                      1946
```

Figure 20. Blisterin A.

Sequence of blisterin A and predicted protein product. The cDNA divergence point is indicated. The protease domain is double underlined. The catalytic amino acid Ser is indicated by an up arrow (↑).

```
GATGTGCTGGAGAAGTAGCTGCACTCGCCGGAAACAATCAATGTGGTGTCGGTGTTGCCTTCAAAGCAAAAATAGGAGGAGTTCGTATGCTTGATGGAGC 100
  C  A  G  E  V  A  A  L  A  G  N  N  Q  C  G  V  G  V  A  F  K  A  K  I  G  G  V  R  M  L  D  G  A  33

TGTTTCGGATTCTGTCGAAGCTGCGTCATTGTCTTTGAATCAAGATCATATTGATATATACTCAGCATCATGGGGACCTGAAGACGACGGAAAGACTTTT 200
  V  S  D  S  V  E  A  A  S  L  S  L  N  Q  D  H  I  D  I  Y  S  A  S  W  G  P  E  D  D  G  K  T  F  66

GATGGTCCAGGTCCACTTGCCCGAGAAGCTTTTTAtcgtggaatcaagaatggcagaggtggaaaaggaaacatttttgtatgggccagtggaaacggtg 300
  D  G  P  G  P  L  A  R  E  A  F  Y  R  G  I  K  N  G  R  G  G  K  G  N  I  F  V  W  A  S  G  N  G   99

gatcaagccaagactcatgttcagctgatggttacacaacttcagtctacacgctttccatatcttcggctacttatgataatcacagaccatggtattt 400
  G  S  S  Q  D  S  C  S  A  D  G  Y  T  T  S  V  Y  T  L  S  I  S  S  A  T  Y  D  N  H  R  P  W  Y  L 133

ggaagaatgcccatcatcaattgcaacaacatacagttctgcggatttccgtcaaccagcaattgtgaccgttgatgttcctggaggatgtactgataag 500
  E  E  C  P  S  S  I  A  T  T  Y  S  S  A  D  F  R  Q  P  A  I  V  T  V  D  V  P  G  G  C  T  D  K  166

catacaggaacttctgcgtctgctccattggcagctggaattattgctcttgctttagaggctaatcctgaattaacatggcgtgatatgcaacatcttg 600
  H  T  G  T  S  A  S  A  P  L  A  A  G  I  I  A  L  A  L  E  A  N  P  E  L  T  W  R  D  M  Q  H  L  199
                 ↑
ttcttcgaaccgccaactggaaaccacttgagaataatcctggatggtcaagaaatggtgttggacgtatggttagcaataaattcggttatggtctaat 700
V  L  R  T  A  N  W  K  P  L  E  N  N  P  G  W  S  R  N  G  V  G  R  M  V  S  N  K  F  G  Y  G  L  I 233

cgatggtggtgcacttgtgcaatatggcaaaaacatgggaagacagtcccagagcagcacatttgcacatatgagatacagacttgctaatccaaatcct 800
  D  G  G  A  L  V  Q  Y  G  K  N  M  G  R  Q  S  Q  S  S  T  F  A  H  M  R  Y  R  L  A  N  P  N  P  266

cggccaattgtgggtcgtttccaacttaatttcactttagatgtgaatggatgcgagtcaggaacccctgttttatatttggaacacgttcaagtgcatg 900
  R  P  I  V  G  R  F  Q  L  N  F  T  L  D  V  N  G  C  E  S  G  T  P  V  L  Y  L  E  H  V  Q  V  H  299

ctactgttcgatatctgaaacgtggcgatctgaagcttacgctcttctcaccatccggcacccgatcggttcttcttccaccacgaccacaagatttcaa 1000
A  T  V  R  Y  L  K  R  G  D  L  K  L  T  L  F  S  P  S  G  T  R  S  V  L  L  P  P  R  P  Q  D  F  N 333

tgctaacggattccacaagtggcctttccttcagtccaacaatggggagaagatccacgtggaacatggcttctcatggtggaatcagtcaccactaat 1100
  A  N  G  F  H  K  W  P  F  L  S  V  Q  Q  W  G  E  D  P  R  G  T  W  L  L  M  V  E  S  V  T  T  N  366

ccagctgctactggtacgttccacgattggactcttcttctttatggaactgctgatccagctcaatcaggtgatcctgtctactcggctactcctgcaa 1200
  P  A  A  T  G  T  F  H  D  W  T  L  L  L  Y  G  T  A  D  P  A  Q  S  G  D  P  V  Y  S  A  T  P  A  399
                                                  ┌► cDNA Divergence point
catcccaaggagTTCTCTCACGCGTTCATCAACTCACTTCTCAGATACTAATCACCATCGCTATCCACCTCGTCGTCAACGCATGAattatttttctgttt 1300
T  S  Q  G  V  L  S  R  V  H  Q  L  T  S  Q  I  L  I  T  I  A  I  H  L  V  V  N  A  *              427

cactgctcacagcttcatagagcttaaaattagatttattgtccctttatccttggttacgatgtgttctgaaacttgtctccccatttttctgtgtcatt 1400

ttccctttgacctcactatgatgaatacttttttacgagaaaaaaaaaaaaaaaaaaaa                                          1459
```

3.2.3 Analysis of *bli-4* Genomic DNA preliminary sequence data

Preliminary sequence data for the genomic DNA including exons common to all three cDNA clones was derived from the plasmid pCeh202, and is presented in Figure 21. This sequence includes exons encoding all of the cDNA sequences 5′ to the cDNA divergence point. The relationship between the genomic sequence and the cDNA clones is presented in Figure 16b. The ATG start codon of the blisterin B open reading frame occurs at nucleotide position 374. The cDNA divergence point occurs at the 3′ end of Exon XII and coincides with an exon/intron boundary. This observation supports the hypothesis that the unique 3′ sequences of the cDNA clones arose through alternative splicing.

Genomic DNA sequence corresponding to the unique 3′ end of blisterin A was derived from pCeh205, and is presented in Figure 22. All of the blisterin A unique 3′ sequence occurs in a single exon. This exon is deleted by the *e937* deletion (Section 3.2.5.2).

Genomic DNA sequence corresponding to the unique 3′ end of blisterin C was derived from pCeh207, and is presented in Figure 23. pCeh207 includes genomic DNA encoding all of the unique 3′ sequence of blisterin C. The blisterin C sequence is encoded by at least five exons. The sequence labeled "Exon XV" in Figure 22 falls into a gap in the pCeh207 sequence data, and therefore may be encoded by more than one exon.

In combination with the hybridization of cDNA clones to genomic DNA presented in section 3.2.2.1 and summarized in Figure 16, the genomic sequence data presented here supports the conclusion that the genomic DNA encoding the three cDNA clones is entirely contained within the cosmid K04F10.

Figure 21. Blisterin common genomic sequence.

DNA sequence encoding the 5' ends of blisterin cDNA clones. Introns are lower case; exons are upper case.

```
tgcttattccccgaatgaaactttcccagcttagtagtgcctccgatagccaattttaaacaaatattctgaaattattaatttttcttgtttttcgcga 100
                                                                    ┌170 Exon I
aaaaactgtcattttctttattattctctttttaaccgctttccaaattaattcggctatttttaaagCGTTCGCTTCGGGTCGATGTCAGAATAAACG 200
                                                                                                  ┌298 Intron I
GAAAAGAATAACCCGTTGCACCAGCGAATCGTCGAACATTTTCAATACTCACCCATATCAGTCACCCAAAACGGATTTATTATTATTATTAGAGCATgtg 300
                                                  ┌346 Exon II                ┌Start of blisterin B ORF
agttcgttgaatttgtgaatttattctataaatctacaatttcagATCATCCACATTTATTCTCAAACGCGTGATGCGTATATCGATAGGCCGGATAGCA 400
                                                        ┌450 Intron II                                  ┌495 Exon III
TGGCAAATTCTGGCAGTTTTAATCGCAGTTGCATTCACTATTGAACATGgttccgtaagttatttcaatcacntctaaaattcntattatacttATTCCA 500

TTTGCGATGAAAGTATAGGTGCCTGTGGGGAACCAATACATACCGTAATACGTTTAGCAAAAAGAGATGATGAGCTTGCACGGCGAATAGCTGCTGATCA 600

TGACATGCATGTAAAAGGTGATCCGTTTTTGGATACTCACTACTTCCTTTATCACTCGGAAACAACAAGGACACGGCGACATAAAAGAGCGATTGTTGAA 700
                                                                                                ┌788 Intron III
CGATTGGATTCACATCCAGCCGTCGAATGGGTTGAAGAACAGCGACCGAAGAAGAGAGTCAAAAGAGATTATATTCTCCTGGATAATgttagttttttag 800
                                                                                                      ┌898
aaatattcaacctacatggatatttatccagaatcattaccgtttttctttttatcttagctttagagtttttttcctgacaattaatatttttccagGAT 900
Exon IV
GTTCATCATTCTAACCCCTTCCGCCGTTCGGTTTTGAACCGTGATGGTACTCGTAGAGCTCAACGACAGCAGCCACAGTCTCCACGAGAAATTCCATCAC 1000
                                                  ┌1041 Intron IV
TTCCATTTCCTGATCCACTTTATAAAGACCAGTGGTATTTGgtgagtttcaaattcaaattttctttaaagagaaaaaaaaaccactggatacatttaca 1100
 ┌1102 Exon V
gCATGGTGGAGCAGTTGGTGGATATGATATGAATGTTCGTCAAGCGTGGCTTCAAGGATATGCAGGCAGAAATGTTTCAGTTTCGATTCTTGACGATGGA 1200

ATTCAAAGAGATCATCCTGATTTGGCAGCGAACTACGATCCACTCGCGTCAACAGATATCAATGATCACGATGATGATCCAACACCACAAAACAATGGAG 1300
       ┌1308 Intron V                                                                          ┌1390 Exon VI
ATAACAAgtaaggatttttaaacacaagttcttcagaaatttttatggncattaantgttttttgggccaaggtacaagnntgtntccgACATGGTACAA 1400

GATGTGCTGGAGAAGTAGCTGCACTCGCCGGAAACAATCAATGTGGTGTCGGTGTTGCCTTCAAAGCAAAAATAGGAGGAGTTCGTATGCTTGATGGAGC 1500

TGTTTCGGATTCTGTCGAAGCTGCGTCATTGTCTTTGAATCAAGATCATATTGATATATACTCAGCATCATGGGGACCTGAAGACGACGGAAAGACTTTT 1600

GATGGTCCAGGTCCACTTGCCCGAGAAGCTTTTTATCGTGGAATCAAGAATGGCAGAGGTGGAAAAGGAAACATTTTTGTATGGGCCAGTGGAAACGGTG 1700
      ┌1708 Intron VI                                  ┌1754 Exon VII
GATCAAGgtatttaggttttgaaaagagaattttgtgcaagaagaatgtttagCCAAGACTCATGTTCAGCTGATGGTTACACAACTTCAGTCTACACGC 1800
                                        ┌1835 Intron VII
TTTCCATATCTTCGGCTACTTATGATAATCACAGgtatgattttatattactataaacccttaatttgtctgaatgaaaattgatattcctgcaagataa 1900
                                                  ┌1948 Exon VIII
aaaatggatttccctactcatcaaattcttattctacttgtctacagACCATGGTATTTGGAAGAATGCCCATCATCAATTGCAACAACATACAGTTCTG 2000

CGGATTTCCGTCAACCAGCAATTGTGACCGTTGATGTTCCTGGAGGATGTACTGATAAGCATACAGGAACTTCTGCGTCTGCTCCATTGGCAGCTGGAAT 2100
                        ┌2124 Intron VIII                                  ┌2176 Exon IX
TATTGCTCTTGCTTTAGAGGCTAAgttagtttattaattccagaaaacttgctcactgtttttttttaatttcagTCCTGAATTAACATGGCGTGATATG 2200
```

Figure 21. Blisterin common genomic sequence. Continued.

```
CAACATCTTGTTCTTCGAACCGCCAACTGGAAACCACTTGAGAATAATCCTGGATGGTCAAGAAATGGTGTTGGACGTATGGTTAGCAATAAATTCGGTT 2300

ATGGTCTAATCGATGGTGGTGCACTTGTGCAATATGGCAAAAACATGGGAAGACAGTCCCAGAGCAGCACATTTGCACATATGAGATACAGACTTGCTAA 2400
      ┌2407 Intron IX                                    ┌2454 Exon X
TCCAAAgtgaggttcaatcatagatttctctggaaaaaatacgatcattacagTCCTCGGCCAATTGTGGGTCGTTTCCAACTTAATTTCACTTTAGATG 2500
                                                                  ┌2564 Intron X
TGAATGGATGCGAGTCAGGAACCCCTGTTTTATATTTGGAACACGTTCAAGTGCATGCTACTGgtcagttttcaattgatactgttctctacaaatctta 2600
               ┌Exon XI
tcaaattgaattcagTTCGATATCTGAAACGTGGCGATCTGAAGCTTACGCTCTTCTCACCATCCGGCACCCGATCGGTTCTTCTTCCACCACGACCACA 2700

AGATTTCAATGCTAACGGATTCCACAAGTGGCCTTTCCTTTCAGTCCAACAATGGGGAGAAGATCCACGTGGAACATGGCTTCTCATGGTGGAATCAGTC 2800
                       ┌2823 Intron XI
ACCACTAATCCAGCTGCTACTGgtgagaatttatttaaatgt...//...taaaatttgaacgaacggaaggattttcgtttttctttttttttttaatg 2892

aaaaatccatttttcccaaaaaaacgacagcnctgatttctatttaaaaaatatgttcgcctttgaagagtactgtaagttcaaacttttgttgctgctg 2992

aattttggttgatttttcatagttttcaataagaccatatttatttatttaaaaaaaacgcaaatttttagcggttaacaaaaatttgtcgtgctgag 3092

accggtttcgatacttctggcgtaaaatttcgtgactaggtaatagtaaaaaattcaaaattgaagcgaaaaatttccctataaatatttganccntata 3192
            ┌3213 Exon XII
tggtgatttcagGTACGTTCCACGATTGGACTCTTCTTCTTTATGGAACTGCTGATCCAGCTCAATCAGGTGATCCTGTCTACTCGGCTACTCCTGCAAC 3292
                                          ┌3344 cDNA divergence point   Intron XII
ATCCCAAGGAGTTCTCTCACGCGTTCATCAACTCACTTCTCAGgtacgcacatttttcggaggattcggatagaatcatattgtnctttgtgtgtgtgtt 3392

gtgttgnctcctttgnttttatttncatatttaggggcccttgaattcaaactgatttacccttcccctctnctcactttctatctaatttggtttcggt 3492

gatagttgtgttttaattttgaaaattcccaacccacagagtgtgtgtccttgcgagtgtgtgagcctggtgtgataatcaattttcaaattgaaaaatt 3592

gaaaaaatcaacaaaatgcctaattttctcataattattcatcgtccttacctctggcacaattttcagtntattcatgtttctattcataaattncttg 3692

tgtatgtatctgtatgtctctatgtaaatgtgtcgctgacttttaatgtttcttgtgaactcgtcttcgtctatgttcttgttgttattgtgatgtgaa 3792

tcaacttttaattaaaaagagaattgattaacagttatatttatttctttatttctttaatttctttgttcaactagacgtccacattgttttctgcag 3892

3' end of pCeh202                                                                                    3894
```

Figure 22. Genomic DNA encoding the unique carboxy terminal of Blisterin A. DNA sequence derived from pCeh205. A single exon encodes the unique 3' end of blisterin A.

Figure 23. Genomic sequence encoding the unique 3' end of blisterin C.
Genomic sequence derived from pCeh207 (top) compared to the unique 3' end of blisterin C (bottom). Gaps in the genomic sequence are indicated by dots (...); introns are indicated by dashes (---); matched positions are indicated by vertical bars (| | |).

```
Genomic cacgacattgnccancccaaacttttttgaatngncgaaacgcagtagatattacggtagatcgttgttttacagGGTGAAGAGTCTACTCGATCTTCA
                                                                                    | |||||||||| ||||||||||
Bli C   ----------------------------------------------------------------------------GTTGAAGAGTCTGCTCGATCTTCA

Genomic TTTCCAGATTTGACGT.CGGCTGGAAATTGTCA..TGATGAATGTAATGGAGGTgaggaattgattttcgcgattactataaatgtttaaaaagtatttt
        |||||||||||||||| |||||||||||||||  ||||||||||||||||||
Bli C   TTTCCAGATTTGACGTCCGGCTGGAAATTGTCATGTGATGAATGTAATGGAGGT----------------------------------------------

Genomic cggtttttttcagaaatttcaaagncaaaaattaaacattgcagaaattncaaaatcaactggaaatattgtccagatgcgaaattttgcgactattgct

Bli C   ----------------------------------------------------------------------------------------------------

Genomic ncaaaaatacggtacccggtctcgacacggccgatttttttcaatccaaaagaatgcgcgcctttaaggaatactgtagtttccacattttccgcgttgct

Bli C   ----------------------------------------------------------------------------------------------------

Genomic cgattttcgtggagctccaattcaccntatagtgagtcntatggcaattcaaa.......Genomic sequencing.gap.....................

Bli C   ----------------------------------------------------------TGCACAGAATCGAGCTCAGCAACATCATGTTTCGCTTGCAAACA

Genomic ..................................................................Genomic sequencing gap ..................

Bli C   TTTAACTCAAACTCTTCGAAATAAAGGAGGAAGTGGATTCAAGCGTGTTCAAAAATGCGATGATACTTACTACTTGGATGGTGACAAGTGTAAAATGTGC

Genomic ..................................................................Genomic sequencing gap.....................

Bli C   TCATCCCATTGTCATACTTGTACGAAAGCTGAGGCGTGTGAAACGTGTCCTGGAAGTCTTTTGCTCATTGATGTGGATAATATGCCACACTATGATCATG

Genomic ..............................................aaaacnggggnattgnntgtgaatgttnagtnctteccctttnnnaggctaaaccc

Bli C   GAAAGTGTGTTGAGTCATGTCCTCCAGGATTGGTTGCTGATT-----------------------------------------------------------

Genomic tacntccnaccacctgaaacttcccccttagaattatttaccatcttgaaatnccttggcctcctaacccattnnccctcatgttacaagtttttttgt

Bli C   ----------------------------------------------------------------------------------------------------

Genomic ttaaaccaatgtgaaatttcagATGAATCGAATCTTGTTCAAGCTAAATGTATCTGGAGAAAAGATCTTTGTGGTGACGGATATTACATCAACGCTGTTG
                           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Bli C   ----------------------ATGAATCGAATCTTGTTCAAGCTAAATGTATCTGGAGAAAAGATCTTTGTGGTGACGGATATTACATCAACGCTGTTG

Genomic GAAAATGTGATCTTTGCGACTCCTCATGCGAAACTTGCACCGCTCCTGGTCCAATGAGCTGTGAGAAATGTTCAAAAGGTTATGGTAAAGGATCGATTGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Bli C   GAAAATGTGATCTTTGCGACTCCTCATGCGAAACTTGCACCGCTCCTGGTCCAATGAGCTGTGAGAAATGTTCAAAAGGTTATGGTAAAGGATCGATTGG
```

Figure 23. Genomic sequence encoding the unique 3' end of blisterin C. Continued.

```
Genomic ATACTGTAGACCGTGTTGTCCTGAAGGATCAACAAAGAGTTGGCAATGTggtaaggaagacactctcaaaatcttccaggtatctcacaatttatttttt
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Bli C   ATACTGTAGACCGTGTTGTCCTGAAGGATCAACAAAGAGTTGGCAATGT----------------------------------------------------

Genomic caaGAGGACTGCTCCAAGCGGGATCCTACACTTTTGATTGATTCTAATAAACCATCTGGATTTGGATTGATGNTCTGGATTGTAGTTAGTTTGATTGCGG
          |||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||
Bli C   ---GAGGACTGCTCCAAGCCGGATCCTACACTTTTGATTGATTCTAATAAATCATCTGGATTTGGATTGATGTTCTGGATTGTAGTTAGTTTGATTGCGG

Genomic CTTGTGGAATCTGTGCCTGTAAAAAGTGTGCAAGTGAGACGAAAAGCTCAACCGTAGAgtaagccttgctagttatcgtctccgtaattcgaatttttaat
        |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Bli C   CTTGTGGAATCTGTGCCTGTAAAAAGTGTGCAAGTGAGACGAAAAGCTCAAACGTAGA----------------------------------------

Genomic atttttaGATATGCGCCGCTTGCTCAATATAATGCCACAAATGGTGCTAACAATTNAGGAGCACACACTGACGATGAAGACGATGATGAGGATGAAGTAT
              |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Bli C   -------GATATGCGCCGCTTGCTCAATATAATGCCACAAATGGTGCTATCAATTTAGGAGCACACACTGACGATGAAGACGATGATGAGGATGAAGTAT

Genomic TTGTGAACCCTCAAATTGTTTAAACCAAACCTTTCAAAATTCATGTTTTTAATTGTAATTTTTCTGCCAACTTCTTTGTGTATGCTTGTGAAATCCGGGT
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Bli C   TTGTGAACCCTCAAATTGTTTAAACCAAACCTTTCAAAATTCATGTTTTTAATTGTAATTTTTCTGCCAACTTCTTTGTGTATGCTTCTGAAATCCGGGT

Genomic AATTGTTTCTTTTCCAAAATTTTAAAACCTGAAACGTTTTCAAGTGGCCTTTTATCATGTGATTGTATTGTTTCTTTTGTCTTATACCGG.....
        |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Bli C   AATTGTTTCTTTTTCAAAATTTTAAAACCTGAAACGTTTTCAAGTGGCCTTTTATCATGTGATTGTATTGTTTCTTTTGTCTTATACCGGGTTTATTTTT

Genomic ........................................

Bli C   ATTATACCGTACTGTTTCCATGTTATAAATAGATTGTTGAATT  3' end of blisterin C.
```

3.2.4 Analysis of the predicted *bli-4* proteins

3.2.4.1 Identification of similarity with kex2-like serine proteases

A search of the translated EMBL DNA data base, SWISSPROT using the FASTA search algorithm with a k-tup value of 2 was conducted using all three predicted proteins. All proteins with an optimized similarity score of 100 or greater are listed in Table 8. All of the proteins listed in Table 8 are subtilisin-type serine endoproteases. The proteins identified in the FASTA search of the SWISSPROT library fell into two distinct groups; those with higher scores, ranging from 846 to 1316, and those with lower scores, ranging from 100 to 159. All of the proteins in the higher scoring group were eukaryotic subtilisin-type serine endoproteases. The scores in the lower scoring group were prokaryotic subtilisins.

The higher scoring group included mouse, rat and human furin, mouse and human prohormone convertase 2 (PC2), mouse prohormone convertase 1 (PC3), the *Saccharomyces cervisiae* KEX2 gene product kex2, and the *Kluvermyces lactis* KEX1 gene product Kex1. Because some of these proteins are more than 95% identical to each other, a representative of each type will be used in sequence comparisons to the blisterins. Human furin (hfurin) will be used to represent human, rat and mouse furin. Human PC2 (hPC2) will be used to represent human and mouse PC2. *S. cerevisiae kex2* will be used to represent kex2 and its *K. lactis* homologue, kex1. The eukaryotic proteins as a group will be referred to as the kex2-like proteins, because kex2 was discovered first.

The identity scores reported by FASTA for the proteins listed in Table 8 are for the region of highest similarity rather than the whole protein. To determine the overall identity scores of the blisterins and the kex2-like

proteins, the ALIGN program was used to compare the proteins listed Table 9. The overall identity scores between blisterin B and the eukaryotic kex2-like enzymes ranged from 25.8% for kex2, to 36.4 % for hfurin. Subtilisin scored much lower, at 11.0% identical amino acid positions. The identity scores between the blisterins and the kex2-like proteins are in the same range as scores obtained when the kex2-like proteins are compared to each other (Table 9). Thus, blisterins are as related to the kex2-like proteins as the kex2-like proteins are to each other.

An alignment of blisterin B and the mature subtilisin *amylosaccharits* revealed a domain of similarity between subtilisin and blisterin B, residues Tyr(191) to Gln(441) (Figure 24). Within this domain, 26% of amino acid residues are identical. This domain includes most of the 275 residue subtilisin protein, and will be referred to as the "Subtilisin-domain" of blisterin.

Alignment of the eukaryotic kex2-like enzymes revealed regions of similarity extending beyond the subtilisin domain (Figure 25). These regions begin near the beginning of the blisterin B and extend to residue Ala(629). The domain of similarity to the amino side of the subtilisin domain is referred to as the "Pre-domain" of the blisterins. The Pre-domain includes residues Met(1) to Gly(190). The subtilisin domain extends from Tyr(191) to Gln(441). The domain of similarity among the kex2-like endoproteases to the carboxyl-side of the Subtilisin Domain has been noted previously (Fuller, Brake and Thorner, 1989b; Smeekins and Steiner, 1990) and is referred to as the "P-domain". The P-domain includes residues His(442) to Ala(629). All three domains, the Pre-domain, the Subtilisin-domain and the P-domain occur within the blisterin common sequence, which ends at Gln(658).

The identity scores obtained by the alignment of whole protein sequences provides a rough index of relative similarity of the sequences. To obtain a more sensitive ranking of protein identities, ALIGN was used to compare the individual blisterin domains to the corresponding segments of the kex2-like proteins (Table 10). In the pre-domain of blisterin B, the highest identities were obtained for mPC3 (23.1%), hfurin, (21.7%) and hPC2 (21.6%). The lowest identity was obtained for kex2 (15.1%). In the subtilisin-domain, the highest scores were for hfurin (67.8%), while the lowest score for the eukaryotic proteins was obtained for kex2 (43.3%). For the P-domain, the highest identity was obtained for hPC2 (38.9%), hfurin (38.3%), and mPC3 (37.2%). Again, the lowest score was obtained for kex2 (23.5%). The only clear pattern that emerged from these alignments was that the identity of blisterins with kex2 was consistently lower than that of blisterins with the proteins of higher eukaryotes. Based on these results, it is not possible to conclude that the blisterins are most like any one of the kex2-like proteins.

To identify small regions of similarity between the blisterins and the kex2-like proteins, the blisterin B sequence was divided into segments of 50 amino acid residues and used individually to search a library of kex2-like sequences with the FASTA program (Figure 26). In the pre-domain, only segment four, residues 151-200, identified a similar region in all of the protein sequences. Thus, most of the identity between the Pre-domain and the kex2-like proteins in Table 10 occurs in the last 25% of the Pre-domain. Exceptions to this observation were similarities identified with segment two to hPC2 and mPC3, and segment three to mPC3. The more extended similarity of these proteins with the blisterin pre-domain is reflected by the high percentage of identical amino acid positions in Table 10. Segments five

to nine span the subtilisin-domain. These segments identified similar regions in all of the proteins, with the exception of segment seven, which did not identify a similarity in subtilisin. Segment seven also had the lowest scores among the subtilisin-domain segments for all of the other proteins, indicating that the region of residues 301 to 350 are less conserved that other parts of the subtilisin domain. Segments 10 to 13 span the P-domain. Segments 10 and 12 identified stronger similarities than segments 11 and 13, with the exception of kex2, which did not have a region similar to segment 10. In general, the P-domain segments had lower similarity scores than did the subtilisin domain segments. The blisterin sequences diverge from each other at blisterin B residue 658, in segment 14. Of the segments in the unique carboxyl-ends of the blisterins, only one identified a similar region in the kex2-like proteins. A segment of blisterin C identified a region of similarity in hfurin (See section 3.2.4.2.4).

Table 8. Summary of Fasta[a] search of SWISSPROT protein database using Blisterin C.

Protein ID	Protein name	Optimized similarity score[b]	Identity[c]	Reference
FURI$MOUSE	Mouse furin (mfurin)	1316	52.8% (511)	Hatsuzawa *et al.*, 1990
FURI$RAT	Rat furin (rfurin)	1322	50.0% (558)	Misumi, Sohda and Ikehara, 1990
FURI$HUMAN	Human furin (hfurin)	1309	52.4% (510)	Roebroek *et al.*, 1986; van den Ouweland *et al*,, 1989
FURH$MOUSE	Mouse prohormone convertase 1 (mPC1 - also known as mPC3)	1301	43.7% (565)	Seidah *et al.*, 1990 Smeekins *et al.*, 1991
KEX2$YEAST	*S. cerevisiae* kex2 (kex2)	917	36.2% (527)	Mizuno *et al.*, 1988
KEX1$KLULA	K. lactis kex1 (klkex1)	846	35.4% (492)	Tanguy-Rougeau, Wesolowski and Fukuhara, 1988
KEX2$HUMAN	Human prohormone convertase 2 (hPC2)	1139	42.5% (570)	Smeekins and Steiner, 1990
KEX2$MOUSE	Mouse prohormone convertase 2 (mPC2)	1139	43.0% (570)	Seidah *et al.*, 1990
SUBT$BACSA	*B. subtilis amylosaccaritis* mature subtilisin	100	34.8% (92)	Yoshimoto *et al.*, 1988

[a]Lippman and Pearson (1985).

[b]Calculated for region of similarity using the PAM250 matrix.

[c]Percentage of identical amino acids within the region of similarity. Numbers in brackets indicate the number of nucleotides in the region of similarity identified by FASTA.

Table 9. Overall identity scores[a] of blisterin B and kex2-like proteins.

Protein	Subtilisin	mPC3	hPC2	kex2	hfurin	blisterin B
blisterin B	11.8%	35.8%	35.0%	24.8%	36.4%	100%
hfurin	9.8%	40.2%	35.0%	26.5%	100%	
kex2	9.5%	28.3%	62.0%	100%		
hPC2	11.6%	38.8%	100%			
mPC3	9.0%	100%				
Subtilisin	100%					

[a]Identity scores were obtained using the ALIGN program.

Figure 24. Alignment of blisterin B and *B. subtilus* subtilisin.

Alignment of blisterin B (top) and *Bacillus subtilus amylosaccharitis* subtilisin. Within the blisterin B interval Gln_{190} to Gln_{441}, 26% of the amino acid positions are identical. Active site amino acids are indicated by asterisks (*); gaps introduced into the alignment are indicated by hyphens (---); identical positions are indicated by vertical bars (:::); conservative positions are indicated by dots (...). The alignment was determined using ALIGN.

```
MRISIGRIAWQILAVLIAVAFTIEHDSICDESIGACGEPIHTVIRLAKRDDELARRIAADHDMHVKGDPFLDTHYFLYHSETTRTRRHKRAIVERLDSHP
 .              .:...:..                                    .:. .
AQ--------------SVPYGISQ-----------------------------------IKAPA----------------------------

                                                                                    ┌190 Subtilisin
AVEWVEEQRPKKRVKRDYILLDNDVHHSNPFRRSVLNRDGTRRAQRQQPQSPAEIPSLPFPDPLYKDQWYLHGGAVGGYDMNVRQAWLQGYAGRNVSVSI
                                                                    ::.              :::.:.::.:..
------------------------------------------------------------------LHS--------------QGYTGSNVKVAV
Domain
LDDGIQRDHPDLAANYDPLASTDINDHDDDPTPQNNGDNKHGTRCAGEVAALAGNNQCGVGVAFKAKIGGVRMLDGAVSDSVEAASLSLNQDHIDIYSAS
.:.::...::::...    .....   ...:    .....:::. ::..:::. :.   .::: .:.. .:..::.. :.         :   : . .
IDSGIDSSHPDLNVR----GGASFVPSETNP---YQDGSSHGTHVAGTIAALN-NSIGVLGVAPSASLYAVKVLDSTGSG----------QYSWIINGIE

WGPEDDGKTFDGPGPLAREAFYRGIKNGRGGKGNIFVWASGNGGSSQDSCSADGYTTSVYTLSISSATYDNHRPWYLEECPSSIATTYSSADFRQPAI-V
:. ... .... .   ....    .  ... ...: : :......: .:.:. ::... :. .:. .. .. ...      .:.... :. : ..:.. .
WAISNNMDVINMSLGGPSGSTALKTVVDKAVSSGIVVAAAAGNEGSSGSSSTVGYPAK-YPSTIAVGAVNSSNQR------ASFSSAGSELDVMAPGVSI
                                                ┌441 End of Subtilisin Domain
TVDVPGGCTDKHTGTSASAPLAAGIIALALEANPELTWRDMQHLVLRTANWKPLENNPGWSRNGVGRMVSNKFGYGLIDGGALVQYGKNMGRQSQSSTFA
  ..:::  . ..::: ..:.::   :: :. .:  :: . :
QSTLPGGTYGAYNGTSMATPHVAGAAALILSKHP--TWTNAQ-------------------------------------------------------

HMRYRLANPNPRPIVGRFQLNFTLDVNGCESGTPVLYLEHVQVHATVRYLKRGDLKLTLFSPSGTRSVLLPPRPQDFNANGFHKWPFLSVQQWGEDPRGTW
 .: ::                                  ..:. ::
-VRDRL-----------------------------------ESTATYL-----------------------------------------------

WLLMVESVTTNPAATGTFHDWTLLLYGTADPAQSGDPVYSATPATSQGVLSRVHQLTSQGDEVVERIRNHWEVTLEESSHWNWEHAREHKSLQELNSSSR
                                  ::. : .                                                :.: .......
--------------------------------GDSFYYG-----------------------------------------------KGLINVQAAAQ

THSFLYSFTKFQPIFLIILVCIFDAIHRQFAV*
```

Figure 25. Alignment of blisterin B and kex2-like proteins.
Amino acid residues in the kex2-like proteins identical to blisterin B are indicated by hyphens (-). Similar amino acid residues are lower case. Dissimilar amino acids are upper case. BliB: blisterin B; hfur: human furin; mPC2: mouse prohormone convertase 2; mPC3: mouse prohormone convertase 3; kex2: yeast kex2. ALIGN was used for protein alignments

```
      ┌► Pre-domian
bliB  MRISIGRIAWQILAVLIAVAFTIEHDSICDESIGACGEPIHTVIRLAKR   DDELARRIAADHDMHVKG DPFLDTHYFLYH SETTRTRRHKRAIVER
hfur  -Elr    p-LlWv-AAtGtlvl La     aDaq -qKvF-NtwAvrIP GgPAv-nsv-RK-gfLNL-QI  FGDY-HFW-RGV-K-SLSPH-PRHS-
mPC2  -kGgCVS q-k     A-ag-Lf     -VMVfasAeR-vF-NHF-veLHKGgeDK--qv--E-gfG- RKL--AEGL-HF--NGLAKAK--RSLHHKQQ
mPC3  -EQr    g-Tl    qCt--af FCVW-aL-sVK akrQFvNEWA-eIP Ggq-A-sA--eELgYdlL-QIGS-EN---GK-K-HPR-S--SALH-TK-
kex2  -kvrKY -tLCFWWAfstS-lvssqqi   PLKdHTsRqYFa-E       SneT-s-LEeMHpnWkYEHDVRG-PN--VFSKELLKLGK-SSLEELQGD

                                                                                                      ┌►
bliB  LDSHPAVEWVEEQRPKKRVKRDYILLDNDV HHSNPFRRSVLNRDGTRRAQRQQPQSPREIPSLPFPDPLYKDQWYLHG    GAVGGYDMNVRQAWLQG
hfur  -QRE-Q-Q-L-Q-VA-r-T---             VYQEPT                          --kFPq----S-      -tQR-l--ka--A--
mPC2  -ERD-R-KMALQ-EGFD-K--G-    R-I NEIDINMN                            ---Ftd----InTGQAdgtP-L-l--Ae--EL-
mPC3  -SDDDR-T-A-Q-YE-E-S--S            VQKD-A-DL                       -N--mWnq----qdTRMTa-lPKL-lh-IPv-Ek-
kex2                    NN-H--SVH-LFPRNDL-K-  -PVPAPPMDSSLL-VKEA- DK-SIN---FeR--h-VnPS    FP-S-i--LdL-Ynn
      Subtilisin domain
bliB  YAGRNVSVSILDDGIQRDHPDLAANYDPLASTDINDHDDDPTPQNN  GDNKHGTRCAGEVAALAGNNQCGVGVAFKAKIGGVRMLDGA VSDSVEAASL
hfur  -t-hgiV--------ekn-----g----G--F-v--q-P--Q-rYtQMn--r----------v-n-gV------fn-r---------e -t-a---R--
mPC2  -t-kg-tIg-m----dYL-----s--nAE--Y-fssn-PY-F-rYtDDWF-s---------s-A-n--I------fns-va-i----QpFmt-ii--s-i
mPC3  It-kg-VItv----Lewn-t-iY-----E--Y-f--n-h--F-rYdLTne----------i-MQ-n-hk------fnsdN--i-----I -t-ai--s-i
kex2  It-Ag-VAa-v---LdYenE--Kd-fCAEG-W-f--ntnL-k-r   Ls-dY--------i--KK---F-----gfn---s-i-i-s-DIttEd -----

bliB  SLNQDHIDIYSASWGPEDDGKTFDGPGPLAREAFYRGIKNGRGGKGNIFVWASGNGGSSQDSCSADGYTTSVYTLSISSATYDNHRPWYLEECPSSIATT
hfur  g--pn--h-------------V---ar--e---f--vsq----L-s----------reh---nC----n-i---------QFgnV---S-a-s-tl---
mPC2  -HMpqL----------t-n---V---RDvtLq-mAd-vnk------s-y-----d---Yd- -nC---as-mw-i--n--in-grTaL-D-s-s-tl-s-
mPC3  gf-pg-v---------n-----Ve---r--qk--EY-v-q--Q---s---N------rQG-n-dC----d-i--i-----sQqgLs---A-k-s-tl--s
kex2  iYgL-vN----C----a---rHlq--sD-vkk-lVk-vte--ds--a-y-f------trG-n-nY----n-m-Sit-saIDhkdLh-P-S-g-saVm-v-

                                                    ┌► P-domain
bliB  YSSADFRQPA  IVTVDVPGGCTDKHTGTSASAPLAAGIIALALEANPELTWRDMQHLVLRTANWKPLENN PG WSRNGVGRMVSNKFGYGLIDGGALV
hfur  ---gnQneKq  ---t-lrQK--es----------------t----Kn----------vq-skPAh-na-  D -aT-----k--hsy----l-a--m-
mPC2  f-ngRK-n-eAGva-t-lY-n--Lr-s----a--E---vf-------Lg---------tvL-skrnq-hdeVhQ -r-----LeFnhL----vl-a--m-
mPC3  ---g-yTdqR  -tsa-lhnd--et-------------f--------n----------vw-seyd--as- -- -kk--a-L--nsr--f--lnaK---
kex2  ---gs    gEY-HsS-iN-R-Sns-g----a------vYT-L-----n-----v-y-SiLs-  VG--k-AD-D-rDsam-kkY-hry-f-K--aHK-I

bliB  QYGKNMGRQSQ SSTFAHMRYRLANPNPRPIVGRFQLNFTLDVNGCESGT PVLYLEHVQVHATVRYLKRGDLKLTLFSPSGTRSVLLPPRPQDFNAN G
hfur  aLaq  nwTtV apQRKCi I DILte-kD-GK-lEvrK-vt  a-LgepNhITR---a-arL-ls-NR----AiH-V--M----t--aa--h-ys-d -
mPC2  kMa-  dwktV peR-HCvGG  sVqd-EK-PsTGK-Vl--ttda--gKeNF-R------aVI--nATR----ninmT--M--k-i--sr--r-DdakV-
mPC3  dLadPrtwrnV pekKeCv VkDn-Fe--alKanGEvIVEiPtra--gQeNaIKS-----Fe--iE-SR----hv--T-aV--st---aE-Er-Tsp- -
kex2  eMs-  twenVNaQ-wFylpTLYvsqstnstEETLESVi-isEds  LQdANFKRi---T-TvdiDTEI--TTtvd-i--a-iI-N-GVV--r-Vsse -

                                                                ┌► Blisterin divergence point
bliB  FHKWPFLSVQQWGEDPRGTWLLMVESVTTNPAATGTFHDWTLLLYGTA DPAQSGDPVYSATPA TSQGVLSRVHQLTSQGDEVVERIRNHWEVTLEESS
hfur  -ND-A-MTTHS-D---S-E-V-EI-NTSE ANNY--LTKF--V----- -EGLPV-PE--gCK-L--SQaCV-CeeGF-LHqkSCVqHCPpGFapQVLdT
mPC2  -D----MTTHT----A----T-EL GFVGSAPQK-VLKE---M-H--Q SAPYIDQV-   RDYQ-k Lam-kKee-eeeL--a---sLkSILNkN
mPC3  -KN-D-M--HT---N-V---T-KITDMSGRMQNE-RIVN-K-I-H--S SQPEHMKQPRVY-sYN-V-nDRrG-ekmvnVVekRPtqKsLnGnlLvPkn-
kex2  -KD-T-M--AH---NGV-D-KIK-K  --ENGHRID--S-R-K-F-ESI-SSKTETF-F          GnDKeevepaat-StVsQYsASstsisI-A
```

Figure 25. Alignment of blisterin B and kex2-like proteins. Continued.

```
bliB  HWNWEHAREHKSLQELNSSSRTHSFLYSFTKFQPIFLIILVCIFDAIHRQFAV
hfur  -ysT-ndV-TIrASVCapchAsCaTCQgPaLTdClSCPsHaslDPvEqTCSRQSQSSRESPPQQQPPRLPPEVEAGQRLRAGLLPSHLPEVVAGLSCAFI
mPC3  SSsNVeg-RdeQv-gTP-kamLrLlQSa-s-NaLsKQsPKKsPSAKlSIpyeSFYEALEKLNKPSKLEGSEDSLYSDYVDVFYNTKPYKHRDDRLLQALM
kex2  TStSsIsIgVetSaIPqtttAsTdPDSdPntPkKlSSPRQaMHyFLtIFLigaTFLVLYFMFFMKSRRRIRRSRAETYEFDIIDTDSEYDSTLDNGTSGI

hfur  VLVFVTVFLVLQLRSGFSFRGVKVYTMDRGLISYKGLPPEAWQEECPSDSEEDEGRGERTAFIKDQSAL
mPC3  DILNEEN
kex2  TEPEEVEDFDFDLSDEDHLASLSSSENGDAEHTIDSVLTNENPFSDPIKQKFPNDANAESASNKLQELQPDVPPSSGRS
```

Table 10. Identity of blisterin domains with kex2-like enzymes.

	Blisterin domain[a]		
Protein	Pro	Subtilisin-like	P
hfurin	21.7%[b]	67.8%	38.3%
kex2	15.1%	43.3%	23.5%
hPC2	21.6%	53.5%	38.9%
mPC3	23.1%	59.6%	37.2%
Subtilisin[c]		26%	

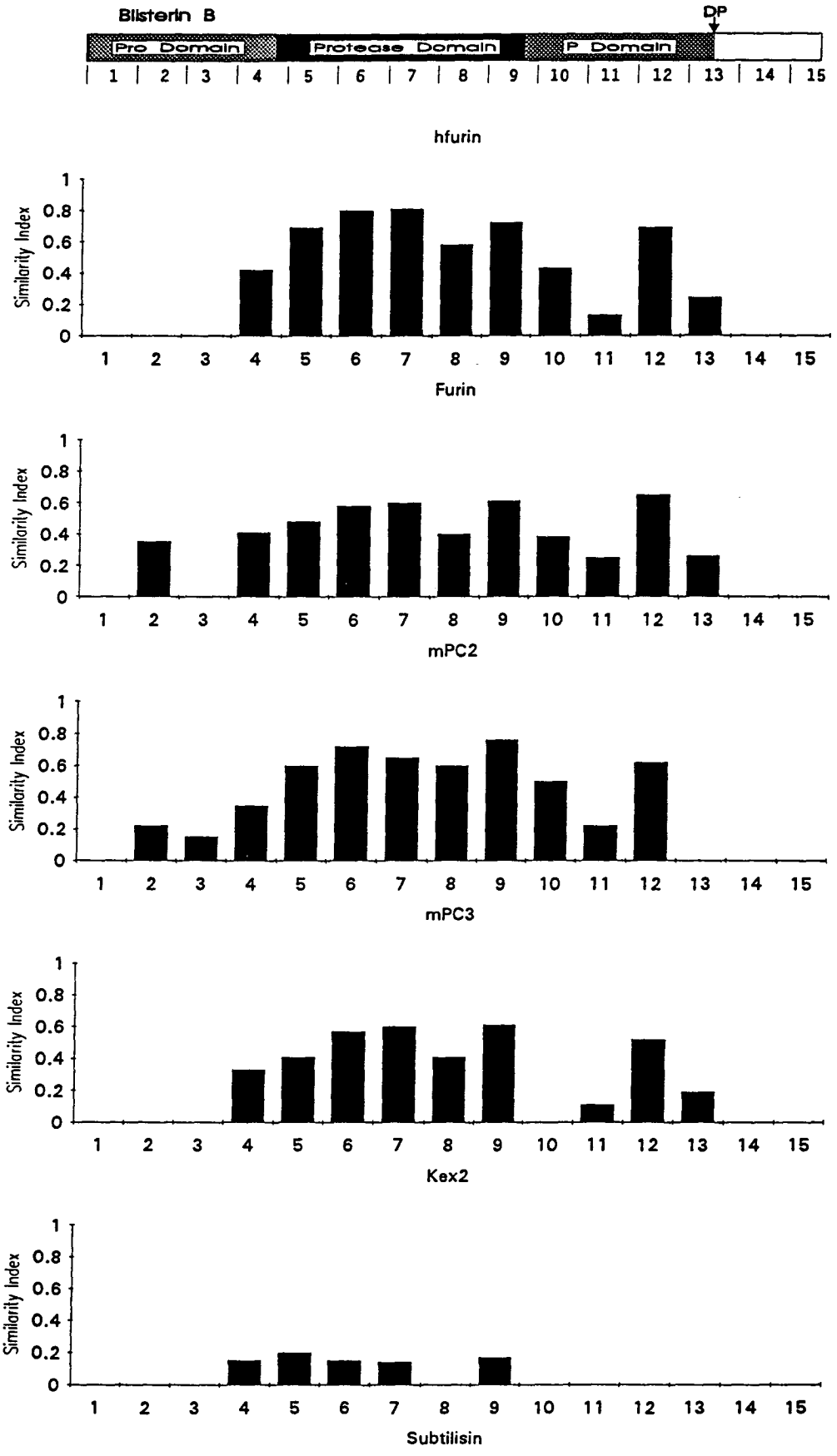
[a]Blisterin domains: Pro = residues 1-190; Subtilisin-like = 191-440; P = 441-629.

[b]Identities were determined using the ALIGN program to align each blisterin domain with the corresponding segment of the kex2-like enzyme.

[c]*Bacillus subtilus amylosaccharitis* subtilisin.

Figure 26. Similarity of blisterin 50 amino acid segments with kex2-like proteins

Blisterin sequences were divided into segments of 50 amino acid residues which were used individually to search a library of kex2-like sequences with the FASTA program. The scores were normalized to the score obtained when a blisterin segment was aligned with itself. A score of 1 indicates completely identical sequence. A score of 0 indicates that the blisterin segment did not align with the corresponding region of the kex2-like protein.


Blisterin B
DP
Pro Domain
Protease Domain
P Domain
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
hfurin
Similarity Index
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Furin
Similarity Index
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
mPC2
Similarity Index
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
mPC3
Similarity Index
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Kex2
Similarity Index
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Subtilisin

3.2.4.2 Protein features

The structural features of the Blisterins are presented in Figure 27.

3.2.4.2.1 Active site

The region of the kex2-like proteins corresponding to the blisterin subtilisin domain includes the serine protease active site. The catalytic amino acids of the serine protease active site of the kex2-like proteins, aspartic acid, histidine and serine are found at amino acid residues 202, 241 and 415 respectively. The alignment of the blisterins and kex2-like proteins presented in Figure 25 shows that the position of these residues within the subtilisin domain site is conserved. The observation that the blisterin subtilisin-domain is the most conserved segment of the protein sequence (Table 10) is consistent with the fact that this region corresponds to the active site of the kex2-like proteins.

3.2.4.2.2 Prediction of potential post-translation modification sites

kex2 is localized in a late compartment of the golgi body (Julius, Schekman and Thorner, 1984; Fuller, Brake and Thorner, 1989a; Franzuoff *et al.*, 1991; Redding *et al.*, 1991; Wilcox and Fuller, 1991). Because it passes through the endoplasmic reticulum, the nascent kex2 protein chain includes a signal peptide that is removed by cleavage by signal peptidase, probably at residue 19 (Fuller, Brake and Thorner, 1989a; Wilcox and Fuller, 1991). The kex2 is both N-glycosylated and O-glycosylated (Wilcox and Fuller, 1991). In addition, a pro-region is removed from the maturing protease by autocatalysis at residues Lys(108)Arg(109) (Brenner and Fuller, 1991). Because of its similarity to kex2, blisterin sequence was examined for potential signal peptidase, glycosylation, and autocatytic sites.

The first 29 amino acid residues of the blisterin common sequence meet the criteria of von Heijne (1983) for a secretion signal peptide. The 24

amino-terminal residues have an overall hydrophobic character, as determined by hydropathy analysis (Figure 28)(Kyte and Doolittle, 1983). A 16 amino acid residue sequence composed of mostly hydrophobic amino acid residues flanked by charged amino acids is found between Arg(7) and Glu(24). A core of ten exclusively hydrophobic amino acid residues is found from between residues Ile(12) and Phe(21). A potential signal peptidase cleavage site was identified between residues Cys(29) and Asp(30) using the technique of von Heijne (1983) (Appendix E). Potential N-linked glycosylation sites with the motif Asn-X-Ser/Thr are found at positions 195 and 695 of the blisterin common sequence. Additional potential N-glycosylation sites are found at amino acid residue 695 of blisterin B, and residue 497 of blisterin C. Potential sites for autocatalytic cleavage of the Blisterins occur after Lys(48)Arg(49), Arg(55)Arg(56), Arg(87)Arg(88), Lys(90)Arg(91), Lys(111)Lys(112), Lys(112)Arg(113), Arg(131)Arg(132), Arg(142)Arg(143).

3.2.4.2.3 A potential transmembrane domain in blisterin C

Kex2 and furin both have a potential hydrophobic transmembrane domain near the carboxyl-terminal. In the case of kex2, this domain is required to maintain the enzyme within the golgi body; deletion of the transmembrane domain results in secretion of the majority (70%) of the enzyme (Fuller, Brake and Thorner, 1989b). The transmembrane domain divides kex2 and hfurin into a luminal portion, which includes the proteolytic domain, and a negatively charged cytosolic tail of 115 amino acid residues in kex2, and 49 amino acid residues in furin. Blisterin C includes a hydrophobic domain of 24 amino acid residues, between residues 500 and 524 (Figure 28). This hydrophobic domain could potentially span a membrane. The hydrophobic domain is followed by a 48 amino acid

carboxyl-terminal region which includes 11 acidic (glutamic or aspartic acid) residues. In particular, nine of the last 17 residues are acidic. By comparison to hfurin and kex2, the hydrophobic domain of blisterin C is likely to be a transmembrane domain.

In contrast to blisterin C, blisterin A unique and blisterin B do not include a transmembrane-like domain. Of the 13 amino acid residues of the blisterin A unique carboxyl-terminal, nine residues are non-polar residues (Ileu, Leu, Val and Ala), two, Thr and Asn, are uncharged polar residues, and one, His, is basic. Thus, the dominant feature of the blisterin A unique sequence is its non-polar nature. The 73 amino acid residues of the blisterin B unique carboxyl-terminal also include a short hydrophobic region of 13 amino acid residues. However, 13 amino acids is not enough to form a membrane spanning domain. Therefore, neither blisterin B nor blisterin A are likely to be membrane spanning proteins; both hydrophobic sequences are too short, not strongly hydrophobic and located right at the carboxyl-terminus.

3.2.4.2.4 A cysteine-rich region in hfurin and blisterin C

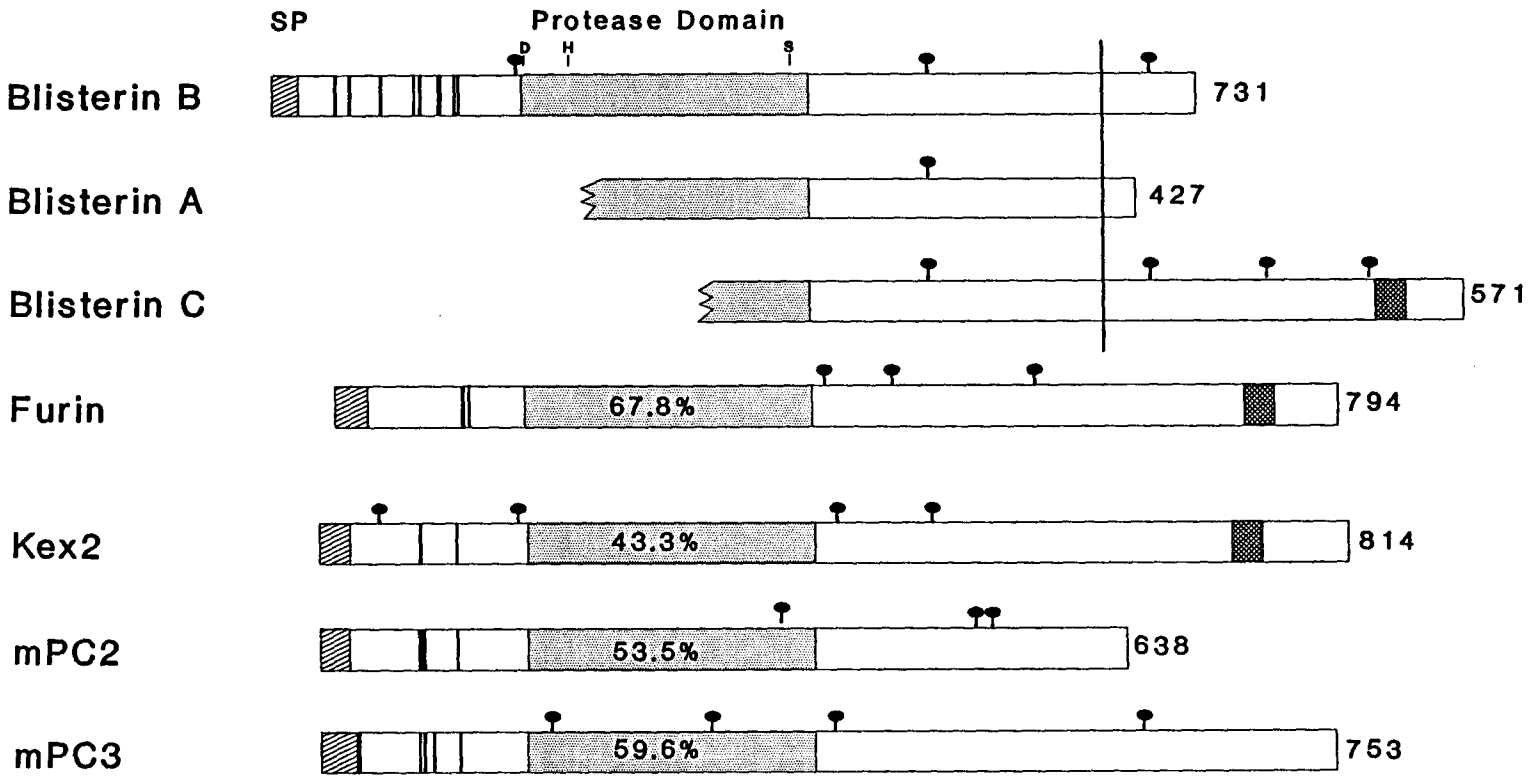
hfurin contains a cysteine-rich region to the amino side of its transmembrane domain. Optimal alignment of this segment of the proteins using ALIGN revealed that eight Cys residues in an interval of 90 amino acids in the blisterin C sequence align with Cys residues in the hfurin sequence (Figure 29). In blisterin C, this interval, Asp(391) to Ser(457), is located 43 residues to the amino side of the potential blisterin C transmembrane domain. In hfurin, the interval containing the conserved Cys residues, His(606) to Pro(663), is located 66 residues to the amino side of the furin transmembrane domain. The conserved Cys residues would, therefore, be located near the membrane in the lumen of the golgi body.

3.2.4.2.5 A conserved Cell attachment site

A search of the blisterins for protein motifs using the PCGENE PROSITE program revealed the presence of a cell attachment site, Arg(551)Gly(552)Asp(553). This sequence is characteristic of proteins such as fibronectin that interact with receptors on the cell surface (reviewed by Rouslahti and Pierschbacher, 1986). This site occurs in the P-domain of all of the kex2-like proteins except kex2, which has the sequence Arg-Gly-Thr at this position. The cell attachment site occurs 110 residues to the carboxyl-side of the blisterin subtilisin domain, and is found in similar positions in the kex2-like proteins (Figure 25).

Figure 27. Blisterin and kex2-like protein structural features

Pairs of basic residues, potential autocatalytic sites are shown as vertical bars on the amino side of the active site. Secretion signal peptides, SP, are shown as hatched boxes; transmembrane domains, TMD, are shown as cross hatched boxes; active sites are shown as dense dots. (ᵠ) indicates potential glycosylation sites. The positions of the catalytically important amino acids Asp, His, and Ser are indicated by D, H and S.


SP
Protease Domain
D
H
S
Blisterin B
731
Blisterin A
427
Blisterin C
571
Furin
67.8%
794
Kex2
43.3%
814
mPC2
53.5%
638
mPC3
59.6%
753

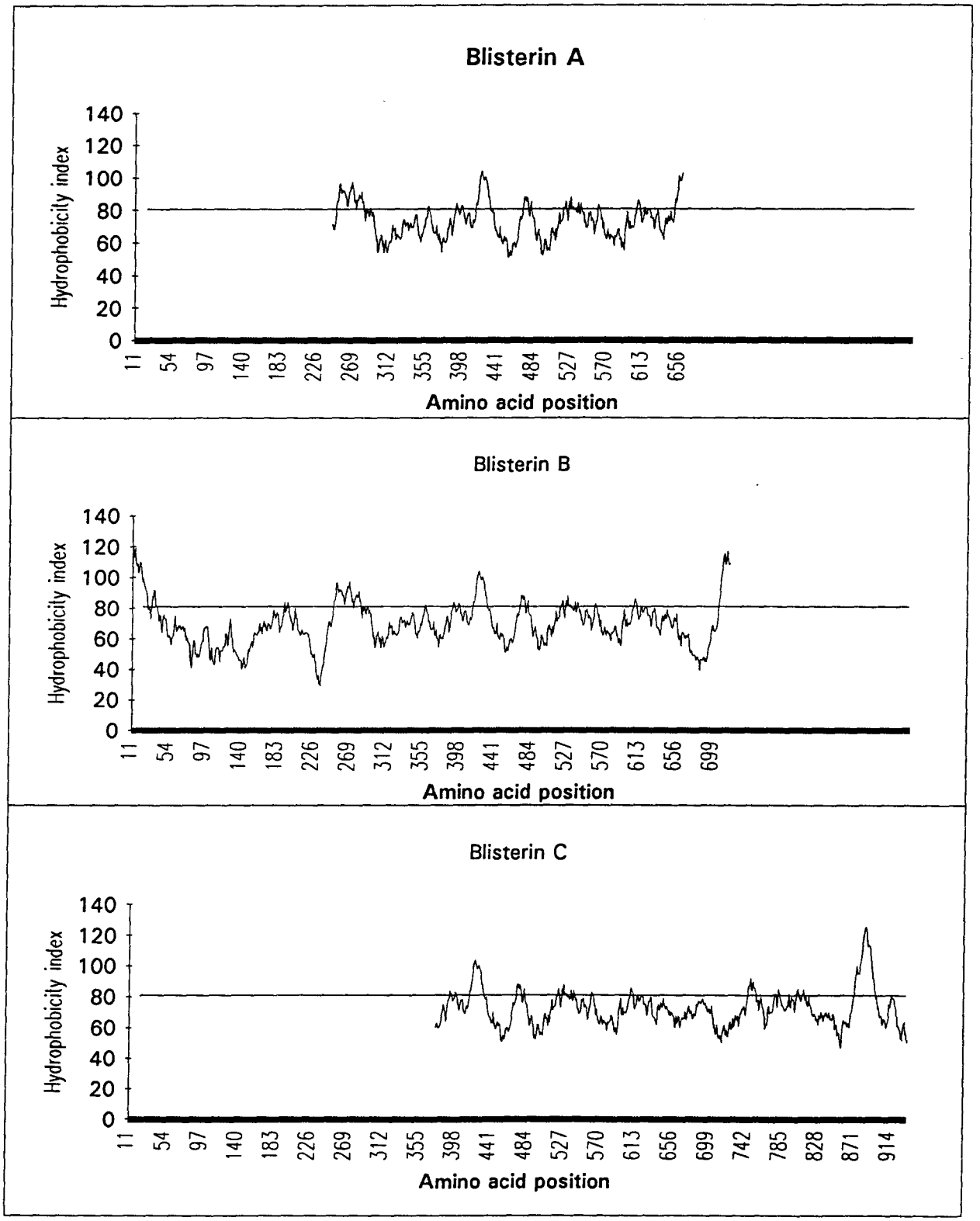


Figure 28. Hydropathy analysis of the blisterins.
Hydropathy values for each amino acid position were calculated using GREASE with a window of 19. Blisterin A and blisterin C were aligned with coresponding positions in blisterin B for comparison.

```
    1   2                                        3  4   5  6       7  8
DHGKCVESCPPGLVADYESNLVQAKCIWRKDLCGDGYYINAVGKCDLCDSSCETCTAPGPMSCEKCS
HQKSCVQHCPPG----FAPQVLDTHYSTENDVET----IRA-SVCAPCHASCATCQGPALTDCLSCP
.. .**. ****    ..........   .*. .    *.* . *. *..**.** .*.  .* .*.
```

Figure 25. Alignment of blisterin C and Furin Cysteine-rich regions.

Alignment of blisterin C (top sequence), Asp(391) to Ser(457), with hfurin (bottom sequence), His606 to Pro(663). 26.9% amino acids are identical in a 67 amino acid residue overlap. Identical amino acids are indicated by "*"; similar amino acids are indicated by ".". Cysteine residues that are present in both sequences are numbered 1 to 8.

3.2.5 Position of *bli-4* mutations in predicted protein sequences

3.2.5.1 *h1010*

Restriction mapping and Southern analysis indicated that the *h1010* Tc1 insertion is in a 1.3 Kb *Eco*RI fragment of pCeh202 (Section 3.2.1.2.1). Analysis of the sequence of pCeh202 revealed that the boundaries of this fragment are *Eco*RI sites at nucleotide 1,199, in Exon V, and at nucleotide 2,608 in Intron X (Figure 18). To confirm my interpretation of the *h1010* restriction data, I used PCR to amplify across the *h1010* Tc1 insertion site (Figure 29). The primers used were p618 and KRp13. p618 is specific to a Tc1 sequence starting 72 bp from one end of the transposable element. KRp13 is specific to a sequence 80 bp to the 5′ side of the 1.3 kb *Eco*R1 fragment containing the insertion. A band of approximately 770 bp was obtained following PCR (Figure 30, lane 2), indicating that the Tc1 element was inserted approximately 620 bp into the 1.3 kb *Eco*RI fragment. The size of the amplification band indicated that the *h1010*::Tc1 insertion occurred within the active site of the kex2-like protease.

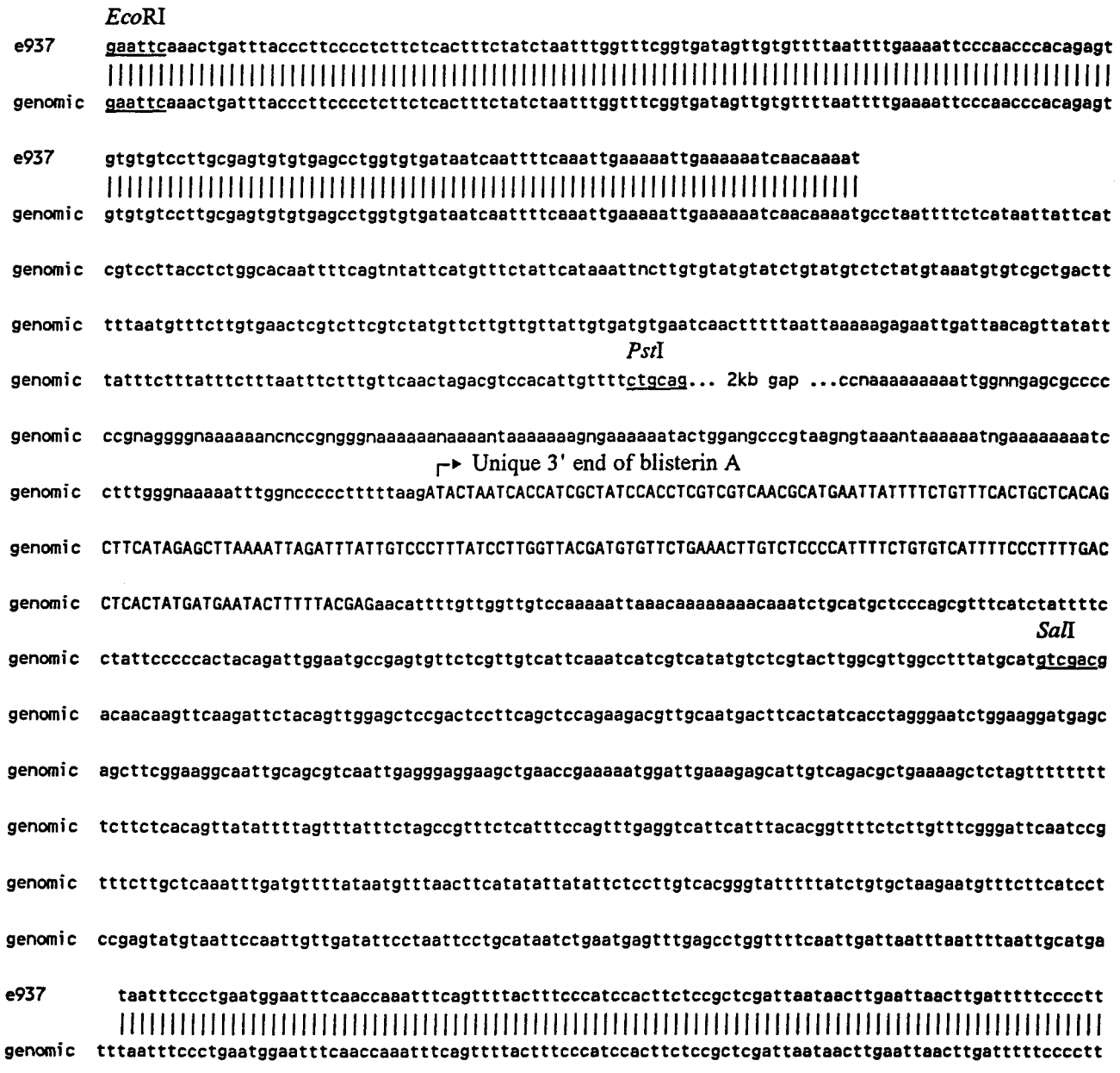
3.2.5.2 *e937*

Restriction mapping and Southern analysis indicated that the *e937* mutation is a 3.5 Kb deletion (Section 3.2.1.2.2). Based on the genomic sequence data, the left break point occurs in Intron XII, to the 3′ side of Exon XII. This interpretation of the left breakpoint was confirmed by sequencing through the deletion breakpoint. This was accomplished by sequencing both ends of the *e937* fusion fragment cloned in pCeh206. As shown in Figure 31, the left breakpoint of the *e937* deletion occurs in a 0.4 kb *Eco*RI/*Pst*I restriction fragment of pCeh181, 170 bp into the fragment. The right deletion breakpoint occurs in the 1.0 kb *Sal*I fragment of pCeh181,

674 bp into the fragment. Thus, the left breakpoint of the *e937* deletion does not affect the blisterin common sequence. Moreover, *e937* does not delete genomic DNA encoding blisterin B or blisterin C. *e937* does, however, delete Exon XIII, which encodes the unique 13 amino acid residue carboxyl-terminal of blisterin A. Thus, *e937* may affect only one of the three *bli-4* gene products. This may explain why the deletion does not result in a lethal phenotype.

Figure 30. Sequence of *e937* deletion breakpoints.

Partial sequence of *e937* fusion fragment in plasmid pCeh206 (Top) compared to genomic sequence (Bottom). Restriction enzyme recognition sites are underlined. The exon encoding the 3' end of blisterin A is in uppercase letters. Identical positions are indicated by vertical bars (|).



CHAPTER FOUR: DISCUSSION

In this thesis, I have presented evidence supporting the following conclusions. First, that the *bli-4* gene of *Caenorhabditis elegans* is a complex locus with at least one essential role in development, as well as a non-essential role in the adult cuticle. Second, the *bli-4* coding region is contained within the cosmid K04F10 Third, the *bli-4* transcripts are alternatively spliced to produce more than one protein product. Fourth, that the gene products produced by *bli-4* are related to the kex2-like mammalian prohormone convertases. In this discussion, I will examine the basis of these conclusions and attempt to demonstrate their validity.

4.1 EVIDENCE THAT *bli-4* IS A COMPLEX LOCUS

I have divided the alleles of *bli-4* into three classes based on phenotype and complementation pattern. Class I was *e937*, which was homozygous viable, and caused adult cuticle blisters. Class II included nine lethal mutations that caused developmental arrest at the end of embryogenesis and failed to complement the blistered phenotype of *e937*. Class III included two lethal mutations, that caused developmental arrest at an early larval stage and complemented the blistered phenotype of *e937*. The blistered phenotype of *e937* was not complemented by class II alleles, but was complemented by class III alleles. Class II and class III alleles did not complement each other. The complementation data can be interpreted in two ways; first, that there is one gene with complementing alleles, and second, that there are two genes, and the class III alleles affect both. I favor the one gene hypothesis. My evidence is discussed in the following paragraphs.

I will first consider the two-gene hypothesis in light of the genetic and molecular data. The two gene hypothesis is that two genes have been

identified, *bli-4*, defined by *e937*, and an essential gene, which I will call *let-77*, defined by the class III alleles. In this hypothesis, the class II alleles fail to complement both *bli-4* and the *let-77* mutations because they affect both genes. Mechanisms for this affect could include either or both of the following. First, each of the class II alleles could be a two mutation event, with mutations in both *bli-4* and *let-77*. Second, each of the class II alleles could be a deletion, affecting both *bli-4* and *let-77*.

The two-hit mechanism requires that all nine of the class II alleles simultaneously acquire mutations in *bli-4* and *let-77*. Each such double hit event is very unlikely. A further problem with the two hit hypothesis is that it does not account for the fact that the class II have a different phenotype than the class III allele: class II alleles arrest development earlier than class III alleles. Thus, in addition to requiring a highly improbable series of events, all of the double hit mutations involving *let-77* would be more severe than *s90* and *h754*. While the double hit hypothesis cannot be formally ruled out on the basis of genetic evidence alone, it is not a reasonable explanation for the complementation data.

The deletion mechanism for the two-gene hypothesis requires that all nine of the class II alleles be deletions of *bli-4* and *let-77*. Deletions are rarely recovered following EMS mutagenesis. McDowall (1990) demonstrated that the class II alleles *h384*, *h427*, *h520*, *h699* and *h799* complement alleles of all other genes in the interval around *bli-4* defined by the duplications *hDp16* and *hDp19* (Figure 1). These mutations, therefore, cannot be large deletions. If the class II alleles were all deletions, we would expect the 10 kb XhoI fragment of K04F10 that detects *h1010*::Tc1 and *e937* to detect restriction fragment length differences in the genomic DNA of at least some of the other class II lethal alleles. No such differences were

detected in the alleles *h42*, *h199*, *h384*, and *h427*. It is possible that these alleles are deletions with breakpoints outside of the 10 kb fragment. However, these alleles complement alleles of all other genes around *bli-4*, and therefore cannot be large deletions. Finally, at least one class II lethal, *h1010*, cannot be a deletion: none of the cosmids spanning 200 kb around *bli-4* that were used to probe *h1010* DNA detected deletions. I conclude, therefore, that the deletion mechanism for the two gene hypothesis is incorrect.

In the one-gene hypothesis, *e937*, class II and class III mutations affect a single genetic locus with complementing alleles. In this hypothesis, which I will call the one-gene hypothesis, the *bli-4* locus has more than one function. *e937* affects one function, required in the adult cuticle, while *s90* and *h754* affect another function. The class II alleles affect all functions of the gene. Several lines of evidence lead me to conclude that the one gene hypothesis is correct. First, the induction frequency of class II alleles is very high. Nine alleles were identified in the *sDp2* lethal set, which was produced by screening 30,000 mutagenized chromosomes. The induction frequency of *bli-4* alleles in the *sDp2* set was, therefore, approximately 1 in 3000. A hypothesis accounting for the nature of the class II alleles must take this frequency into account. The event resulting in a class II allele must be common, not rare. Mutations involving double hits, EMS induced deletions or mutations in a regulatory region shared by two genes are all likely to be very rare. In contrast, point mutations in a large gene that result in the loss of function of that gene are likely to be common. Second, I have shown the that insertion of Tc1 into the active site of the kex2-like gene in K04F10 resulted in a concurrent loss of the ability of the *h1010* allele to complement the function disrupted by *e937*. The latter observation demonstrates that

the kex2-like protease active site is required to prevent blisters. Finally, the genomic DNA encoding at least one transcript that includes the kex2-like active site, blisterin A, is affected by *e937*. I conclude, that *e937* and all of the *bli-4* lethal alleles are alleles of one gene.

4.2 EVIDENCE THAT *bli-4* IS THE kex2-LIKE CODING ELEMENT ON K04F10

As a first step in identifying the *bli-4* coding region, I used a probe detecting a strain-specific restriction fragment length difference and a deletion to define the limits of *bli-4* within the *C. elegans* physical map. In doing so, I restricted the possible range of the *bli-4* locus to about 200 kb of genomic DNA.

My next step was to use the cosmids in this interval to probe the mutation *h1010*, which was isolated in a screen specifically designed to generate alleles that would be detectable by southern analysis. Only one of the cosmids in the interval containing *bli-4* detected rearrangements in *h1010* DNA. The pattern of restriction fragments detected by K04F10 was consistent with the insertion of a Tc1 element. Subsequent PCR analysis demonstrated that the band shifts detected by K04F10 were indeed a result of the insertion of Tc1.

The genomic fragment that the Tc1 element was inserted into detected bands on Northern blots, demonstrating that the fragment containing the Tc1 insertion site is transcribed into RNA. The genomic restriction fragments flanking the fragment containing the Tc1 insertion site detected the same bands. Based on these data, I concluded that the Tc1 element detected by K04F10 was inserted into a coding region.

PCR amplification using a Tc1-specific primer, p618, and a primer that directed DNA synthesis into the 1.3 kb EcoRI fragment containing the Tc1 insertion in *h1010* produced a band of approximately 770 bp. based on the

position of the primers, I estimated that the Tc1 insertion site was about 580 nucleotides into the 1.3 kb EcoRI fragment. Based on the sized of the amplification band, examination of preliminary sequence data revealed that the Tc1 insertion site was near the center of the genomic DNA encoding the predicted active site of the kex2-like protease sequence identified in K04F10.

Two questions are of critical importance to my contention that the kex2-like coding region in K04F10 is *bli-4*. First, did the lethal phenotype of *h1010* result from the Tc1 element insertion detected by K04F10? Second, if the lethal phenotype of *h1010* did result from the Tc1 element insertion detected by K04F10, was it the disruption of the kex2-like coding element by the Tc1 insertion that was responsible for the *h1010* phenotype, or was it some other effect of the Tc1 insertion? For example, the Tc1 insertion could be inserted into an enhancer required for the expression of a coding element other than the kex2-like gene.

KR1822, the parent strain of KR1858 did not contain a lethal mutation linked to flanking markers *unc-63* and *unc-13*. Nor did the parental strain contain a Tc1 insertion at this position. Therefore, *h1010* and the Tc1 insertion arose at the same time. The answer to the first question, does the *h1010* phenotype result from the Tc1 element insertion, must be consistent with this observation. The hypothesis that the *h1010* phenotype does not result from the Tc1 insertion, therefore, requires two separate but simultaneous mutation events. Since the Tc1 insertion was the only genetic alteration detected in southern analysis using the cosmids of the *bli-4* interval as probes, the mutation causing the *h1010* phenotype must be either a very small rearrangement or point mutation. Since K04F10 rescues *bli-4* lethal mutations, these events must have both occurred in the genomic DNA cloned in K04F10.

The hypothesis that the Tc1 insertion is responsible for the *h1010* phenotype requires only one mutation event. The position of the Tc1 insertion is consistent with the physical mapping and rescue data, and with the observation that the *e937* deletion mutation affects the same kex2-like coding region as the Tc1 insertion affects. Definitive proof that the *h1010* phenotype resulted from the Tc1 insertion could have been provided by the recovery of a spontaneous revertant of the *h1010* mutation and demonstration that the Tc1 element had excised in the revertant strain. In the absence of such proof, I conclude that the hypothesis that the *h1010* phenotype is results from the Tc1 insertion is the most probable.

Given that the lethal phenotype of *h1010* did result from the Tc1 element insertion detected by K04F10, was it the disruption of the kex2-like coding element by the Tc1 insertion that was responsible for the *h1010* phenotype, or was it some other effect of the Tc1 insertion? Two points support the hypothesis that it is the disruption of the kex2-like gene that is responsible for the *h1010* phenotype. First, *h1010* is inserted into the active site of the kex2-like gene. The Tc1 element is therefore very likely to affect the function of the kex2-like gene. Second, the *h1010* mutation cannot complement the blistered phenotype of *e937*, which affects the same kex2-like gene. These observations, combined with the fact that K04F10 rescues *bli-4* lethal mutations, lead us to conclude that the Tc1 insertion is responsible for the mutant phenotype of *h1010*.

The *h1010* data, however, does not prove that the kex2-like protease disrupted by the Tc1 insertion is *bli-4*. It is possible that the *h1010* phenotype is a result of the Tc1 insertion in the kex2-like protease active site interfering in some way with the expression of the 'real' *bli-4*, a gene other than the kex2-like gene. Given that K04F10 rescues *bli-4* lethal mutations,

bli-4 would also be on K04F10 in this hypothesis. For example, the Tc1 could be inserted into an intron containing an enhancer required for the expression of the *bli-4*. Therefore, the other gene hypothesis cannot be eliminated on the basis of the *h1010* data alone.

The same 10 kb XhoI K04F10 restriction fragment that detects *h1010*::Tc1 also detects a chromosomal rearrangement associated with the *bli-4* allele, *e937*. The *e937* rearrangement was shown to be a 3.5 kb deletion by restriction mapping and sequencing of the deletion breakpoints. *e937* was induced using ^{32}P as a mutagen (Brenner, 1974). ^{32}P causes chromosomal breaks, consistent with the finding that *e937* is a deletion. *e937* deletes genomic DNA encoding the last exon of blisterin A. Blisterin A includes at least part of the kex2-like gene active site. Therefore, like *h1010*::Tc1, *e937* disrupts the kex2-like protease. The fact that both *h1010*::Tc1 and *e937* affect the kex2-like protease leads us to conclude that the kex2-like protease is the *bli-4* gene.

In summary, K04F10 rescues *bli-4* mutants, and detects rearrangements in two *bli-4* alleles. Both rearrangements affect the same coding region. Together, these observations provide compelling evidence for the conclusion that the affected coding region is *bli-4*.

4.3 *bli-4* MUTATIONS

That *e937* is a deletion was initially surprising given that *e937* is the least severe of the *bli-4* mutations. This can be explained by the observation that is an internal deletion affecting genomic DNA encoding only one of the three messages identified, blisterin A; *e937* does not delete genomic DNA encoding blisterin B or blisterin C. The fact that the *e937* deletion is homozygous viable indicates that the function of blisterin A is not essential, consistent with the hypothesis that disruption of blisterin A is reponsible for

the blistered phenotype. It is possible, however, that *e937* could also delete other exons that have not yet been identified. Thus, blisterin A should be considered as a candidate only for the protein required in the adult cuticle.

Given that *e937* affects DNA encoding only one of the three transcripts identified, a mechanism for intragenic complementation of *bli-4* alleles might be that the two complementing alleles, *s90* and *h754*, affect only one of the other two transcript classes. In this hypothesis, the alleles complement because the *e937* allele produces some *bli-4* transcripts, including, blisterin B and C, but not others, such as blisterin A. *s90* and *h754* would affect different transcripts from those affected by *e937*. Class II alleles would eliminate all *bli-4* functions. This idea is supported by the observation that the complementing alleles arrest development slightly later than the class II alleles, which is evidence that the complementing alleles retain some function. This hypothesis could be tested by sequencing the blisterins from *s90* and *h754* homozygotes.

Of the three *bli-4* mutation classes, class II mutations occur most frequently, and have the most severe phenotype. It is therefore likely that the class II phenotype is the null phenotype. Class II mutants do not arrest development until the end of embryogenesis. This could indicate that *bli-4* is not required until the end of embryogenesis. More likely, the blisterins are required earlier in embryogenesis, and a maternal endowment of the *bli-4* transcripts provides the null mutants with sufficient blisterin function to survive to this stage. Examples of this are null mutations of the *ama-1* and *ama-2* genes in *C. elegans*, which encode the large and small subunits of RNA polymerase II (Rogalski and Riddle, 1988; Rogalski, Bullerjahn and Riddle, 1988). In these mutants, worms survive to the end of embryogenesis with no zygotic RNA polymerase II-dependent transcription at all.

Of the three cDNA clones sequenced, only blisterin B contained a complete open reading frame. This conclusion is based on the presence of an ATG beginning at position 176 of the cDNA, and upstream in frame stop codons. It is possible that the blisterin B cDNA clone does not contain the entire 5′ untranslated region. Particularly since the length of the blisterin B message detected in northern analysis is 3.5 kb, much longer that the 2.4 kb cDNA insert. The Blisterin A and Blisterin C cDNA clones are likely to contain incomplete open reading frames. This conclusion is based on the observations that both are shorter than the bands they detect in northern analysis, and that they start within the open reading frame of blisterin B. An attractive hypothesis is that blisterin A and Blisterin C are derived from mRNA transcripts with the same 5′ end as blisterin B. The fact that three of the pCeh181 *Eco*RI genomic fragments encoding all of the 5′ end sequence of blisterin A detected all of the messages identified supports this conclusion. However, it is possible that the 5′ ends of these molecules differs from blisterin B through alternative splicing. The fact that a blisterin C unique probe detected a smaller band than did a blisterin B unique probe supports the latter possibility. Thus, no conclusions on the nature of the 5′ ends of blisterin A and C can be drawn based on this data.

4.4 *bli-4* GENE PRODUCTS ARE RELATED TO kex2-LIKE PROTEASES

I have shown that the blisterin sequences have a high degree of identity and similarity with the kex2-like serine endoproteases. Sequence similarity can originate in three ways. Coincidence, convergence, and evolutionary conservation. The identity of the predicted blisterin sequences with the kex2-like proteins was as great as 60% over more than 300 amino acid residues. Evolutionary conservation is the only likely explanation for this degree of identity. On this basis, I conclude that *bli-4* is a homologue

of the kex2-like genes: *bli-4* and the kex2-like genes evolved from a common ancestral sequence. All of the members of the kex2-like proteins that have been tested have been shown to be serine endoproteases that cleave substrates after pairs. Blisterin B, the only *bli-4* cDNA that appeared to contain a complete open reading frame, was found to be as similar to members of the kex2-like serine endoproteases as the kex2-like proteases were to each other. It is therefore likely that the sequence conservation of the blisterins and the kex2-like proteases is reflected in functional conservation: the blisterins are likely to be serine endoproteases.

The similarity of the blisterin B primary amino acid sequence to other members of the kex2-like enzyme family is greatest in the 300 amino acid sequence including the protease domain. Outside of this region, other structural features are conserved, including a signal peptide, potential autocatalytic sites, and, in blisterin C, a Cys-rich region and potential transmembrane domain followed by an acidic carboxy tail. By analogy to the kex2-like proteases, therefore, the blisterins are likely to be proprotein processing enzymes that cleave substrates in the golgi body during secretion.

Structurally, the kex2 protein family can be divided into two classes based on the presence or absence of a transmembrane domain (TMD) near the carboxy terminus. PC1(PC3), PC2, blisterin A and blisterin B lack a TMD. Kex2, hfurin, and blisterin C contain a TMD. *bli-4* is the first example of both types of enzyme arising through alternative splicing. Although alternative splicing in other KEX2-like genes has not been reported, alternative splicing cannot be ruled out. PC1(PC3) and PC2 detect multiple bands on Northern blots (Seidah *et al.*, 1990; Smeekens and Steiner, 1990).

It is possible that alternatively spliced transcripts from these genes exist but have not yet been identified.

The TMD is required for protein retention within the golgi body: deletion of the TMD in kex2 results in the secretion of the protein (Fuller, Brake and Thorner, 1989b). It has been suggested that PC1(PC3) and PC2 might be membrane associated through an amphipathic helix structure (Smeekins *et al.*, 1991). No direct evidence to support this hypothesis has been produced. It is possible that the different proteins encoded by the *bli-4* locus fulfill functions in the nematode that are fulfilled in mammals by different genes.

Of the other members of the KEX2-like gene family, only KEX2 was identified genetically: null mutations in KEX2 result in the inability to process the alpha-factor mating pheromone and the K1 killer toxin (Wickner and Leibowitz, 1976; Leibowitz and Wickner, 1976). KEX2 is not essential to viability in yeast, indicating that kex2 does not play an essential intracellular role (that is, kex2 is not an essential metabolic gene). In contrast, null mutations in *bli-4* result in developmental arrest near the end of embryogenesis prior to hatching. Such evidence demonstrates that *bli-4* is essential to development. Why is *bli-4* essential where KEX2 is not? Unlike yeast, *C. elegans* is multicellular, and has a complex development process. The observation that the blisterins are essential to development implies the existence of blisterin substrates required for embryogenesis. Such substrates could include both structural molecules and peptide signals.

4.5 CONCLUDING COMMENTS

No essential cellular interaction involving a secreted protein derived from a proprotein has yet been described in *C. elegans*. The disruption of development by *bli-4* mutations implies the existence of such interactions.

This work provides the first link between the molecular analysis of kex2-like enzymes and genetic analysis of mutations in a multicellular eukaryote.

References

Ambros, V. and H.R. Horvitz. 1984. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226**, 409-416.

Ambros, V. and H.R. Horvitz. 1987. The *Lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific post-larval events. *Genes Dev.* **3**, 399-414.

Baillie, D.L., K.A. Beckenbach, and A.M. Rose. 1985. Cloning within the *unc-43* to *unc-31* interval (linkage group IV) of the *Caenorhabditis elegans* genome using Tc1 linkage selection. *Can. J. Genet. Cytol.* **27**, 457-466.

Barstead, R.J. and R.H. Waterston. 1989. The basal component of the nematode dense-body is vinculin. *J. Biol. Chem.* **264**, 10177-10185.

Benjannet, S., N. Rondeau, R. Day, M. Chretien, and N.G. Seidah. 1991. PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues. *Proc. Natl. Acad. Sci.* **88**, 3564-3568.

Brenner, C. and R.S. Fuller. 1992. Structural and enzymatic characterization of a purified prohormone-processing enzyme: secreted, soluble Kex2 protease. *Proc. Natl. Acad. Sci.* **89**, 922-926.

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

Bresnahan, P.A., R. Leduc, L. Thomas, J. Thorner, H.L. Gibson, A.J. Brake, P.J. Barr, and G. Thomas. 1990. Human *fur* gene encodes a yeast KEX2-like endoprotease that cleaves pro-beta-NGF in vivo. *J. Cell Biol.* **111**, 2851-9.

Bullock, W.O., J.M. Fernadez, and J.M. Short. 1987. XL1-Blue: a high efficiency plasmid trasforming recA *Escherichia coli* strain with beta-galactosidase selection. *BioTechniques* **5**, 376.

Cassada, R.C., and R.L. Russell. 1975. The dauer larva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **46**, 326-342.

Chance, R.E., R.M. Ellis, and W.W. Brommer. 1968. Porcine proinsulin: characterization and amino acid sequence. *Science* **161**, 165-167.

Chretien, M. and C.H. Li. 1967. Isolation, purification, and characterization of gamma-lipotropic hormone from sheep pituitary glands. *Can. J. Biochem.* **45**, 1164-1174.

Clark, D.V., Rogalski, T.M., L.M. Donati, and D.L. Baillie. 1988. The *unc-22(IV)* region of *Caenorhabditis elegans*: genetic analysis of lethal mutations. *Genetics* **119**, 345-353.

Coulson, A., J. Sulston, S. Brenner, and J. Karn. 1986. Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821-7825.

Coulson, A., R. Waterston, J. Kiff, J. Sulston, and Y. Kohara. 1988. Genome linking with yeast artificial chromosomes. *Nature* (London) **355**, 184-186.

Cox, G.N., J.S. Laufer, M. Kusch and R.S. Edgar. 1980. Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans*. *Genetics* **95**, 17-339.

Cox, G.N., M. Kusch and R. S. Edgar. 1981. The cuticle of *Caenorhabditis elegans*: Its isolation and partial characterization. *J. Cell Biol.* **90**, 7-17.

Cox, G.N., S. Staprans and R.S. Edgar. 1981. The cuticle of *Caenorhabditis elegans* II. stage specific changes in ultrastructure and protein composition during post-larval development. *Dev. Biol.* **86**, 456-470.

Cox, G.N., J.M. Kramer and D. Hirsh. 1984. Number and organization of collagen genes in *Caenorhabditis elegans*. *Mol. Cell. Biol.* **4**, 2389-2395.

Cox, G.N. and D. Hirsh. 1985. Stage specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. *Mol. Cell. Biol.* **5**, 363-372.

Craxton, M. 1991. Methods: A companion to Methods in Enzymology 3, 20-26.

Cross, S.H., and P. Little. 1986. A Cosmid vector for systematic chromosome walking. *Gene* **49**, 9-22.

Crow, E.L. and R. S. Gardener. 1959. Confidence intervals for the expectation of a poisson variable. *Biometrika* **46**, 441-453.

Davis, R.W., D. Botsein and J.R. Roth. 1980. A manual for genetic engineering: Advanced bacterial genetics.. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Docherty, K. and D.F Steiner. 1982. Post-translational proteolysis in polypeptide hormone biosynthesis. *Ann. Rev. Physol.* **44**, 625-638.

Edgley, M.L. and D.L. Riddle. 1990. The nematode *Caenorhabditis elegans*. In *Genetic maps*, vol **5**.

Eidie, D., and P. Anderson. 1985a. The gene structures of spontaneous mutations affecting a *Caenorhabditis elegans* myosin heavy chain gene. *Genetics* **109**, 67-79.

Eidie, D., and P. Anderson. 1985b. Transposition of Tc1 in the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **82**, 1756-1760.

Emmons, S.W., M.R. Klass, and D. Hirsh. 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *C. elegans*. *Proc. Natl. Acad. Sci.* **76**, 1333-1337.

Emmons, S.W., L. Yesner, K.-S. Ruan, and D. Katzenberg. 1983. Evidence for a transposon in *Caenorhabditis elegans*. *Cell* **32**:55-65.

Feinberg, A.P. and B. Vogelstein. 1984. A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266-267.

Fodor, A. and P. Deak. 1985. The isolation and genetic analysis of a *Caenorhabditis elegans* translocation *(szT1)* strain bearing an X-chromosome balancer. *J. Genet.* **64**, 143-1673.

Foster, D.C., R.D. Holly, C.A. Sprecher, K.M. Walker, and A.A. Kumar. 1991. Endoproteolytic processing of the human protein C precursor by the yeast Kex2 endopeptidase coexpressed in mammalian cells. *Biochem.* **30**, 367-72.

Franzuoff, A. and R. Scheckman. 1989. Functional compartments of the yeast golgi apparatus are defined by the *sec7* mutation. *EMBO J* **8**, 2695-2702.

Franzuoff, A., K. Redding, J. Crosby, R.S. Fuller, and R. Scheckman. 1991. Localization of components involved in protein transport and processing through the yeast Golgi apparatus. *J. Cell Biol.* **112**, 27-37.

Fuller, R.S., A. Brake, and J. Thorner. 1986. The *Saccharomyces cerevisiae* KEX2 gene, required for processing prepro-alpha-factor, encodes a calcium dependent endopeptidase that cleaves after Lys-Arg sequences. *In* Microbiology, L. Leive, editor. American Society for Microbiology, Washington, DC. pp 273-278.

Fuller, R.S., R.E Sterne, and J. Thorner. 1988. Enzymes required for yeast prohormone processing. *Ann. Rev. Physiol.* **50**, 345-362.

Fuller, R.S., A. Brake, and J. Thorner. 1989a. Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2+-dependent serine protease. *Proc. Natl. Acad. Sci.* **86**, 1434-1438.

Fuller, R.S., A.J. Brake, and J. Thorner. 1989b. Intracellular targeting and structural conservation of a prohormone-processing endoprotease. *Science* **246**, 482-486.

Gibson, T.J., A.R. Coulson, J.E. Sulston, and P. Little. 1987. Lorist 2, a cosmid with transcriptional terminators insulating vector genes from interference by promoters within the insert: effect of DNA yield and cloned insert frequency. *Gene* **53**, 275-281.

Greenwald, I. 1985. *lin-12*, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. *Cell* **43**, 583-590.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557

Hatsuzawa, K., M. Hosaka, T. Nakagawa, M. Nagase, A. Shoda, K. Murakami, and K, Nakayama. 1990. Structure and expression of mouse furin, a yeast kex2-related protease. *J. Biol. Chem. 265*, 22075-22078.

Heine, U., and T. Blumenthal. 1986. Characterization of regions of the *C. elegans* X Chromosome containing vitellogenin genes. *J. Mol. Biol.* **188**, 301-312.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III in DNA sequence analysis. *Gene* **28**, 351-259.

Higgins, B.J. and D. Hirsh. 1977. Roller mutants of *Caenorhabditis elegans*. *Mol. Gen. Genet.* **150**, 63-72.

Hodgkin, J. 1980. Primary sex determination in the nematode C. elegans. *Genetics* **46**, 649-664.

Hodgkin, J., H.R. Horvitz, and S. Brenner. 1979. Nondisjunction mutants of the nematode *Caenorhabditis elegans*. *Genetics* **91**, 67-94.

Horvitz, H.R., S. Brenner, J. Hodgkin, and R. Herman. 1979. A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. *Mol. Gen. Genet.* **175**, 129-133.

Howell, A.M. 1989 Essential genes in a region of chromosome *I* in *Caenorhabditis elegans*. Ph.D. Thesis, University of British Columbia, Vancouver, B.C., Canada.

Howell, A.M., S.G. Gilmore, R.A. Mancebo, and A.M. Rose. 1987. Genetic analysis of a large autosomal region in *Caenorhabditis elegans* by the use of a free duplication. *Genet. Res.* **49**, 207-213.

Ish-Horowicz, D. and J.F. Burke. 1981. Rapid and efficient cosmid cloning. *Nucleic Acids Res.* **9**, 1323-1338.

Johnsen, R.C. and D.L. Baillie. 1991. Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. *Genetics* **129**, 735-752.

Julius, D., R. Schekman, and J. Thorner. 1984. Glycosylation and processing of pre-pro-alpha-factor in the yeast secretory pathway. *Cell* **36**, 309-318.

Julius, D., A. Brake, L. Blair, R. Kunisawa, and J. Thorner. 1984. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. *Cell* **37**, 1075-1089.

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

Kramer, J.M., G.N. Cox and D. Hirsh. 1982. Comparisons of the complete sequences of two collagen genes of *Caenorhabditis elegans*. *Cell* **30**, 599-606.

Kramer, J.M., G.N. Cox and D. Hirsh. 1985. Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. *J. Biol. Chem.* **260**, 1945-1951.

Kramer, J.M., J.J. Johnson, R.S. Edgar, C. Basch and S. Roberts. 1988. The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* **55**, 555-565.

Kramer, J.M., R.P. French, E.C. Park, and J. J. Johnson. 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**, 2081-9.

Kusch, M. and R. S. Edgar. 1986. Genetic studies of unusual loci that affect body shape of the nematode *Caenorhabditis elegans* and may code for cuticle structural proteins. *Genetics* **113**, 621-639.

Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Liebowitz, M.J. and R.W. Wickner. 1976. A chromosomal gene required for killer plasmid expression, mating and sporulation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **73**, 2061-2065.

Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular cloning (a laboratory manual). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

McDowall, J. 1990. Essential genes in the *hDp16/hDp19* region of LG*I* in *Caenorhabditis elegans*. M.Sc. Thesis, University of British Columbia, Vancouver, B. C., Canada.

McKim, K.S., A.M. Howell, and A.M. Rose. 1988. The effects of translocations on recombination frequency in *Caenorhabditis elegans*. *Genetics* **120**: 987-1001.

McKim, K.S. and A.M. Rose. 1990. Chromosome I duplications in *Caenorhabditis elegans*. *Genetics* **124**, 115-132.

McKim, K.S., T.V. Starr, and A.M. Rose. 1992. Genetic and molecular analysis of the *dpy-14* region in *Caenorhabditis elegans*. *Mol. Gen. Genet.* **233**, 241-251.

McKim, K.S., K. Peters, and A.M. Rose. 1992. Genetic analysis of homologue pairing in *Caenorhabditis elegans*. *Genetics*, in press.

Misumi, Y., M. Sohda and Y. Ikehara. 1990. Sequence of the cDNA encoding rat furin, a possible propeptide-processing enzyme. *Nucleic Acids Res.* **18**, 6719.

Mizuno, K., T. Nakamuta, K. Takada, S. Sakakibara, and J. Matsuo. 1987. A membrane-bound, calcium-dependent protease in yeast alpha-cell cleaving on the carboxyl side of paired basic residues. *Bioch. Biophys. Res. Comm.* 144, 807-814.

Mizuno, K., T. Nakamura, T. Ohshima, S. Tanaka, and H. Matsuo. 1988. Yeast KEX2 gene encodes an endopeptidase homologous to subtilisin-like serine proteases. *Biochem. Biophys. Res. Comm.* **156**, 246-54.

Moerman, D.G. and D.L. Baillie. 1979. Genetic organization in *Caenorhabditis elegans*: Fine-structure analysis of the *unc-22* gene. *Genetics* **91**, 95-103.

Moerman, D.G. and D.L. Baillie. 1981. Formaldehyde mutagenesis in the nematode *Caenorhabditis elegans*. *Mutat. Res.* 80, 273-279.

Moerman, D.G., and R.H. Waterston. 1984. Spontaneous unstable *unc-22 IV* mutants in *Caenorhabditis elegans* var. Bergerac. *Genetics 108*, 859-877.

Moerman, D.G., G.M. Benian and R.H. Waterston. 1986. Molecular cloning of the muscle gene unc-22 in *C. elegans* by Tc1 transposon tagging. *Proc. Natl. Acad. Sci.* **83**, 2579-2583.

Mori, I., D.G. Moerman, and R.H. Waterston. 1988. Analysis of a mutator activity necessary for germline transposition and excision of Tc1 transposable elements in *Caenorhabditis elegans*. *Genetics* **120**, 397-407.

Muller, H.J. 1937. Further studies on the nature and causes of gene mutations, pp 213-255 in *Proceedings of the Sixth International Congress of Genetics*, edited by D. Jones. Brooklyn Botanical Gardens, Menasha, Wisconsin.

Novick, P., and R. Scheckman. 1979. Secretion and cell-surface growth are blocked in a temperature-sensitive mutant of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **76**, 1858-1862.

Payne, G.S. and S. Schekman. 1989. Clathrin: a role in the intracellular retention of a Golgi membrane protein. *Science* **245**, 1358-1365.

Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85**, 2444-2448.

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

Prasad, S.S., and D.L. Baillie. 1989. Evolutionarily conserved coding sequences in the *dpy-20* to *unc-22* region of *Caenorhabditis elegans*. *Genomics* **5**, 185-198.

Redding, K., C. Holcomb, and R.S. Fuller. 1991. Immunolocalization of Kex2 protease identifies a putative late golgi compartment in the yeast *Saccharomyces cerevisiae*. *J. Cell Biol.* **113**, 527-538.

Riddle, D. 1988. The Dauer Larva. In: The nematode *Caenorhabditis elegans*. W.B. Wood, Ed. Cold Spring Harbor Laboratories, Cold Spring Harbor.

Roebroek, A.J.M., J.A. Schalken, J.A.M. Leunissen, C. Onnekink, H.P.J. Bloemers, and W. Van de Ven. 1986. Evolutionary conserved close linkage of the *c-fes/fps* proto-oncogene and genetic sequences encoding a receptor-like protein. *EMBO J.* **5**, 2197-2202.

Rogalski, T.M. and D.L. Baillie, 1985. Genetic organization of the *unc-22* region of *Caenorhabditis elegans*. *Mol. Gen. Genet.* **201**, 409-414.

Rogalski, T.M. and D.L. Riddle. 1988. A *Caenorhabditis elegans* RNA polymerase II gene, *ama-1 IV*, and nearby essential genes. *Genetics* **118**, 61-74.

Rogalski, T.M., A.M.E. Bullerjahn, and D.L. Riddle. 1988. Lethal and amanitin resistance mutations in the *ama-1* and *ama-2* genes of *Caenorhabditis elegans*. *Genetics* **120**, 409-422.

Rose, A.M. and D.L. Baillie. 1979. The effects of temperature and parental age on recombination and nondisjuction in *Caenorhabditis elegans*. *Genetics* **92**, 409-418.

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

Rose, A.M., D.L. Baillie, E.P.M. Candido, K.A. Beckenbach, and D. Nelson. 1982. The linkage mapping of cloned restriction fragment length differences in *Caenorhabditis elegans*. *Mol. Gen. Genet.* 188, 286-291.

Rose, A.M., D.L. Baillie, and J. Curran. 1984. Meiotic pairing behavior of two free duplications of linkage group I of *Caenorhabditis elegans*. *Mol. Gen. Genet.* **195**, 52-56.

Rosenzweig, B., L.W. Liaom and D. Hirsh. 1983. Sequence of the *Caenorhabditis elegans* transposable element Tc1. *Nucleic Acids Res.* **11**: 4201-4209.

Rouslahti, E. and M.D. Pierschbacher. 1986. Arg-Gly-Asp: a versatile cell recognition signal. *Cell* **44**, 517-518.

Saki, R.K., D.H. Gelfand, S. Stoffel, S.H. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Sanger, F., and A.R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 444-448.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74, 5463-5467.

Seidah, N.G., L. Gaspar, P. Mion, M. Marcinkiewicz, M. Mbikay, and M. Chretien. 1990. cDNA sequence of two distinct pituitary proteins homologous to Kex2 and furin gene products: tissue-specific mRNAs encoding candidates for pro-hormone processing proteinases. *DNA and Cell Biol.* **9**, 415-424.

Singh, R.N. and J. Sulston. 1978. Some observations on moulting in *Caenorhabditis elegans*. *Nematologica* **24**, 63-71.

Smeekens, S.P., A.S. Avruch, J. LaMendola, S.J. Chan, and D.F Steiner. 1991. Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. *Proc. Natl. Acad. Sci.* **88**, 340-344.

Smeekens, S.P. and D.F. Steiner. 1990. Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease Kex2. *J. Biol. Chem.* **265**, 2997-3000.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Starr, T., A.M. Howell, J. McDowall, K. Peters, and A.M. Rose. 1989. Identification of DNA probes within the linkage group I gene cluster of *Caenorhabditis elegans*. *Genome* **32**, 365-372.

Steiner, D.F., D. Cunningham, L. Spigelman, and B. Aten. 1967. Insulin biosynthesis: evidence for a precursor. *Science* **157**, 697-700.

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

Sulston, J.E. and H.R. Horvitz. 1981. Post-embryonic cell lineages of the nematode, *C. elegans*. *Dev. Biol.* **56**, 110-156.

Sulston, J.E., E. Schiernberg, J.G. White, and J.N. Thomson. 1983. The embryonic cell lineage of the nematode *C. elegans*. *Dev. Biol.* **100**, 64-119.

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

Tanguy-Rougeau, C., M. Wesolowski-Louvel and H. Fukuhara. 1988. The *Kluyveromyces lactis* KEX1 gene encodes a subtilisin-type serine proteinase. *FEBS Letters* **234**, 464-470.

Thomas, G., B.A. Thorne, and D.E. Hurby. 1988. Gene transfer techniques to study neuropeptide processing. *Ann. Rev. Physiol.* **50**, 323-332.

Thomas, G., B.A. Thorne, L. Thomas, R.G. Allen, D.E. Hruby, R. Fuller, and J. Thorner. 1988. Yeast *KEX2* endopeptidase correctly cleaves a neuroendocrine prohormone in mammalian cells. *Science* **241**, 226-230.

Thomas L., R. Leduc, B.A. Thorne, S.P. Smeekens, D.F. Steiner, and G. Thomas. 1991. Kex2-like endoproteases PC2 and PC3 accurately cleave a model prohormone in mammalian cells: evidence for a common core of neuroendocrine processing enzymes. *Proc. Natl. Acad. Sci.* **88**, 5297-5301.

van den Ouweland, A.M., H.L. van Duijnhoven, G.D. Keizer, L.C. Dorssers, and W.J.M. Van de Ven. 1990. Structural homology between the human *fur* gene product and the subtilisin-like protease encoded by yeast KEX2. *Nucleic Acids Res.* **18**, 664.

van de Ven, W.J.M., J. Voorberg, R. Fontijin, H. Pannekoek, A.M.W. van den Oweland, H.L.P. van Duijnhoven, A.J.M. Roebroek, and R.J. Siezen. 1990. Furin is a subtilisin-like proprotein processing enzyme in higher eukaryotes. *Mol. Biol. Rep.* **14**, 265-275.

von Heijne, G. 1983. Patterns of amino acids near signal sequence cleavage sites. *Eur. J. Biochem.* **133**, 17-21.

von Mende, N., D. Bird, P.S. Albert and D.L. Riddle, 1988 *dpy-13*: a nematode collagen that affects body shape. *Cell* **55**, 567-576.

Wickner, R.W. and M.J. Liebowitz. 1976. Two chromosomal genes required for killing expression in killer strains of *Saccharomyces cerevisiae*. *Genetics* **82**, 429-442.

Williams, B.D., B. Schrank, C. Huynh, R. Shownkeen, and R.H. Waterston. 1992. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* **131**, 609-624.

Wise, R.J., P.J. Barr, P.A. Wong, M.C. Kiefer, A.J. Brake, and R.J. Kaufman. 1990. Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site. *Proc. Natl. Acad. Sci.* **87**, 9378-82.

Wilcox, C.A. and R.S. Fuller. 1991. Posttranslational processing of the prohormone-cleaving Kex2 protease in the *Saccharomyces cerevisiael* secretory pathway. *J. Cell Biol.* **115**, 297-307.

Wood, W.B., R. Hecht, S. Carr, R. Vanderslice, N. Wolf and D. Hirsh. 1980. Paternal effects and phenotypic characterization of mutations that affect early development in *Caenorhabditis elegans*. *Dev. Biol.* **74**, 446-469.

Wood, W.B.(Ed). 1988. The *Nematode Caenorhabditis* elegans. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Yoshimoto, T., H. Oyama, T. Honda, H. Tone, T. Takeshita, T. Kamiyama, and D. Tsuru. 1988. Cloning and expression of subtilisin Amylosacchariticus gene. *J. Biochem.* 103, 1060-1065.

APPENDIX A. ISOLATION AND CHARACTERIZATION OF *hT2*

Screen for a *bli-4* balancer: Isolation of *hT2*. Chromosomal translocations in *C. elegans* suppress recombination in the translocated DNA and cause pseudolinkage between markers on the affected chromosomes (Rose and Baillie, 1981). To develop a tool for the isolation of new *bli-4* mutations, I designed a screen to identify translocations of the left half of LGI and the right half of LGIII that would include the *bli-4* allele *e937* on the translocated portion of LGI. Such a translocation would permit rapid precomplemention screens for new *bli-4* alleles. A screen to identify potential translocations of the left arm of LGI and the right arm of LGIII was performed as follows. Homozygous *e937* males were treated with 1500R of gamma radiation using an Atomic Energy Commission of Canada Gamma Cell and crossed to hermaphrodites (KR169), which were homozygous for *unc-13 (e51) I; dpy-18 (e364) III*. In the absence of a translocation, the F1 progeny from this cross would have the genotype *bli-4 + / + unc-13 (I); + / dpy-18*. Segregation from hermaphrodites of this genotype would be expected to produce progeny in the following ratios: 6 wild-type: 3 Bli-4: 3 unc-13: 2 Dpy-18: 1 Bli-4 Dpy-18 : 1 Dpy-18 Unc-13 (ignoring suppression of the blistered phenotype by Dpy-18). However, a hermaphrodite heterozygous for a translocation involving chromosomes I and III would segregate 4 wild-type: 1 Bli-4: 1 Dpy-18 Unc-13. This apparent linkage, or psuedo-linkage of Unc-13 and Dpy-18 was used to identify candidates for I;III translocations. The progeny of 1400 F1 hermaphrodites were screened for Dpy-18 Unc-13 pseudo-linkage. One strain, KR1235, was identified as having a candidate translocation designated *hT2*. Wild type segregants of KR1235 had the genotype *unc-13/hT2 I; dpy-18/ hT2[bli-4(e937)] III*. Homozygous individuals segregated from KR1235 were blistered, with the genotype *hT2 I; hT2[bli-*

4(e937)] III, and were used to establish the strain KR1234. The strain names reflect the order in which the strains were frozen, not isolated.

To eliminate adult translocation homozygotes among the progeny of KR1235 hermaphrodites, KR1235 worms were treated with 0.012 M EMS for 4 hours and the F1 progeny screened for hermaphrodites that did not segregate Bli-4. One such hermaphrodite was isolated and used to establish the strain KR1274. This hermaphrodite presumably had a lethal mutation, *h661*, on one of the translocation arms, either linked or pseudo-linked to *hT2[bli-4(e937)] III* with the following genotype: *unc-13/ hT2 I; dpy-18 (III)/ hT2[let- (h661) bli-4(e937)] III*. h661 was not mapped, and could be on either *hT2 I* or *hT2 III*.

Characterization of crossover suppression by *hT2*. To determine the extent of the translocation, the boundries of recombination suppression were determined. Recombination was determined from *hT2* heterozygotes of the genotypes listed in Table 11. Recombination was eliminated or greatly reduced in the intervals tested between *let-386* and *unc-101* on LGI, and *unc-64* to *unc-32* on LGIII (Figure 31). This work has been refined and expanded by K. McKim (McKim, Peters and Rose, 1992).

Isolation of *bli-4* alleles using *hT2*

To identify new *bli-4* alleles, three screens were conducted, summarized in Table 12. First, KR1274 worms were treated with formaldehyde. Hermaphrodites were suspended in 0.1% formaldehyde for four hours (Moerman and Baillie, 1981). 5000 F1 progeny were screened for blisters. No Bli-4 F1 progeny were identified. Second, KR1274 was treated with 1500 rads of gamma radiation. 30,000 F1 progeny were screened, and four Bli-4 worms were recovered in the F1 progeny; two died without progeny, and two survived. The two that survived were designated *h665*

and *h666*. Third, KR1274 was treated with EMS. 40,000 F1 progeny were screened, and four putative *bli-4* alleles isolated, *h667, h668, h669 and h670*. None of these putative *bli-4* alleles have been characterized or complementation tested with other *bli-4* alleles. However, based on preliminary *hT2* screen results, the *hT2* system provides a rapid method for the isolation of new *bli-4* mutations.

One additional putative *bli-4* mutation arose spontaneously in KR1407, a strain of the genotype *unc-74 unc-13 I/ hT2 I; + III / hT2[let- (h661) bli-4(e937)] III*. This strain segregated a homozygous Unc-74 Bli-4 Unc-13 hermaphrodite, designated strain KR1803. Complementation analysis indicated that this strain carried a viable allele, *h862*, of *bli-4*. However, when *EcoRI* digested DNA prepared from this strain was probed using pCeh181, the band pattern observed for the *e937* deletion was observed (Figure 32). *h862*, therefore, is likely to be a reisolate of *e937*.

Table 11. Recombination in *hT2* heterozygotes.

Maternal Genotype	Wild-type	Bli-4	DpyUnc	Dpy	Unc
LG I					
hT2/ dpy-5 unc-54	238	36	10	44	41
hT2 / dpy-5 unc-101	194	40	51	0	0
hT2 / let-362 dpy-5 unc-13	132	21	0	0	1[a]
dpy-5 unc-54 / + +	389	-	53	89	62
let 362 dpy-5 unc-13 / + + +	511	-	35	0	5
LG III					
hT2/ dpy-18 unc-64	447	95	87	1[a]	1[a]
hT2/ unc-36 dpy-18	623	117	115	0	0
hT2 / dpy-1 unc-32	127	21	11	25	14

[a]Recombinants were not confirmed by progeny testing.

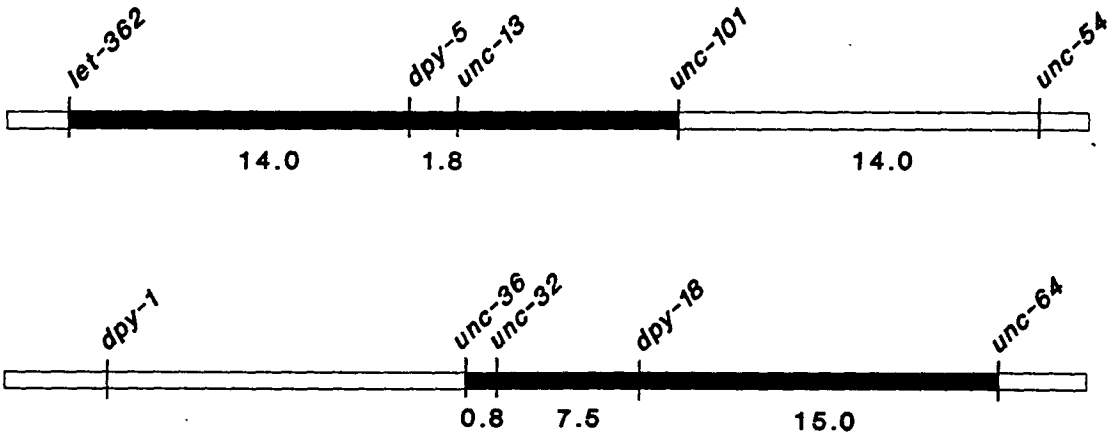


Figure 31. Recombination suppression in *hT2* heterozygotes.
Partial genetic map of LGI (top) and LGIII (bottom). Regions that do not recombine are shown in black. Wild-type genetic map distances shown below each chromosome are from Edgley and Riddle, 1990.

Table 12. Genetic screens using *hT2*.

Maternal Strain	Mutagen	Chromosomes screened	Putative *bli-4* alleles identified
KR1274	Formaldehyde	5,000	None
KR1274	Gamma radiation	31,000	*h665, h666*
KR1274	EMS	40,000	*h667, h668, h669, h670*
KR1401	Spontaneous	Spontaneous	*h862*

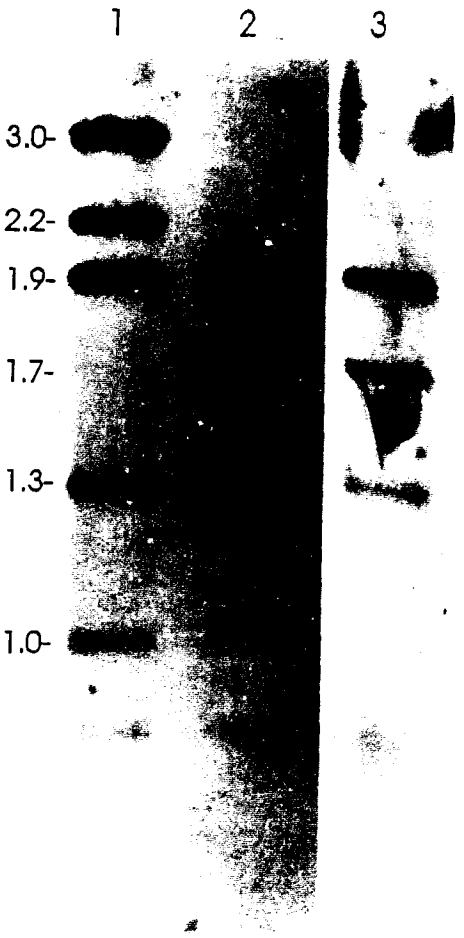


Figure 32. Southern analysis of *bli-4(h862)*
*Eco*RI digested genomic DNA probed with pCeh181. Lane 1: N2; lane 2: CB937 *bli-4(e937)*; lane 3: KR1803 *bli-4(h862)*. pCeh181 detects bands at 3.0 kb and 2.2 kb in N2 DNA that are absent in both CB937 and KR1803. A novel 1.7 kb band is present in both CB937 and KR1803.

APPENDIX B. STRAINS USED IN THIS STUDY

Table 13. Strains[a] used in this study (excluding *hP5* mapping strains, presented in Table 5)

Strain	Genotype
KR2301	*bli-4(e937) I; lin-14 (n179ts) X*
KR1922	*bli-4(e937) I lin-29(n1440) II*
KR1000	*unc-11(e47) dpy-5(e61) dpy-14 (e188) unc-13(?)/ (hDf8)*
BC0069	*dpy-14(e188) unc-13(e51) I*
BC0189	*unc-13 (e450) I*
BC0835	*unc-13 (e51) I; dpy-18(e364) III*
CB0001	*dpy-1(e1) III*
CB0014	*dpy-6(e14) X*
CB0024	*sqt-3(e24) V*
CB0027	*dpy-3(e27) X*
CB0128	*dpy-10(e128) II*
CB0130	*dpy-8(e130) X*
CB0164	*dpy-17(e164) III*
CB0184	*dpy-13(e184) IV*
CB0364	*dpy-18(e364)*
CB0424	*dpy-9(e424) IV*
CB0933	*unc-17(e933) IV*
CB0973	*bli-4(e937) I*
CB1214	*unc-15 (e1214) I*
KR0001	*Bristol (N2)*
KR0003	*Bergerac (BO)*
KR0081	*unc-74(e883) dpy-5(e61) I*
KR0214	*unc-74(e883) unc-13(e51) I*
KR0290	*sDp2/dpy-5(e61)(e51) bli-4(h42) unc-13(e450)*
KR0513	*sDp2/dpy-5(e61) bli-4(h199) unc-13(e450) I*
KR0573	*sDp2/let-77 dpy-14(e188) unc-13(e450) I*
KR0709	*sDp2/dpy-5(e61) bli-4(h384) unc-13(e450) I*
KR0752	*sDp2/dpy-5(e61) bli-4(h427) unc-13(e450) I*
KR0841	*sDp2/dpy-5(e61) bli-4(h520) unc-13(e450) I*
KR0870	*let-77 dpy-14(e188) unc-13(e450) I/ dpy-5(e61) unc-13(e450) I*
KR0912	*dpy-5(e61) bli-4(e937) unc-13(e450) I*
KR1025	*dpy-5(e61) bli-4(e937) I*
KR1123	*bli-4(e937) unc-13(e450)*
KR1234	*hT2 I; hT2[bli-4(e937)]*
KR1235	*unc-13(e51)/hT2 I; dpy-18(e364)/hT2[bli-4(e937)]III*
KR1274	*unc-13(e51)/hT2 I; dpy-18(e364)/ hT2[let-X(h661) bli-4(e937)]III*
KR1407	*unc-74(e883) unc-13(e51)/hT2 I;+ / hT2[let-X(h661) bli-4(e937)]III*
KR1340	*sDp2/dpy-5(e61) bli-4(h699) unc-13(e450) I*
KR1395	*sDp2/dpy-5(e61) bli-4(h754) unc-13(e450) I*
KR1395	*sDp2/dpy-5(e61) bli-4(h754) unc-13(e450) I*
KR1434	*sDp2/dpy-5(e61) bli-4(h791) unc-13(e450) I*
KR1822	*unc-63(e384) unc13(e51) (I); mut-6*
KR1844	*lin-14(n129ts) him-5*
KR1845	*lin-29 (n1440)/C1*
KR1858	*szT1/unc-63(384) bli-4(h1010) unc-13(e51) I*
RW7096	*mut-6(st702) unc-22(st192::Tc1)*
RW7097	*mut-6(st702)*

[a]Strains and allele designations:

KR	*h*	A.M. Rose, University of British Columbia, Vancouver, BC.
CB	*e*	MRC-LMB, Cambridge, England
BC	*s*	D.L. Baillie, Simon Fraser University, Burnably, BC.
RW	*st*	R. Waterston, Washington University, St.Louis, MO

APPENDIX C. BACTERIAL STRAINS.

Table 14. Bacterial Strains

E. coli Strain	Use	Origin
OP50	Nematode food	Brenner, 1974
BB4	λ-Zap host cells	Bullock, Fernandez and Short, 1987
DH5α	Plasmid host cells	Hanahan, 1983; Brethesda Research Laboratories

APPENDIX D. COSMID AND PLASMID CLONES

Table 15. Cosmid Clones

Clone[a]	Vector[b]	Antibiotic resistance
BO480	pJB8	Ampicillin
C27D2	pJB8	Ampicillin
C32G12	pJB8	Ampicillin
C40A4	pJB8	Ampicillin
C44D11	pJB8	Ampicillin
C44D11	pJB8	Ampicillin
C48E7	pJB8	Ampicillin
K04F10	Lorist 2	Kanamycin
K06E6	Lorist 2	Kanamycin
T22C4	Lorist 2	Kanamycin
ZC308	pJB8	Ampicillin

[a]All cosmids were a gift from J. Sulston and A. Coulson of the Medical Research Council, Cambridge, England (Coulson *et al.*, 1986; Coulson et al. 1988).

[b]pJB8 is described in Ish-Horowicz and Burke (1981). Lorist 2 is described in Cross and Little (1986) and Gibson *et al.* (1987).

Table 16. Plasmid clones

Plasmid clone	Insert size	Insert site	Insert source	Vector[a]
pCeh180	11.6 Kb	*Xho*I	K04F10	BS KS (M13+)
pCeh181	11.0	*Xho*I	K04F10	BS KS (M13+)
pCeh182	2.4	*Xho*I	K04F10	BS KS (M13+)
pCeh195		*Eco*RI	cDNA[b]	BS SK (M13-)
pCeh196		*Eco*RI	cDNA	BS SK (M13-)
pCeh197		*Eco*RI	cDNA	BS SK (M13-)
pCeh198		*Eco*RI	cDNA	BS SK (M13-)
pCeh199		*Eco*RI	cDNA	BS SK (M13-)
pCeh200		*Eco*RI	cDNA	BS SK (M13-)
pCeh201	6.4	*Pst*I/*Xho*I	pCeh181	BS KS (M13+)
pCeh202	4.6	*Pst*I/*Xho*I	pCeh181	BS KS (M13+)
pCeh205	1.7	*Eco*RI	pCeh181	BS KS (M13+)
pCeh206	0.8	*Eco*RI/*Sal*I	CB937 genomic library[c]	BS KS (M13+)
pCeh207	3.3	*Eco*RI	pCeh180	BS KS (M13+)
pCeh210	0.6	*Eco*RI	pCeh180	BS KS (M13+)

[a]Bluescript vectors are abrievated BS. SK and KS refer to the orientation of the cloning site. "+" and "-" indicates the orientation of the M13 intragenic region.

[b]cDNA clones were isolated from the Barstead and Waterston (1988) lamda-Zap library using the 1.3 kb *h1010* probe.

[c]pCeh206 was isolated from a lamda-Zap library of *Eco*RI digested CB937 genomic DNA.

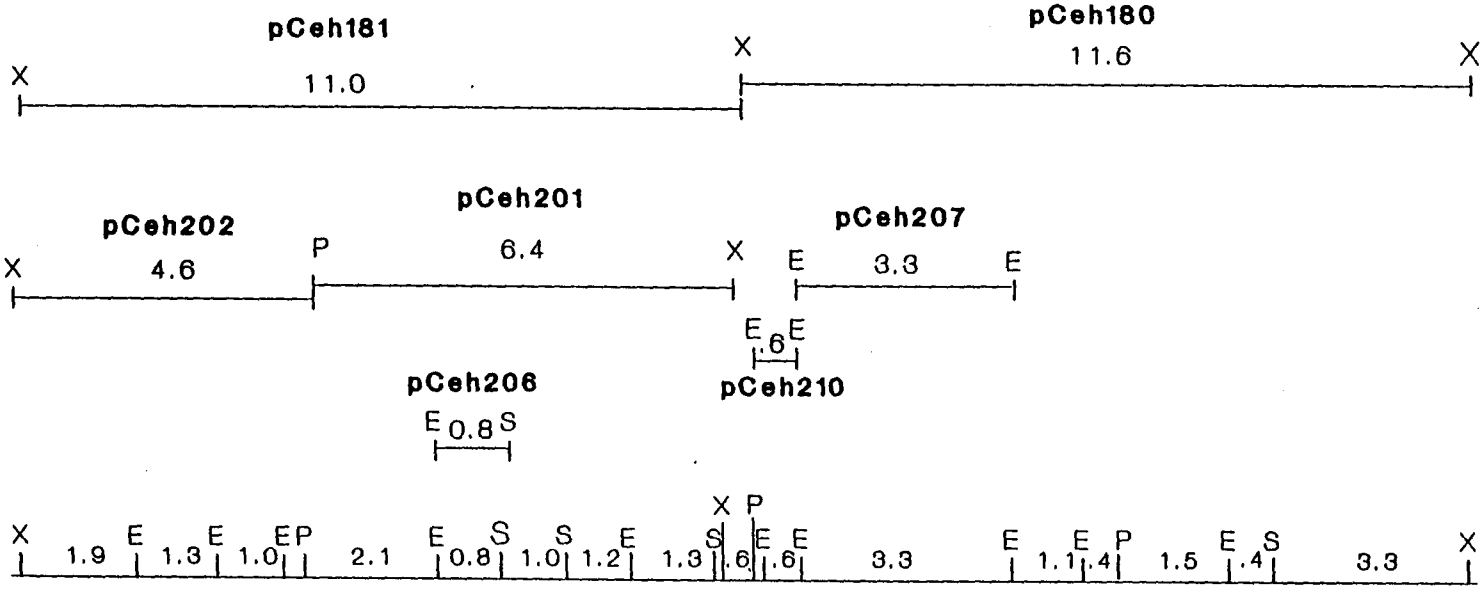


Figure 33. Genomic subclones of the *bli-4* region
Genomic restriction map (bottom) of 21 kb of K04F10. Subclones are shown above the map and are labeled with plasmid names labeled. Numbers above lines indicate restriction fragment sizes in kb. E=*Eco*RI; S=*Sal*I; P=*Pst*I; X=*Xho*I.

APPENDIX E. VON HEIJNE PREDICTION OF BLISTERIN SIGNAL PEPTIDASE SITE.

39 amino acids at the amino terminal of blisterin B are presented at the top of this figure. A probable site for cleavage by signal peptidase was detected using the method of VonHeijne (1983). In this method, the first four consecutive hydrophobic amino acid residues are identified starting at position i, and used to define a window, i +12 to i + 20. The probability of any the amino acids within this window occuring at positions +1 to -5 in the signal peptidase recognition site was obtained from von Heijne's table. The mutiple of the probabilities for each amino acid at each possible cleavage site within the window was determined, and is presented as a probability in the table below. Using this method, the signal peptidase was predicted to cleave after Cys_{29}.

```
                                    i+12      i+20
MRISIGRIAWQILAVLIAVAFTIEHDSICDESIGACGEP
```

+1	-1	-2	-3	-4	-5	Probabiltiy
E_{24}	I	T	F	A	V	
1.0	0.0	0.6	1.0	1.0	0.7	0.0
H_{25}	E	I	T	F	A	
1.0	0.0	1.0	3.0	1.0	1.0	0.0
D_{26}	H	E	I	T	F	
1.0	0.0	1.0	1.0	1.0	0.7	0.0
S_{27}	D	H	E	I	T	
1.0	0.0	2.0	0.0	1.0	1.0	0.0
I_{28}	S	D	H	E	I	
1.0	1.0	1.0	0.0	1.0	0.7	0.0
C_{29}	I	S	D	H_{25}	E	
1.0	0.0	0.6	0.0	1.0	1.0	0.0
D_{30}	C	I	S	D	H	
1.0	1.0	1.0	3.0	1.0	1.0	3.0
E_{31}	D	C	I	S	D	
1.0	0.0	0.6	1.0	1.0	1.0	0.0
S_{32}	E	D	C	I	S	
1.0	0.0	1.0	3.0	1.0	1.0	0.0