CLONING AND CHARACTERIZATION OF THE *CAENORHABDITIS ELEGANS* HISTIDYL-tRNA SYNTHETASE GENE.

by

Yousef Grera Amaar

B.Sc., University of Victoria, 1989

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Biological Sciences

© Yousef Grera Amaar

Simon Fraser University

March, 1992

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

APPROVAL

Name: **YOUSEF GRERA AMAAR**

Degree: **Master of Science**

Title of Thesis:

CLONING AND CHARACTERIZATION OF THE CAENORHABDITIS ELEGANS HISTIDYL-tRNA SYNTHETASE GENE.

Examining Committee:

Chairman: Dr. N. Haunerland

Dr. D.L. Baillie, Professor,
Department of Biological Sciences, SFU

Dr. B.P. Brandhorst, Professor,
Department of Biological Sciences, SFU

Dr. T. Borgford, Assistant Professor,
Department of Chemistry, SFU

Dr. B. Honda, Associate Professor,
Department of Biological Sciences, SFU
Public Examiner

Date Approved April 3 1992

PARTIAL COPYRIGHT LICENSE

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

Title of Thesis/Project/Extended Essay

Cloning and Characterization of the Caenorhabditis elegans Histidyl-tRNA Synthetase Gene.

Author: ______________
(signature)

Yousef Amaar
(name)

April 2 '92
(date)

Abstract

The *C. elegans* histidyl-tRNA synthetase gene has been cloned and sequenced. The complete genomic sequence and most of the cDNA sequence of this gene is now determined. The gene size including flanking and coding regions is 2200 nucleotides. Three small introns (45-50 bp) are found to interrupt the open reading frame. The open reading frame translates to 523 amino acids. This putative protein sequence shows extensive homology with the human and yeast histidyl-tRNA synthetases. The *C. elegans* histidyl-tRNA synthetase gene maps on Linkage Group IV (LGIV) to 0.3 map units interval between *mec-3* and *lin-3* within the deficiency *sDf2* in the *unc-22* region. On the molecular map this gene falls in a region defined by four overlapping yacs in the *unc-22* region. The histidyl-tRNA synthetase gene is a single copy gene. Hence, it is very likely that it encodes both the cytoplasmic and the mitochondrial histidyl-tRNA synthetases. It may also be trans-spliced since it contains a putative trans-splice site in its 5' untranslated region.

Acknowledgment

I would like to thank my senior supervisor Dr. Baillie for giving me the chance to work in his laboratory and also for his guidance and support. I would like to thank all of my lab mates and friends for their help and support. I would like to thank Dr Kim McKim for his help with the PCR experiment and Dr. B. Waterston for kindly providing the genomic mapping filter.

TABLE OF CONTENTS

Approval page....ii

Abstract....iii

Acknowledgement....iv

List of tables and figures....vii

Introduction....1

Materials and Methods.

Maintaining Nematodes on NGM plates....10

Picking random clones from a *C. elegans* Lambda zap cDNA library....10

Screening the Lambda zap cDNA library....10

Genetic and molecular mapping of the gene....10

Preparation of plasmid DNA....11

Preparation of cosmid DNA....11

Restriction enzyme digestion....11

Agarose electrophoresis....12

Southern blots....12

Preparation of DNA fragments for oligolabelling....12

Labelling of DNA probes....12

Hybridization of probes to DNA filters....13

Subcloning....13

Oligonucleotides....14

Sequencing Methods....14

Results.

Recovery of a cDNA fragment corresponding to the *C. elegans* histidyl-tRNA synthetase gene....16

Screening a *C. elegans* cDNA library....16

Hybridization of the 0.42 kb cDNA fragment to a *C. elegans* *Eco*R1 genomic DNA digest....18

Mapping of the *C. elegans* histidyl-tRNA synthetase gene on LGIV....18

Cosmid CO4G1 contains the *C. elegans* histidyl-tRNA synthetase gene....23

Partial cDNA sequence of the histidyl-tRNA synthetase gene....26

The genomic DNA sequence of the histidyl-tRNA synthetase gene....26

The putative protein sequence of the histidyl-tRNA synthetase gene....30

Discussion.

The *C. elegans* histidyl-tRNA synthetase gene is a single copy gene....34

Mapping the *C. elegans* histidyl-tRNA synthetase gene....35

Analysis of the genomic DNA sequence of the *C. elegans* histidyl-tRNA synthetase gene....36

Protein Sequence analysis....38

Summary....41

Proposals for future work....42

References....43

List of Tables.

Table 1. Intron positions and sequence......29

Table 2. Pairwise comparisons of *C. elegans*, *E. coli*, yeast, and human histidyl-tRNA synthetase protein sequences......31

Table 3. Motif positions and sequences......32

List of Figures.

Figure 1. Hybridization of the 0.42 kb cDNA to two cDNA fragments......17

Figure 2. Hybridization of the 0.42 kb cDNA fragment to C. elegans *Eco*R1 genomic digest......19

Figure 3a. Hybridization of the 0.42 kb *Eco*R1 cDNA fragment to a genomic mapping filter......20

Figure 3b. yac map in the *unc*-22 region......21

Figure 4. Genetic map of the region around *unc*-22......22

Figure 5. Hybridization of the 0.42 kb cDNA fragment to an *Eco*R1 digest of cosmid CO4G1......24

Figure 6. Ethidium bromide stained gel pattern *Eco*R1 digest of the 5.5 kb genomic fragment subcloned from cosmid CO4G1......25

Figure 7. The overlap point between the 0.42 kb and the 1.5 kb cDNA fragments......27

Figure 8. Nucleotide and protein sequences of the *C. elegans* histidyl-tRNA synthetase gene......28

Figure 9. Multiple alignment of the *C. elegans*, yeast, and human histidyl-tRNA synthetases......33

INTRODUCTION

Aminoacyl-tRNA synthetases are a family of enzymes which make up a vital component of the protein synthesis machinery. These enzymes transfer amino acids to their cognate tRNA through a two step reaction. First, the enzyme catalyzes the binding of the appropriate amino acid to adenosine triphosphate (ATP) to produce an amino acid adenylate complex. Secondly, the enzyme catalyzes the transfer of the amino acid to its cognate tRNA (Neidhardt *et al*, 1975). There are 20 distinct aminoacyl-tRNA synthetases in every living cell, one for each naturally occurring amino acid. However, an exception to this has been reported in *Bacillus subtilis* and barley chloroplasts where $tRNA^{Glu}$ synthetase has been found *in vivo* to misacylate $tRNA^{Gln}$ (tRNA for Glutamine) with the amino acid glutamic acid (Glu) (Lapointe *et al*, 1986; Schon *et al*, 1988). The conversion of the intermediate Glu-$tRNA^{Gln}$ to Gln-$tRNA^{Gln}$ requires a transamidation step (Neidhardt *et al*, 1975).

All synthetases from *Escherichia coli* as well as most of those from lower and higher eukaryotes have been purified and characterized (Marinde, 1990). On the basis of biochemical function and similarities, synthetases have been classified into four groups according to amino acid codons as follows; XAX-(TYR, HIS, GLN, LYS, ASN, ASP, and GLU), XGX-(CYS, TRP, ARG, and GLY), XCX-(SER, PRO, THR, and ALA), and XUX-(LEU, ILE, VAL, MET, and PHE) groups. It has been proposed that synthetases which belong to XUX- and XCX-codon groups have evolved more recently than those synthetases which belong to XAX- and XGX-codon groups (Watzel, 1978).

Cloning and sequencing of synthetases.

The primary sequences of more than 20 synthetases from prokaryotes have been determined via gene cloning and sequencing (Marinde, 1990). These synthetases are encoded by single distinct genes with the exception of the *E. coli* lysyl- and the *Bacillus*

subtilis threonyl-tRNA synthetases. The *E. coli* lysyl-tRNA synthetase has been found to be encoded by two distinct genes; one gene expresses the enzyme constitutively (Lys S) and the other gene has been found to be heat-inducible (Lys U). The Lys S and Lys U genes have been mapped at 62.1 and 92 min on the *E. coli* chromosome, respectively (Emmerich *et al*, 1987; Van Bogelen *et al*, 1983). In *Bacillus subtilis* two genes, Thr Sy and Thr S2, encode functional threonyl-tRNA synthetases (Putzer *et al*, 1990).

The primary sequence of more than 18 synthetases from lower and higher eukaryotes have been reported (Marinde, 1990). Eukaryotes, unlike prokaryotes, have three sets of synthetases; a cytoplasmic, a mitochondrial, and a chloroplast set (Schneller *et al*, 1978). Most of the animal eukaryotic aminoacyl-tRNA synthetases have been found to be encoded by distinct genes for cytoplasmic and mitochondrial synthetases (Schneller *et al*, 1978). However, in yeast, single nuclear genes have been found to code for both the cytoplasmic and the mitochondrial histidyl- and valyl-tRNA synthetases (Natsoulis *et al*, 1986; Chatton *et al*, 1988). The cytoplasmic and the mitochondrial synthetases are encoded by two classes of transcripts initiating at distinct in-frame start codons (Natsoulis *et al*, 1986; Chatton *et al*, 1988).

Exon / Intron organization of synthetases.

Aminoacyl-tRNA synthetase genes sequenced from lower eukaryotes to date have been found to be devoid of introns except for the cytoplasmic leucyl-tRNA synthetase gene and the mitochondrial tyrosyl and leucyl-tRNA synthetase genes from *Neurospora crassa*. Each of these genes has been found to contain one small intron (60-64 bp) (Chow *et al*, 1989a; Chow *et al*, 1989b; Akins and Lambowitz, 1987). In contrast, higher eukaryotic synthetase genes have been found to contain several exons flanked by large introns. Two genomic sequences corresponding to hamster histidyl- and human aspartyl-tRNA synthetases have been reported (Tsui and Siminovitch, 1987; Jacob-Molina *et al*, 1989). The hamster histidyl-tRNA synthetase gene has been found

to contain 13 exons and 12 introns which span 18 kb of the genome; and the human aspartyl-tRNA synthetase gene has been found to contain 16 exons which span 40 kb of the genome (Marinde, 1990).

Structure of synthetases.

Aminoacyl-tRNA synthetases have been found to exhibit structural variety. Prokaryotic and eukaryotic synthetases exist as monomers, homologous dimers, and tetramers of alpha2, alpha2 beta2 types (Marinde, 1990). Throughout prokaryotic and eukaryotic evolution, the subunit structure of homologous enzymes appears to be conserved. Exceptions are glycyl-, methionyl-, and alanyl-tRNA synthetases which in prokaryotes exist as alpha2 beta2 (tetramers), alpha2 (dimers), and alpha2 (dimers), respectively, while in eukaryotes they exist as alpha2 (dimers), alpha (monomers), and alpha (monomers), respectively (Marinde, 1990).

In addition to quaternary structure differences, aminoacyl-tRNA synthetases have been found to differ in subunit molecular weight. Prokaryotic enzymes have lower subunit masses than their homologs from eukaryotes (Schimmel, 1987). The difference in the subunit masses has been attributed to the fact that eukaryotic synthetases have a chain extension at the amino terminus. Evidence in support of this hypothesis has been obtained from experiments in which synthetases from lower and higher eukaryotes were exposed to limited proteolysis (Cirakoglu and Waller, 1985). Limited proteolysis of yeast lysyl- and aspartyl-tRNA synthetases, mammalian tryptophanyl-, methionyl-, lysyl-, and arginyl-tRNA synthetases gave fully active truncated proteins that have similar molecular weights to their homologs from prokaryotes (Lemaire *et al*, 1975; Kellermann *et al*, 1978; Siddiqui *et al*, 1985; Cirakoglu and Waller, 1985; Vellekamp *et al*, 1985). The differences in subunit masses can also be explained in part by the presence of internal sequences that link functional domains of the eukaryotic synthetases. Deleting an internal sequence that

joins two parts of the nucleotide fold in isoleucyl-tRNA synthetase has been shown to have no effect on the enzyme's activity (Starzyk *et al*, 1987).

Partitioning of synthetases into two classes

Primary sequence comparisons among aminoacyl-tRNA synthetases from prokaryotes and eukaryotes have resulted in dividing synthetases into two classes based on the presence or absence of certain motifs (Eriane *et al*, 1990). Class I synthetases possess the consensus sequence HIGH (His-Ile-Gly-His) and KMSKS (Lys-Met-Ser-Lys-Ser) found on the amino and carboxy termini of the enzyme, respectively (Schimmel, 1987). The polarity of class I synthetases has been shown to be as follows NH2---HIGH----KMSKS---COOH. The catalytic domain of class I synthetases is based on a Rossman fold comprised of parallel beta strands and connecting helices with both the HIGH and KMSKS motifs close to the ATP binding site (Marinde, 1990). Class I synthetases include the enzymes for methionine, isoleucine, arginine, cystine, leucine, tyrosine, tryptophan, valine, glutamine, and glutamic acid (Eriani *et al*, 1990).

Class II synthetases do not possess either the HIGH or the KMSKS sequence motifs. They are characterized by three signature motifs denoted 1, 2, and 3 that have been identified in the enzymes for aspartic acid, lysine, histidine, asparagine, phenylalanine, proline, and threonine. Enzymes for alanine and glycine have been found to contain only motif 3 (Eriani *et al*, 1990). Each motif has been found to contain a highly conserved core with an invariant residue: goxxoxxPoo for motif 1; (F/Y/H)Rx(E/D) followed 4 to 14 residues away by (R/H)xxxFxxx(D/E) for motif 2; and *xogogoeRooooo for motif 3. The definition of symbols is as follows; x is for any amino acid, single letter amino acid with lower case is for residues conserved in at least 11 out of 17 possible sequences, * is for small amino acids (P,G,S, and T), o is for hydrophobic residues (F, Y, I, L, W, V, M, and A) (Eriani *et al*, 1990). These three

motifs are found to be involved in building the catalytic domain which is based on an antiparallel fold (Cusack *et al*, 1990).

In addition, class II synthetases aminoacylate on the 3' OH of the ribose, unlike class I synthetases which aminoacylate on the 2' OH of the ribose on the 3' end of tRNA (Eriani *et al*, 1990). Class I and class II synthetases hence can be distinguished at the structural (having different catalytic domains) and functional levels. Eriani (1990) points out that these two classes have arisen from two different ancestors. Class II synthetases also differ in the order of their functional domains from the amino terminus to the carboxy terminus from class I synthetases as well as from one another (Burbaum and Schimmel, 1991).

Homology among synthetases

Comparisons of homologous enzymes from a number of prokaryotes have shown extensive primary structure homologies. For instance, tyrosyl-tRNA synthetase from *B. caldotenax* and *B. stearothermophilus* are distinguished by only four amino acid substitutions (Jones *et al*, 1986; Winter and Hartely, 1978). Comparisons of homologous enzymes from different eukaryotes have shown significant protein sequence similarities, up to 46% identity in the case of histidyl-tRNA synthetase (Tsui and Siminovitch, 1987). Indeed, similarity between the same enzyme from prokaryotes and eukaryotes have also been found. *E. coli* and yeast glutaminyl-tRNA synthetases are found to share 47.9% identity at the amino acid level (Schimmel, 1987; Hoben *et al*, 1982). It has also been shown that *E. coli* tyrosyl-tRNA synthetase can substitute for *Saccharomyces cerevisiae* mitochondrial tyrosyl-tRNA synthetase provided that the import leader sequence is fused to the *E. coli* enzyme (Edwards and Schimmel, 1987). This implies that the *E. coli* and yeast mitochondrial synthetases are functionally equivalent.

In contrast, cytoplasmic synthetases from bacteria and yeast differ in size. For instance, yeast glutamine-tRNA synthetase is 309 amino acids larger than its counter part from *E. coli*. The size difference is due to an extension of the amino terminus (Schimmel, 1987). Hence, the extension sequence may have an in vivo role other than aminoacylation. It has been suggested that the extension sequences may function as a constraint for specific recognition of tRNA sequences characteristic of a certain organism (Schimmel, 1987).

Eukaryotic synthetases bind polyanionic carriers.

Eukaryotic synthetases, unlike prokaryotic synthetases, display the property of binding to polyanionic carriers through ionic interactions. The polypeptides of eukaryotic synthetases are shown to have a basic isoelectric point (pI) (Mirande and Waller, 1988, Jordana *et al*, 1987). This alkaline pI is thought to be responsible for the binding of synthetases to polyanionic carriers. This polyanionic binding property may have been acquired to promote the compartmentalization of eukaryotic synthetases near the site of protein synthesis through electrostatic interactions with negatively charged cellular compartments (Cirakoglu *et al*, 1985).

Evidence in support of the polyanionic property has been obtained from experiments in which the same synthetases were subjected to controlled proteolysis. Yeast lysyl- and aspartyl-tRNA synthetases have been shown to interact with immobilized heparin. While controlled proteolysis of the two synthetases decreased their affinity to interact with immobilized heparin, their aminoacylation activity was not affected. Thus, the polyanionic binding property is conferred by a domain distinct from the catalytic domain of eukaryotic synthetases (Cirakoglu *et al*, 1985). It has also been suggested that the polyanionic domain is located on one side and the catalytic domain on the other side of the protein. This would enable the enzyme to attach to the

polyanionic component using one side without impairing the enzyme from interacting with its cognate tRNA (Marinde, 1990).

Furthermore some eukaryotic synthetases are found to participate in other roles such as splicing of mRNA. For instance, the yeast mitochondrial leucyl-tRNA synthetase has been found to be involved in splicing of the introns bI4 and aI4 of the cytochrome b gene and subunit 1 of the cytochrome oxidase gene (Tzagoloff *et al*, 1988, Labouesse and Herbert, 1987, Herbert *et al*, 1988). The mitochondrial tyrosyl-tRNA synthetase from *Neurospora crassa* has also been found to function in splicing intron I of the cytochrome b gene (Majumder *et al*, 1989). The mechanism of mRNA splicing by these two synthetases is as yet unknown. Eukaryotic synthetases have also been found to function in transcriptional and translational control (Putney and Schimmel, 1981; Chow and RajBhandary, 1989; Mechulam *et al*, 1985; Cheung *et al*, 1985).

The purpose of this thesis was to clone and characterize the *C. elegans* histidyl-tRNA synthetase gene. This work is of interest is because (i) although aminoacyl-tRNA synthetases have been investigated in a number of prokaryotes and eukaryotes, almost nothing is known about them in *C. elegans*, (ii) cloning and characterizing of this gene will contribute to the ongoing effort in our laboratory to determine the genomic organization of a region on LGIV defined by the deficiency *sDf2*, (iii) cloning and characterizing this gene will also be useful for identifying conserved domains and hence gaining a better understanding of the evolution of synthetases.

C. elegans is a multicellular organism which has a small genome size of 1×10^8 bp (five autosomes and one sex chromosome) of which 83% is single copy (Sulston and Brenner, 1974). *C. elegans* is a self-fertilizing hermaphrodite. Males are also found which are capable of fertilizing hermaphrodites and hence stocks can be produced by genetic crosses (Brenner, 1974). Mutations can be easily generated and mutants can be

analyzed to identify the genes involved. In addition, its complete cell lineage of approximately 1,000 cells has been described (Sulston and Horvitz, 1981 and Sulston *et al*, 1983). Its physical map is approaching completion very rapidly (Coulson *et al*, 1986 and 1988). The nematode *C. elegans* will make a good model for studying mutants of aminoacyl-tRNA synthetase genes.

In this thesis, I report the complete genomic sequence as well as most of the cDNA sequence of the *C. elegans* histidyl-tRNA synthetase gene. The size of the gene including the flanking and the coding sequences is 2200 nucleotides long. Three small introns (45-50 bp long) have been found to interrupt the open reading frame. This gene encodes a putative protein of 523 amino acids. The deduced protein sequence shows extensive homology with the human and yeast histidyl-tRNA synthetases. In addition, the histidyl-tRNA synthetase gene has been mapped genetically on LGIV within the deficiency *sDf2* between *mec-3* and *lin-3*. It has also been mapped molecularly in a region defined by four overlapping YACs. The gene most likely encodes both the cytoplasmic and mitochondrial synthetases since it is a single copy gene. It may also be trans-spliced since it contains a trans-splice site in its 5' untranslated region.

MATERIALS AND METHODS

Caenorhabditis elegans var Bristol, strain N2 was the organism used in this study. The nematodes were maintained on NGM plates supplemented with *E. coli* strain OP50 as a food source (Brenner, 1974).

Picking random clones from a *C. elegans* Lambda Zap cDNA library.

A *C. elegans* Lambda Zap cDNA library (Stratagene) containing reversed transcribed poly (A) RNA of mixed stages was provided by R. Barstead and R. Waterston, Washington University, Missouri; a titre of about 2×10^9 was used. About ten thousand phage were plated according to the Stratagene protocol. Fifty phage were randomly picked and plasmids of twenty-nine of them were excised, again using the Stratagene protocol. Twenty-three of these plasmids contained cDNA inserts ranging from 400 to 1600 bp. Plasmid # 29 which contained a 420 bp cDNA insert was chosen for sequencing.

Screening The Lambda Zap cDNA library.

A *C. elegans* Lambda Zap cDNA library was screened with the 420 bp cDNA fragment contained in plasmid # 29. Three rounds of screening were carried out and the plasmids from two phage isolated were excised according to the Stratagene protocol. The two plasmids obtained were called # 3-1 and # 6-1 and contained 1.5 and 1.4 kb cDNA inserts, respectively.

Genetic and molecular mapping of the gene.

A genomic mapping filter provided by R. Waterston was probed with the 420 bp cDNA fragment. This allowed mapping of the gene genetically and molecularly. It also identified the cosmid which contains the gene.

Preparation of plasmid DNA.

Plasmid DNA was prepared using a modified alkaline lysis protocol (Pharmacia miniprep kit plus). The DNA extracted with this method was used for both restriction digests and sequencing reactions.

Preparation of cosmid DNA.

Cosmid DNA was prepared using a protocol obtained from J. Sulston adapted from Maniatis *et al* (1982). This protocol is the same as described in Maniatis up to step 15. Then 50 ul of 4.4 M LiCl was added to 50 ul of DNA solution, mixed, and left at 4°C for two hours. It was then spun for 5 minutes and 90 ul of the supernatant was transferred to another Eppendorf tube. Then 180 ul of 95% ethanol was added, mixed, and left at -20°C for one hour. This was spun for 5 minutes, and the pellet was washed with 70% ethanol and drained. The pellet was resuspended in 100 ul of 0.1 M potassium acetate (KoAc) (pH 7.4) and 200 ul of 95% ethanol was added and stored at -20°C for one hour. The DNA was spun down for 5 minutes and the pellet was washed with 70% ethanol and vacuum dried, then the pellet was resuspended in 50 ul of 1X RNase free TE.

Restriction enzyme digestion.

All restriction enzymes were purchased from either Bethesda Research Laboratories (BRL) or from Pharmacia. The amount of restriction enzyme used per reaction was about 2 units of enzyme per 1 ug of DNA. The digestion reactions were carried out at 37°C for at least one hour. The reaction buffers used were those recommended by the manufacturer.

Agarose electrophoresis.

Genomic and cloned DNA digested to completion was loaded on 0.7% agarose gels and ran in 0.5X TBE buffer(10X TBE is 89 mM Tris-borate, 89 mM boric acid and 2.5 mM EDTA pH 8.3) as described by Maniatis (1982). DNA was visualized under ultraviolet light by addition of ethidium bromide, 1 ul (10_{ng}\ul) per 100 ml of agarose solution, before pouring the gel. Gels were photographed using GRAFLEX camera.

Southern blots.

The gel containing the DNA to be transferred was soaked in four volumes of 0.25 M HCl for one half hour, then washed with distilled water, and soaked in four volumes of 0.5 M NaOH, 1.5 M NaCl for one half hour. The gel was subsequently soaked in transfer buffer (1 M ammonium acetate, 0.02 M NaOH) for one hour. The transfer of DNA to a nitrocellulose (Schleicher and Schuell) or nylon (GENE SCREEN from Pharmacia) filter was carried out using the bidirectional method described by Smith and Summers (1978). When the transfer was complete, the filters were soaked in 2X SSPE(1X SSPE is 0.18 M NaCl, 0.01 M NaH_2PO_4, 1 mM disodium EDTA, pH 7.4; Davis *et al* (1980) and baked at 80^oC in a vacuum oven for 1 to 2 hours.

Preparation of DNA fragments for oligolabelling.

Cloned DNA for oligolabelling was prepared by applying the low melting Seaplaque GTG agarose gel protocol as described by FMC Bioproducts.

Labelling of DNA probes.

All of the probes used in this study were labelled with ^{32}P-dATP using the oligolabelling technique of Feinberg and Vogelstein (1983). Random hexamer primers were obtained from Pharmacia. The labelling reaction was stopped by the addition of 6 ul of 10X stop solution (200 mM EDTA, 5% SDS). Then the probe was put through a G-25 Sephadex spin column (obtained from Pharmacia). The specific activity of the probe was measured using liquid scintillation counter (Beckman) and the required amount of probe was used (one million counts per ml of hybridization buffer). The probe was filtered with the hybridization buffer through 0.45 um nitrocellulose filters (Schleicher and Schuell).

Hybridization of probes to DNA filters.

DNA filters were first prehybridized overnight in the hybridization buffer (5X SSPE, 0.3% SDS, 2.5% Denhardt's solution (1X Denhardt's is 0.02% BSA, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone) at 68°C. Then fresh filtered hybridization buffer with the probe was added and the hybridization was carried out at 68°C for overnight. The next day the filters were washed in 1X SSPE and 0.02% SDS buffer at 68°C. Four 15 minute washes were carried out and the filters were left to dry at room temperature. Kodak X-OMAT films were placed on the air dried filters with Dupont lighting plus intensifying screen and kept at -70°C overnight. The autoradiographs were developed the next day.

Subcloning.

Subcloning was conducted by utilizing the shotgun method. The ligation reactions were carried out as previously described by Snutch (1984). The

transformation procedure was carried out as described in the BRL transformation protocol. Transformed cells were detected as white colonies on agar plates containing ampicillin, X-gal(5-bromo-4 chloroindolyl-Beta-D-galactoside) and IPTG(isopropyl thiogalactoside).

Oligonucleotides.

Oligonucleotides were designed to obtain the cDNA and the genomic sequences. Nine oligonucleotides were designed from the cDNA sequence as it became available. The oligonucleotides were designed to obtain sequences corresponding to regions upstream and downstream from the 5' end of the gene. All nine oligonucleotides were designed using the Oligo program (Rychlik and Rhoads, 1989). The oligonucleotides were on average 18 nucleotides in length and were obtained from the Institute of Molecular Biology and Biochemistry, SFU. The oligonucleotides were synthesized using an ABI 391 DNA synthesizer according to the phosphormidite method of oligonucleotide synthesis. Before use, the dry oligonucleotides were suspended in 200 ul of distilled water and the concentration was measured by spectrophotometry. The appropriate dilutions were made and about 2.5 pmol were used for each sequencing reaction.

Sequencing method.

Nucleotide sequencing was done using Terry Snutch's modified version of the dideoxy termination method of Sanger *et al* (1977) (personal communication). This method was used in conjunction with the Sequenase version 2.0 kit (USB). First, 2-4 ug of dsDNA and 2.5 pmol of primer were mixed in a total volume of 8 ul, then 1 ul of DMSO (dimethyl sulfoxide) was added to the mixture tube. The mixture tube was incubated at 95°C for 3 minutes, and then placed in dry ice-95% ethanol bath for 5

minutes. The tube was subsequently quickly thawed, spun for 1 second, and 2 ul of 5X reaction buffer were added (200 mM Tris.HCl pH 7.5, 100mM $MgCl_2$, and 250mM NaCl). Then, the sample was placed at room temperature for 5 minutes. Second, 6.3 ul of labelling mix were added to the sample tube and the sample placed at room temperature for 5 minutes. The labelling mix was made up of 1 ul of DDT, 2 ul of 1X labelling mix (5X= 15uM dGTP, 7.5uM dCTP, 7.5uM dTTP), 0.8 ul of ^{35}S-dATP, 2 ul of diluted sequenase (1 ul of Sequenase in 6 ul of dilution buffer, 10mM Tris.HCl pH 7.5, 5mM DDT, 0.5/ml BSA), 0.5 ul DMSO. Third, 3.5 ul aliquots of the sample were added to the termination tubes containing 2.5 ul of ddNTPs. The mixture of ddNTPs is described in the Sequenase version 2 protocol. Then, the termination tubes were incubated at 37-42°C for 5 minutes. Fourth, the termination reactions were stopped by adding 4 ul of stop solution (95% Formamide, 20mM EDTA, 0.05% Bromophenol blue, 0.05% Xylene Cyanol FF). Finally, the reactions were incubated at 95°C for 5 minutes before loading onto a 6% polyacrylamide urea gel.

All DNA and protein sequence analyses were conducted using the following computer programs: (I) FASTA (W.R. Pearson and D.J. Lipman. (1988), (II) PC/Gene (the nucleic acid and protein sequence analysis software system: A. Bairoch/ University of Geneva; (TM) IntelliGenetics Inc. and Genofit SA.), (III) ESEE (The Eyeball Sequence Editor: created and implemented by E. Cabot/ Simon Fraser University).

RESULTS

(1) Recovery of a cDNA fragment that encodes a portion of the *C. elegans* his-tRNA synthetase gene.

A 420 bp *Eco*R1 fragment was serendipitously recovered from a *C. elegans* Lambda zap cDNA library of poly (A) RNA of mixed stages (provided by B. Barstead and R. Waterston, Washington University, Missouri) which was found to encode a portion of the *C. elegans* histidyl-tRNA synthetase. This fragment was sequenced; the 420 bp coding nucleotide sequence was translated and the predicted peptide sequence was used to search the EMBL and SWISS PROT 13 data banks. The predicted peptide showed extensive homology with the human histidyl-tRNA synthetase peptide sequence. The fragment showed 70.9% amino acid identity in a 177 amino acid overlap and 88% similarity when conservative changes were taken into consideration. The high degree of similarity suggested that this 420 bp cDNA fragment represents a portion of the coding sequence for the *C. elegans* histidyl-tRNA synthetase gene.

(2) Screening a *C. elegans* cDNA library for a larger cDNA fragment(s) using the 420 bp cDNA as a probe. .

The recovered 420 bp ^{32}P labelled cDNA fragment was used as a probe to screen the *C. elegans* Lambda Zap cDNA library in the hope of recovering a full length cDNA fragment(s). Approximately twenty thousand phage were screened resulting in the recovery of two cDNAs of 1.4 kb and 1.5 kb (figure 1). Both cDNA fragments were found to contain coding sequences corresponding to the *C. elegans* histidyl-tRNA synthetase gene, however neither fragment was a full length cDNA. The 1.5 kb cDNA fragment was chosen and used to continue the sequencing process.

Figure 1.

1(a) shows the ethidium bromide stained pattern of two cDNA fragments, recovered from a *C. elegans* lambda zap cDNA library, electrophoresed on a 0.7% agarose gel; M = 12 kb lambda ladder, A = *Eco*R1 digest of plasmid # 3-1 containing an insert of 1.5 kb, and B = *Eco R1* digest of plasmid # 6-1 containing an insert of 1.4 kb. 1(b) shows the hybridization of the ^{32}P labelled 0.42 kb cDNA fragment to the two cDNAs; the 1.4 kb (C) and 1.5 kb (D).

a
b
c
d
A
B
M
2.95
1.40
1.50

The nucleotide sequence from the 1.5 kb fragment was subsequently found to overlap with the sequence of the 420 bp cDNA fragment.

(3) Probing a *C. elegans* *Eco*R1 genomic DNA digest with the 420 bp cDNA *Eco*R1 fragment.

The search for the genomic DNA fragment containing the *C. elegans* histidyl-tRNA synthetase gene necessitated determination of the size of the genomic fragment(s) containing all or part of the gene. A Southern blot containing *C. elegans* *Eco*R1 digested genomic DNA, probed with the 420 bp ^{32}P labelled cDNA fragment resulted in probe hybridization to a 5.5 kb *Eco*R1 genomic fragment (figure 2), and suggests that the *C. elegans* histidyl-tRNA synthetase gene is probably present in a single copy.

(4) Mapping the *C. elegans* histidyl-tRNA synthetase gene on LGIV.

Results of a probe of the YAC genomic mapping filter allowed genetic and molecular mapping of the *C. elegans* histidyl-tRNA synthetase gene. The probe hybridized to four YACs; Y71D6, Y47F2, Y59E6, and Y69E1 (figure 3(a) and (b)). This result indicates the *C. elegans* histidyl-tRNA synthetase gene maps between *mec-3* and *lin-3* within the deficiency *sDf2* on LGIV (figure 4). The *C. elegans* histidyl-tRNA synthetase gene maps to a region which has been investigated in great detail in our laboratory (Clark *et al*, 1988; Clark, 1990; Moerman, 1980; Rogalski *et al*, 1982; Rogalski and Baillie, 1985).

Figure 2.

Hybridization of the ^{32}P labelled *Eco*R1 0.42 kb cDNA fragment to *Eco*R1 digested *C. elegans* genomic DNA (N2). The probe detected a 5.5 kb *Eco*R1 genomic fragment. The filter was washed and exposed to a Kodak film at -70°C for one day.

N2

—5.5

Figure 3a.

Hybridization of the ^{32}P labelled 0.42 kb *Eco*R1 cDNA to a *C. elegans* genomic mapping filter. The probe detected four adjacent YACs; (1) Y171D6, (2) Y47F2, (3) Y59E6, and, (4) Y69E1. The filter was washed and exposed to Kodak film at -70°C overnight.

1 2 3 4

Figure 3b.

The position of each of the detected four YACs; Y171D6, Y47F2, Y59E6, and Y69E1 is shown on the YAC map around *unc-22* on LGIV in *C. elegans*. The cosmid CO4G1 which contains the histidyl-tRNA synthetase gene is also shown.

Y47F2

Y69E1

Y71D6

Y59E6

Y21D8

CO8C5

BO364

F13E9

RO4H8

CO6G8

T11G6

BO521

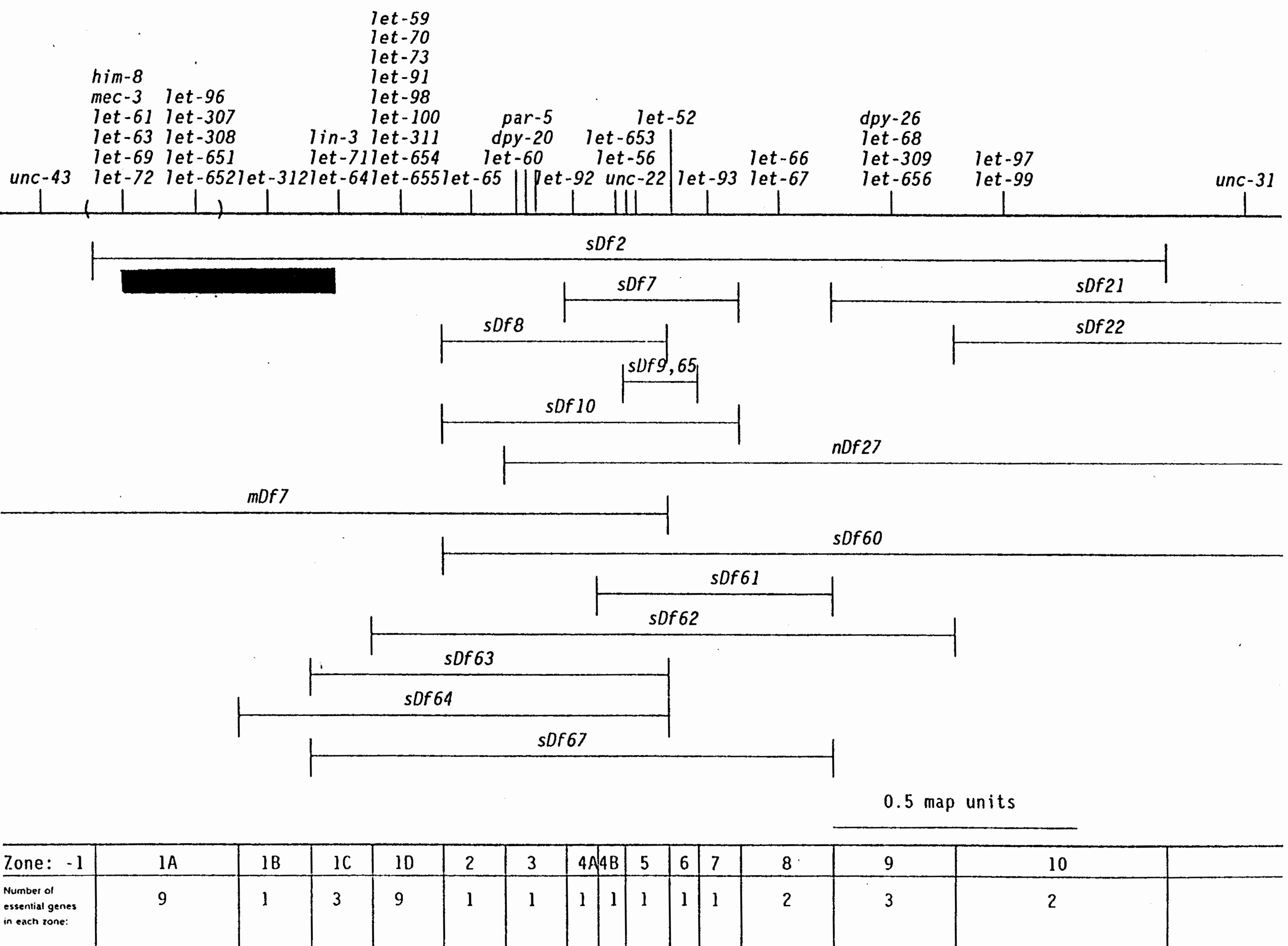
C14B5

CO4G1

(Y68A12)

Figure 4.

The genetic map of the region around *unc-22* on LGIV in *C. elegans* (Clark, 1990). The *C. elegans* histidyl-tRNA synthetase gene falls between *mec-3* and *lin-3* within the deficiency *sDf2*. Below this genetic map is the physical map showing the deficiencies used to position essential genes in zones. The histidyl-tRNA synthetase gene must be in zones 1A, 1B or 1C.



him-8
mec-3
let-61
let-63
let-69
let-72
let-96
let-307
let-308
let-651
let-652
let-312
lin-3
let-71
let-64
let-59
let-70
let-73
let-91
let-98
let-100
let-311
let-654
let-655
let-65
par-5
dpy-20
let-60
let-92
let-653
let-56
unc-22
let-52
let-93
let-66
let-67
dpy-26
let-68
let-309
let-656
let-97
let-99
unc-43
unc-31
sDf2
sDf7
sDf21
sDf8
sDf22
sDf9,65
sDf10
nDf27
mDf7
sDf60
sDf61
sDf62
sDf63
sDf64
sDf67
0.5 map units
Zone: -1
1A
1B
1C
1D
2
3
4A
4B
5
6
7
8
9
10
Number of essential genes in each zone:
9
1
3
9
1
1
1
1
1
1
1
2
3
2

(5) Cosmid CO4G1 contains the genomic fragment corresponding to the *C. elegans* histidyl-tRNA synthetase gene..

The overlapping region identified by the four YACs on the physical map was best covered by the cosmid CO4G1 (figure 3b). Therefore, this cosmid was suspected to contain the genomic fragment corresponding to the histidyl-tRNA synthetase gene. A Southern blot of *Eco*R1 digested cosmid CO4G1 DNA probed with the 420 bp ^{32}P labelled cDNA fragment resulted in hybridization to a 5.5 kb *Eco*R1 genomic fragment (figure 5). This result was consistent with that obtained from probing *C. elegans* *Eco*R1 digested genomic DNA with the same probe. Thus, the cosmid CO4G1 contains the genomic DNA corresponding to the *C. elegans* histidyl-tRNA synthetase gene. The 5.5 kb *Eco*R1 genomic fragment containing the gene was subsequently subcloned into pBluescript for further analysis (figure 6).

Figure 5

5(a) shows the ethidium bromide stained pattern of an *Eco*R1 digest of cosmid CO4G1 electrophoresed on a 0.7% agarose gel; M = 12 kb lambda ladder and A = *Eco*R1 digest of cosmid CO4G1. 5(b) shows hybridization of ^{32}P labelled *Eco*R1 0.42 kb cDNA fragment to *Eco*R1 digest of cosmid CO4G1. The probe detected a 5.5 kb *Eco*R1 genomic fragment. This genomic fragment contains the *C. elegans* histidyl-tRNA synthetase gene. The filter was washed and exposed to a Kodak film at -70°C overnight.

B
5.5

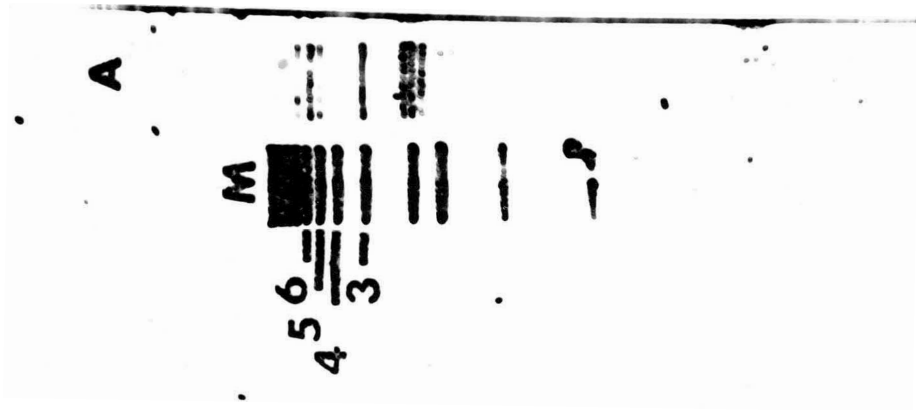

A
M
6
5
4
3

Figure 6.

6(a) shows the ethidium bromide stained pattern of a 5.5 kb *Eco* R1 genomic fragment subcloned from cosmid CO4G1 electrophoresed on a 0.7% agarose gel; M = 12 kb lambda ladder and A = the *Eco*R1 5.5 kb genomic fragment subclone and the 2.95 kb band corresponds to the vector (pBluescript). 6(b) shows hybridization of the ^{32}P labelled *Eco R1* 0.42 kb cDNA fragment to the 5.5 kb *Eco*R1 genomic fragment subcloned from cosmid CO4G1. The filter was washed and exposed to a Kodak film at -70°C overnight.

M A B

6
5
4
3

5.50
2.95
5.5

(6) The cDNA sequence obtained from the two cDNA fragments.

The sequences from both the 420 bp and the 1.5 kb cDNA fragments were found to overlap with each other (figure 7). The 1.5 kb cDNA fragment was found to contain a poly A tail marking the 3' end of the gene. Combining the nucleotide sequence from the two cDNA fragments, a sequence of 1770 nucleotides corresponding to the *C. elegans* histidyl-tRNA synthetase gene was obtained. This sequence, however, is missing the 5' initiating methionine.

(7) Obtaining the genomic DNA sequence corresponding to the *C. elegans* histidyl-tRNA synthetase gene.

Synthetic oligonucleotides designed from the cDNA sequence were used to obtain the genomic sequence of interest from the 5.5 kb *Eco*R1 genomic fragment subclone (figure 6). Genomic sequence corresponding to the 5' untranslated region, coding region, and 3' untranslated region of the *C. elegans* histidyl-tRNA synthetase gene was obtained (figure 8). Examination of the 5' end untranslated region of the gene shows a putative TATA box, (TATTA), CCAAT box, (CCAAAT), and a trans-splice site (TTTTTCAG). The genomic sequence corresponding to the coding region of the gene exactly matches the cDNA sequence except when interrupted by introns. The genomic gene sequence contains an open reading frame interrupted by three small introns (45-50 bp long) (Table 1). The coding sequence of the *C. elegans* histidyl-tRNA synthetase translates to a putative protein sequence of 523 amino acids.

Figure 7

This figure shows the overlap point between the 0.42 kb (plasmid # 29) and the 1.5 kb (plasmid # 3-1) cDNA fragments corresponding to the *C. elegans* histidyl-tRNA synthetase gene coding elements. The T7 and T3 primers used to sequence the ends of the cDNA fragments are shown. The position of each the synthetic oligonucleotides designed to complete the cDNA sequence and to obtain the missing 5' end and the rest of the genomic sequence of the gene are shown..

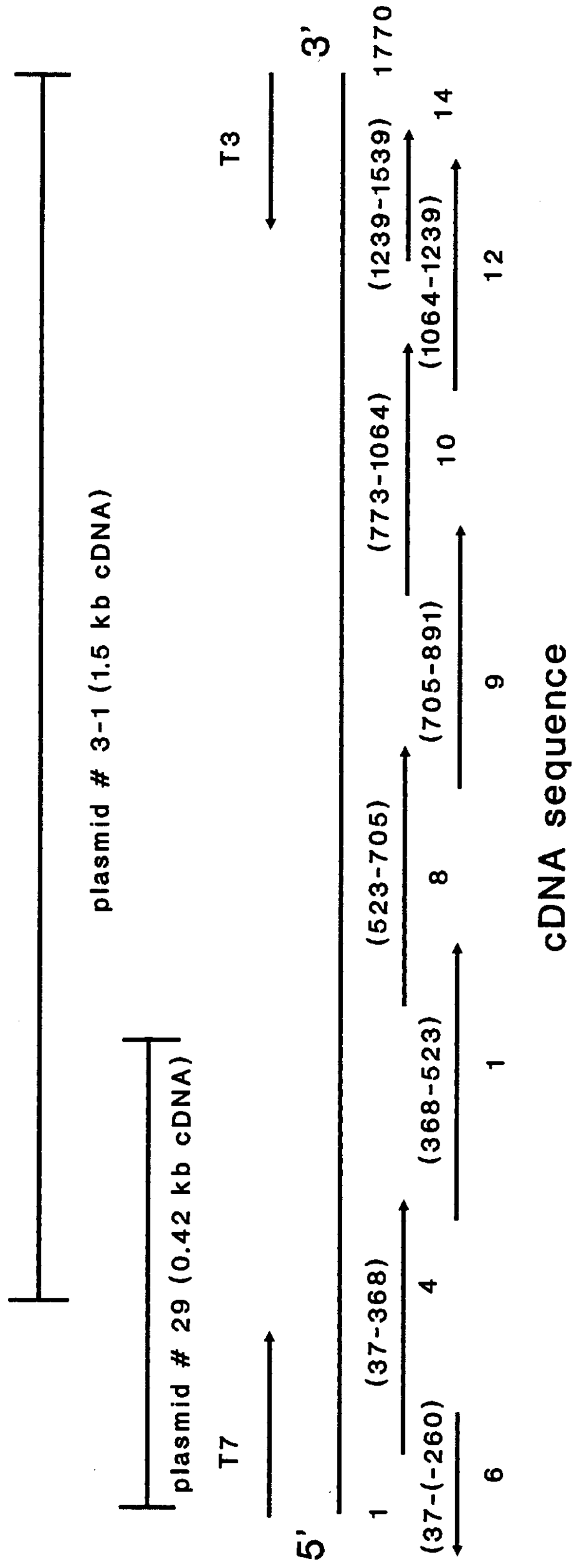

plasmid # 3-1 (1.5 kb cDNA)
plasmid # 29 (0.42 kb cDNA)
T7
T3
5'
3'
1
1770
(37-(-260)
(37-368)
(368-523)
(523-705)
(705-891)
(773-1064)
(1064-1239)
(1239-1539)
6
4
1
8
9
10
12
14
cDNA sequence

Figure 8.

The nucleotide sequence and the deduced amino acid sequence of the *C. elegans* histidyl-tRNA synthetase gene. The numbering of nucleotides starts at the first residue of the ATG codon coding for the initiating methionine. Single-letter amino acid code is used. Lower case letters indicate the 5' untranslated region and the location of the three introns identified. The putative TATA box, (tatta), the CCAAT box, (ccaaat), the trans-splice site (tttttcag), and the three motifs (M1, M2, M3) are underlined.

```
-428 cgagctaacgaattgtaggtttactagtttttcgtacattgttaatttaaacgaaattca

gtcagaaacaaacaaataatcgtttaattccaaaattagatagactttttacatgtatcg

ttttttttcgcattttgaacaaataaaggatacccggttcatgtctgatcttctatttt

aatgaaacgcgtttgcaatccgcatagttttctcaccaaattcatgaaaattgtgacgaa

ctttcacatcttcagtaaaatgtttcgaagcctgtattacgagtaggctggctttctcag

ttcggcaggctgaagaaactgtaaaaccgcggattgcatatgcacagagaagtggacgac

gagtattcttgaataacatgaaattccaccaatacatgctttttcagatcaaagagaaag

        1
tcgcactgATGCAAGCGAAAAGAAAGGAGGCAGGCGAGACCGGTGCACCCGAAAAACCCG 52
        M  Q  A  K  R  K  E  A  G  E  T  G  A  P  E  K  P   17

GAAAGTTTGTTCTGAAAACAGGAAAGGGAACTCGTGACTACGGACCAGCTCAATCGGCTC 112
G  K  F  V  L  K  T  G  K  G  T  R  D  Y  G  P  A  Q  S  A   37

TTCGCAATTCTGTGTTACAGACAGTCACTGAAACATTCAATAGATATGGTGCAGAGACCA 172
L  R  N  S  V  L  Q  T  V  T  E  T  F  N  R  Y  G  A  E  T   57
                                                   M1
TTGATACACCAGTATTCGAGCTTCGTGATGTTCTGATGGGAAAATATGGCGAGGAAGGAG 232
I  D  T  P  V  F  E  L  R  D  V  L  M  G  K  Y  G  E  E  G   77

GAAAACTTGTATATGATCTTCAAGATCAAGGAGGAGAGCTTCTATCATTGCGTTATGATC 292
G  K  L  V  Y  D  L  Q  D  Q  G  G  E  L  L  S  L  R  Y  D   97

TCACTGTTCCATTTGCACGATACTTGGCAATGAACAAAATCACGAATATTACAAGATATC 352
L  T  V  P  F  A  R  Y  L  A  M  N  K  I  T  N  I  T  R  Y   117

AAATTGCGAAAGTATACAGAAGAGATCAACCTGTAATGTCTCGTGGACGGTACCGCGAGT 412
Q  I  A  K  V  Y  R  R  D  Q  P  V  M  S  R  G  R  Y  R  E   137
                                M2
TCTATCAATGCGATTTTGATATTGCCGGACAGTATGACCTTATGCTTCCGGAGCTGAGTG 472
F  Y  Q  C  D  F  D  I  A  G  Q  Y  D  L  M  L  P  E  L  S   157

TCTGGGAATTGTTGATGAGTTGCTCACGAAACTGGAAATTGGAGAGTTTTTCATTAATgt 532
V  W  E  L  L  M  S  C  S  R  N  W  K  L  E  S  F  S  L  I   177

aaggctatttgttgaacaatcattttaattcgaattgtgttgtttcagCTCAATCACCGT 592
                                                 S  I  T  V  181

CTTGATCCTCGAAGGAATGTTTGCCGTTTCCGGTATTCCTGCTAAAGATTTCAAAACAAT 652
 L  I  L  E  G  M  F  A  V  S  G  I  P  A  K  D  F  K  T  I 201
```

```
TTGCTCATCGGTCGATAAGCTTGACAAAACTCCATGGAAGATGTTGAACAAGAAATGAT 712
 C  S  S  V  D  K  L  D  K  T  P  W  E  D  V  E  Q  E  M  I 221

TAATGAAAAGTTCCTAACTAAAGAGCAGACTGGAAAGCTGGGAGAACTTGTTCGATTCGA 772
 N  E  K  F  L  T  K  E  Q  T  G  K  L  G  E  L  V  R  F  E 241

GAGCTCAATAGCGATCTCAACAATTTGGAACTTCTTGAAAAAATGTCACAACCTCCCAGA 832
 S  S  I  A  I  S  T  I  W  N  F  L  K  K  C  H  N  L  P  D 261

TCTCGGGCAAAATGACAAGTTTAAAAAGGGAGCCGAAGAACTGAAGGTGTTAATTGAGTA 892
 L  G  Q  N  D  K  F  K  K  G  A  E  E  L  K  V  L  I  E  Y 281

CCTCAATGTCGATGGAGTGACCACTGTTCGATACGAACCATCCCTTGCCAGAGGTCTTGA 952
 L  N  V  D  G  V  T  T  V  R  Y  E  P  S  L  A  R  G  L  D 301

TTATTATACAGGTGCAATCTATGAAGCAGTCGCTCCTAgtatgtttttaaatattgaatc 1012
 Y  Y  T  G  A  I  Y  E  A  V  A  P                          313

gactatcaaattgtttaatttcagAAGCTCTTGAAGGAACTGCTGTTGAAAACTCAGAGG 1072
                          K  A  L  E  G  T  A  V  E  N  S  E 325

ATACTGCTGGACAACCAGTCGGAGTAGGATCAGTAGCTGCCGGTGGACGATACGATGGAT 1132
D  T  A  G  Q  P  V  G  V  G  S  V  A  A  G  G  R  Y  D  G  345

TGGTTAAAATGTTCGACTCAAAAGCCAATGTTCCATGTTGCGGTGTTTCTTTTGGTATCG 1192
L  V  K  M  F  D  S  K  A  N  V  P  C  C  G  V  S  F  G  I  365
                                  ____________________________
                                               M3
AACGTCTGTTTGCTATCATGGAAGCTCGACAGAAAGTTGCCATTCGTACGACGCAAACCG 1252
E  R  L  F  A  I  M  E  A  R  Q  K  V  A  I  R  T  T  Q  T  385
____________________

AAGTCTATGTTGCATCTGCTCAAAAGAATTTGGTCAGGGATCGCAAGAAGTTGGTAAAGA 1312
E  V  Y  V  A  S  A  Q  K  N  L  V  R  D  R  K  K  L  V  K  405

TGCTTCGATCTGCTGGAACTCAAGGACTGAAAATGGCACTGAAGGCTAATCCGAAATTGT 1372
M  L  R  S  A  G  T  Q  G  L  K  M  A  L  K  A  N  P  K  L  425

TAACGTCAATTCCAGTATGCCTGAGGAAAGACCGTATTCCGTCTTGCTATTGTTATTATG 1432
L  T  S  I  P  V  C  L  R  K  D  R  I  P  S  C  Y  C  Y  Y  445

AGAGCAGAGCTAAGATGAGTGTCAGTGCGAAATGTGTCACCAGAGACGAACAGgtaattt 1492
E  S  R  A  K  M  S  V  S  A  K  C  V  T  R  D  E  Q        463

ttctaatatatttttttccgatatcaccaatttaaaatattttcagACTATTAAACTCGAC 1552
                                                T  I  K  L  D  468

CAGCTGATCACTGCCGTTCGTGACACTCTTGCGCCCTCTAAACAATCGATAATTTCTTGT 1612
 Q  L  I  T  A  V  R  D  T  L  A  P  S  K  Q  S  I  I  S  C 488

TTTCTTAGCTTTCTAATTCTCTGCGTTCACCTCACTTCAGCTTTTATGTCCCTCTTTCCA 1672
 F  L  S  F  L  I  L  C  V  H  L  T  S  A  F  M  S  L  F  P 508

GTCATCATAATTCATATTTCCTGTTTTATTGATTTTTGCATAGTCGGTGTTTGACTTGTT 1732
 V  I  I  I  H  I  S  C  F  I  D  F  C  I  V  G  V  *       525

TGTTTTTCTCCTTTTTTTGTTCCAGATTTCATATTTATTTTTGTTGGTTTGAAAATTCAG 1792
```

Table 1

Shows the size, position, and sequence of each of the three introns as derived from the genomic sequence of the *C. elegans* histidyl-tRNA synthetase gene.

Intron	Size	Sequence		
1	50	(958)	TgtaaggctatttgttgaacaatcattttaattcgaattgtgttgtttcagC	(959)
2	47	(1368)	AgtatgtttttaaatattgaatcgactatcaaattgtttaatttcagA	(1369)
3	53	(1897)	GgtaattttctaatatatttttccgatatcaccaatttaaaatattttcagA	(1898)

The 3' untranslated region of the cDNA sequence contains a poly(A) tail. However, a concensus poly(A) addition site was not found in the obtained cDNA and genomic sequences corresponding to the 3' untranslated region.

(8) The putative protein sequence of the *C. elegans* histidyl-tRNA synthetase.

The primary protein sequence of the *C. elegans* histidyl-tRNA synthetase has been found to contain the three motifs characteristic of class II synthetases which include histidyl-tRNA synthetases (Table 2). These three motifs are found in eight of the class II synthetases (Eriani, 1990). The predicted protein sequence for the *C. elegans* histidyl-tRNA synthetase was compared to those of the human (Tsui and Siminovitch, 1987), yeast (Natsoulis, 1986), and *E. coli* (Freedman *et al*, 1985) histidyl-tRNA synthetases. Pairwise alignment of the four primary protein sequences reveals regions of homology between them. The alignment shows a 43.9% identity in a 476 amino acid overlap, 44.1% identity in a 412 amino acid overlap, and 23% identity in an 173 amino acid overlap in humans, yeast, and *E. coli*, respectively (Table 3). Figure 9 shows multiple aligments of *C. elegans,* yeast, and human histidyl-tRNA synthetases.

Table 2.

Shows the percentage of identity obtained from performing pairwise comparisons between the *C. elegans* histidyl-tRNA synthetase protein sequence with *E. coli*, yeast, and human histidyl-tRNA synthetase protein sequences. It also shows the number of consecutive amino acid identities between each pair.

Histidyl-tRNA synthetase	*C. elegans*	Sequence of consecutive matches
E. coli	23.1% identity in 173 A.A overlap	AGGRYDGLV
Yeast	43.2% identity in 412 A.A overlap	SLRYDLTVPFARY
Human	44.1% identity in 476 A.A overlap	DQGGELLSLRYDLTVPFARYLAMNK

Table 3.

Shows the position and sequence of each of the three motifs characteristic of class II synthetases identified in the *C. elegans* histidyl-tRNA synthetase protein sequence.

Motif	Position	Sequence
1	A.A 54 to 63	GAETIDTPVF
2	A.A 124 to 142	YRRDQPVMSRGRYREFYQCD
3	A.A 357 to 372	PCCGVSFGIERLFAIM

Figure 9

Shows multiple alignment of the yeast (ScHIS), human (HsHIS), and *C. elegans* histidyl-tRNA synthetase (CeHIS) protein sequences. The (*) shows identical amino acids and the (.) shows conserved amino acids. The dashes are inserted to maximize the homology. Stretches of 1-25 amino acid consecutive identity are seen between the human and the *C. elegans* protein sequences and 1-13 amino acid consecutive identity are seen between the *C. elegans* and yeast.

```
ScHIS  MLSRSLNKVVTSIKSSSIIRMSSATAAAT-SAPTANAANALKASKAPKKG   49
HsHIS  MAERAPLEELVKLQGERVRGLKQQKASAELIEEEVAKLLKLKAQLGPDES   50
CeHIS  MQAK-------------------RKEAGETGAPEKPG-------------   18
       * ..                    .....  . .

ScHIS  KLQVSLKTPKGTKDWADSDMVIREAIFSTLSGLFKKHGGVTIDTPVFELR   99
HsHIS  KQKFVLKTPKGTRDYSPRQMAVREKVFDVIIRCFKRHGAEVIDTPVFELK  100
CeHIS  --KFVLKTGKGTRDYGPAQSALRNSVLQTVTETFNRYGAETIDTPVFELR   66
         .  *** ***.*..  . ..*. .. ..   *...*. .********.

ScHIS  EILAGKYGEDS-KLIYNLEDQGGELCSLRYDLTVPFARYVAMNNIQSIKR  148
HsHIS  ETLMGKYGEDS-KLIYDLKDQGGELLSLRYDLTVPFARYLAMNKLTNIKR  149
CeHIS  DVLMGKYGEEGGKLVYDLQDQGGELLSLRYDLTVPFARYLAMNKITNITR  116
       ..* *****.. **.*.*.****** *************.***.. .*.*

ScHIS  YHIAKVYRRDQPAMTKGRMREFYQCDFDVAGTFESMVPDSECLSILVEGL  198
HsHIS  YHIAKVYRRDNPAMTGGRYPNSITVDFDIAGQFDPMNPDAESLKIMCEIL  199
CeHIS  YQIAKVYRRDQPVMSRGRYREFYQCDFDIAGQYDLMLPELSVWELLMSCS  166
       *.********.*.*. ** ..    ***.** .. * *. .  ... .

ScHIS  TSLGIKDFKIKLNHRKILDGIFQIAGVKDEDVRKISSAVDKLDKSPWEAV  248
HsHIS  SSLQIGNFLVKVNDRRILDGMFAVCGVPDSKFRTICSSVDKLDKVSWEEV  249
CeHIS  RNWKLESFSLISITVLILEGMFAVSGIPAKDFKTICSSVDKLDKTPWEDV  216
        .  . .* .      **.*.*.. *. ... ..*.*.****** .**.*

ScHIS  KKEMTEEKGQSEETADKIGEYVKLNGSLKE--IHAVLS---ADANITSNE  293
HsHIS  KNEMVGEKGLAPEVADRIGDYVQQHGGVSL--VEQ-LV---QDPKLSQNK  293
CeHIS  EQEMINEKFLTKEQTGKLGELVRFESSIAISTIWNFLKKCHNLPDLGQND  266
       ..**..**  . * ....*. *. ....    . . *    . .... *.

ScHIS  KAKQGLDDIATLMKYTEAFDIDSFISFDLSLARGLDYYTGLIYEVVTSAS  343
HsHIS  QALEGLGDLKLLFEYLTLFGIDDKISFDLSLARGLDYYTGVIYEAVL---  340
CeHIS  KFKKGAEELKVLIEYLNVDGVTT-VRYEPSLARGLDYYTGAIYEAVAPKA  315
       .  .* ...  *..*  .  .... .... *********** ***.*

ScHIS  APPENASELKKKAKSAEDASEFVGVGSIAAGGRYDNLVNMFSEASGKKST  393
HsHIS  --------LQTPAQEGEEPW----CGQCGCWRRYDGLVGMFDPQRRKVAM  378
CeHIS  L---EGTAVENSEDTAGQP---VGVGSVAAGGRYDGLVKMFDSK-----A  354
               ... .......      *  .   ***.** **.

ScHIS  QIPCVGISFGVERIFSLIKQRINS-STTIKPTATQVFVMAFGGGKDWTGY  442
HsHIS  ---CGAQHWG-GRIFSIVEQRLEALEEKIRTTETQVLVAS--AQKKLA--  420
CeHIS  NVPCCGVSFGIERLFAIMEARQKV---AIRTTQTEVYVAS--AQKNL---  396
          * .  .* .*.*.....* .     *..*.*.* * .  . *.

ScHIS  LPERMKVTKQLWDAGIEA-EYVYKAKANPRKQFDTTKKAGCHIAVILGKE  491
HsHIS  -RGKTKACLRLWDAGIKA-ELLYKKNPKLLNQLQYCEEAGIPLVAIIGEQ  468
CeHIS  VRDRKKLVKMLRSAGTQGLKMALKANPKLLTSIPVCLRKD---RIVIGEQ  443
        ... *    *..**... .   * ...  ..       .     ..*..

ScHIS  EYLEGKL----------RVKRLGQEFADDDGELVSAADIV----------  521
HsHIS  ELKDGVI----------KLRS------------VTSREEV----------  486
CeHIS  ELKDRQLSSAKCVTRDEQTIKLDQLITAVRDTLAPSKQSIISCFLSFLIL  493
       *  .  .          .  .            ..... .
```

DISCUSSION

(I) The *C. elegans* histidyl-tRNA synthetase gene is a single copy gene.

Southern experiments have shown that the *C. elegans* histidyl-tRNA synthetase gene is present in a single copy in the genome (figure 2). This result is interesting because in animal eukaryotic cells two forms of synthetases must exist. Cytoplasmic and mitochondrial synthetases are required for normal development and viability of the organism (Schneller *et al*, 1978). Usually, the cytoplasmic and the mitochondrial synthetases are encoded by two structurally distinct nuclear genes (Schneller *et al*, 1978; Meyers and Tzagoloff, 1985; Pape and Tzagoloff, 1985; Pape *et al*, 1985). Some of these enzymes have been reported to be specific for either the cytoplasmic or the mitochondrial tRNAs. That is, a cytoplasmic synthetase will only aminoacylate the cytoplasmic tRNA and not a tRNA from the mitochondrion of the same organism and vice versa (Barnett *et al*, 1967; Schneller *et al*, 1978). This is exemplified by leucine, methionine, and phenylalanine. Their cytoplasmic synthetases have been shown to differ from their mitochondrial counterparts (Diatewa *et al*, 1980, Diatewa and Stahl, 1981, Schneller *et al*, 1978).

In yeast histidyl- and valyl-tRNA synthetases, it has been shown that a single structural nuclear gene can encode both the cytoplasmic and the mitochondrial synthetases (Natsoulis *et al*, 1986; Chatton *et al*, 1988). In both cases the polypeptides corresponding to the cytoplasmic and the mitochondrial synthetases are encoded by two classes of transcripts initiating at distinct in-frame translation start codons. A short transcript encodes the cytoplasmic synthetase and a long transcript encodes the mitochondrial synthetase including the mitochondrial import leader sequence (Natsoulis *et al*, 1986; Chatton *et al*, 1988). It has been shown that either form of yeast histidyl-tRNA synthetases can aminoacylate the histidyl-tRNAs from the cytoplasm and the mitochondria despite their structural differences (Natsoulis *et al*, 1986).

In the case of the *C. elegans* histidyl-tRNA synthetase gene, it is very likely that this gene, being a single copy gene, encodes both the cytoplasmic and the mitochondrial synthetases. Primer extension analysis on mature mRNA of this gene could be used to determine if this gene encodes both forms of the histidyl-tRNA synthetases.

(II) Molecular and genetic mapping of the *C. elegans* histidyl-tRNA synthetase gene.

The *C. elegans* histidyl-tRNA synthetase gene has been mapped with respect to the physical and genetic maps. On the molecular map, this gene falls in a region defined by four overlapping yacs; Y71D6, Y47F2, Y59E6, and Y69E1 (figure 3b). The region of overlap is best covered by the cosmid CO4G1. This cosmid has been found to contain the genomic DNA corresponding to the *C. elegans* histidyl-tRNA synthetase gene. Genetically, this gene has been mapped on LGIV between *mec-3* and *lin-3* within the deficiency *sDf2* (figure 4). The result was interesting because one of our laboratory interests is studying the genomic organization of a region on LGIV defined by the deficiency *sDf2*. The area into which this gene maps is in a region which has been investigated in great detail in our laboratory (Clark *et al*, 1988; Clark, 1990; Moerman, 1980; Rogalski *et al*, 1982; Rogalski and Baillie, 1985). The *mec-3* gene is in region 1A and *lin-3* is in region 1C (figure 4) (Clark, 1990, Clark and Baillie, 1992 MGG In press). There are 12 essential genes that have been identified in this region. Clark has estimated that this region is 70% saturated for essential genes. Thus, it is possible that mutations in the histidyl-tRNA synthetase gene may already exist.

The location of the *C. elegans* histidyl-tRNA synthetase gene has been more precisely defined with respect to the break points of four deficiencies; *sDf62*, *sDf63*, *sDf64*, and *sDf67* (Kim McKim, personal communication). It has been found that none of these deficiencies deletes this gene and hence the gene falls into zone 1A (figure 4).

This result allows the elimination of 3 of the 12 essential genes (*let-71*, *let-641*, and *let-312*) (Rogalski et al,1982; Clark et al, 1988) that map between *mec-3* and *lin-3* as potential candidates for the *C. elegans* histidyl-tRNA synthetase gene (figure 4). I assume that null mutations in this gene will result in death of the organism. It has been demonstrated that a mutation affecting the initiation codon of the yeast cytoplasmic histidyl-tRNA synthetase results in death of the organism, while a mutation affecting the initiation codon of the mitochondrial form results in a respiratory deficient phenotype (Natsoulis, 1986). It would be very interesting to carry out microinjection experiments to determine if one of the remaining nine lethals (*let-61*, *let-63, let-69, let-72, let-96, let-307, let-615,* and *let-652*) (Clark, 1990; Clark *et al*, 1988; Moerman, 1980; Rogalski *et al*, 1982) can be rescued by the DNA corresponding to the *C. elegans* histidyl-tRNA synthetase gene. Three alleles of *let-69*,(*s172*, *s1085*, and *s1111*), and one allele of *let-651* (*s1185*) arrest development in an early larval stage when homozygous (Clark *et al*, 1988; Clark, 1990; Moerman, 1980). It appears that *let-69* and *let-651* are strong candidates for the *C. elegans* histidyl-tRNA synthetase gene since worms carrying null mutations of this gene are expected to arrest development just after the maternally contributed mRNA is exhausted. Rogalski and Riddle (1988) have concluded that the RNA polymerase II null phenotype is a first larval stage (LI) lethal and have suggested that maternal mRNA of RNA polymerase II permits embryonic development to reach completion. It has also been suggested that the synthesis of RNA polymerase II does not begin until the LI stage (Rogalski and Riddle, 1988).

(III) Analysis of the genomic DNA sequence of the *C. elegans* histidyl-tRNA synthetase gene.

The complete DNA sequence of the first aminoacyl-tRNA synthetase gene (histidyl-tRNA synthetase) from the nematode *C. elegans* has been determined (figure

8). The size of this gene including the flanking and coding sequences is 2220 nucleotides. This gene contains an open reading frame interrupted by three small introns (47-53 bp) (Table 1). The position of each of the three introns has been identified by comparing the genomic sequence to the partial cDNA sequence. The introns contain the same consensus sequences found in other *C. elegans* introns (Wood, 1988).

Examination of the 5' untranslated region of the gene shows a putative TATA box, (TATTA), and a putative CCAAT box, (CCAAAT). In all eukaryotic genes, these two consensus sequences indicate the promoter region of the gene (Watson, 1987). Thus, the presence of the TATTA and CCAAAT boxes provides evidence that the 5' end untranslated region contains the promoter region of the *C. elegans* histidyl-tRNA synthetase gene. The 5' untranslated region has also been found to contain a splice site (TTTTTCAG). This suggests that this gene may be trans-spliced. The trans-splicing phenomena is well known in the protozoan *Trypanosome bruci* where all mature mRNAs contain a 35 nucleotide leader sequence at their 5' end (Borst, 1980).

In *C. elegans* about 10 to 15% of the genes have been estimated to be trans-spliced (Bektesh *et al*, 1988). Some of the genes that have been shown to be trans-spliced include three actin genes (Krause and Hirsh, 1987), three heat shock genes; hsp-1 (Snutch *et al*, 1989), hsp-3 and hsp-6 (Heschl and Baillie, 1988), ubiquitin (Graham *et al*, 1988), and two ribosomal genes (Bektesh *et al*, 1988). In all of these genes, the initiation codon has been located between 6 to 21 nucleotides downstream of the trans-splice site. In the case of the *C. elegans* histidyl-tRNA synthetase gene, an initiation codon has been identified exactly 21 nucleotides downstream of the trans-splice site. Thus, this initiation codon is very likely to be the codon for the initiating methionine. Most of the *C. elegans* genes that have been shown to be trans-spliced contain the same 22-nucleotide leader sequence (SL1) on their mature mRNAs 5' ends. The leader sequence is acquired in a trans-splice reaction (Bektesh *et al*, 1987). It is

encoded by a gene residing in the 5s rDNA in *C. elegans* (Kruase and Hirsh, 1987, Nelson, 1988). Recently, Huang and Hirsh (1989) have identified a second trans-spliced leader (SL2) on the 5' end of the mRNA encoded by one of the four glyceraldehyde-3-phosphate dehydrogenase genes. The 5' end of the mRNAs encoded by the other three genes are found to contain the SL1. The SL2 is also 22 nucleotide long; and it shares 17 nucleotide identity with SL1 (Huang and Hirsh, 1989).

Since the two cDNAs corresponding to the *C. elegans* histidyl-tRNA synthetase were not full length cDNAs, it was not possible to determine whether or not a splice leader sequence was present on the 5' end of the mature mRNA of this gene. Therefore, it is important to carry out primer extension analysis on the 5' end of the mature mRNA of this gene. Perhaps this will determine if the mature mRNA from the *C. elegans* histidyl-tRNA synthetase gene contains a 22-nucleotide leader sequence at its 5' end. It would be also interesting to determine if the mature mRNA contains the SL1 or SL2 at its 5' end.

(IV) Protein sequence analysis.

The *C. elegans* histidyl-tRNA synthetase gene encodes a putative protein sequence of 523 amino acids (figure 7). This putative protein sequence shares extensive homology with both the human and yeast histidyl-tRNA synthetase protein sequences, 44.1% identity in 476 amino acids, and 43.2% identity in 412 amino acids, respectively (Table 2 and figure 9). It also shares a significant homology, 23% identity in 173 amino acids, with the *E. coli* histidyl-tRNA synthetase protein sequence (Table 2). The comparison shows regions of conservation ranging from 3 to 25 consecutive amino acid identities between the *C. elegans* and human protein sequences, 3 to 13 consecutive amino acid identities between *C. elegans* and the yeast protein sequences,

and 1 to 9 consecutive amino acid identities between the *C. elegans* and *E. coli* protein sequences (Table 2). The homology among the four species of histidyl-tRNA synthetase appears to be restricted to a middle region of the enzyme. Thus, the conserved regions among these four species probably have similar roles in terms of structure and function of the enzyme.

The *C. elegans* putative protein sequence exhibits the three signature motifs characteristic of class II synthetases (Eriani *et al*, 1990). In *C. elegans*, motif 1 is found at amino acid number 54 to 63 (GAETIDTPVE) of the protein sequence, motif 2 is found at amino acid number 125 to 145 (YRRDQPVMSRGRYREFYQCD), and motif 3 is found at amino acid number 367 to 382 (PCCGVSFGIERLFAIM) (Table 3). Motifs 1 and 2 are separated by 60 amino acids, and motifs 2 and 3 are separated by 154 amino acids. The presence of these motifs in the *C. elegans* protein sequence further confirms the notion that class II synthetases have arisen from a common ancestral gene (Eriani *et al*, 1990). These three motifs have been shown to be involved in building the catalytic domain of all members of class II synthetases (Cusack *et al*, 1990). The *E. coli* seryl-tRNA synthetase three dimensional (3-D) structure has been determined, and it has been found to have a C-terminal catalytic domain of 330 residues based on a seven-stranded antiparallel beta sheet (Cusack *et al*, 1990). Recently, it has been reported that the yeast aspartyl-tRNA synthetase 3-D structure has a similar antiparallel beta sheet to that of the *E. coli* seryl-tRNA synthetase (Ruff *et al*, 1991). Based on the 3-D structures of the *E.coli* seryl and the yeast aspartyl-tRNA synthetases, motif 1 has been shown to be involved in the formation of the dimer interface, and motifs 2 and 3 have been shown to be involved in the recognition of the CCA end of the tRNA and ATP (Cusack *et al*, 1991, and Ruff *et al*, 1991).

The 3-D structure of each member of class II synthetases has been predicted in light of the *E.coli* seryl-tRNA synthetase primary sequence and 3-D structure (Cusack *et al*, 1991). The histidyl-tRNA synthetases are predicted to possess 6 out of the 7

antiparallel beta strands found in the *E. coli* seryl-tRNA synthetase 3-D structure (Cusack *et al*, 1991). Alignments of the *C. elegans* protein sequences with the *E. coli*, yeast, and human-tRNA synthetases have also revealed the conservation of the same residues (RGLDY and RYDG) among the four species. These conserved residues may be specific for this enzyme and hence may be involved in amino acid binding (Cusack *et al*, 1990). The protein sequence analysis discussed above provides solid evidence that the *C. elegans* putative protein sequence corresponds to the *C. elegans* histidyl-tRNA synthetase gene.

SUMMARY

1. The *C. elegans* histidyl-tRNA synthetase gene has been cloned and sequenced.
2. The *C. elegans* histidyl-tRNA synthetase gene is a single copy gene.
3. On the molecular map, it falls in a region defined by four overlapping yacs; Y171D6, Y47F2, Y59C6, and Y69C1.
4. Cosmid CO4G1 contains the *C. elegans* histidyl-tRNA synthetase gene.
5. On the genetic map, it falls between *mec-3* and *lin-3* within the deficiency *sDf2* on LGIV.
6. The gene contains a putative trans-splice site in its 5' untranslated region which suggests that it is likely trans-spliced.
7. The putative protein sequence shows extensive homology with the human (44.1%) and yeast (43.2%) histidyl-tRNA synthetases.
8. The putative protein sequence exhibits the three characteristic motifs of class II synthetases.

PROPOSALS FOR FUTURE WORK

1. Primer extension analysis could be conducted to determine if the mature mRNA of the *C. elegans* histidyl-tRNA synthetase contains a trans-splice leader sequence. This would also determine if two transcripts of the histidyl-tRNA synthetase exist. This would indicate if this gene codes for both the cytoplasmic and the mitochondrial forms.
2. The expression profile of this gene should be established to determine at which stage of development it is most abundant.
3. Microinjection experiments could be carried out using the genomic DNA corresponding to the histidyl-tRNA synthetase. This perhaps will determine if any of the nine lethals between *mec-3* and *lin-3* can be rescued.
4. If a lethal is rescued, then the lethal and its alleles can be analyzed to find the nature of the mutation(s).
5. The histidyl-tRNA synthetase gene from the nematode *C. briggsae* could be cloned and sequenced for evolutionary comparisons and also to determine 5' regulatory or additional coding sequences.

REFERENCES

Akins, Robert A and Lambowitz, Alan M. (1987) A Protein for Splicing Group I Introns in Neurospora Mitochondria is Mitochondrial Tyrosyl-tRNA Synthetase or a Derivative Thereof. Cell 50: 331-345.

Barnett, W Edgar, Brown, David H, and Epler, J.L.(1967) Mitochondrial- Specific Aminoacyl-tRNA Synthetases. Proc. Natl. Acad. Sci. *USA*. 57: 1775-1781.

Bektesh, Susan, Van Doren, Kevin, and Hirsh, David. (1988) Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes and Development, 2:1277-1283.

Borst, P. (1986) Discontinuous transcription and antigenic variations in trypanosomes. Annu. Rev. Biochem. 55: 701-732.

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

Burbaum, Jonathan J and Schimmel, Paul. (1991) Strucural Relationship and the Classification of Aminoacyl-tRNA Synthetases. J. Biol.Chem. 266: 16965-16968.

Chatton, Bruno, Walter, Philippe, Ebel, Jean-Pierre, Lacroute, Francois, and Fasiolo, Franco. (1988) The Yeast VAS1 Gene Encodes Both Mitochondrial and Cytoplasmic Valyl-tRNA synthetases. J. Biol. Chem. 263: 52-57.

Cheung, Alice Y, Watson, Lynley, and Soll, Dieter. (1985) Two Control Systems Modulate the Level of Glutaminyl-tRNA Synthetase in *Escherichia coli*. J. Bacter. 161: 212-218.

Chow, C. Ming and Rajbhandary, Uttam L. (1989a) Regulation of the Nuclear Genes Encoding the Cytoplasmic and Mitochondrial Leucyl-tRNA Synthetases of *Neurospora crassa*. Mol. Cell. Biol. 9: 4645-4652.

Chow, C. Ming, Metzenberg, Robert L, and Rajbhandary, Uttam L. (1989b) Nuclear Gene for Mitochondrial Leucyl-tRNA Synthetase of *Neurospora crassa*: Isolation, Sequence, Chromosomal Mapping, and Evidence that the *leu*-5 Locus Specifies Structural Information. Mol. Cell. Biol. 9: 4631-4644.

Cirakoglu, Beyazit and Waller, Jean-pierre. (1985) Leucyl-tRNA and lysyl-tRNA synthetases, derived from the high-Mr complex of sheep liver, are hydrophobic proteins. (1985) Eur. J. Biochem. 151: 101-110.

Clark, D.V. (1990) The *unc-22* Region of *Caenorhabditis elegans*: Genetic Analysis and Molecular Mapping. Ph.D. thesis, Simon Fraser University, Burnaby, B.C., Canada.

Clark, D.V. and Baillie, D.L. (1992) Mol. Gen. Genet, in press.

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

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

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

Cusack, Stephen, Berthet-Colominas, Carmen, Hartlein, Michael, Nassar, Nicolas, and Leberman, Reuben. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 A. Nature 347: 249-255.

Cusack,Stephen, Hartlein, Micheal, and Leberman, Reuben. (1991) Sequence, structural, and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. Nucl. Acid. Res. 19: 3489-3498.

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

Diatewa, M and Stahl, A. J. C. (1980) Purification and Subunit Structure of Mitochondrial Phenylalanyl-tRNA Synthetase from Yeast. Biochem. Biophys. Res. Commun. 94: 189-198.

Diatewa, M. and Stahl, A. (1981) Biosynthesis and transport of yeast mitochondrial phenylananyl-tRNA synthetase. Nucl. Acid. Res. 9: 6293-6304.

Edward, Helen and Schimmel, Paul. (1987) An *E. coli* Aminoacyl-tRNA Synthetase Can Substitute for Yeast Mitochondrial Enzyme Function In Vivo. Cell 51: 643-649.

Emmerich, Rosemarie V and Hirshfield, Irvin N. (1987) Mapping of the Constitutive Lysyl-tRNA Synthetase Gene of *Escherichia coli* K-12. J. Bacter. 169: 5311-5313.

Eriani, Gilbert, Delarue, Marc, Poch, Olivier, Gangloff, Jean, and Moras, Dino. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347: 203-*347*.

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

Freedman, Richard, Gibson, Bradford, Donovan, Dorothy, Biemann, Klaus, Eisenbeis, Scott, Parker Jack, and Schimmel, Paul. (1985) Primary Structure of Histidine-tRNA Synthetase and characterization of hisS transcripts. J. Biol. Chem. 260: 10063-10068.

Graham, Roger W, Van Doren, Kevin, Bektesh, Susan, and Candido, E. Peter M. (1988) Maturation of Major Ubiquitin Gene Transcripts in *Ceanorhabditis elegans* Involves the Acquisition of a Trans-spliced Leader. J. Biol. Chem. 263: 10415-10419.

Herbert, C, and Labouesse, M. (1988) The NAM2 protein from *S.cerevisiae* and *S.douglasii* are mitochondrial leucyl-tRNA synthetase, and are involved in mRNA splicing. The EMBO J 7: 473-483.

Heschl, M.P. and Baillie, D.L. (1989) Characterization of the *hsp70* Multigene Family of *Caenorhabditis elegans*. DNA 8: 233-243.

Hoben, P., Royal, N., Cheung, A., Yamao, F., Biemann, K., and Soll, D. (1982) *Escherichia coli* Glutaminyl-tRNA Synthetase. J. Biol. Chem. 257: 11644-11650.

Huang, X. and Hirsh, D. (1989) A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. *USA*. 86: 8640-8644.

Jacob-Molina, Alfredo, Peterson, Ronald, and Yang, David C. H. (1989) cDNA Sequence, Predicted Primary Structure, and Evolving Amphiphilic Helix of Human Aspartyl-tRNA synthetase. J. Biol. Biochem. 264: 16608-16612.

Jones, Michael D, Lowe, Denise M, Borgford, Thor, and Fersht, Alan R. (1986) Natural Variation of Tyrosyl-tRNA Synthetase and Comparison with Engineered Mutants. Biochemistry 25: 1887-1891.

Jordana, Xavier, Chatton, Bruno, Paz-Weisshaar, Maria, Buhler, Jean-Marie, Cramer, Fritz, Ebel, Jean Pierre, and Fasiolo, Franco. (1987) Structure of the Yeast Valyl-tRNA Synthetase Gene (VASI) and the Homology of its Translated Amino Acid Sequence with *Escherichia coli* Isoleucyl-tRNA Synthetase. J. Biol. Chem. 262: 7189-7194.

Killermann, O, Viel, Claude, and Waller, Jean-Pierre. (1978) Methionyl-tRNA Synthetase from Sheep Mammary Gland: Purification of Fully Active Monomeric Enzyme Derived from High-Molecular-Weight Complexes by Controlled Proteolysis. Eur. J. Biochem. 88: 197-204.

Krause, Micheal and Hirsh, David. (1987) A Trans-spliced Leader Sequence on Actin mRNA in *C. elegans*. Cell 49: 753-761.

Labouesse, M., Herbert, C, Dujardin, G., and Slonimski, P.P. (1987) Three suppressor mutations which cure a mitochondrial RNA maturase deficiency occur at the same codon in the open reading frame of the nuclear NAM2 gene. The EMBO J 6: 713-721.

Lapointe, Jacques, Duplain, Louise, and Proulx, Mario. (1986) A single Glutamyl-tRNA Synthetase Amioacylates Glu-tRNA and Gln-tRNA in *Bacillus subtilis* and Efficiently Misacylates *Escherichia coli* Gln-tRNA1 In Vitro. J. Bacter. 165: 88-93.

Lemaire, G, Gros, Claude, Epely, Sylvie, Kaminski,Marie, and Labouesse. (1975) Multiple Forms of Tryptophanyl-tRNA Synthetase from Beef Pancreas. Eur. J. Biochem. 51: 237-252.

Majumder, Arun Lahiri, Akins, Robert A, Wilkinson, Julie G, Kelley, Robert L, Snook, Amy J, and Lambowitz, Alan M. (1989) Involvment of Tyrosyl-tRNA Synthetase in Splicing of Group I Introns in *Neurospora crassa* Mitochondria: Biochemical and immunochemical Analysis of Splicing Activity. Mol. Cell. Biol. 9:2089-2104

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

Marinde, Marc and Waller, Jean-Pierre. (1988) The Yeast Lysyl-tRNA Synthetase Gene. J. Biol. Chem. 263: 18443-18451.

Marinde, Marc. (1991) Aminoacyl-tRNA Synthetase Family from Prokaryotes and Eukaryotes: Structural Domains and Their Implications. Progress in Nucl. Acid. Res and Mol. Biol. 40: 96-142.

Marinde, Marc, Le Corre, Daniel, Louvard, Daniel, Reggio, Hubert, Pailliez, Jean-Philippe, and Waller, Jean-Pierre. (1985) Associating of an Aminoacyl-tRNA Synthetase Complex and of Phenylalanyl-tRNA Synthetase with the Cytoskeletal Framework Fraction from Mammalian Cells. Exp. Cell Res. 156: 91-102.

Mechulam, Yves, Fayat, Guy, and Blanquet, Sylvain. (1985) Sequence of the *Escherichia coli* pheST Operon and Identification of the himA Gene. J. Bacter. 163: 787-791.

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

Myers, Alan M, and Tzagoloff, Alexander. (1985) *MSW*, a Yeast Gene Coding for Mitochondrial Tryptophanyl-tRNA Synthetase. J. Biol. Chem. 260: 15371-15377.

Natsoulis, Georges, Hilger, Francois, and Fink, Gerald. (1986) The HTS1 Gene Encodes Both the Cytoplasmic and Mitochondrial Histidine tRNA Synthetase of S. Cerevisiae. Cell 46: 235-243.

Neidhardt, Frederick C, Parker, Jack, and McKeever, W. Galen. (1975) Function and Regulation of Aminoacyl-tRNA Synthetases in Prokaryotic and Eukaryotic Cells. Ann. Rev. Bacter. 216-234.

Nelson, Donald W. (1988) Structure, Organization, and Expression of the 5s RNA Genes of *C. elegans* and *C. briggsae*. PhD. Thesis, Simon Fraser University, Burnaby, B.C, Canada.

Pape, Louis K, Koerner, T.J, and Tzagoloff, Alexander. (1985) Characterization of a Yeast Nuclear Gene (MST1) Coding for the Mitochondria Threonyl-tRNA Synthetase. J. Biol. Chem. 260: 15362-15370.

Pape, Louis K and Tzagoloff, Alexander. (1985) Cloning and characterization of the gene for the yeast cytoplasmic threonyl-tRNA synthetase. Nucl. Acid. Res. 13: 6171-6183.

Pearson, William R and Lipman, David J. (1988) Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. *USA*. 85: 2444-2448.

Putney, Scott D and Schimmel, Paul. (1981) An aminoacyl tRNA synthetase binds to a specific DNA sequence and regulates its gene transcription. Nature 291: 632-635.

Putney, Scott D, Sauer, Robert T, and Schimmel, Paul. (1981) Purification and Properties of Alanine tRNA Synthetase from *Escherichia coli*. J. Biol. Chem. 256: 198-204.

Putzer, Harald, Brakhage, Axel A, and Manago, Grunberg-Manago. (1990) Independent Genes for Two Threonyl-tRNA Synthetase in *Bacillus subtilis*. J. Bacter. 172: 4593-4602.

Rogalski, T.M, Moerman, D.G, and Baillie, D.L. (1982) Essential genes and deficiencies in the *unc-22* IV region of *C. elegans*. Genetics 102: 725-736.

Rogalski, T.M. and Baillie, D.L. (1985) Genetic organization of the *unc-22* (IV) gene and the adjacent region in *C. elegans*. Mol. Gen. Genet. 201: 409-414.

Rogalski, Teresa M and Riddle, Donald. (1988) A *Caenorhabditis elegans* RNA Polymerase II Gene, *ama-1* IV, and Nearby Essential Genes. Genetics 118: 61-74.

Rogalski, Teresa M, Bullerjahn, Anne M. E and Riddle, Donald. (1988) Lethal and Amanitin-Resistance Mutations in the *Caenorhabditis elegans ama-1* and *ama-2* Genes. Genetics 120: 409-422.

Rosenberg, Martin and Court, Donald. (1979) Regulatory Sequences Involved in The Promotion and Termination of RNA Transcription. Ann. Rev. Genetics. 13: 319-353.

Ruff, M, Krishnaswamy, S, Boeglin, M, Poterszman, A, Mitschler, Podjarny, A, Rees, B, Thierry, D, and Moras, D. (1991) Class II Aminoacyl Transfer RNA Synthetases: Crystal Structure of Yeast Aspartyl-tRNA Synthetase Complexed with Aspartyl-tRNA. Science 252: 1682-1689.

Rychlik, W and Rhoads, Robert. (1989) Computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in the amplification of DNA. Nucl Acid Res. 17: 8543-8551.

Schimmel, Paul. (1987) Aminoacyl-tRNA Synthetases: General scheme of Structure-Function relationship in the Polypeptides and Recognition of Transfer RNA. Ann. Rev. Biochem. 56: 125-158.

Schneller, Jean-Marie, Scheider, Claude, and Stahl, A.J. (1978) Distinct Nuclear Genes for Yeast Mitochondrial and Cytoplasmic Methionyl-tRNA Synthetases. Biochem. Biophys. Res. Commun. 85: 1392-1399.

Schon, Astrid, Gamini, C, Gough, Simon, and Soll, Dieter. (1988) Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature 331: 187-190.

Siddiqui, Farooq F and Yang, David C. H. (1985) Generation of multiple forms of methionyl-tRNA synthetase from multi-enzyme complex of mammalian aminoacyl-tRNA synthetase by endogenous proteolysis. Biochimica et Biophysica Acta 828: 177-187.

Smith, G.E. and Summers, M.O. (1979) The bidirectional transfer of DNA and RNA to nitrocellulose or DBM paper. Anal. Biochem. 109: 123-129.

Snutch, T.P. (1984) A molecular and genetic analysis of the heat shock response of *Caenorhabditis elegans*. PH.D thesis, Simon Fraser University, Burnaby, B.C., Canada.

Snutch, P., Heschl, M., and Baillie, D.L. (1988) The *Caenorhabditis elegans hsp70* gene family: a molecular genetic characterization. Gene 64: 241-255.

Starzyk, Ruth M, Webster, Teresa A, and Schimmel, Paul. (1987) Evidence for Dispensable Sequence Inserted into a Nucleotide Fold. Science 237: 1614-1617.

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

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

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

Tsui, W.L and Siminovitch, L. (1987) Cloning, Structure, and Expression of Mammalian Genes for Histidyl-tRNA Synthetase. Nucl. Acid. Res. 15: 3349-3367.

Tzagoloff, Alexander, Akai, Anna, Kurkulos, Maryellen and Repetto, Barbara. (1988) Homology of Yeast Mitochondrial Leucyl-tRNA Synthetase and Isoleucyl- and Methionyl-tRNA Synthetases of *Escherichia coli*. J. Biol. Chem. 263: 850-856.

VanBogelen, Ruth A, Vaughn, Vicki, and Neidhardt, F.C. (1983) Gene for Heat-Inducible Lysyl-tRNA Synthetase (Lysyl) maps near CadA in *E. coli*. J. Bacter. 153: 1066-1068.

Vellekamp, Gary, Sihag, Ram K, and Deutscher, Murray P. (1985) Comparison of the Complexed and Free Forms of Rat Liver Arginyl-tRNA Synthetase and Origin of the Free Form. J. Biol. Chem. 260: 9843-9847.

Watson, J.D, Hopkins, N.H, Roberts, J.W, Steitz, J.A, and Weiner, A.M. (1987) Molecular Biology of The Gene. The Benjamin/Cummings publishing company, Inc. Menlo park, California.

Watzel, R. (1978) Aminoacyl-tRNA synthetase families and their significance to the origin of the genetic code. Origin of Life. 9: 39-50.

Winter, G. P and Hartley, B. S. (1977) The Amino Acid Sequence of Tryptophanyl tRNA Synthetase From *Bacillus stearothermophilus*. FEBS LETTERS. 80: 340-342.

Wood, W.B., ed. (1988) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Monograph series 17, p 58. Cold Spring Harbor Laboratory, New York.