

STRUCTURE, ORGANIZATION AND EXPRESSION
OF THE 5S RNA GENES OF
Caenorhabditis elegans and Caenorhabditis briggsae

by

Donald William Nelson

B.Sc.(Biochem), Simon Fraser University, 1981

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Biological Sciences

© Donald W. Nelson, 1988.

Simon Fraser University

March, 1988

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: Donald William Nelson

Degree: Doctor of Philosophy

Title of Thesis:

The structure, organization and expression of the 5S RNA genes of the
nematodes Caenorhabditis elegans and C. briggsae

Examining Committee:

Chairman:

Dr. B.M. Honda, Associate Professor,
Senior Supervisor

Dr. D.L. Baillie, Professor

Dr. M.J. Smyth, Professor

Dr. D.G. Moerman, Assistant Professor,
Dept. of Zoology, U.B.C., Public
Examiner

Dr. K.C. Eastwell, Assistant
Professor, Dept. of Biology, S.F.U.,
Public Examiner

Dr. Karen U. Sprague, Associate
Professor, Institute of Molecular
Biology, University of Oregon,
External Examiner

Date Approved Mar 21, 1988

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

The structure, organization and expression of the 5S RNA genes of the

nematodes Caenorhabditis elegans and C. briggsae

Author:

(signature)

Donald William Nelson

(name)

21/03/88

(date)

ABSTRACT

The objective of this work was to study the mechanisms of transcription in the nematode C. elegans, using homologous 5S RNA genes as model templates.

The C. elegans genome contains approximately 110 homogeneous copies of a 1 kb tandem repeat which encodes 5S RNA. A strain-specific restriction fragment length difference tightly linked to C. elegans 5S DNA was used to genetically map the tandem cluster between unc-76 and dpy-21 on the right arm of linkage group V.

Two approaches were used to aid in the identification of functionally relevant features of the C. elegans 5S DNA repeat: the first considers evolutionary sequence conservation as an indicator of functional importance, the second relies on a direct functional assay.

The 5S RNA genes of C. briggsae, a closely related nematode species, consist of approximately 65 copies of a 1 kb repeat and 20 copies of an 0.7 kb repeat. Each 5S DNA repeat class is organized in a separate tandem cluster in the genome.

Comparison of the C. elegans and C. briggsae 5S DNA repeat sequences shows that all three share perfectly conserved 5S RNA coding sequences. The 1 kb repeats share a second highly conserved sequence which represents an oppositely oriented transcription unit. This region was independently shown to encode an abundant 100 nt transcript (5'L) which participates in a trans-splicing process in C. elegans. The 1 kb C. briggsae 5S

DNA repeat encodes an analogous 100 nt cellular transcript. The 0.7 kb C. briggsae 5S DNA repeat contains an imperfectly conserved 5'L coding sequence in the complementary orientation and does not produce a detectable cellular transcript.

All three nematode 5S DNA repeats are efficient templates for the RNA polymerase III-specific transcription of 5S RNA (but not 5'L RNA) in a cell-free extract of C. elegans embryos. The transcription of deletion/substitution derivatives of the C. elegans and C. briggsae 1 kb 5S DNA repeats shows that 5' flanking sequences are required for efficient 5S RNA transcription. A 12 bp conserved sequence element, found upstream of the 5S RNA coding region of the 1 kb repeats, is implicated in the control of transcription efficiency. Only internal 5S RNA coding sequences are required for a template to compete for limiting components of the cell-free extract.

QUOTATION AND DEDICATION

"This is about my dad's work.

This is about pigs.

Big Black Pigs."

To Daniel and Doormouse,
two pals who know
a dirty blot when they see one.

ACKNOWLEDGEMENTS

This work was made possible through the support and patience of my senior supervisor, Dr. B.M. Honda. I would like to thank Dr. D.L. Baillie and Dr. M.J. Smith for constant encouragement and guidance; Dr. R. Waterston for the C. elegans 5S DNA phage; R. Rosenbluth for advise on genetic manipulations; Dr. J. Curran, Dr. T. Snutch, Dr. R. Devlin, J. Boom, J. Hewitt, S. Prasad, E. Cabot and K. Beckenbach for valuable discussions; Dr. A. Beckenbach for the use of the laser printer; Dr. V. Bourne for help with figures; K. McNeil and I. Scott for Cb5S1 sequence; M. Rogalski for excellent technical assistance; Gwen for reality and Daniel for plasmids and laughs.

TABLE OF CONTENTS

Approval	ii
Abstract	iii
Dedication	v
Acknowledgements	vi
List of Tables	x
List of Figures	xi
A. General Introduction	1
References	16
B. The First Part	
Genes coding for 5S ribosomal RNA of the nematode	
<u>Caenorhabditis elegans</u>	26
I. Introduction	27
II. Materials and Methods	28
Strains	28
DNA Isolations	28
Plasmid Subcloning	29
Restriction Mapping and DNA Sequencing	29
Genomic Blots and Hybridizations	30
Determination of 5S DNA Repeat Copy Number	30
III. Results	31
The 1 kb Bam HI Repeat Encodes 5S RNA	31
Genomic Organization and Repeat Homogeneity	
of <u>C. elegans</u> 5S DNA	34
IV. Discussion	37
V. References	40

C.	The Second Part	
	Genetic Mapping of the 5S RNA Gene Cluster of the Nematode <u>Caenorhabditis elegans</u>	44
I.	Introduction	45
II.	Materials and Methods	46
	Reagents	46
	Nematodes	46
	Southern Transfers and Hybridizations	49
	Linkage of the 5S RNA Gene Cluster with the RFLD ...	50
III.	Results and Discussion	50
	Identification and Description of the RFLD	50
	The Bam HI N2/BO RFLD is Linked to the 5S RNA Gene Cluster	52
	Genetic Mapping of the 5S RNA Gene Cluster	55
IV.	References	63
D.	The Third Part	
	Comparisons of <u>C. briggsae</u> and <u>C. elegans</u> 5S DNA Repeat Sequences Identify Two Highly Conserved Transcribed Regions	65
I.	Introduction	66
II.	Materials and Methods	67
	Reagents	67
	Nucleic Acid Isolations	68
	Isolation of <u>C. briggsae</u> 5S DNA Repeats	68
	Blots and Hybridization Conditions	69
	Plasmid Sequencing and Sequence Data Analysis	70
	Transcription Reactions	70
III.	Results	71
	Isolation and Genomic Organization of <u>C. briggsae</u> 5S DNA	71

Genomic Organization of <u>C. briggsae</u> 5S RNA genes	...74
Structural Features of the <u>C. briggsae</u> 5S DNA repeats77
Comparison of <u>C. elegans</u> and <u>C. briggsae</u> 5S DNA Repeat Sequences80
Are Both <u>C. briggsae</u> 5S DNA Repeat Families Functional?84
IV. Discussion88
V. References90
E. The Fourth Part	
Sequence Requirements for Efficient 5S RNA Transcription in a <u>C. elegans</u> cell-free extract92
I. Introduction93
II. Materials and Methods94
Template Constructions94
Transcription Reactions96
III. Results97
Template Requirements for 5S RNA Transcription from pCe5S197
Competition of pCe5S1 5'D/S Templates for Limiting Extract Components100
5S RNA Transcription From the <u>C. briggsae</u> 1 kb 5S DNA Repeat Requires a Conserved 5' Flanking Sequence Element104
How does the 5' Flanking Sequence Promote 5S RNA Transcription?108
IV. Discussion108
V. References113
F. General Discussion115
References128

List of Tables

Table

2.1 Summary of Genetic Mapping Data59

List of Figures

FIGURE

1.1	Structure of <u>C. elegans</u> 5S DNA isolates	32
1.2	Sequence of the <u>C. elegans</u> 5S DNA Unit Repeat	33
1.3	Genomic Organization of <u>C. elegans</u> 5S DNA	35
1.4	Quantitation of the <u>C. elegans</u> Haploid 5S DNA content	36
1.5	Other genomic sequences hybridizing to the <u>C. elegans</u> 5S DNA repeat	38
2.1	Genetic Protocol for Three-factor Crosses	48
2.2	Identification of an RFLD Associated with the <u>C. elegans</u> 5S RNA Gene Cluster	51
2.3	Linkage of the RFLD and the 5S RNA Gene Cluster	54
2.4	Genetic Mapping of the 5S RNA Gene Cluster	58
3.1	Identification of <u>C. briggsae</u> Genomic Sequences Encoding 5S RNA	72
3.2	Restriction Maps of <u>C. briggsae</u> Genomic 5S DNA Repeat Families	73
3.3	Genomic Organization of <u>C. briggsae</u> 5S DNA	75
3.4	Quantitation of the <u>C. briggsae</u> haploid 5S DNA Content	76
3.5	Sequence of the <u>C. briggsae</u> 5S DNA repeats	79
3.6	Comparisons of <u>C. elegans</u> and <u>C. briggsae</u> 5S DNA Repeat Sequences	81
3.7	Transcription of <u>C. briggsae</u> 5S DNA in a <u>C. elegans</u> cell-free transcription system	85
3.8	<u>in vivo</u> Expression of the 5' L Coding Regions of the <u>C. briggsae</u> 5S DNA Repeats	87
4.1	Construction and Structure of pCe5S1 5' and 3' Deletion/Substitution Templates, and the Internal Reference Template, pCe5S1MAXI	99

4.2	Transcription of pCe5S1, 5' and 3' Deletion/Substitution Templates	102
4.3	Competition of pCe5S1 5'D/S Templates for Limiting Extract Components	103
4.4	Construction and Structure of pCbe5S1 5' D/S Templates	105
4.5	Transcription of pCbe5S1 5'D/S Templates	107
4.6	Construction and Transcription of pCe5S1 5'D/S+MAXI and pCbe5S1 5'D/S+MAXI Templates	109

GENERAL INTRODUCTION

All of the information required for cell growth and development is encoded in the cell's genomic DNA. The expression of this stored information is outlined by the central dogma (Crick, 1970):

DNA -----> RNA -----> protein.

In the first step in this expression pathway, genomic DNA serves as a template for the synthesis of RNA, a process termed transcription. The RNA products of transcription either function directly (ie: RNA components of the ribosome), or serve as mRNA templates for the synthesis of proteins in the process of translation. The protein products of translation represent the effectors of the information stored in the genome.

It is clear that only a fraction of the information stored in the genome is expressed at any time; many gene products are expressed only in specific differentiated cell types or when required for cell growth (Alberts et.al., 1983). This regulation of gene expression provides the conceptual foundation of our understanding of development and the cellular response to environmental stimuli.

Gene expression can be regulated at any step in the transfer of genetically encoded information to functional product (Brown, 1981; Darnell, 1982; Alberts et.al., 1983). The process of transcription represents the major target of coarse regulatory mechanisms; generally, genes whose products are not required by a cell are not transcribed. In E. coli, the induction of the lac operon results in a thousand fold increase in lac transcription

(Miller and Reznikoff, 1978). In eukaryotes, this transcription differential between repressed and induced states is even more extreme (Alberts et al., 1983). If we are to gain an understanding of the mechanistic underpinnings of a complex biological phenomenon such as eukaryotic development, we must first establish the molecular mechanisms by which transcription is regulated.

Two complementary approaches, based on classical in vivo genetic and in vitro biochemical analyses, have been developed for the study of transcription. The classical genetic approach relies on the in vivo identification and characterization of randomly induced regulatory mutations to deduce the interactions between components of the transcription apparatus. The biochemical approach focuses on the manipulation of individual components of the transcription apparatus in vitro to define their function in the overall process. The limitations of these approaches, imposed by difficulties in identifying the appropriate regulatory mutant classes in vivo and the numerous artifacts associated with biochemical reconstruction of complex processes in vitro, are overcome by combining the construction of defined mutant templates in vitro with their in vivo characterization following transfection into a suitable host.

The application of these approaches to the analysis of prokaryotic transcription, particularly in E. coli, has been very productive in defining the molecular mechanics of this complex process (McClure, 1985; Reznikoff and McClure, 1986; Travers,

1987). The essential components of the prokaryotic transcription apparatus are the complex enzyme, RNA polymerase, and the template DNA which is to be transcribed. Transcription in E. coli is catalysed by a single class of RNA polymerase, a complex enzyme containing several subunits. The core enzyme, $\alpha_2\beta\beta'\omega_2$, which is responsible for transcript elongation, requires an additional factor, σ , to accurately initiate transcription. Template sequences located upstream of the transcription initiation site are also required for accurate initiation. The comparison of the 5' sequences flanking many prokaryotic transcription units have identified two highly conserved sequences centered 35 (-35 box, TTGACT) and 10 (-10 box, TATAAT) bp upstream of the initiation site. These promoter sequences serve as a recognition site for the binding of the RNA polymerase holoenzyme, $\alpha_2\beta\beta'\omega_2\sigma$ to form a pre-initiation or closed complex. The holoenzyme in this complex is directionally oriented; the σ factor interacts directly with the -35 box and the core enzyme with the -10 box. This closed complex undergoes a conformational change which results initially in the unwinding of 10 bp or 1 turn of the template helix centered over the -10 box. Migration of the holoenzyme and the unwound region to the initiation site completes the transition to the open initiation complex which is poised for rapid initiation of transcription.

Superimposed on these initiation events are additional protein-template interactions which positively or negatively regulate the frequency of transcription initiation. The lac

operon of E. coli provides a well-characterized model system illustrating these regulatory interactions (Miller and Reznikoff, 1978). The negative regulator of lac transcription, the lac repressor, binds to specific template sequences located between the promoter and initiation site. This repressor-operator complex allows the RNA polymerase holoenzyme to bind to the promoter, but blocks isomerization to the open transcription complex, thus repressing lac transcription (Straney and Crothers, 1987). In contrast, the cAMP-CAP (catabolite activator protein) complex binds template sequences upstream of the lac promoter and stimulates the initiation of transcription by facilitating the entry of RNA polymerase holoenzyme into the open initiation complex by either protein-protein interaction, or by altering the template conformation (ie: inducing negative supercoiling which may promote unwinding of the -10 box or its migration to the initiation site) (Travers, 1987).

Additional regulatory mechanisms affect the transcription process in prokaryotes. Alternate RNA polymerases (bacteriophage T7 transcription (Lewin, 1977)) and σ factors (E. coli heat shock response (Grossman et.al., 1984), B. subtilis vegetative and sporulation-specific transcription (Johnson et.al., 1983) and bacteriophage T4 late gene expression (Lewin, 1977)) are employed to coordinately regulate different classes of genes.

Transcription termination is also the target of some regulatory systems (Platt, 1986), such as the antitermination functions of λ N and Q gene products (Roberts, 1987) and the coupling of

transcription and translation in the attenuation of the E. coli trp operon (Yanofsky, 1987).

Progress in characterizing the transcription process in eukaryotes has been slow, primarily because of difficulties in identifying the appropriate regulatory mutant classes by classical genetic analysis. This problem was overcome with the development of recombinant DNA techniques with which isolated templates could be altered at will. The subsequent development of transcriptionally active cell-free extracts provided the requisite tools for the biochemical approach to the study of eukaryotic transcription (Heintz and Roeder, 1982).

Both the template DNA and the protein components of the eukaryotic transcription apparatus are significantly more complex than their prokaryotic counterparts. The eukaryotic genome is typically one to three orders of magnitude larger than the prokaryotic genome, and is organized into a nucleic acid-protein complex referred to as chromatin (reviewed by McGhee and Felesenfeld, 1980; Igo-Kemenes et.al., 1982), in contrast to prokaryotic 'bare DNA' template. The basic repeat unit of chromatin, the nucleosome, consists of about 200 bp of DNA wrapped twice around a histone octamer core containing two copies of each histone H2A, H2B, H3 and H4. This basic repeat unit is organized into a hierarchy of higher order structures by interaction with histone H1 and other non-histone chromosomal proteins (Weisbrod, 1982). These higher order structures are

thought to reflect the transcriptional potential of the genomic sequences found therein (Eissenberg et.al., 1985).

The protein components of the transcription apparatus of eukaryotes are also more complex than in prokaryotes. Eukaryotes use three distinct classes of RNA polymerase for transcription (reviewed by Roeder, 1976; Sentenac, 1985). These three classes of polymerase can be distinguished on the basis of their subunit composition, sensitivity to the fungal toxin α -amanitin and the class of genes they transcribe. RNA polymerase I is typically insensitive to high levels of α -amanitin and transcribes the large rRNA precursor which is subsequently processed into 18S, 5.8S and 26S rRNAs. RNA polymerase II is typically sensitive to low levels of α -amanitin and transcribes the hnRNAs which are subsequently processed into protein coding mRNAs. RNA polymerase III is sensitive to intermediate levels of α -amanitin and transcribes 5S RNA, tRNAs and other small RNAs (adenovirus VA RNAs, U6 snRNA, 7SL RNA and 7SK RNA).

Like the E. coli core enzyme, these eukaryotic RNA polymerases are unable to accurately initiate transcription in the absence of other cellular factors (Heintz and Roeder, 1982). Recent studies have revealed the existence of a multitude of cellular factors which interact with template sequences to direct or modulate the initiation of transcription by these enzymes (Dyran and Tjian, 1985; McKnight and Tjian, 1986).

The 5S RNA genes of X. laevis represent the most extensively characterized example of the eukaryotic transcription process.

The accurate initiation of 5S RNA transcription from cloned 5S DNA templates in vitro is directed by template sequences located within the transcribed region (Sakonju et.al., 1980; Bogenhagen et.al., 1980). The sequence of this internal control region (ICR, positions +50 to +80 of the 5S RNA coding sequence) is recognized by a 5S-specific (TFIIIA) (Engelke et.al., 1980) and at least two general (TFIIC and TFIIIB) RNA polymerase III-specific transcription factors (Segall et.al., 1980), which bind sequentially to form a stable transcription complex (Bogenhagen et.al., 1982; Lassar et.al., 1983). This complex is recognized by RNA polymerase III, and remains in place for many rounds of transcription (Wolffe et.al., 1986).

Xenopus 5S RNA genes form several distinct gene families whose transcription is developmentally regulated (reviewed by Korn, 1982); the predominant families encode the major oocyte-specific 5S RNA and the somatic cell-specific 5S RNA. The developmental expression of these genes appears to be regulated by both the availability of the 5S-specific transcription factor TFIIIA (Pelham and Brown, 1980; Honda and Roeder, 1980; Shastry, Honda and Roeder, 1984; Andrews and Brown, 1987) and the stability of the TFIIIA-ICR complex (Brown, 1984; Brown and Schlissel, 1985; Wolffe and Brown, 1987). Variation in the nucleotide sequence of the ICR of the oocyte-specific 5S RNA genes is thought to preferentially direct stable transcription complex formation on the somatic cell-specific 5S RNA genes at low TFIIIA levels, while the mass excess of TFIIIA present during

oogenesis serves to activate both oocyte-specific and somatic cell-specific genes.

The basic features of 5S RNA transcription observed in X. laevis are conserved in other species. Recent results however indicate that in several other systems, template sequence upstream of the transcription initiation site are also required for 5S RNA transcription in vitro (Morton and Sprague, 1984; Selker et.al., 1986; Tyler, 1987; Garcia et.al., 1987).

The transcription of many other well-characterized RNA polymerase III genes, particularly those encoding tRNAs (Sharp et.al., 1986) and adenovirus VA1 RNA (Wu et.al., 1987), also follow the basic features of 5S RNA transcription. These genes do not however require TFIIIA, which is 5S-specific, and possess a different conserved split internal promoter which interacts with TFIIIC (recently shown to consist of two separate proteins (Ottonello et.al., 1987; Dean and Berk, 1987; Boulanger et.al., 1987) and TFIIIB to form a stable transcription complex (Lassar et.al., 1983). As noted above, the transcription of these genes often requires, or is modulated by, 5' flanking and/or 3' flanking template sequences (Sharp et.al., 1986).

Recent studies on the transcription of U6 snRNA, 7SL RNA and 7SK RNA further emphasize the involvement of 5' flanking sequences in the regulation of RNA polymerase III transcription (Sollner-Webb, 1988). While all three gene families appear to contain internal promoter sequences, transcription also requires intact 5' flanking sequences (Carbon et.al., 1987; Ullu and

Weiner, 1985; Murphy et.al., 1986). Surprisingly, RNA polymerase III transcription can be programmed by the sequences upstream of the 7SK RNA coding region even in the absence of the 7SK RNA coding sequence (Murphy et.al., 1987).

In contrast to RNA polymerase III genes, the promoter sequences required for transcription by RNA polymerases I and II are usually located in the template sequences upstream of the transcription initiation site.

RNA polymerase I promoters (Sollner-Webb and Tower, 1986) appear to consist of two separate blocks, one from -40 to +10 which interacts directly with RNA polymerase I-specific transcription factors, and one from -160 to -80 which participates in some unknown way to facilitate in vitro transcription under stringent assay conditions.

RNA polymerase II transcription represents the most complex eukaryotic transcription system, both in terms of the variety of genes transcribed and the regulation of their transcription. Several template sequence elements have been implicated in the regulation of RNA polymerase II transcription. 'Housekeeping' genes which are required for general cell metabolism appear to employ a 'GC'-rich promoter region for efficient transcription (Dynan, 1986). Developmentally regulated genes use a 'TATA' box near position -30 for specification of the transcription initiation site, as well as 'upstream elements' for efficient and inducible expression (Sassone-Corsi and Borrelli, 1986; Maniatis et.al., 1987).

The formation of a stable transcription complex which remains in place for many rounds of transcription appears to be a common feature in all three transcription systems. In each case, putative regulatory sequence elements appear to interact with RNA polymerase class-specific transcription factor(s) to form stable complexes which program transcription by the appropriate class of RNA polymerase.

The biochemical dissection of eukaryotic transcription has proven valuable in identifying some of the interactions required for regulated gene expression. The demonstration that these interactions also occur in vivo relies on transfection experiments where modified templates are transfected into host cells and their in vivo expression characterized. These studies have largely confirmed the in vitro expression studies, but revealed an additional regulatory sequence element, termed an enhancer (Sassone-Corsi and Borrelli, 1986; Maniatis et.al., 1987), not easily detected in cell-free extracts. These are relatively short sequence elements which act in a position- and orientation-independent fashion to enhance transcription from a cis-linked promoter by up to two orders of magnitude. These elements are thought to act by altering the conformation of the template (directing a change in supercoiling or altering general chromatin configuration), or by facilitating the interaction of other regulatory regions to direct the initiation of transcription. Such enhancer elements have been shown to be important factors in regulating transcription by RNA polymerases

I and II. Enhancers also appear to modulate the transcription of U6 snRNA (Bark et.al., 1987; Carbon et.al., 1987) by RNA polymerase III .

While the biochemical analysis of eukaryotic transcription has focused on cell-free extracts derived from genetically inaccessible sources (X. laevis oocytes and mammalian tissue culture cells), recently cell-free systems have been described from genetically well-characterized species, most notably the unicellular yeast S. cerevisiae (Klekamp and Weil, 1982; Koski et.al., 1982) and the complex metazoan D. melanogaster (Dingermann et.al., 1981). The genetic dissection of the transcription apparatus in these species is also underway, with mutations affecting a variety of components under examination. While both these species represents a powerful genetic system, it would be advantageous to have a third system of intermediate complexity in which to pursue these studies.

The free-living nematode Caenorhabditis elegans represents an attractive model system for the genetic and biochemical analysis of transcription in higher eukaryotes. C. elegans is a simple metazoan, possessing less than 1000 somatic cells as an adult. Despite its anatomical simplicity, C. elegans contains the major classes of differentiated cell types including hypodermis, gut, muscle, neurons and gonad. The small number of cell divisions during development and transparency of the developing worm have allowed the determination of the complete cell lineage of C. elegans (Sulston and Horvitz, 1977; Kimble and

Hirsh, 1979; Sulston et.al., 1980; Sulston, et.al., 1983), making it an attractive model system for studying eukaryotic development (Wilkins, 1986).

Genetic analysis in C. elegans (Brenner, 1974) is facilitated by a short generation time (3.5 days) and large progeny yield (approximately 300 progeny per adult). In addition, C. elegans reproduces as either a self-fertilizing hermaphrodite (X/X), simplifying the establishment and maintenance of homozygous mutant strains, or by mating with rare males (X/O) spontaneously arising by X chromosome non-disjunction, allowing for the transfer of genetic markers between strains. The exploitation of these features by many investigators has produced an extensive linkage map of mutations exhibiting visible and lethal phenotypes. These mutations are spread over five autosomal linkage groups (LG I through V) and a sex (X) chromosome.

C. elegans is also well suited to molecular analysis. The relatively small size of the haploid genome, 8×10^7 bp, and high proportion of single copy sequence (83%) (Sulston and Brenner, 1974), facilitates the isolation of specific genomic sequences from recombinant libraries. The development of techniques for the genetic mapping of strain-specific restriction fragment length differences (RFLDs) (Hirsh et.al., 1979; Emmons et.al., 1979; Rose et.al., 1982; Files et.al., 1983; Cox et.al., 1985) and in situ hybridization (Albertson, 1984) allow random DNA fragments to be positioned on the genetic map. Genetically

defined loci can be identified by transposon-tagging with Tc-1 (Emmons et.al., 1983; Liao et.al., 1983), a C. elegans transposable element (Moerman, Waterston, 1984; Greenwald, 1985; Moerman, Benian and Waterston, 1986; Herman and Shaw, 1987; Collins et.al., 1987). Recently, John Sulston and Alan Coulson have undertaken the construction of a physical map of the C. elegans genome which will further simplify the isolation and characterization of biochemically undefined genes (Coulson et.al., 1986). Finally, techniques for introducing exogenous DNA into the genome of Caenorhabditis elegans are being developed to facilitate the analysis of cloned genes in vivo (Stinchcomb et.al., 1985; Fire, 1986; Jefferson et.al., 1987).

The ease with which C. elegans is cultured facilitates the isolation of large amounts of starting material required for biochemical fractionation. Several enzymes and structural proteins have been characterized in C. elegans including myosin heavy chains (Dibb et.al., 1985; Miller et.al., 1986), a developmentally regulated gut esterase (McGhee, 1987), glyceraldehyde-3-phosphate dehydrogenase (Yarbrough and Hecht, 1984), choline acetyltransferase (Rand and Russell, 1985) and vitellogenin (Sharrock, 1984). Recently, the biochemical and genetic analysis of RNA polymerase II from C. elegans has been reported (Sanford et.al., 1983, 1985; Rogalski and Riddle, 1988).

The object of this thesis was to establish the use of cell-free extracts for the study of RNA polymerase III-specific transcription in C. elegans. The thesis has been divided into

four parts as follows. First, in order to obtain a well-characterized template for the development of a transcriptionally active C. elegans cell-free extract, I undertook the identification and characterization of the C. elegans 5S RNA gene family. In the second part, I identify a strain-specific restriction fragment length difference tightly linked to the C. elegans 5S RNA gene cluster, and genetically map it to the right arm of linkage group V. In the third part, I extend my analysis of 5S DNA structure and organization to the closely related species C. briggsae. Finally, I use the cell-free extract of C. elegans embryos developed in this lab to examine the template requirements of C. elegans 5S RNA transcription in vitro.

REFERENCES

- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1983).
Molecular Biology of the Cell.
Garland publishing, NY, NY.
- Albertson, D.G. (1984).
Localization of the ribosomal genes in Caenorhabditis elegans chromosomes by in situ hybridization using biotin-labeled probes.
EMBO J. 3 1227-1234.
- Andrews, M.T. and Brown, D.D. (1987).
Transient activation of oocyte 5S RNA genes in Xenopus embryos by raising the level of the trans-acting factor TFI_{IIIA}.
Cell 51 445-453.
- Bark, C., Weller, P., Zabielski, J., Janson, L., and Pettersson, U. (1987).
A distant enhancer element is required for polymerase III transcription of a U6 RNA gene.
Nature 328 356-359.
- Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
II. The 3' border of the region.
Cell 19 27-35.
- Bogenhagen, D.F., Wormington, W.M. and Brown, D.D. (1982).
Stable transcription complexes of Xenopus 5S RNA genes: a means to maintain the differentiated state.
Cell 28 413-421.
- Boulanger, P.A., Yoshinaga, S.K. and Berk, A.J. (1987).
DNA-binding properties and characterization of human transcription factor TFI_{IIIC2}
J. Biol. Chem. 262 15096-15105.
- Brenner, S. (1974).
The genetics of Caenorhabditis elegans.
Genetics 77 71-94.
- Brown, D.D. (1981).
Gene expression in eukaryotes.
Science 211 667-674.

Brown, D.D. (1984).

The role of stable complexes that repress and activate eukaryotic genes.

Cell 37 359-365.

Brown, D.D. and Schlissel, M.S. (1985).

The molecular basis of differential gene expression of two 5S RNA genes.

Cold Spring Harbor Symposia of Quantitative Biology 50
p. 549-553. Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY.

Carbon, P., Murgo, S., Ebel, J-P., Krol, A., Tebb, G. and Mattaj, I.W. (1987).

A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II.

Cell 51 71-79.

Collins, J., Saari, B. and Anderson, P. (1987).

Activation of a transposable element in the germ line but not the soma of Caenorhabditis elegans.

Nature 328 726-728.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986).

Toward a physical map of the genome of the nematode C. elegans.

Proc. Natl. Acad. Sci. 83 7821-7825.

Cox, G.N., Carr, S., Kramer, J.M. and Hirsh D. (1985).

Genetic mapping of Caenorhabditis elegans collagen genes using DNA polymorphisms as phenotypic markers.

Genetics 109 513-528.

Crick, F.H.C. (1970)

Central dogma of molecular biology.

Nature 227 561-563.

Darnell, J.E. (1982).

Variety in the level of gene control in eukaryotic cells.

Nature 297 365-371.

Dean, N. and Berk, A.J. (1987).

Separation of TFIIIC into two functional components by sequence specific DNA affinity chromatography.

Nuc. Acid. Res. 23 9895-9907.

Dibb, N.J., Brown, D.M., Karn, J., Moerman, D.G., Bolten, S.L. and Waterston, R.H. (1985).

Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of the unc-54 myosin heavy chain of C. elegans.

J. Mol. Biol. 183 543-551.

Dingermann, T., Sharp, S., Appel, B., Defrance, D., Mount, S., Heiermann, R., Pongs, O. and Soll, D. (1981).

Transcription of cloned tRNA and 5S RNA genes in a *Drosophila* cell free extract.

Nuc. Acid. Res. 9 3904-3918.

Dynan, W.S. (1986).

Promoters for housekeeping genes.

TIG 2 196-197.

Dynan, W.S. and Tjian, R. (1985).

Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins.

Nature 316 774-778.

Eissenberg, J.C., Cartwright, I.L., Thomas, G.H. and Elgin, S.C.R. (1985).

Selected topics in chromatin structure.

Ann. Rev. Genet. 19 485-536.

Emmons, S.W., Klass, M.R. and Hirsh, D. (1979).

Analysis of the constancy of DNA sequences during development and evolution of the nematode Caenorhabditis elegans.

Proc. Natl. Acad. Sci. USA 76 1333-1337.

Emmons, S.W., Yesner, L., Ruan, K-S. and Katzenberg, D. (1983).

Evidence for a transposon in Caenorhabditis elegans.

Cell 32 55-65.

Engelke, D.R., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1980).

Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes.

Cell 19 717-728.

Files, J.G., Carr, S. and Hirsh, D. (1983).

Actin gene family of Caenorhabditis elegans.

J. Mol. Biol. 164 355-375.

Fire, A. (1986).

Integrative transformation of Caenorhabditis elegans.

EMBO J. 5 2673-2680.

- Garcia, A.D., O'Connell, A.M. and Sharp, S.J. (1987).
Formation of an active transcription complex in the
Drosophila melanogaster 5S RNA gene is dependent on an
upstream region.
Mol. Cell. Biol. 7 2046-2051.
- 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.
- Grossman, A.D., Erickson, J.W. and Gross, C.A. (1984).
The htpR gene product of E. coli is a sigma factor for heat
shock promoters.
Cell 38 383-390.
- Heintz, N. and Roeder, R.G. (1984).
Transcription of eukaryotic genes in soluble cell-free
systems.
in: Genetic Engineering, Principles and Methods, Vol. 4.
edited by J.K. Setlow and A. Hollaender. Plenum
Publishing, NY, NY. pp. 57-89.
- Herman, R.K. and Shaw, J.E. (1987).
The transposable genetic element Tc1 in the nematode C.
elegans.
TIG 3 222-225.
- Hirsh, D., Emmons, S.W., Files, J.G. and Klass, M.R. (1979).
Stability of the C. elegans genome during development and
evolution.
in: Eukaryotic gene expression. edited by R. Axel,
T. Maniatis and C.R. Fox. Academic press, NY.
p 205-218.
- Honda, B.M. and Roeder, R.G. (1980).
Association of a 5S gene transcription factor with 5S RNA
and altered levels of the factor during cell
differentiation.
Cell 22 119-126.
- Igo-Kemenes, T., Horz, W. and Zachau, H.G. (1982).
Chromatin.
Ann. Rev. Biochem. 51 89-121.
- Jefferson, R.A., Klass, M., Wolf, N. and Hirsh, D. (1987).
Expression of chimeric genes in C. elegans.
J. Mol. Biol. 193 41-46.

- Johnson, W.C., Moran, C.P. and Losick, R. (1983).
Two RNA polymerase sigma factors from Bacillus subtilis
discriminate between overlapping promoters for a
developmentally regulated gene.
Nature 302 800-804.
- Kimble, J. and Hirsh, D. (1979).
The postembryonic cell lineages of the hermaphrodite and
male gonads in Caenorhabditis elegans.
Dev. Biol. 70 396-417.
- Klekamp, M.S. and Weil, P.A. (1982).
Specific transcription of homologous class III genes in
yeast-soluble cell-free extracts.
J. Biol. Chem. 257 8432-8441.
- Korn, L.J. (1982).
Transcription of Xenopus 5S ribosomal RNA genes.
Nature 295 101-105.
- Koski, R.A., Allison, D.S., Worthington, M. and Hall, B.D.
(1982).
An in vitro RNA polymerase III system from S. cerevisiae:
effects of deletions and point mutations upon SUP4 gene
transcription.
Nuc. Acid. Res. 10 8127-8143.
- Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983).
Transcription of class III genes:
formation of preinitiation complexes.
Science 222 740-748.
- Lewin, B.M. (1977).
Gene Expression-3: Plasmids and Phages.
John Wiley and Sons, Inc., NY, NY.
- Liao, L.W., Rosenzweig, B. and Hirsh, D. (1983).
Analysis of a transposable element in Caenorhabditis
elegans.
Proc. Natl. Acad. Sci. 80 3585-3589.
- Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987).
Regulation of inducible and tissue-specific gene expression.
Science 236 1237-1245.
- McClure, W.R. (1985).
Mechanism and control of transcription initiation in
prokaryotes.
Ann. Rev. Biochem. 54 171-204.

- McGhee, J.D. (1987).
Purification and characterization of a carboxylesterase from the intestine of the nematode C. elegans.
Biochem. 26 4101-4107.
- McGhee, J.D. and Felsenfeld, G. (1980).
Nucleosome structure.
Ann. Rev. Biochem. 49 1115-1156.
- McKnight, S. and Tjian, R. (1986).
Transcriptional selectivity of viral genes in mammalian cells.
Cell 46 795-805.
- Miller, J.H. and Reznikoff, W.S. (1978).
The operon
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Miller, D.M., Stockdale F.E. and Karn, J. (1986).
Immunological identification of the genes encoding the four myosin heavy chain isoforms of Caenorhabditis elegans.
Proc. Natl. Acad. Sci. 83 2305-2309.
- Moerman, D.G., Benian, G.M. and Waterston, R.H. (1986).
Molecular cloning of the muscle gene unc-22 in Caenorhabditis elegans by Tc1 transposon tagging.
Proc. Natl. Acad. Sci. 83 2579-2583.
- Moerman, D.G. and Waterston, R.E. (1984).
Spontaneous unstable unc-22 IV mutations.
Genetics 108 859-877.
- Morton, D.G. and Sprague, K.U. (1984).
in vitro transcription of a silkworm 5S RNA gene requires an upstream signal.
Proc. Natl. Acad. Sci. 81 5519-5522.
- Murphy, S., Tripodi, M. and Melli, M. (1986).
A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes.
Nuc. Acid. Res. 14 9243-9260.
- Murphy, M., Liegro, C.D. and Melli, M. (1987).
The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter.
Cell 51 81-87.

Ottonello, S., Rivier, D.H., Doolittle, G.M., Young, L.S. and Sprague, K. (1987).

The properties of a new polymerase III transcription factor reveal that transcription complexes can form by more than one pathway.

EMBO J. 6 1921-1927.

Pelham, H.R.B. and Brown, D.D. (1980).

A specific transcription factor that can bind either the 5S RNA gene or 5S RNA.

Proc. Natl. Acad. Sci. 77 4170-4174.

Platt, T. (1986).

Transcription termination and the regulation of gene expression.

Ann. Rev. Biochem. 55 339-372.

Rand, J.B. and Russell, R.L. (1985).

Properties and partial purification of choline acetyltransferase from the nematode C. elegans.

J. Neurochem. 44 189-200.

Reznikoff, W.S. and McClure, W.R. (1986).

E. coli promoters.

p 1-33 in: Maximizing Gene Expression. W. Reznikoff and L. Gold (ed), Butterworth Publishing, Stoneham, MA.

Roberts, J.W. (1987).

Phage lambda and the regulation of transcription termination.

Cell 52 5-6.

Roeder, R.G. (1976).

Eukaryotic nuclear RNA polymerases.

in: RNA Polymerase, R. Losick and M. Chamberlin eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

p. 285-329.

Rogalski, T.M. and Riddle, D.L. (1988).

A Caenorhabditis elegans RNA polymerase II gene, ama-1 IV, and nearby essential genes.

Genetics 118 61-74.

Rose, A.M., Baillie, D.L., Candido, E.P.M., Beckenbach, K.A. and Nelson, D. (1982).

The linkage mapping of cloned restriction fragment length differences in Caenorhabditis elegans.

Mol. Gen. Genet. 188 286-291.

- Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
I. The 5' border of the region.
Cell 19 13-25.
- Sanford, T., Golomb, M. and Riddle, D.L. (1983).
RNA polymerase II from wild type and amanitin-resistant strains of Caenorhabditis elegans.
J. Biol. Chem. 258 12804-23809.
- Sanford, T., Prenger, J.P. and Golomb, M. (1985).
Purification and immunological analysis of RNA polymerase II from Caenorhabditis elegans.
J. Biol. Chem. 260 8064-8069.
- Sassone-Corsi, P. and Borelli, E. (1986).
Transcriptional regulation by trans-acting factors.
TIG 2 215-219.
- Segall, J., Matsui, T., and Roeder, R.G. (1980).
Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III.
J. Biol. Chem. 255 11986-11991.
- Selker, E.U., Morzycka-Wroblewska, E., Stevens, J.N. and Metzenberg, R.L. (1986).
An upstream signal is required for in vitro transcription of Neurospora 5S RNA genes.
Mol. Gen. Genet. 205 189-192.
- Sentenac, A. (1985).
Eukaryotic RNA polymerases.
CRC Crit. Rev. in Biochem. 18 31-90.
- Sharp, S.J., Schaack, J., Cooley, L., Burke, D.J. and Soll, D. (1986)
Structure and transcription of eukaryotic tRNA genes.
CRC Crit. Rev. in Biochem. 19 107-144.
- Sharrock W.J. (1984).
Cleavage of two yolk proteins from a precursor in Caenorhabditis elegans.
J. Mol. Biol. 174 419-431.
- Shastry, B.S., Honda, B.M. and Roeder, R.G. (1984).
Altered levels of a 5S gene-specific transcription factor (TFIIIA) during oogenesis and embryonic development of X. laevis.
J. Biol. Chem. 259 11373-11382.

- Sollner-Webb, B. (1988).
Surprises in polymerase III transcription.
Cell 52 153-154.
- Sollner-Webb, B. and Tower, J. (1986).
Transcription of cloned eukaryotic ribosomal RNA genes.
Ann. Rev. Biochem. 55 801-830.
- Stinchcomb, D.T., Shaw, J.E., Carr, S.G. and Hirsh, D. (1985).
Extrachromosomal DNA transformation of Caenorhabditis elegans.
Mol. Cell. Biol. 5 3484-3496.
- Straney, S.B. and Crothers, D.M. (1987).
Lac repressor is a transient gene-activating protein.
Cell 51 699-707.
- Sulston, J.E., Albertson, D.G. and Thomson, J.N. (1980).
The Caenorhabditis elegans male:
postembryonic development of non-gonadal structures.
Dev. Biol. 78 542-576.
- Sulston, J.E. and Brenner, S. (1974).
The DNA of Caenorhabditis elegans.
Genetics 77 95-104.
- Sulston, J.E. and Horvitz, H.R. (1977).
Post-embryonic cell lineages of the nematode Caenorhabditis elegans.
Dev. Biol. 56 110-156.
- Sulston, J.E., Scheirenberg, E., White, J.G. and Thomson, J.N. (1983).
The embryonic cell lineage of the nematode Caenorhabditis elegans.
Dev. Biol. 100 64-119.
- Travers, A.A. (1987).
Structure and function of E. coli promoter DNA.
CRC Crit. Rev. in Biochem. 22 181-219.
- Tyler, B.M. (1987).
Transcription of Neurospora crassa 5S RNA genes requires a
TATA box and three internal elements.
J. Mol. Biol. 196 801-811.
- Ullu, E. and Weiner, A.M. (1985).
Upstream sequences modulate the internal promoter of the
human 7SL RNA gene.
Nature 318 371-374.

- Weisbrod, S. (1982).
Active chromatin.
Nature 297 289-295.
- Wilkins, A.S. (1986).
Genetic analysis of animal development.
John Wiley and Sons, Inc., NY, NY.
- Wolffe, A.P. and Brown, D.D. (1987).
Differential 5S RNA gene expression in vitro.
Cell 51 733-740.
- Wolffe, A.P., Jordan, E. and Brown, D.D. (1986).
A bacteriophage RNA polymerase transcribes through a Xenopus
5S RNA gene transcription complex without disrupting it.
Cell 44 381-389.
- Wu, G-j., Railey, J.F. and Cannon, R.E. (1987).
Defining the functional domains in the control region of the
adenovirus type 2 specific VA RNA1 gene.
J. Mol. Biol. 194 423-442.
- Yanofsky, C. (1987).
Operon-specific control by transcription attenuation.
TIG 3 356-360.
- Yarbrough, P.O. and Hecht, R.M. (1984).
Two isoenzymes of glyceraldehyde-3-phosphate dehydrogenase
in C. elegans: Isolation, properties and immunochemical
characterization.
J. Biol. Chem. 259 14711-14720.

THE FIRST PART

GENES CODING FOR 5S RIBOSOMAL RNA OF THE NEMATODE

Caenorhabditis elegans

Introduction

We have chosen the nematode C. elegans as a model eukaryotic system for the study of transcription and its regulation. Molecular and recombinant DNA analysis is readily done but, more importantly, genetic methods (reviewed in Zuckerman, 1980; Wood, 1988) provide a powerful complementary tool to dissect the transcriptional apparatus. We have therefore undertaken the characterization of the 5S RNA genes of C. elegans, as a first step towards studying the transcription of these and other genes in this organism.

Eukaryotic 5S RNA genes are ideal templates for the study of transcription and its regulation. These genes are transcribed by RNA pol III; such expression in Xenopus requires internal DNA sequences, a gene-specific protein TFIIIA, as well as other components (reviewed in Heintz and Roeder, 1984; Brown, 1984). These genes have also been used to study the structure, organization and evolution of multigene families. 5S RNA genes are generally organized in tandemly repeated units, containing a gene and spacer DNA (Long and Dawid, 1980). There are, however, variations on this basic organizational theme, including examples of dispersed 5S RNA genes (Rosenthal and Doering, 1983), variations in repeat size (Carroll and Brown, 1976), pseudogenes (Miller et.al., 1978) and even different development-specific gene types (Korn, 1982) in the well-characterized 5S RNA genes of Xenopus. The rapid changes in spacer sequence relative to 5S

RNA-coding regions, and the high degree of homogeneity among individual repeats have raised important questions about the evolution and maintenance of repeat sequences (Fedoroff and Brown, 1978).

In this paper, we report on the nucleotide sequence and genomic organization of the 5S RNA genes of C. elegans. In addition, we have identified a molecular marker linked to these genes which will allow chromosomal localization of the 5S RNA gene cluster and permit the study of the surrounding region.

Materials and Methods

Strains

The two wild-type strains of C. elegans, Bristol N2 and Bergerac B0, used in this study have been described (Hirsh et al, 1979; Emmons et.al., 1979; Rose et.al., 1982). These strains were kindly provided by Dr. D.L. Baillie.

DNA isolations

Nematodes were grown on wild-type E. coli plated on high peptone-agar media (Rose et.al., 1982) and genomic DNA was prepared as described (Emmons et.al., 1979).

Recombinant phages λ 2070-RW2 and λ 2070-RW4 (provided by R. Waterston; Bolten et.al., 1984) were grown on E. coli C-600 in liquid culture (Maniatis et.al., 1982) and purified on CsCl

gradients. Phage DNA was prepared by formamide extraction (Davis et.al., 1980).

Recombinant plasmids were purified from cleared lysates (Maniatis et.al., 1982) on CsCl-EtBr gradients (Davis et.al., 1980).

Restriction fragments were isolated by fractionating restriction digests on 1% agarose (Sigma type I) and recovering the desired fragment by electroelution (Maniatis et.al., 1982).

Plasmid subcloning

Gel purified restriction fragments were ligated into appropriately cleaved pBR322 (Bolivar et.al., 1977) or pUC13 (Viera and Messing, 1982) with T4 DNA ligase at a fragment to vector free-end ratio of 5:1. Ligation reactions were used to transform CaCl₂-treated (Maniatis et.al., 1982) *E. coli* RR-1 or JM-83 to Ap^rTc^s, or Ap^rLac⁻ phenotypes respectively. The identity and orientation of the inserted fragments were determined by restriction mapping.

Restriction mapping and DNA sequencing

Restriction maps were constructed by comparison of single- and double-digest fragment patterns. Fragments were sized using Eco RI plus Hind III digested λ DNA and Hpa II digested pBR322 as markers.

Restriction fragments for sequencing were 3' end-labelled using the fill-in reaction of DNA pol I (Klenow fragment)

(Maniatis et.al., 1982). Uniquely labelled fragments were gel-purified and sequenced by the chemical cleavage method of Maxam and Gilbert (1980).

Genomic blots and hybridization

C. elegans genomic DNA was digested to completion and fractionated on agarose gels. Genomic digests were transferred to nitrocellulose (Schleicher & Schuell, BA85) (Smith and Summers, 1980), or nylon filters (Dupont/GeneScreen, suppliers, protocol) and baked at 80 °C for 2 hours. Gel-purified restriction fragments were nick-translated (Davis et.al., 1980) to a specific activity of 5×10^7 cpm/ug, denatured and hybridized to genomic blots in 5XSSPE + 0.3% SDS (Davis et.al., 1980) at 65 °C. Filters were extensively washed in 2XSSPE + 0.3% SDS at 55 °C, and autoradiographed on Kodak BB-1 or XAR-5 film with Dupont Cronex Lightning Plus intensifying screens at -70 °C.

Determination of 5S DNA repeat copy number

Bam HI digests of C. elegans genomic DNA (0.5 ug = 5.7×10^6 haploid genomes), recombinant phages λ 2070-RW2 (2.14 ng = 4×10^7 genomes) and λ 2070-RW4 (3.33 ng = 6.1×10^7 genomes), and recombinant plasmid pCe5S1 (0.41 - 4.1 ng = 20 to 200 haploid genome equivalents) were electrophoresed in 1% agarose, transferred to nitrocellulose and hybridized to a nick-translated 5S DNA probe. The relative hybridization intensities were determined by plotting number of cpm's bound against number of

genome equivalents blotted.

RESULTS

Two recombinant λ phages (λ 2070-RW2 and λ 2070-RW4) were obtained from Dr. R. Waterston. These phages, isolated from a library containing partially Sau3A-digested *C. elegans* genomic DNA inserted into λ 1059 (Karn et.al., 1980), show multiple copies of a tandemly repeated 1 kb sequence upon digestion with Bam HI or Hind III (Figure 1.1a). In addition to this repeat, λ 2070-RW4 also carries 4-5 kb of flanking *C. elegans* DNA (Figure 1.1b). The 1 kb Bam HI repeats were subcloned into pBR322 or pUC13 for further analysis.

The 1 kb Bam HI repeat encodes 5S RNA

When the 1 kb Bam HI repeat was used to probe Northern blots of total *C. elegans* RNA fractionated on polyacrylamide gels, a single hybridizing band corresponding to 5S RNA was observed (data not shown). We next localized the 5S RNA coding 120 bp sequence within the repeat by constructing a restriction map of the repeat and comparing it to the sites predicted from the published *C. elegans* 5S RNA sequence (Butler et.al., 1981). As shown in figure 1.1a, the sites characteristic of 5S RNA place the coding sequence centrally within the fragment. Final confirmation was provided by direct sequence analysis of the repeat (Figure 1.2). The underlined coding sequence matches that

Figure 1.1: Structure of C. elegans 5S DNA isolates.

a) Restriction map of the C. elegans 5S DNA repeat.

The 1 kb Bam HI fragment shown is tandemly repeated in both λ 2070-RW2 (10 - 14 copies) and λ 2070-RW4 (6 - 8 copies). The restriction pattern diagnostic of the C. elegans 5S RNA coding sequence is also shown.

Restriction sites:

B = Bam HI; A = Alu I; H = Hind III;

Ha = Hae III; S = Sau IIIA

b) Restriction map of the genomic sequences flanking the 5S DNA tandem repeat carried by λ 2070-RW4.

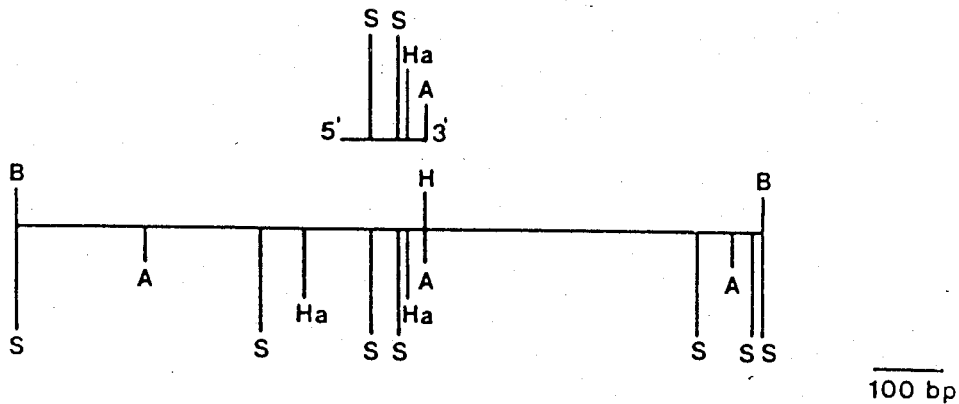
This entire region shows homology to the 1 kb 5S DNA repeat shown in (a). The location of the 5S DNA tandem repeat is shown in brackets.

Restriction sites:

B = Bam HI; H = Hind III; RI = Eco RI;

S = Sal I.

a)



b)

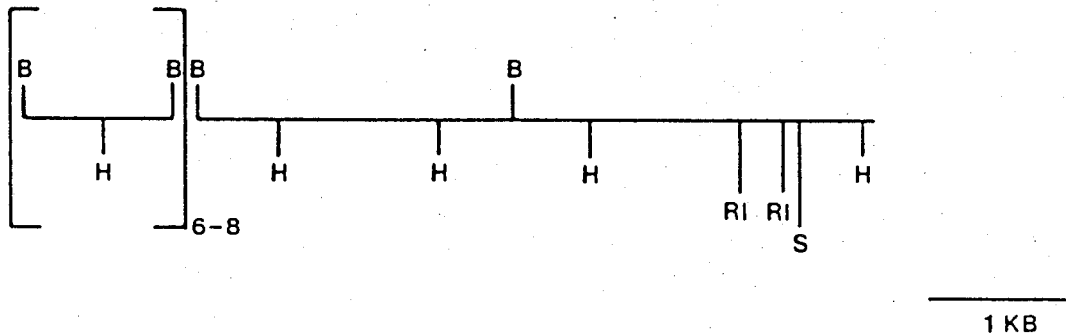


Figure 1.2: Sequence of the C. elegans 5S DNA unit repeat

The entire nucleotide sequence of the 1 kb 5S DNA repeat carried by pCe5S1 is shown. The 5S RNA coding region is indicated by the heavy arrow. The putative RNA polymerase III transcription termination signal (dT₄) immediately flanking the coding sequence is boxed. The sequence shown has Bam HI ends and is numbered relative to the 5S RNA coding region.

Ce5S1

-391 -381 -371 -361 -351 -341 -331
 GGATCC CGTTCCCAA TCAATATCAT CACCTTCCA CACTCCACA TAGGACAGC CATGTCCTT CATCTCCACA
 -291 -271 -261 -251 -241 -231 -221
 TATTTTCGAA TGGACCGTTC TCCAACATGG TGRAACCGATT TTCGAATTC TTAGTATTT TGTCTCCCG AGAGGAGACG TTCCAAAATT TATAGCTAAC GCAAATTTCT
 -211 -201 -191 -181 -171 -161 -151 -141 -131 -121 -111 -101
 TTGGGTCAGT TTCAAATGTTT ACCTCAAACT TGGGTAATTA AACCGACTTG AAGGGCGGTC TCTGGCCCTA ACACCAACCT ACAGCAGACA GCGAGCGAGG TCTACGATCG
 GCAACCCCC TCTCCACACA ACACACAGCC AGCCAGTAGC ACTGACGAAG GGGTAGAGG CTAGTGGGT GGTGGGGGT TTGTAACCTGA CATGGCACTT GTCCATAGAC

GCTTACGACC ATATCACGTT GAATGCACGC CATCCCGTCC GATCTGGCAA GTTAAGCAAC GTTGAGTCCA GTTAGTACTT GGATCGGAGA CGGCCTGGGA ATCCTGGATG
 10 20 30 40 50 60 70 80 90 100 110 120
 TTGTAAGC TTTGGCATTTT TTTAGAAATGA CTAATAAATTT GTTTTGTGTTG CAACATCATA TGAATAATGA AACGATTTTA AAATAATTTT CAAAAAATTT TTCCCCAGCC
 130 140 150 160 170 180 190 200 210 220
 TTCCCGGAGAG TCGTATTTTC GAGTGTAAAT AAGAGACTGC TATTTTTTTTT TGGTGTGTA TGTGTGGAC TTTATTGCGT GTACCCTAATG AGTGAATATT AAAATGACAC
 230 240 250 260 270 280 290 300 310 320 330
 GATGTGTCGG GTCCGATTTT GAAACATGAA AAATTTAAGT TCTTTTGAAA ATTTTGTGGA ACAATGTCAT TTTTCAGTTG AGAAGCGCAC GACACCCATA GATTATACTG
 340 350 360 370 380 390 400 410 420 430 440
 TATAATTTT TTGCATCACA TAGATCTATT TCCATATATG TATGTACCAG TAGCAAGRAAT GAAAAGCACT TTTGAAGAGT GTAGATGAGC TGAAGGSCAA CATTAGAAATC
 450 460 470 480 490 500 510 520 530 540 550

560 570
 CATCATATTT AGATCGGGA TCC
 33
 B

of C. elegans 5S RNA exactly (Butler et.al., 1981). This sequence identity and the adjacent 3' run of dT residues, a putative signal for the termination of transcription by RNA pol III, suggests that the 1 kb repeat is a functional member of the 5S RNA gene family of C. elegans.

Genomic organization and repeat homogeneity of C. elegans 5S DNA

In λ 2070-RW2 and λ 2070-RW4, 5S DNA is organized in tandem arrays of the 1 kb Bam HI repeats. To show that these phages are representative of genomic organization, we used the 1 kb repeat to probe various restriction digests of the C. elegans genome. The results shown in Figure 1.3a confirm the head-to-tail tandem organization of C. elegans 5S RNA genes. Quantitation of the repeat on gels, using known amounts of plasmid clones for standards, shows that there are roughly 100 to 110 copies per haploid genome (Figure 1.4), in good agreement with the solution hybridization results of Sulston and Brenner (1974).

Are the 1 kb repeat homogeneous in size and sequence? They appear to migrate as a single-size class in a 1% agarose gel. We have examined this apparent size homogeneity more closely by cutting genomic DNA with several enzymes that cut within the repeat (Figure 1.3b), followed by probing with labelled 1 kb 5S DNA repeat. If a variant size class were present in significant amounts (5-10%), it would be detectable, given the good separation possible if all the repeats were identical to those shown in Figures 1.1a and 1.3b. We conclude that the 1 kb

Figure 1.3: Genomic organization of C. elegans 5S DNA.

a) C. elegans genomic 5S DNA is tandemly organized.

0.2 ug of C. elegans genomic DNA was digested to completion with the indicated enzymes, fractionated on 1% agarose, transferred to nitrocellulose and hybridized to a nick-translated C. elegans 5S DNA repeat probe.

Note that each enzyme either generates a 1 kb 5S DNA repeat (Bam HI and Hind III) or does not cut within the tandem cluster (Eco RI, Hpa I, Kpn I, Pst I, Sal I, and Xba I).

b) Homogeneity of the C. elegans 5S DNA repeat.

0.5 ug of C. elegans genomic DNA was digested to completion with the indicated enzymes, fractionated on 2% agarose, transferred to a nylon filter and hybridized to a nick-translated C. elegans 5S DNA repeat probe.

The sizes of the observed fragments and their locations are shown in the accompanying restriction map of a 5S DNA tandem repeat (see text). Note that the Hpa II fragment shown as being 39 bp long is actually 390 bp in length.

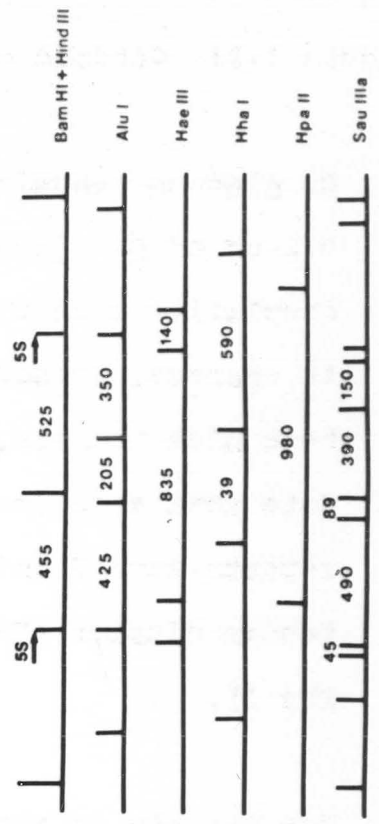
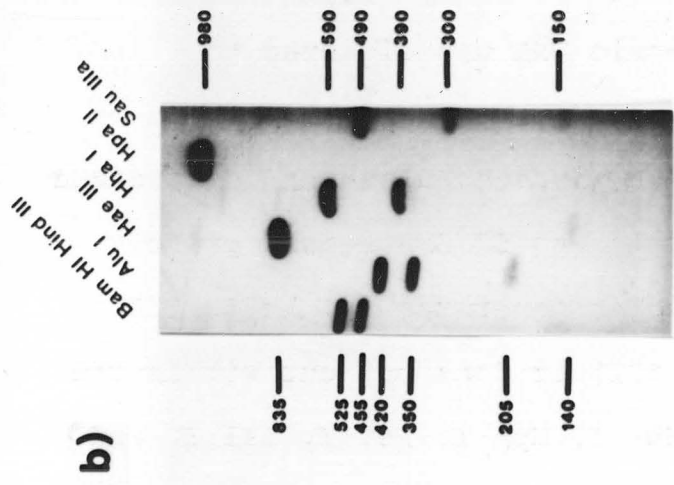
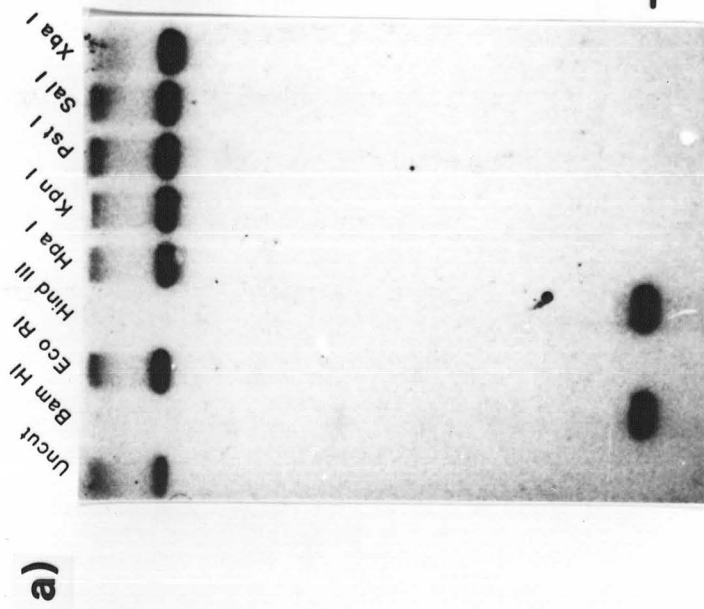
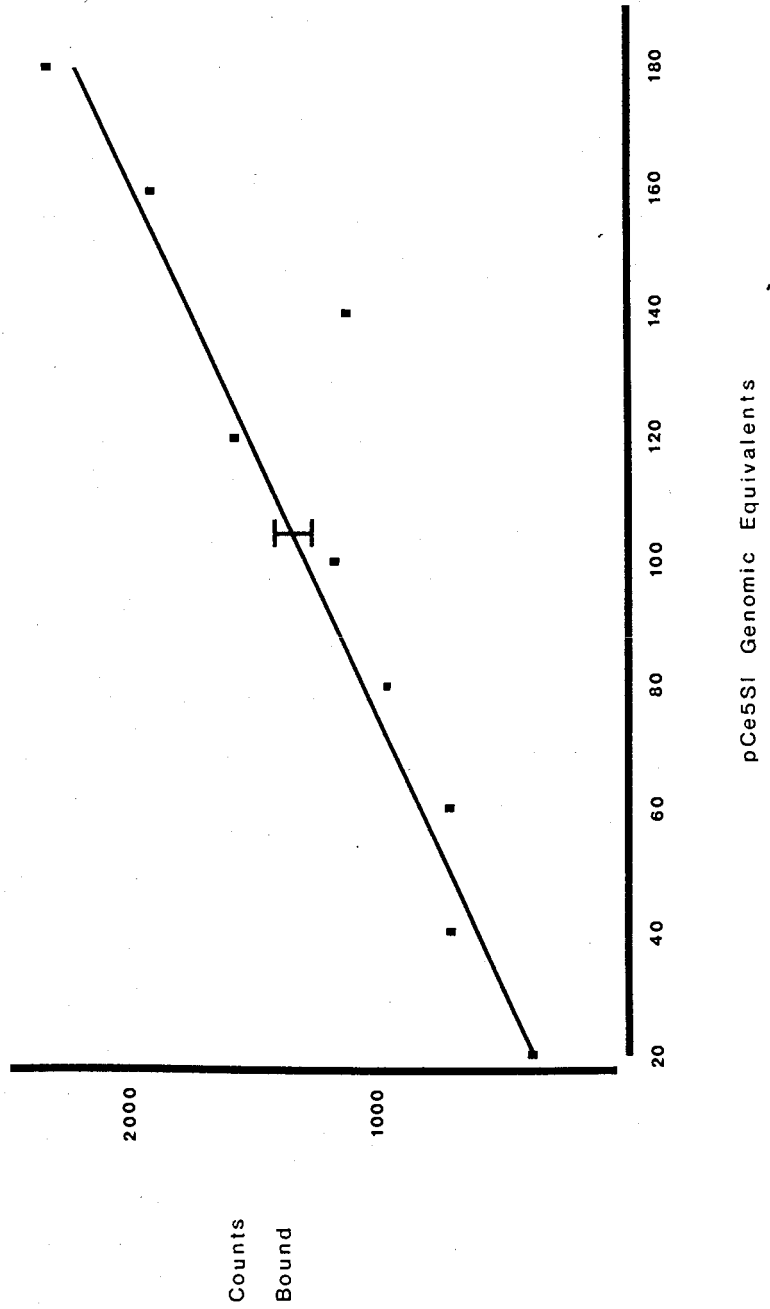


Figure 1.4: Quantitation of the C. elegans haploid 5S DNA content.

C. elegans genomic DNA (0.5 ug) and plasmid standards were digested with Bam HI to release the 5S DNA unit repeat, briefly electrophoresed through 1% agarose, transferred to nitrocellulose and hybridized against a nick-translated 1 kb 5S DNA repeat probe. Bands were visualized by autoradiography, excised and Cerenkov counted. The plasmid (pCe5S1) standard curve is shown as Cerenkov counts bound vs. number of genomic equivalents blotted. The genomic hybridization signal (vertical bar in Figure) represents the average of three measurements for the Bristol (N2S) and Bergerac (BO) strains of C. elegans.



repeats are representative of the major 5S DNA repeat class in the C. elegans genome (although we cannot rule out minor species within the tandem arrays).

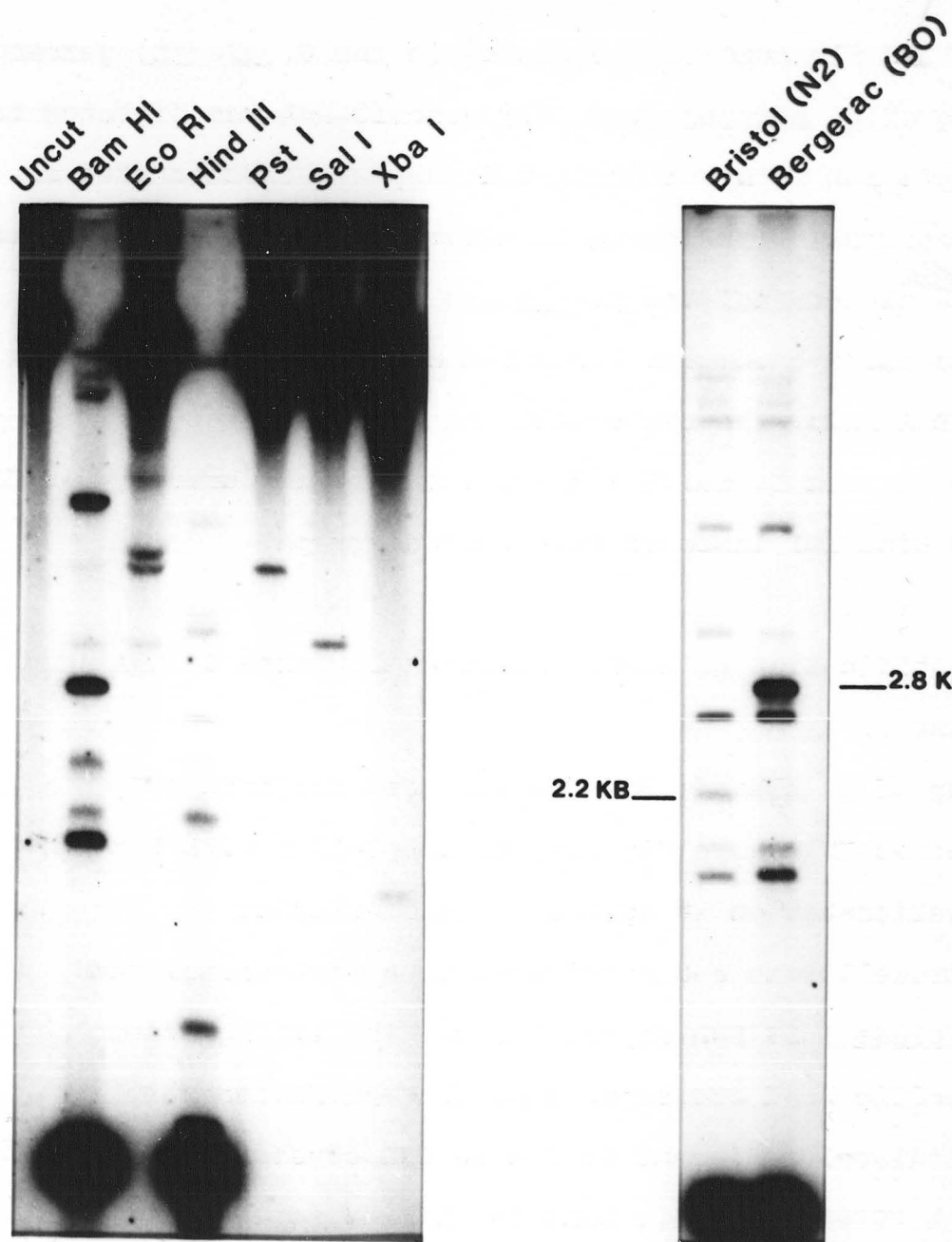
DNA sequence homologous to 5S DNA but not in the major 1 kb repeat class can also be observed in genomic Southern blots (Figure 1.5). These sequences may represent dispersed copies of 5S DNA (Rosenthal and Doering, 1983), or genomic sequences that flank 5S DNA clusters in the chromosome. One such flanking sequence is the 2.2 kb Bam HI fragment carried by λ 2070-RW4 (Figure 1.1b). Restriction fragment length differences (RFLDs) are frequently observed when comparing the genomes of two interbreeding strains of C. elegans (N2 (Bristol) and BO (Bergerac)) (Emmons et.al., 1979). These RFLDs can be used as genetic markers to localize cloned DNA to chromosomal locations. There is an RFLD associated with the 2.2 kb Bam HI end fragment, with a shift to 2.8 kb observed in the BO strain (Figure 1.3b). This RFLD will allow us to place the 5S DNA cluster on the C. elegans genetic map.

DISCUSSION

We have isolated and characterized a 1 kb repeat sequence which contains one copy of the C. elegans 5S RNA coding sequence. These repeats are homogeneous in size and are tandemly organized (approximately 110 copies) in the genome, an arrangement typical of higher eukaryotes. We believe this 1 kb repeat is a functional member of the C. elegans 5S RNA gene family because

Figure 1.5: Other genomic sequences hybridizing to the C. elegans 5S DNA repeat.

- a) Nontandemly repeated sequences in the C. elegans genome. 2 ug of C. elegans (var. N2) genomic DNA was digested to completion with the indicated enzymes, fractionated on 1% agarose, transferred to nitrocellulose and hybridized to a nick-translated C. elegans 5S DNA repeat probe. Note the hybridizing bands not migrating with the major 5S DNA band: the overexposed bands at the top (uncut, Eco RI, Pst I, Sal I and Xba I lanes) and bottom (Bam HI and Hind III lanes) of the autoradiograph.
- b) Identification of a RFLD associated with a 5S DNA cluster. 2 ug of C. elegans Bristol (N2) and Bergerac (BO) genomic DNAs were digested to completion with Bam HI, fractionated on 1% agarose, transferred to nitrocellulose and hybridized to a nick-translated C. elegans 5S DNA repeat probe. The 2.2 kb Bristol specific band corresponds to the Bam HI fragment immediately adjacent to the 5S DNA cluster carried by λ 2070-RW4. This band is shifted to 2.8 kb in the Bergerac genome.



the identified coding sequence exactly matches the published 5S RNA sequence (Butler et.al., 1981). Moreover, this DNA is transcriptionally active in both heterologous (Weil et.al., 1979) and homologous (B.M.H. and D.W.N., unpublished results) cell-free extracts. We have examined the sequence of the 1 kb 5S DNA repeat looking for 5S RNA pseudogenes or internally repetitious sequences. While several short (10-12 bp) direct and inverted repeats were identified, no 5S RNA pseudogenes or internal simple sequence repeats were found. The apparent homogeneity of the C.elegans 5S RNA gene family may be related to the absence of such internal simple sequence repeats.

We have described sequences that are homologous to 5S DNA but which are not in the major 1 kb repeat class. These could represent dispersed 5S RNA genes (either active or pseudogenes). This raises important questions as to the nature of dispersal, expression and evolution of these sequences. In addition to looking at these sequences, we are also looking at the organization of 5S RNA coding genes across the phylum Nematoda, to follow the evolution and maintenance of this multigene family.

Alternatively, some of these 5S RNA-related sequences may represent the ends of 5S DNA clusters. One such sequence, carried by λ 2070-RW4, exhibits an RFLD when the genomes of the two C. elegans strains, N2 and B0, are compared. Preliminary data suggest that this RFLD (and associated 5S DNA cluster) maps to the right arm of chromosome V. This result has provided an important marker for in situ labelling and identification of

individual C. elegans chromosomes (Albertson, 1984). These in situ studies also show that 5S RNA genes are localized in this one region, marked by a Hoechst 33258 dark band on chromosome V. Detailed molecular and genetic analyses are now underway to map these genes more precisely and to provide a secure anchor from which to walk into this potentially interesting, relatively 'gene-poor' region of the chromosome.

REFERENCES

- Albertson, D.G. (1984).
 Localization of the ribosomal genes in Caenorhabditis elegans chromosomes by in situ hybridization.
 EMBO J. 3 1227-1234.
- Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach, M.C., Hynener, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977).
 Construction and characterization of new cloning vehicles. II: A multipurpose cloning system.
 Gene 2 95-113.
- Bolten, S.L., Powell-Able, P., Frischmoff, D.A. and Waterston, R.H. (1984).
 The sup-7 (st5) X gene of Caenorhabditis elegans encodes a Trp-tRNA(UAG) amber suppressor.
 Proc. Natl. Acad. Sci. 81 6784-6788.
- Brown, B.B. (1984).
 The role of stable complexes that repress and activate eukaryotic genes.
 Cell 37 359-367.
- Butler, M.H., Wall, S.M., Luehrsen, K.R., Fox, G.E. and Hecht, R.M. (1981).
 Molecular relationships between closely related strains and species of nematodes.
 J. Mol. Evol. 18 18-23.

- Carroll, D. and Brown, D.D. (1976).
Repeating units of Xenopus laevis oocyte-type 5S DNA are heterogeneous in length.
Cell 7 467-475.
- Davis, R.W., Botstein, D. and Roth, J.R. (1980).
Advanced Bacterial Genetics.
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Emmons, S.W., Klass, M.R. and Hirsh, D. (1979).
Analysis of the constancy of DNA sequences during development and evolution of the nematode Caenorhabditis elegans.
Proc. Natl. Acad. Sci. 76 1333-1337.
- Fedoroff, N.V. and Brown, D.D. (1978).
The nucleotide sequence of oocyte 5S DNA in Xenopus laevis.
I: The AT-rich spacer.
Cell 13 701-716.
- Heintz, N. and Roeder, R.G. (1984).
Transcription of eukaryotic genes in soluble cell-free systems.
in: Genetic Engineering, Principles and Methods, Vol. 4.
edited by J.K. Setlow and A. Hollaender. Plenum Publishing, NY, NY. pp. 57-89.
- Hirsh, D., Emmons, S.W., Files, J.G. and Klass, M.R. (1979).
Stability of the C. elegans genome during development and evolution.
in: Eukaryotic Gene Expression. edited by R. Axel, T. Maniatis and C.F. Fox., Academic Press, NY, NY. pp. 205-218.
- Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980).
A novel bacteriophage lambda cloning vector.
Proc. Natl. Acad. Sci. 77 5172-5176.
- Korn, L.J. (1982).
Transcription of Xenopus 5S ribosomal RNA genes.
Nature 295 101-105.
- Long, E.O. and Dawid, I.B. (1980).
Repeated genes in eukaryotes.
Ann. Rev. Biochem. 49 727-764.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982).
Molecular Cloning. A Laboratory Manual.
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

- Maxam, A.M. and Gilbert, W. (1980).
Sequencing end-labeled DNA with base-specific chemical cleavages.
Methods in Enzymol. 65 499-559.
- Miller, J.R., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V. and Brown, D.D. (1978).
The nucleotide sequence of oocyte 5S DNA in X. laevis
II: The GC-rich region.
Cell 13 717-725.
- Rose, A.M., Baillie, B.L., Candido, E.P.M., Beckenbach, K.A. and Nelson, D. (1982).
The linkage mapping of cloned restriction fragment length differences in Caenorhabditis elegans.
Mol. Gen. Genet. 188 286-291.
- Rosenthal, D.S. and Doering, J.L. (1983).
The genomic organization of dispersed tRNA and 5S RNA genes in Xenopus laevis.
J. Biol. Chem. 258 7402-7410.
- Sakonju, S. and Brown, D.D. (1982).
Contact points between a positive transcription factor and the Xenopus 5S RNA gene.
Cell 31 395-405.
- Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription
I: The 5' border of the region.
Cell 19 13-25.
- Smith, G.E. and Summers, M.D. (1980).
The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzylxymethyl-paper.
Anal. Biochem. 109 123-129.
- Sulston, J.E. and Brenner, S. (1974)
The DNA of Caenorhabditis elegans.
Genetics 77 95-104.
- Vieira, J. and Messing, J. (1982).
The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers.
Gene 19 259-268.
- Weil, P.A., Segall, J., Harris, B., Ng, S.Y. and Roeder, R.G. (1979).
Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates.
J. Biol. Chem. 254 (1979) 6163-6173.

Wood, W.B. (1988).

The nematode Caenorhabditis elegans.

Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Zuckerman, B.M. (Editor) (1980).

Nematodes as Biological Models.

Academic Press, NY, NY.

THE SECOND PART

Genetic mapping of the 5S RNA gene cluster
of the nematode *Caenorhabditis elegans*

Introduction

The genomic sequences encoding the 5S RNAs of higher eukaryotes are typically repetitive and generally consist of a repeat unit containing both coding and spacer sequences. These repeats can be organized in tandem arrays and (or) as dispersed copies in which the flanking sequences are unique. The highly repetitive nature of these genes makes them intractable to classical genetic analysis, but facilitates their localization using in situ hybridization techniques. Such studies have shown that the tandemly repeated 5S RNA genes of X. laevis are dispersed to form clusters at the tips of most chromosomes (Pardue et.al., 1973), while in D. melanogaster they are organized into a single cluster at polytene band 56F on chromosome 2 (Wimber and Steffensen, 1970).

We recently reported on the structure and organization of 5S RNA genes in the nematode C. elegans (Nelson and Honda, 1985). The 5S RNA coding sequence is located in a 1 kb genomic repeat, which is tandemly reiterated approximately 110 times per haploid genome. In situ hybridization data (Albertson, 1984) indicates that most, if not all, of these repeats are located at a single chromosomal site, therefore serving as an important landmark for in situ localization of other genes. However, the small size of the C. elegans metaphase chromosomes precludes the unambiguous assignment of this repetitive gene cluster to a specific linkage group by in situ hybridization.

To locate the 5S RNA gene cluster on the C. elegans genetic

map, we have looked for a restriction fragment length difference (RFLD) associated with this cluster in the genomes of two interbreeding strains of C. elegans. Such RFLDs, once identified, can be used as molecular markers in two- and three-factor interstrain crosses to establish genetic linkage and position relative to known visible phenotypic markers (Hirsh et.al., 1979; Rose et.al., 1982; Files et.al., 1983; Cox et.al., 1985).

In this paper, we describe such an RFLD, demonstrate its linkage to the 5S RNA gene cluster and determine its position on the C. elegans genetic map.

Materials and Methods

Reagents

All restriction enzymes were purchased from Bethesda Research Labs or Pharmacia P-L Biochemicals and were used as described by Maniatis et.al. (1982). α -³²P-dATP and α -³²PdCTP were obtained from Amersham.

Nematodes

Two wild-type strains, C. elegans var. Bristol (N2) and C. elegans var. Bergerac (BO), and the following N2-derived mutant alleles were used in this study: unc-46(e177)V; dpy-11(e224)V; unc-42(e270)V; sma-1(e30)V; unc-76(e911)V; dpy-21(e428) V; and

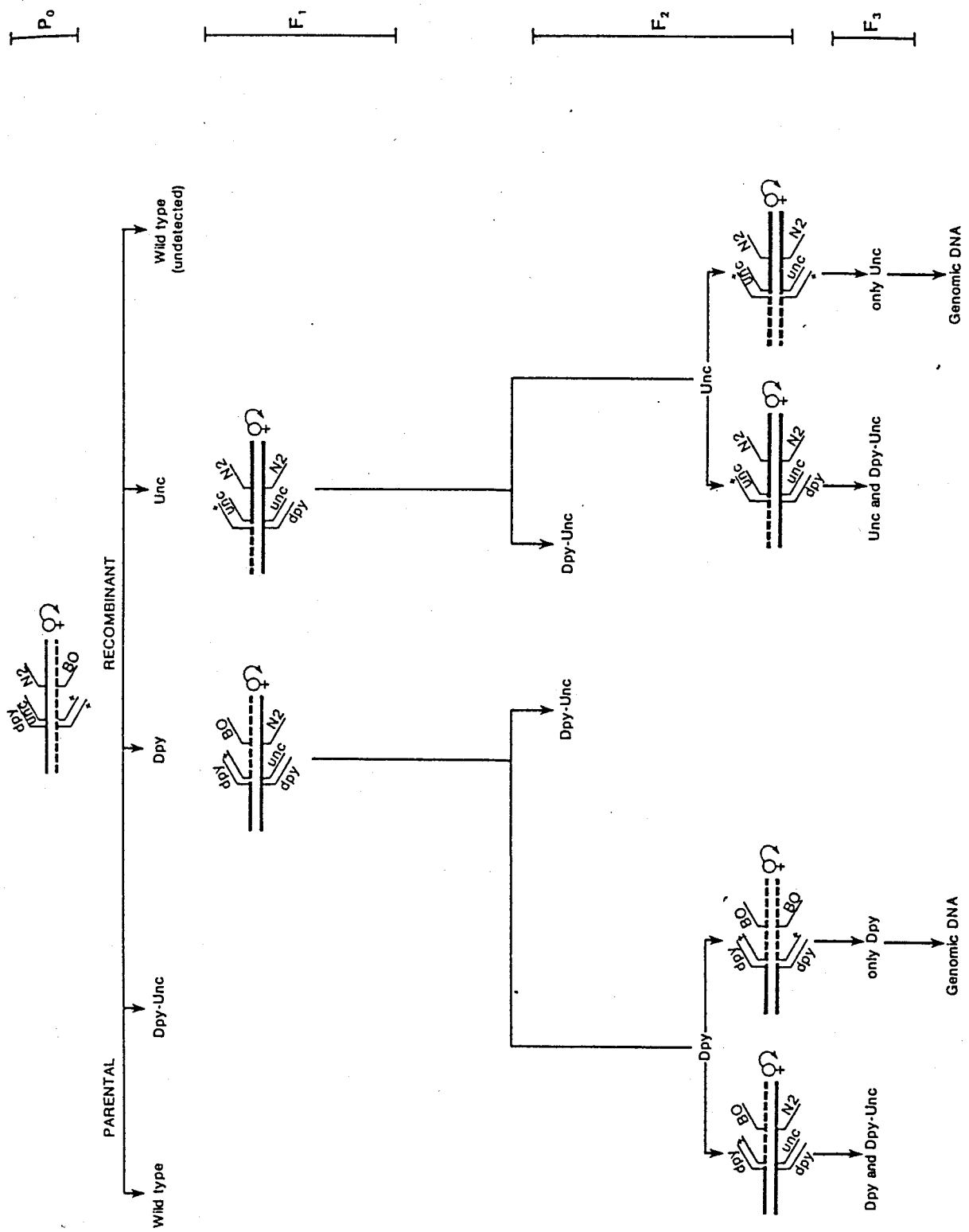
unc-51(e369)V. All nematode strains, kindly provided by D.L. Baillie, were originally obtained from the MRC stock collection in Cambridge, England, or the Caenorhabditis Genetics Center at the University of Missouri in Columbia, Missouri.

Nematodes were maintained and mated on NGM plates streaked with Escherichia coli OP50 as described by Brenner (1974). The rationale behind RFLD mapping and the procedures for interstrain crosses have been described (Hirsh et.al., 1979; Rose et.al., 1982; Files et.al., 1983; Cox et.al., 1985). Figure 2.1 illustrates the genetic protocol used for three-factor RFLD mapping. Briefly, N2 males, heterozygous for two cis-linked recessive markers (one dpy and one unc), were mated with B0 hermaphrodites. Wild-type N2/B0 heterozygous hermaphrodites (P₀ generation in Figure 1) from this cross were isolated and allowed to self-fertilize. From those P₀ heterozygotes segregating 3:1 wild type: Dpy-Unc, all recombinant Dpy or Unc progeny (F₁ generation in Figure 1) were picked and allowed to self-fertilize. In some cases (unc-46, dpy-11; dpy-11, unc-42; unc-42, sma-1) only the Unc recombinant class was used. The self-progeny of these recombinants were scored. Several Unc (or Dpy) progeny (F₂ generation in Figure 1) were isolated from F₁ individuals segregating 3:1 Unc (or Dpy):Dpy-Unc and allowed to self-fertilize. One of these, which segregated only Unc (or Dpy) progeny, served to establish the recombinant strains from which genomic DNA was isolated.

Genomic DNA stocks representing the five homozygous linkage

Figure 2.1: Genetic protocol for three-factor crosses.

A series of crosses were performed as described in the text; segregation of the linkage group carrying the cis-linked recessive markers is shown. The N2-derived chromosome is represented by the solid line and the B0-derived chromosome by the dashed line. The mutant alleles shown here represent dpy-11 and unc-42, but any of the pairs described in the text can be substituted. Each generation is indicated on the right of the figure.



groups in a heterozygous N2/BO background (e.g., N2/N2 LG I; N2/BO Lg II through V) were generously provided by D.L. Baillie (Rose et.al., 1982). Recombinant individuals isolated as described above were grown by clonal expansion of the recombinant on high peptone agar plates streaked with wild-type E. coli. Worms were collected from the plates, extensively washed with 0.04 M NaCl, dissolved by proteinase K (Sigma) digestion, and genomic DNA prepared by the procedure of Emmons et.al. (1979). For large scale isolations, genomic DNA was further purified by banding in CsCl-ethidium bromide equilibrium gradients (Rose et.al., 1982).

Southern transfers and hybridizations

Two micrograms of C. elegans genomic DNA was digested to completion with the indicated enzymes, size fractionated on 1% agarose (Sigma type I) gels, and transferred to nitrocellulose filters (Schleicher & Schuell, BA85) using the method of Smith and Summers (1980). Filters were baked at 80 °C for 2 hours before hybridization in 5XSSPE, 0.3% SDS (SSPE is 0.18 M NaCl, 10 mM Na phosphate at pH 7, 1 mM EDTA) using gel purified restriction fragments nick-translated (Rigby et.al., 1977) to a specific activity of approximately 10^7 cpm/ug. Filters were hybridized for 18 hours at 62 °C, then extensively washed at 60 °C in 1XSSPE, 0.3% SDS, and exposed on Kodak BB-1 film.

Linkage of the 5S RNA gene cluster with the RFLD

Fifty micrograms of Bristol and Bergerac genomic DNA was digested sequentially with Eco RI, Sal I, Hpa I, Kpn I and Pst I, and fractionated on a 0.5% agarose gel. High molecular weight genomic DNA was recovered by electroelution (Maniatis et.al., 1982). This genomic fraction was digested with Bam HI, Hind III, or Xba I, fractionated on 1% agarose, transferred to nitrocellulose and hybridized against nick-translated probes to detect the RFLD fragments described in the text.

Results and Discussion

Identification and description of the RFLD

Our previous studies have shown that the C. elegans haploid genome contains approximately 110 copies of a tandemly organized, homogeneous 1 kb repeat family that encodes 5S RNA (Nelson and Honda, 1985). C. elegans genomic blots probed with the 1 kb 5S DNA repeat reveal the presence of a few additional hybridizing fragments that do not belong to the major repeat class. These genomic fragments may represent aberrant copies of the clustered repeat, dispersed repeat copies, or the ends of the genomic repeat cluster(s).

RFLDs are frequently observed when specific sequences are used to probe genomic blots of the Bristol (N2) and Bergerac (B0) wild-type strains of C. elegans (Emmons et.al., 1979; Hirsh

Figure 2.2: Identification of an RFLD associated with the C. elegans 5S RNA gene cluster.

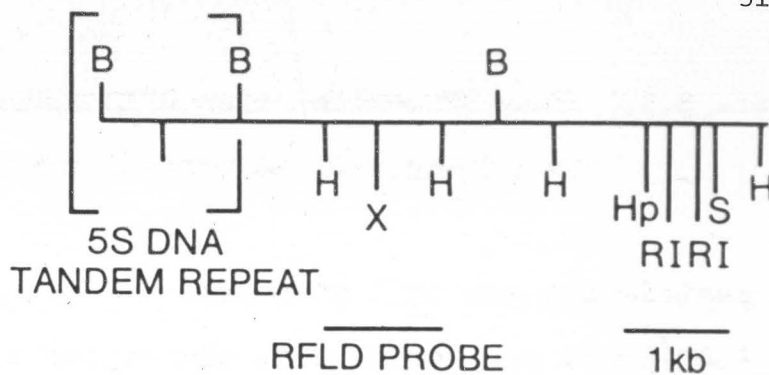
a) Restriction map of λ 2070-RW4.

λ 2070-RW4 carries four to six copies of the C. elegans 1 kb 5S DNA tandem repeat (only one copy is shown) and 5 kb of N2 genomic sequence flanking this tandem repeat cluster. The origin of the 1.2 kb Hind III fragment used as the RFLD probe is indicated. The RFLD itself is located to the left of the Xba I site.

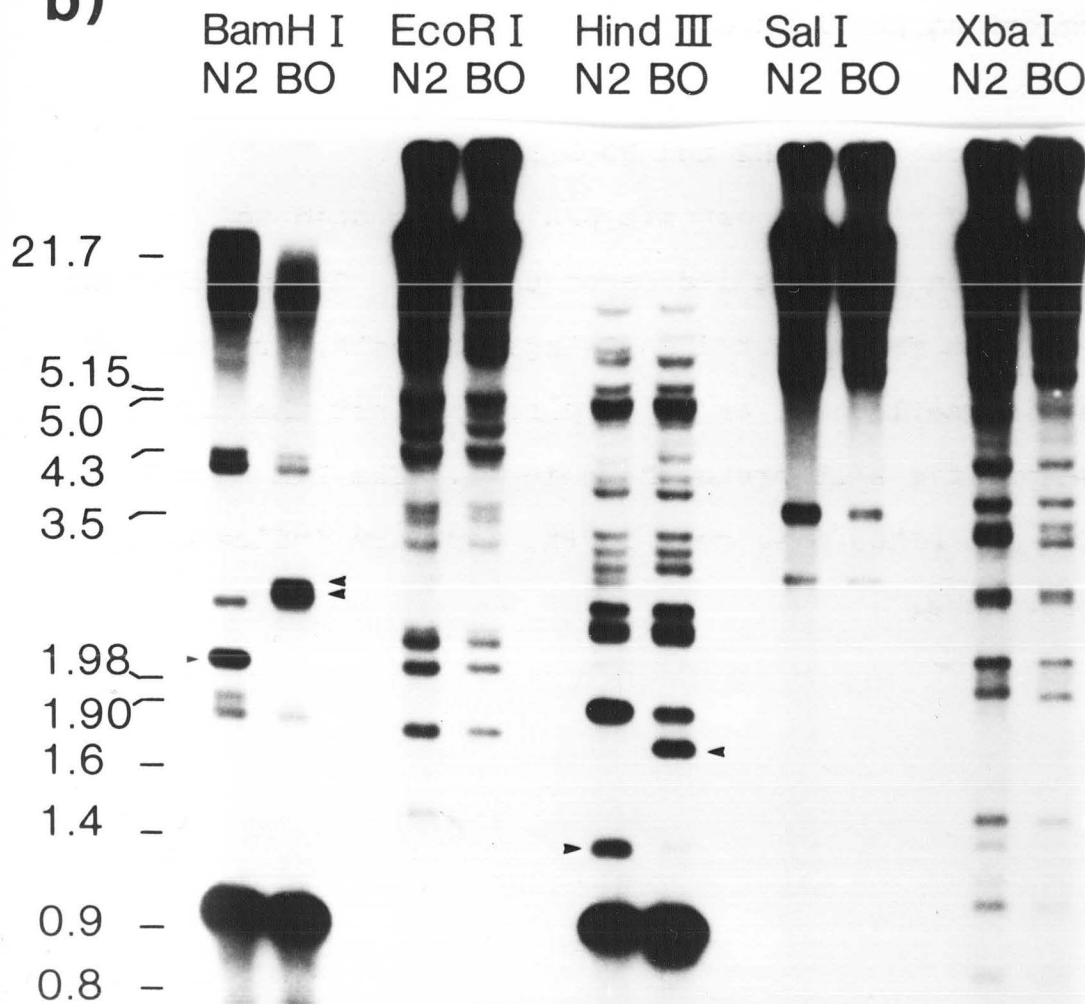
b) Identification of N2 and BO RFLDs.

Two μ g of N2 or BO genomic DNA was digested to completion with the indicated enzymes. Genomic digests were size fractionated on 1% agarose gels, transferred to nitrocellulose, and hybridized against the 1.2 kb N2-specific RFLD probe shown in 2a. The Bam HI and Hind III RFLDs discussed in the text are indicated by arrowheads.

a)



b)



et.al., 1979; Rose et.al., 1982; Files et.al., 1983; Cox et.al., 1985). In an attempt to identify an RFLD that would allow us to place the 5S RNA gene cluster on the C. elegans genetic map, we compared the hybridization patterns generated by probing various restriction digests of N2 and BO genomic DNAs with the 1 kb 5S DNA repeat. The results (data not shown) identified N2/BO RFLDs in Bam HI (a 2.2 kb N2-specific fragment is shifted to 2.8 kb in the BO genome) and Hind III (a 1.2 kb N2-specific fragment is shifted to 1.7 kb in the BO genome) digests.

The Bam HI N2/BO RFLD is linked to the 5S RNA gene cluster

Most of the RFLDs used for genetic mapping in C. elegans are located in unique genomic sequence. The RFLDs we have identified on the other hand are associated with highly repeated genomic sequences. It is therefore essential to demonstrate that the observed RFLD fragments are linked to the 5S RNA gene cluster before using the RFLD as a molecular marker in genetic crosses.

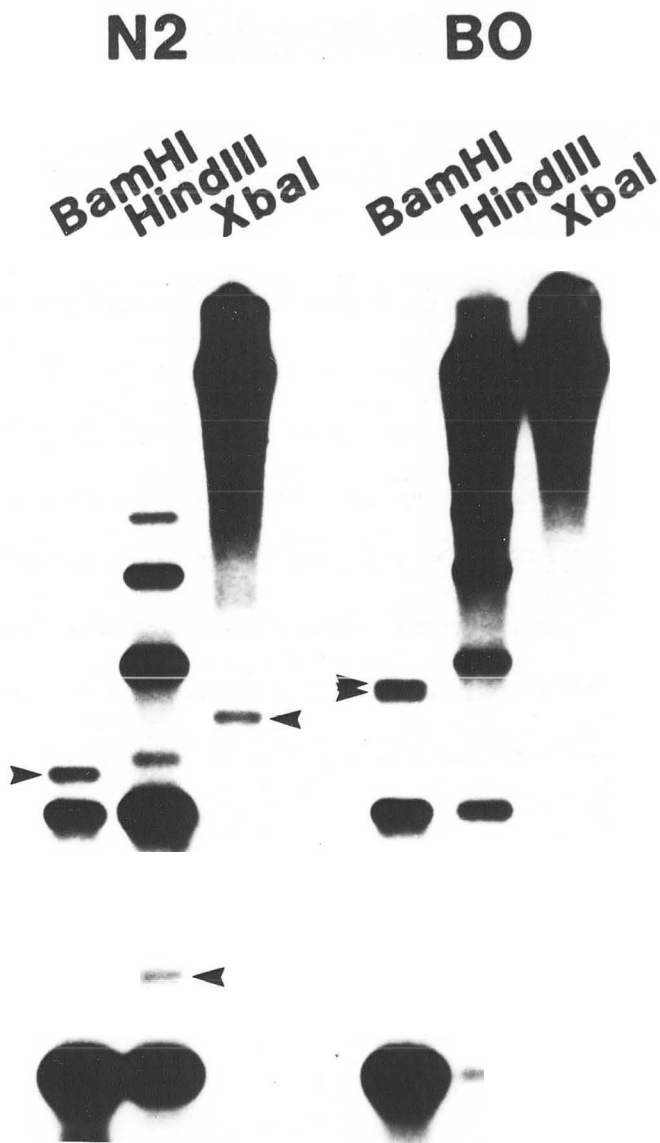
We have previously described two recombinant phages, λ 2070-RW2 and λ 2070-RW4, that carry multiple tandem copies of the N2 genomic 5S DNA repeat (Nelson and Honda, 1985). One of these phages, λ 2070-RW4, carries 4 to 5 kb of N2 genomic sequence in addition to the 5S DNA tandem repeat. A restriction map of the N2 genomic insert carried by λ 2070-RW4 is shown in Figure 2.2a. The 2.2 kb Bam HI fragment immediately adjacent to the 5S DNA tandem repeat and its internal 1.2 Hind III fragment correspond to the N2-specific RFLD fragments discussed above.

The N2- and BO-specific genomic fragments are not simple derivatives of the 5S DNA repeat unit generated by loss of a restriction site or deletion of some repeat sequences; nor are they due to the insertion of the 1.6 kb transposable element Tc1 (Emmons et.al., 1983). Because these fragments may be only partially homologous to the 5S DNA repeat, we used the 1.2 kb Hind III fragment flanking the 5S DNA tandem repeat (see Figure 2.2a) to probe the N2 and BO genomic digests described above. The results (see Figure 2.2b) show that this RFLD probe hybridizes to the 5S DNA repeat released by Bam HI or Hind III digestion as well as another sequence which is present in approximately 20 dispersed copies in the C. elegans genome (see Figure 2.2b, Hind III and Xba I lanes). The RFLD probe reveals several N2/BO RFLDs in addition to those detected using the 5S DNA repeat probe. In particular, note that the 2.2 kb Bam HI and 1.2 kb Hind III fragments carried by λ 2070-RW4 are N2-specific. They apparently shift to a 2.7 plus 2.8 kb Bam HI doublet and to a 1.8 kb Hind III fragment in the BO genome. These results suggest that the genomic sequences immediately flanking the 5S RNA gene cluster differ in the N2 and BO strains of C. elegans.

Although the sequence organization of the genomic insert carried by λ 2070-RW4 suggests that the N2-specific RFLD fragments are tightly linked to the genomic 5S RNA gene cluster, this organization may be an artifact of library construction. We have taken advantage of the tandem organization of C. elegans 5S DNA to confirm the genomic linkage. N2 and BO genomic DNAs were

Figure 2.3: Linkage of the RFLD and the 5S RNA gene cluster.

Fifty ug of N2 and BO genomic DNAs were digested to completion with Eco RI, Kpn I, Hpa I, Pst I and Sal I. The genomic digests were size fractionated on 0.5% agarose, and high molecular weight DNA recovered by electroelution. The recovered DNA was digested with either Bam HI, Hind III or Xba I, fractionated on 1% agarose, transferred to nitrocellulose, and hybridized against the RFLD probe. Arrowheads indicate the N2- and BO-specific RFLD fragments. Note that the N2 and BO Hind III digests are partials and that the BO-specific 1.8 kb Hind III fragment cannot be detected. The 2.8 kb band indicated in the Xba I lane corresponds to the Xba I-Hpa I fragment to the right of the 5S DNA tandem repeat carried by λ 2070-RW4.



digested to completion with five enzymes (Eco RI, Sal I, Hpa I, Kpn I and Pst I) known to lack recognition sites within the 5S DNA repeat. High molecular weight genomic DNA containing the 5S DNA repeat cluster was recovered following size fractionation on 0.5% agarose and redigested with Bam HI, Hind III or Xba I. These digests were then fractionated on a 1% agarose gel, transferred to nitrocellulose, and hybridized to the RFLD probe. The results in Figure 2.3 show that the N2- and BO-specific RFLD fragments remain linked to the high molecular weight fraction, after the null digests. Subsequent digestion of this high molecular weight fraction with Bam HI, Hind III, or Xba I releases the flanking fragments predicted from the restriction map of the genomic sequences carried by λ 2070-RW4. While these results do not unambiguously demonstrate that the RFLD fragments are tightly linked to the C. elegans 5S DNA cluster, it is very unlikely that they would reside in a separate genomic region which, like the 5S DNA cluster, is devoid of the sequences recognized by all of these enzymes.

Genetic mapping of the 5S RNA gene cluster

The N2/BO RFLD we have identified appears to be tightly linked to the C. elegans 5S RNA gene cluster. This RFLD can therefore be used to position the cluster relative to known genetic markers by following the segregation of the RFLD through an appropriate set of genetic crosses. The first step in mapping this cluster is to determine which linkage group (LG) the RFLD

segregates with.

A set of genomic DNA stocks has been described (Rose *et.al.*, 1982) that represents nematode populations that are homozygous (N2/N2 or BO/BO) for a single autosome and randomly heterozygous (N2/BO) for the rest. Bam HI digests of these DNA stocks were hybridized against the RFLD probe to determine their RFLD genotype. Genomic DNA stocks homozygous for N2 LG I, II, and IV all contain roughly equal amounts of the N2- and BO-specific RFLD fragments (data not shown), indicating that the RFLD segregates independently of these autosomal markers. DNA stocks homozygous for N2 LG V or BO LG V on the other hand, contain primarily the N2-specific or BO-specific RFLD fragments, respectively (Figure 2.4b). In addition to the strong homologous RFLD bands, weak heterologous RFLD bands (BO-specific in N2 LG V DNA and N2-specific in BO LG V DNA) are also present in these DNA stocks (see Figure 2.4b). These results indicate that the 5S RNA gene cluster is linked to dpy-11 V, but that a significant amount of recombination has occurred in the dpy-11, RFLD interval during the construction of the nematode population from which these DNAs were isolated.

To locate the 5S RNA gene cluster more precisely, we undertook a series of three-factor crosses using the protocol outlined in Figure 2.1. The recombinant individuals isolated from such a cross contain a homozygous, mosaic LG V that is made up of N2 sequences on one side and BO sequences on the other side of a selected crossover point. By determining which RFLD

fragments segregate with which allele, we can determine the RFLD position relative to these genetic markers. Pairs of markers were chosen to cover most of LG V. The results presented in Figure 2.4 and summarized in Table 2.1 show that the RFLD and the linked 5S RNA gene cluster maps between unc-76 and dpy-21 on the right arm of LG V. Specific details of the results shown in Figure 2.4 are discussed below.

The first two pairs of genetic markers, unc-46, dpy-11 and dpy-11, unc-42, were chosen to orient the RFLD with respect to the dpy-11 allele originally used to mark LG V. The presence of the BO-specific RFLD in all unc-46, +(BO) recombinants (Figure 2.4c), and of the N2-specific RFLD in all +(BO), unc-42 recombinants (Figure 2.4d), shows that the RFLD is located to the right of dpy-11. Note that the N2-specific RFLD fragment is also present in the unc-46, +(BO) recombinants. The presence of this fragment can be explained in terms of the protocol used to isolate the recombinant individuals (Figure 2.1). The recombinant chromosome, in this case unc-46, +(BO), RFLD(BO), is recovered over the doubly marked N2 LG V homolog, unc-46, dpy-11, RFLD(N2). A second recombination event can occur in this heterozygote. If this event occurs between the dpy-11 allele and the RFLD site, a doubly recombinant LG V results with the genotype unc-46, +(BO), RFLD(N2). This double recombinant LG V is isolated over the unc-46, +(BO), RFLD(N2) parental LG V in the next step of the protocol. The genotype of this recombinant individual will therefore be heterozygous with respect to the

Figure 2.4: Genetic mapping of the 5S RNA gene cluster.

Two ug of genomic DNA from each of the recombinant strains isolated as outlined in figure 1 was digested with Bam HI, size fractionated on 1% agarose, transferred to nitrocellulose, and hybridized against the RFLD probe. The position of the N2- and B0-specific RFLD fragments is indicated by the lines between the figures. The putative composition of each recombinant LG V is shown above the appropriate set of digests.

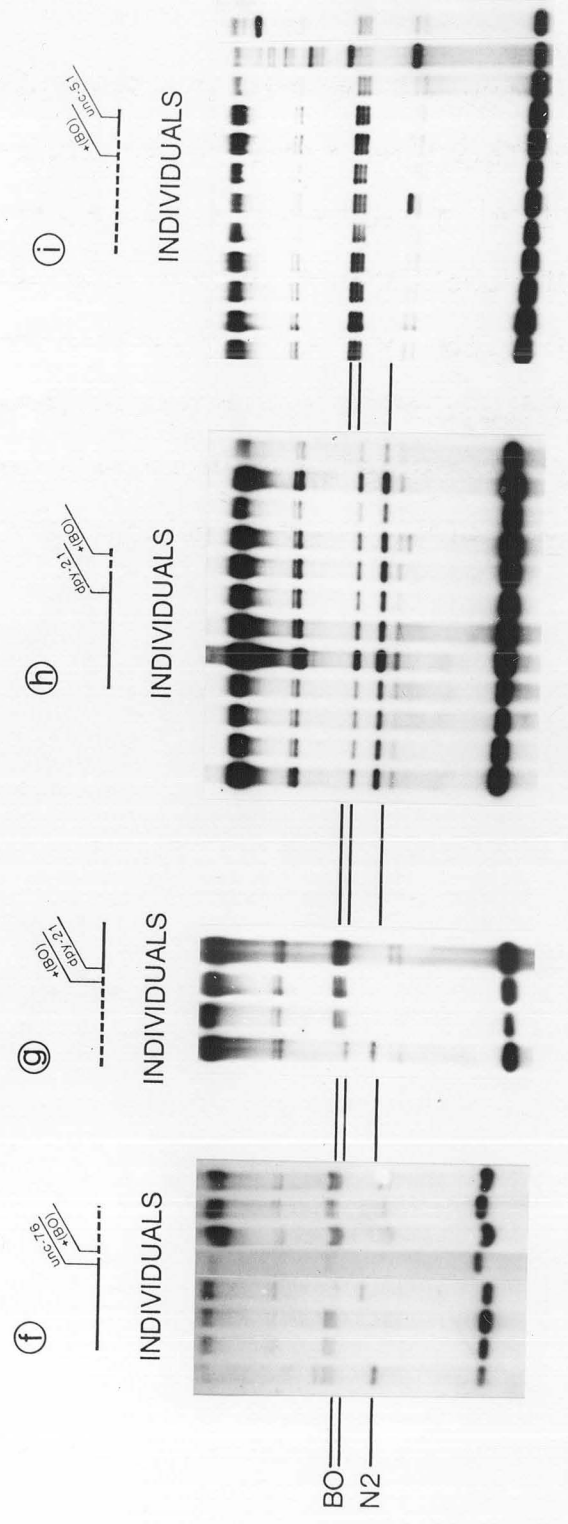
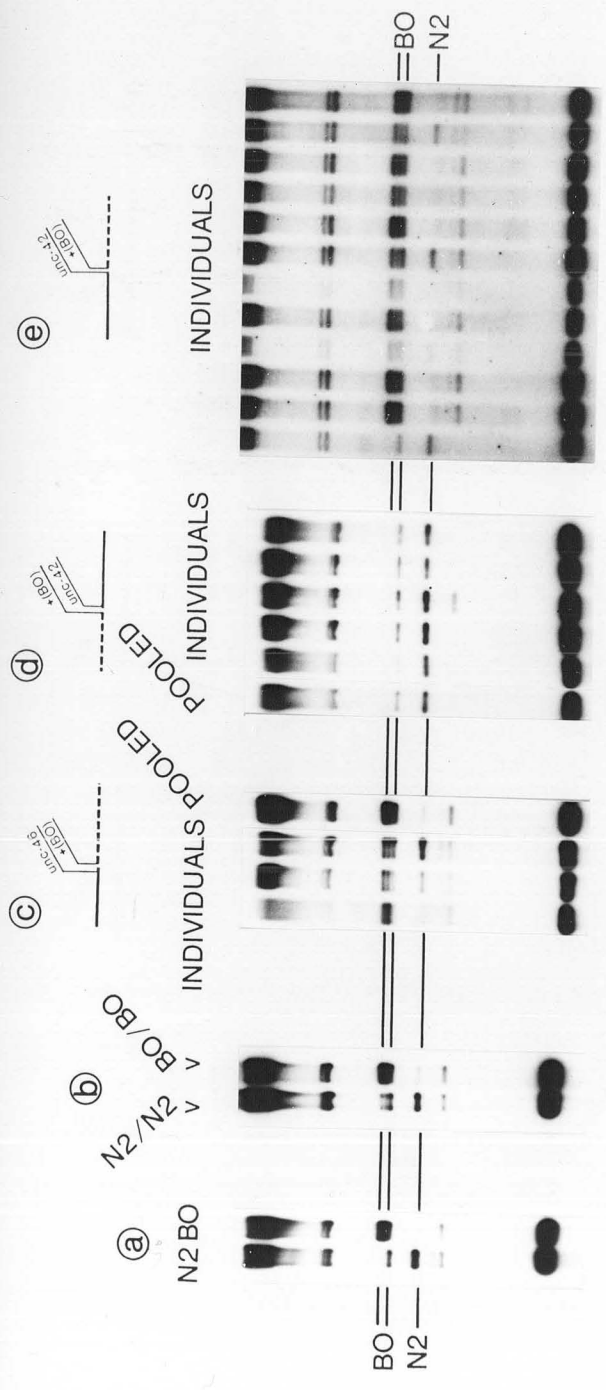
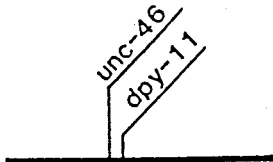
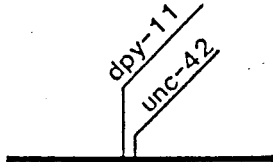
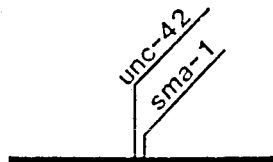
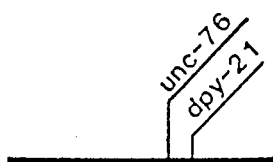
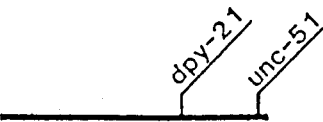


Table 2.1: Summary of genetic mapping data.

LG V MARKERS — 10 map units	RECOMBINANT PHENOTYPE	DATA FROM FIG. 4	RFLD GENOTYPE		
			N2/N2	N2/BO	BO/BO
	Unc +(BO)	c		2	1
	+(BO) Unc	d	5		
	Unc +(BO)	e	1	3	8
	Unc +(BO)	f	1		7
	+(BO) Dpy	g	1		3
	Dpy +(BO)	h	12		
	+(BO) Unc	i			12

RFLD, as observed in two of the three individual recombinants shown in Figure 2.4c. That this event is not as common as suggested by the frequency of individuals recombinants displaying it, is indicated by the weak N2-specific RFLD band in the genomic DNA from the pooled recombinants. The absence of the BO-specific RFLD fragments in +(BO), unc-42 recombinants supports this argument. The +(BO), unc-42, RFLD(N2) recombinant chromosome is recovered over its N2 dpy-11, unc-42, RFLD(N2) homolog and any subsequent recombination event would not affect the recombinant RFLD genotype.

These results place the RFLD to the right of dpy-11, but are ambiguous with respect to its position relative to unc-42. The unc-42, sma-1 cross resolves this ambiguity, placing the RFLD some distance to the right of sma-1. This conclusion is based on the following observations (see Figure 2.4e). Eleven of 12 unc-42, +(BO) individual recombinants exhibit the BO-specific RFLD, placing the RFLD to the right of unc-42. Three of these 11 are heterozygous for the RFLD. Because the unc-42, d+(BO), RFLD(BO) recombinant chromosome is recovered over the N2 unc-42, sma-1, RFLD(N2) homolog, the secondary recombination event described above may occur, generating the RFLD heterozygotes observed. The frequency with which this secondary recombination event occurs (3/11) suggests that the RFLD must lie a considerable distance to the right of sma-1. The final unc-42, +(BO) recombinant appears to be homozygous for the N2 RFLD. This recombinant may have been produced by the fusion of two gametes each carrying a doubly

recombinant LG V generated as described above. Alternatively, a double crossover event may have occurred in the original N2/BO heterozygote to produce a recombinant chromosome with the unc-42, +(BO), RFLD(N2) genotype observed.

To complete the genetic mapping of the 5S RNA gene cluster, we followed the RFLD segregation through recombination events occurring in the intervals unc-76, dpy-21, and dpy-21, unc-51. The results shown in Figure 2.4f and 2.4g unambiguously show that the RFLD maps between unc-76 and dpy-21. Of seven unc-76, +(BO), recombinant individuals examined, six show the BO-specific RFLD and one the N2-specific RFLD. Similarly, of four +(BO), dpy-21 recombinants, three show the BO-specific RFLD and one the N2-specific RFLD (see Figures 2.4f and 2.4g). Although the number of recombinant individuals examined is small, these results serve to place the RFLD between these two markers. The RFLD genotypes of the dpy-21, +(BO) and +(BO), unc-51 recombinants (Figures 2.4h and 2.4i) confirm the above results. All of the dpy-21, +(BO) recombinants show only the N2-specific RFLD, while all of the +(BO), unc-51 recombinants show only the BO-specific RFLD. No secondary recombination events generating RFLD heterozygotes were observed in these crosses.

The tandem organization and repetitive nature of 5S RNA genes encourages their characterization using the technique of in situ hybridization. The in situ hybridization results of Albertson (1984) identified a single site of 5S DNA hybridization to a Hoescht 33258 dark band one-third of the way down an

otherwise featureless C. elegans metaphase chromosome. The RFLD mapping results presented here identify this chromosome as LG V and genetically place the 5S RNA gene cluster on the right arm between unc-76 and dpy-21. This repetitive gene cluster can therefore be used as a cytological LG V landmark with which to localize other cloned sequences using in situ hybridization techniques.

The genetic analysis of eukaryotic 5S RNA genes has only been feasible in D. melanogaster. The localization of these genes to a single cluster at band 56F on chromosome 2 by in situ hybridization (Wimber and Steffensen, 1970) permitted subsequent genetic analysis using cytologically distinct rearrangements of this region (Procunier and Tartoff, 1975; Procunier and Dunn, 1978). Our results, and the in situ hybridization results of Albertson (1984), show that C. elegans 5S RNA genes are localized to a single cluster on the right arm of LG V. This localization will facilitate genetic analysis of this repetitive gene family and other unidentified loci in this region of the C. elegans genome.

REFERENCES

- Albertson, D.G. (1984).
Localization of the ribosomal genes in Caenorhabditis elegans chromosomes by in situ hybridization using biotin-labeled probes.
EMBO J. 3 1227-1234.
- Brenner, S. (1974).
The genetics of Caenorhabditis elegans.
Genetics 77 71-94.
- Cox, G.N., Carr, S., Kramer, J.M. and Hirsh, D. (1985).
Genetic mapping of Caenorhabditis elegans collagen genes using DNA polymorphisms as phenotypic markers.
Genetics 109 513-528.
- Emmons, S.W., Klass, M.R. and Hirsh, D. (1979).
Analysis of the constancy of DNA sequences during development and evolution of the nematode Caenorhabditis elegans.
Proc. Natl. Acad. Sci. 76 1333-1337.
- Emmons, S.W., Yesner, L., Ruan, K. and Katzenberg, D. (1983).
Evidence for a transposon in C. elegans.
Cell 32 55-66.
- Files, J.G., Carr, S. and Hirsh, D. (1983)
Actin gene family of Caenorhabditis elegans.
J. Mol. Biol. 164 355-375.
- Hirsh, D., Emmons, S.W., Files, J.G. and Klass, M.R. (1979)
Stability of the C. elegans genome during development and evolution.
in: Eukaryotic Gene Expression. Edited by R. Axel, T. Maniatis and C.F. Fox. Academic press, NY. p. 205-218.
- Maniatis, T., Fritsch, E.R. and Sambrook, J. (1982).
Molecular cloning, a laboratory manual.
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Nelson, D.W. and Honda, B.M. (1985).
Genes encoding 5S ribosomal RNA of the nematode Caenorhabditis elegans.
Gene 38 245-251.
- Pardue, M.L., Brown, D.D. and Birnstiel, M.L. (1973).
Location of the genes for 5S ribosomal RNA in Xenopus laevis.
Chromosoma 42 191-203.

- Procunier, J.D. and Dunn, R.J. (1978).
Genetic and molecular organization of the 5S locus and mutants in D. melanogaster.
Cell 15 1087-1093.
- Procunier, J.D. and Tartoff, K.D. (1975).
Genetic analysis of the 5S RNA genes in Drosophila melanogaster.
Genetics 81 515-523.
- Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977).
Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase.
J. Mol. Biol. 113 237-251.
- Rose, A.M., Baillie, D.L., Candido, E.P.M., Beckenbach, K.A. and Nelson, D.W. (1982).
The linkage mapping of cloned restriction fragment length differences in Caenorhabditis elegans.
Mol. Gen. Genet. 188 286-291.
- Smith, G.E. and Summers, M.D. (1980).
The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper.
Anal. Biochem. 109 123-129.
- Wimber, D.E. and Steffensen, D.M. (1970).
Localization of 5S RNA genes on Drosophila melanogaster chromosomes by RNA-DNA hybridization.
Science 170 639-641.

THE THIRD PART

COMPARISONS OF C. briggsae AND C. elegans 5S DNA REPEAT .
SEQUENCES IDENTIFY TWO HIGHLY CONSERVED TRANSCRIBED REGIONS

Introduction

The genomic sequences encoding 5S RNA in eukaryotes are typically highly reiterated, forming multigene families which are organized as tandem clusters of homogeneous repeat units (Long and Dawid, 1980). The processes of unequal crossing-over and gene conversion are thought to maintain the structural homogeneity of the 5S DNA repeat unit (Arnheim, 1983); both 5S RNA coding and non-transcribed spacer sequences are highly conserved within a repeat family. In contrast, comparison of different 5S DNA repeat families shows that only the 5S RNA coding sequences are highly conserved; spacer sequences appear to diverge freely (Fedoroff, 1979; Samson and Wegnez, 1984).

We have previously described the structure and genomic organization of the 5S RNA gene family of C. elegans (Nelson and Honda, 1985). The C. elegans haploid genome contains approximately 110 copies of a single homogeneous repeat family encoding 5S RNA, organized as a tandem cluster on the right arm of LG V (Nelson and Honda, 1986). We would like to identify and characterize the functionally important regions of the C. elegans 5S DNA repeat.

A general approach to this problem is based on the assumption that functionally important sequences are maintained under selective pressure and will therefore be conserved between closely related species. C. elegans and C. briggsae are morphologically indistinguishable, but genetically separate nematode species. Previous studies show that the genomes of

these species have diverged extensively, sharing approximately 10% genomic sequence identity (Hirsh et.al., 1979; Emmons et.al., 1979; Butler et.al., 1981). Recent work suggests that this sequence identity is confined to those genomic sequences which appear in the cellular RNA pool (T.P. Snutch, S. Prasad, M. Heschl and D.L. Baillie, personal communication).

In this paper we report on the structure and organization of the 5S RNA genes of C. briggsae. We show that C. briggsae 5S RNA is encoded by two related repeat families which are organized in mutually exclusive tandem clusters. Comparison of C. elegans and C. briggsae 5S DNA repeat sequences reveals two blocks of highly conserved sequence, one corresponding to the 5S RNA coding region and the other encoding a short primary transcript which has been implicated in a trans-splicing process in C. elegans (Krause and Hirsh, 1987).

MATERIALS AND METHODS

Reagents

Restriction enzymes, S1 nuclease and large fragment DNA polymerase (Klenow fragment) were obtained from Pharmacia and used as described (Maniatis et.al., 1982). Exonuclease III was obtained from New England Biolabs.

Nucleic acid isolations

C. briggsae hermaphrodites were grown on hi-peptone agar plates streaked with wild-type E. coli as described previously (Brenner, 1974). Nematodes were harvested and washed with 0.04M NaCl, and lysed with proteinase K (200 ug/ml, Sigma) in 0.1M Tris pH 8.5, 0.05M EDTA, 0.2M NaCl, 1% SDS, at 65 °C. Following three phenol and two chloroform-isoamyl alcohol (24:1) extractions, genomic DNA was precipitated at room temperature with 2 volumes 95% ethanol, dissolved in 10mM Tris pH7, 1mM EDTA and purified on CsCl-ethidium bromide equilibrium density gradients. RNA remaining in the ethanol supernatant was precipitated on further addition of 0.5 volumes 95% ethanol at -20 °C, and used without further purification.

Recombinant plasmids were grown in NZYM media, and isolated following alkaline lysis as described (Maniatis et.al., 1982).

Restriction fragments were isolated from restriction digests fractionated on 1% agarose by electroelution (Maniatis et.al., 1982).

Isolation of C. briggsae 5S DNA repeats

25 ug of C. briggsae genomic DNA was digested to completion with Hind III and size fractionated on 1% agarose. Genomic repeat bands were visualized with ethidium bromide; 1 kb and 0.7 kb repeat bands were excised and recovered by electroelution. Isolated repeats were ligated into Hind III linearized pUC13 and transformed into E. coli JM83 (Vieira and Messing, 1982).

Recombinants carrying 5S DNA were identified by colony hybridization to a fragment of pCe5S1 encoding *C. elegans* 5S RNA (positions -8 to +120) (see Figure 3.1a legend).

Blots and hybridization conditions

C. briggsae genomic DNA was digested to completion with the indicated enzymes and size fractionated on 1% agarose (1 ug/lane). Restriction digests were transferred to nylon filters (Dupont GeneScreen, manufacturers protocol), and hybridized overnight in 5XSSPE (SSPE = 0.18 M NaCl, 10 mM NaPO₄ pH7.0, 1 mM Na₂EDTA) at 65 °C, to isolated restriction fragments nick-translated (Rigby *et.al.*, 1977) to a specific activity of 10⁸ cpm/ug. Filters were washed extensively in 0.5XSSPE at 65 °C and autoradiographed wet on XAR-5 or XK-1 (Kodak) film. After autoradiography, filters were stripped by boiling in distilled water, and reprobed as described above.

5S DNA repeat copy number was determined by comparing the genomic hybridization signal to plasmid standard curves. Hind III digests of *C. briggsae* genomic DNA (1 ug = 1.1 X 10⁷ haploid genomes), pCb5S1 and pCb5S0.7 (0.5 ng to 8.8 ng = 10 to 175 haploid genome equivalents) were size fractionated on 1% agarose, transferred to nitrocellulose and hybridized to nick-translated Cb5S1 and Cb5S0.7 probes. Hybridizing bands were excised, Cerenkov counted and cpm's bound plotted against number of genome equivalents blotted.

C. briggsae RNA was fractionated on 10% acrylamide 7M urea

gels and electroblotted to nylon filters (Dupont GeneScreen, suppliers protocol). Filters were prehybridized for 5 hours in 5XSSPE, 5X Denhardt's (0.5% Ficoll, 0.5% PVP, 0.5% BSA), and 0.3% SDS at 45 °C. Oligonucleotide probes complementary to the 5' splice leader transcript (5'L) derived from either the 1 kb (oligo 1-1) or 0.7 kb repeat (oligo 0.7-1) (see figure 6 for details) were end-labeled using T4 polynucleotide kinase and γ -³²P-ATP (Amersham) as described (Maniatis et.al., 1982). Probes were hybridized to RNA filters in 5XSSPE, 5X Denhardt's, 0.3% SDS overnight at 45 °C, washed briefly in 1XSSPE, 0.3% SDS at 45 °C and autoradiographed on Kodak XK-1 film at room temperature.

Plasmid sequencing and sequence data analysis

Cb5S1 and Cb5S0.7 were cloned in both orientations in the Hind III site of pUC13. Overlapping deletions from both ends of each repeat were constructed using the exo III-S1 protocol of Hennikoff (1984). DNA sequences were determined from denatured plasmid templates as described (Hattori and Sakaki, 1986). Sequence analysis was done using the SEQNCE program (Delaney Software).

Transcription reactions

C. elegans embryonic extracts were prepared and transcription reactions performed essentially as previously reported (Honda et.al., 1986). Plasmid templates were added in the covalently closed circular form at 0.1 nM. Final reaction

volumes were 25 ul and incubations were for 1 hour at 25 °C. Reactions were terminated by the addition of SDS to 1% and phenol-CHCl₃ extracted. Nucleic acids were ethanol precipitated, resuspended in formamide and loaded onto 10% sequencing gels. After drying, in vitro transcripts were visualized by autoradiography on Kodak XK-1 film.

RESULTS

Isolation and genomic organization of *C. briggsae* 5S DNA

The 5S RNAs of *C. elegans* and *C. briggsae* are identical in nucleotide sequence (Butler et.al., 1981). In order to visualize the *C. briggsae* genomic sequences encoding 5S RNA, we probed Southern blots of *C. briggsae* genomic restriction digests with a 120 bp fragment of pCe5S1 (Nelson and Honda, 1985) corresponding to the *C. elegans* 5S RNA coding region (5S coding probe, Figure 3.1 legend). The results (Figure 3.1b) identified two genomic repeat families (1 kb and 0.7 kb Hind III fragments) which potentially encode *C. briggsae* 5S RNA.

These putative 5S DNA repeats were isolated from a size fractionated *C. briggsae* Hind III genomic digest and cloned into pUC13 for further characterization. Two recombinant plasmids, pCb5S1 and pCb5S0.7, were chosen as representative clones and used for all subsequent analyses. The restriction maps of these two representative repeats are shown in Figure 3.2. The putative

Figure 3.1: Identification of C. briggsae genomic sequences encoding 5S RNA.

a) Origin of the 5S RNA coding sequence probe.

The 5S RNA coding sequence probe was isolated as a 128 bp Hind III fragment (positions -8 to +120, Nelson and Honda, submitted) from a 5'deletion/substitution derivative of pCe5S1 (Nelson and Honda, 1985) obtained using the exo III-SI protocol of Henikoff (1984).

b) Hybridization of the 5S RNA coding probe to C. briggsae genomic restriction digests.

One ug of C. briggsae genomic DNA was digested to completion with the indicated enzymes, fractionated on 1% agarose, transferred to a nylon membrane and hybridized to the nick-translated 5S RNA coding probe shown above. The position of the 1 kb and 0.7 kb genomic repeat bands are indicated.

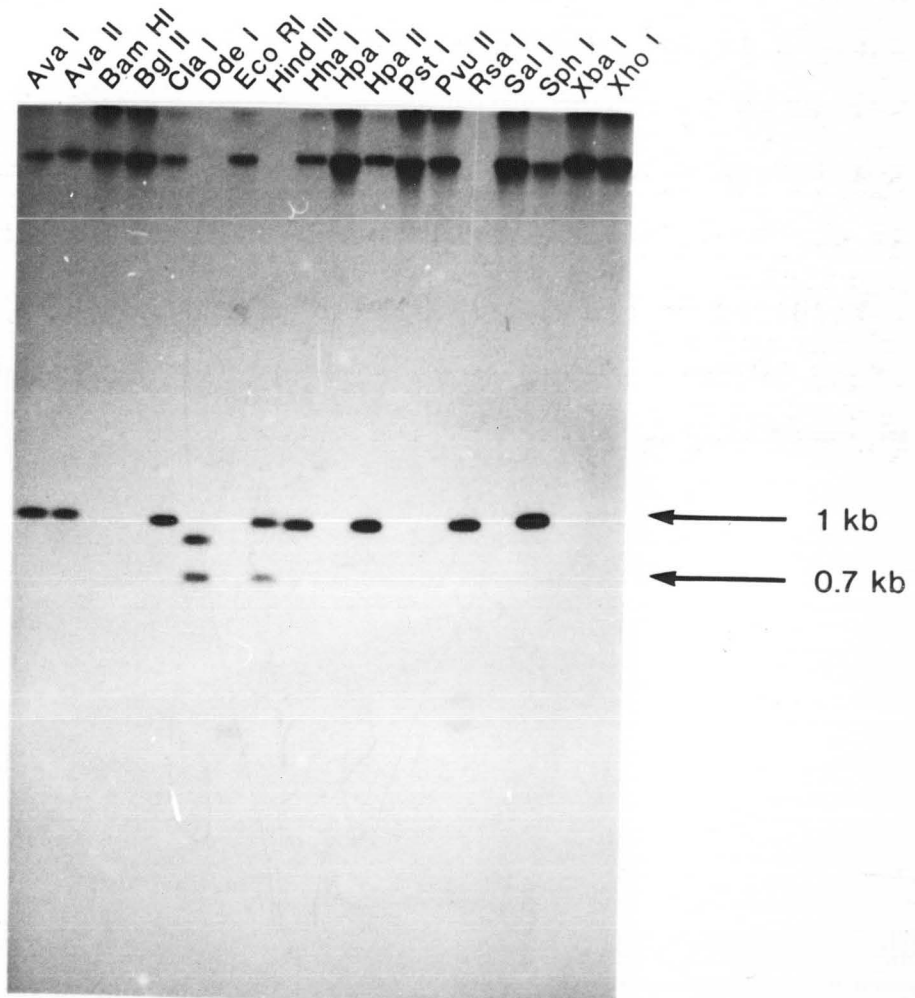
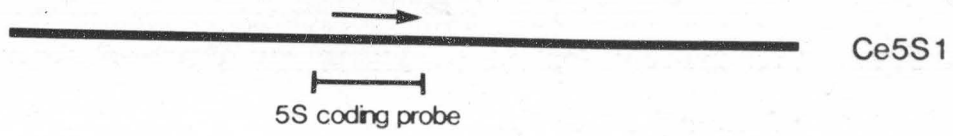
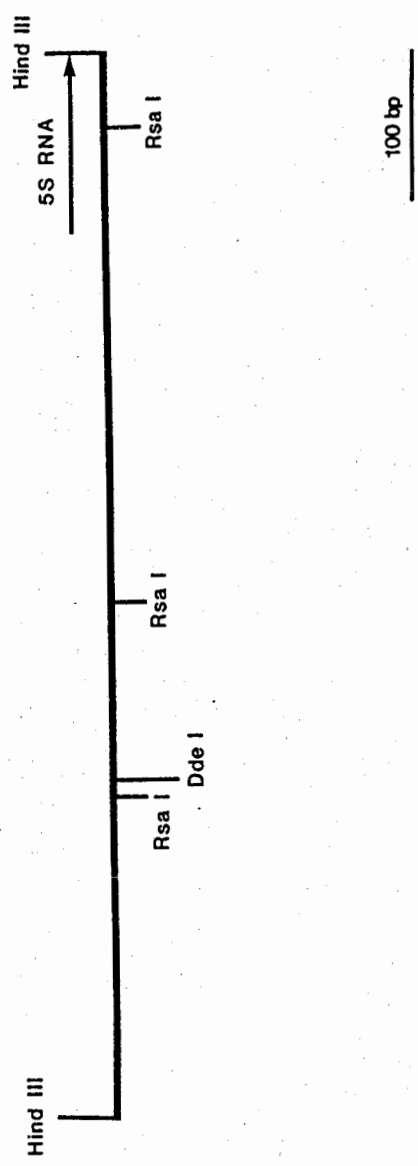


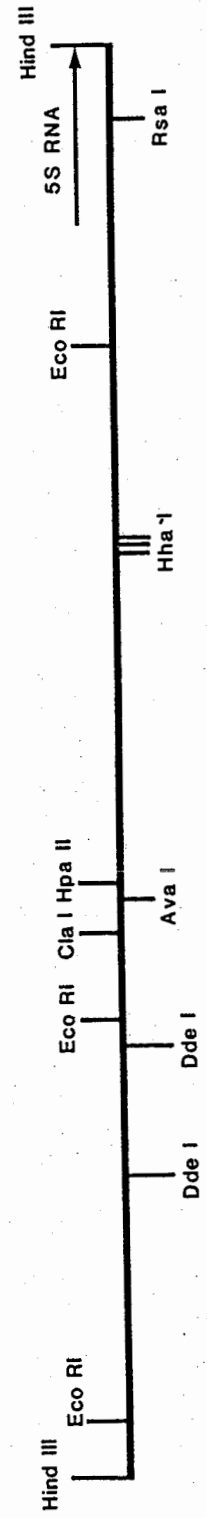
Figure 3.2: Restriction maps of *C. briggsae* genomic 5S DNA repeat families.

Both repeats are drawn to scale and the locations of various restriction sites are shown. The location of the putative 5S RNA coding region is indicated by the arrow.

a) Cb5S0.7



b) Cb5S1



5S RNA coding sequences of both repeats (identified by hybridization to the 5S coding probe) are located immediately adjacent to the right-most Hind III site as drawn in Figure 3.2. The remainder of these repeats share no common restriction sites and cross-hybridize poorly.

Genomic organization of *C. briggsae* 5S RNA genes

Our preliminary genomic Southern blots indicated that the 1 kb and 0.7 kb 5S DNA repeats of *C. briggsae* are organized in tandem clusters. These results could not distinguish between the mixing of families within a cluster and their separation into independent tandem clusters. In order to resolve this ambiguity, we used spacer sequences specific to each repeat class (1 kb spacer and 0.7 kb spacer probes, Figure 3.3) to probe a battery of *C. briggsae* genomic restriction digests. The results presented in figure 3 show that each genomic repeat family lies in an independent tandem cluster. Enzymes which cut in the 1 kb but not in the 0.7 kb repeat all leave the 0.7 kb repeat family as a high MW tandem array, while reducing the 1 kb repeat family to its unit length. The Cb5S1 and Cb5S0.7 repeat families are therefore not interspersed in tandem clusters; each family appears to be organized into a separate tandem cluster of homogeneous repeat units.

The hybridization of the 5S coding probe to the 1 kb repeat in genomic Hind III digests appears to be stronger than to the 0.7 kb repeat. A comparison of genomic hybridization signal to a

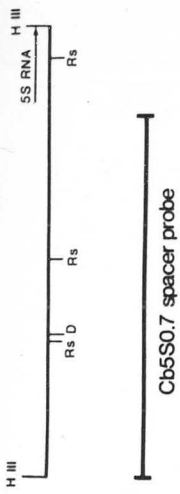
Figure 3.3: Genomic organization of C. briggsae 5S DNA.

a) Organization of the 1 kb 5S DNA repeat family.

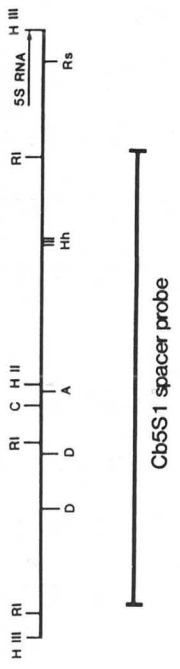
The filter shown in Figure 1 was stripped by boiling in distilled water and rehybridized to the 1 kb repeat spacer probe shown above the autoradiograph.

b) Organization of the 0.7 kb 5S DNA repeat family.

The same filter was again stripped by boiling in distilled water and hybridized to the 0.7 kb repeat spacer probe shown above the autoradiograph.



b)



a)

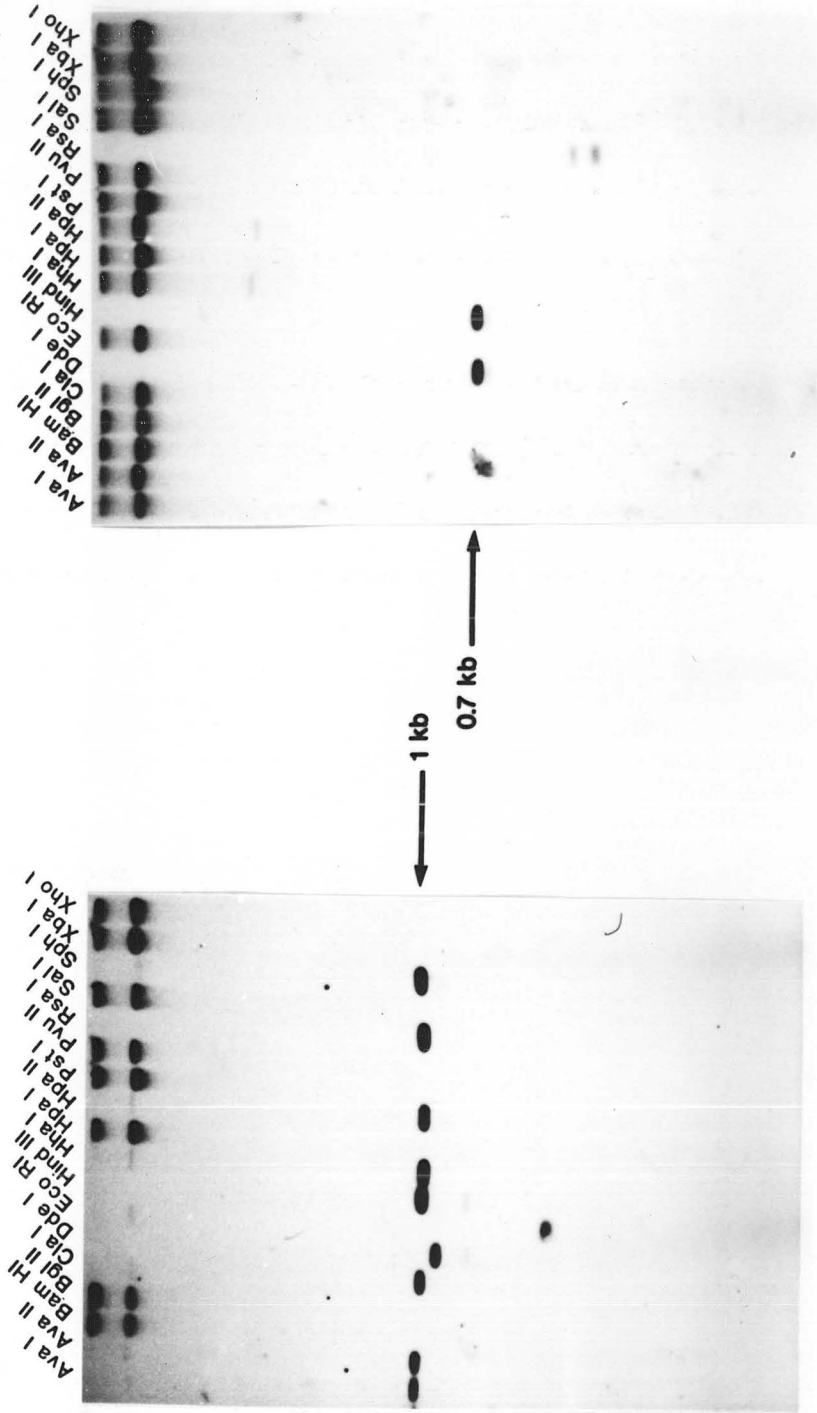
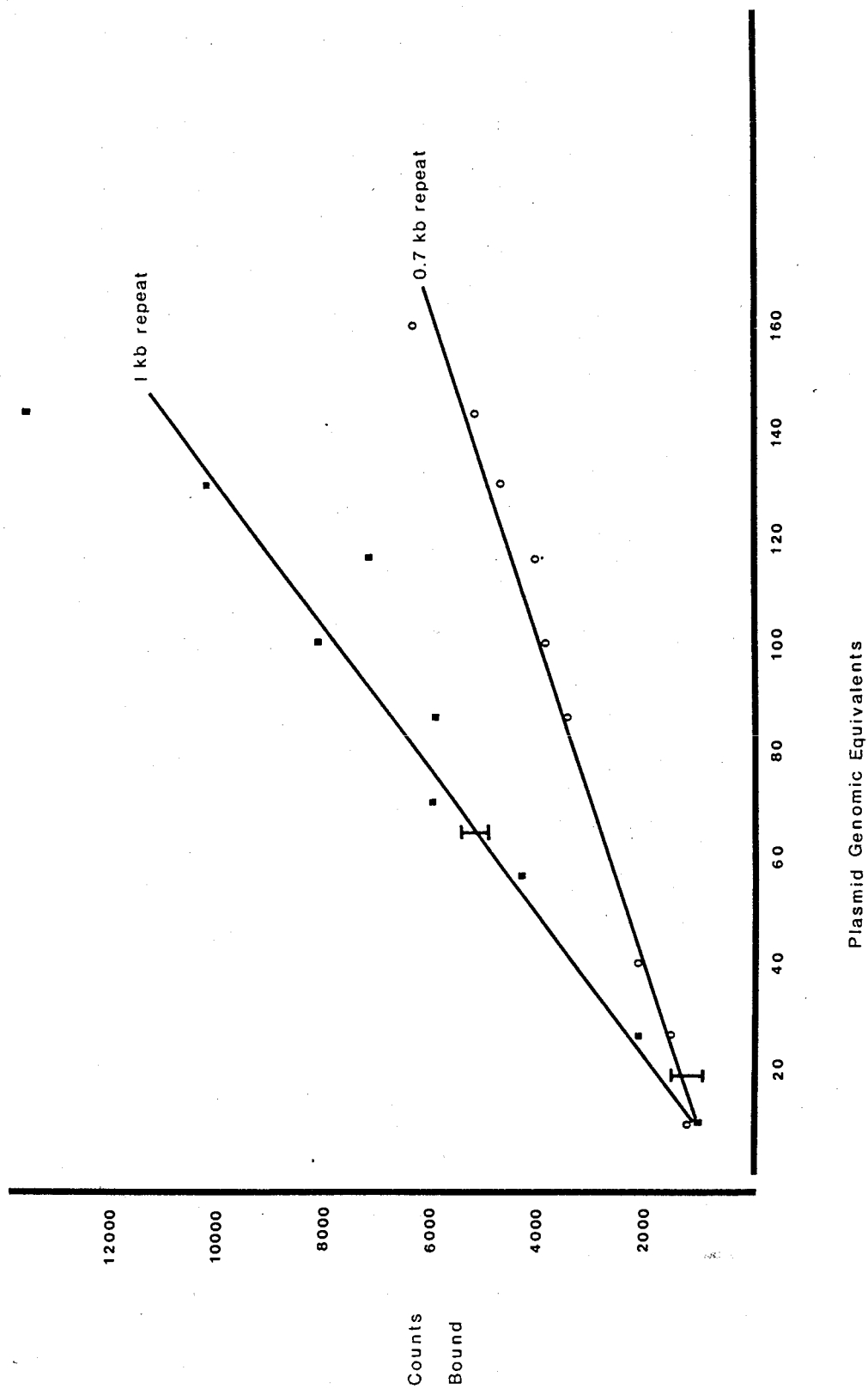


Figure 3.4: Quantitation of the C. briggsae haploid 5S DNA content.

C. briggsae genomic DNA (1 ug) and plasmid standards were digested with Hind III to release the 5S DNA unit repeats, briefly electrophoresed through 1% agarose, transferred to nitrocellulose and hybridized against the nick-translated 1 kb 5S DNA repeat probe or the 0.7 kb 5S DNA repeat probe. Bands were visualized by autoradiography, excised and Cerenkov counted. The plasmid (pCb5S1 or pCb5S0.7) standard curves are shown as Cerenkov counts bound vs. number of genomic equivalents blotted. The genomic hybridization signals (vertical bars in Figure) represent the average of three measurements for the 1 kb and 0.7 kb repeats of C. briggsae.



plasmid standard curve indicates that there are approximately 65 copies of the 1 kb repeat and 20 copies of the 0.7 kb repeat per C. briggsae haploid genome (Figure 3.4).

Many eukaryotes possess dispersed copies of 5S DNA in addition to the tandem clusters. The lack of additional hybridizing bands observed in Figure 3.3 suggests that the C. briggsae genome contains few (note the weak bands revealed with the 0.7 kb spacer probe in Figure 3.3), if any dispersed copies of either repeat family.

Structural features of the C. briggsae 5S DNA repeats

In order to examine the relationship between the C. briggsae and C. elegans 5S DNA repeat families, we determined the complete nucleotide sequences of Cb5S1 and Cb5S0.7. These sequences are shown in figure 3.5. Both C. briggsae 5S DNA repeats share general sequence features previously described for the C. elegans 5S DNA repeat (Nelson and Honda, 1985). In each case, an internal Hind III site separates the 5S RNA coding sequence from putative RNA pol III transcription termination signals. The 3' flanking sequences are AT rich with numerous runs of As and Ts. We find no extensive simple sequence repeats in any of these 5S DNA repeats, although a 35 bp sequence is non-tandemly repeated perfectly twice and partially a third time in Cb5S0.7 (Figure 3.5).

Comparison of Cb5S1 and Cb5S0.7 sequences confirms the lack of extensive sequence conservation suggested by the previous

Figure 3.5: Sequence of the C. briggsae 5S DNA repeats.

Each sequence is numbered relative to the 5S RNA coding region, following the numbering of the C. elegans 5S DNA repeat. 5S RNA (positions +1 to +120) and 5' L coding regions (positions -218 to -312 in Cb5S1 and -265 to -171) (see text) are indicated by the heavy arrows. The putative RNA polymerase III transcription termination signals flanking the 5S RNA coding regions are boxed (positions +119 to +122 or +123). The non-tandemly repeated sequence in Cb5S0.7 (positions -265 to -241, -410 to -386 and +210 to +245) described in the text is also underlined.

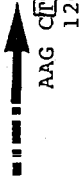
Cb5S0.7

150 TAAATATGAA TTGTGATCC TGATATGTGAT TTTTGATTTT TTTCCGGTTTT GAATAAAAA TTTCTATATT TTTGCCAGGT CCGTTTAAAT ACCCAAGTTT GAGTATTAAT
 160 170 180 190 200 210 220 230 240 250
 260 270 280 290 300 310 320 330 340 350
 ATAGGTATT GTCATCATT TTGCCAGGTC GGTTTAATTA CCCAAGTTTG AGGTAGACGG GCGGAGTGTT GTACCCAATT CTCAGATAAT ACACAGTAGC AFAGTGGCCA
 360 370 380 390 400 410 420 430 440 450
 ACGGGGCCT CGATCGGTGT CCGGTGCTGG GAAATGTGAA GAGGGGAGAC AACGCTTTCA GCTTTGGTTT AATTACCCAA GTTTGAGGTA CTTTGGGTTT TCGACCTAAA
 460 470 480 490 500 510 520 530 540 550
 GAAATTTAGA GTTGATAACA AATTTGGAA CGTCCCCTTT CCGGGGAGACA AAAAATCTAA AGAAATTAAG AATTGGTCAA CTATGAACCG AGTTATGATT TTTTGAAAT
 560 570 580 590 600 610 620 630 640 650
 TTGAAAAAT CCTCTTTTT TCAATTTCAA ATTCTCAATT TTCATCGAGG TTATTTTTTT CTACACAGG AAGAGAAAAT TTATATATGA GAAAAGGAGG TGAGAAAAAT
 660 670 680 690 700 710 720 730 740 750
 GCTTACGACC ATATCACGTT GAATGCACGC CATCCCGTCC GATCTGGCAA GTTAAGCAAC GTTAGTACTT GGATCGGAGA CCGCCTGGGA ATCCTGGATG
 760 770 780 790 800 810 820 830 840 850
 TTGTAAGCCT
 120

120
 AAGCTTTT
 130
 GATTTTTT
 140
 ATTCATTGTA

120
 TTGTAAGCCT
 120

Cb5S1



130 140 150 160
AAG CTTTGGCTTT TTTTGTATTCA ATTTTCATAAA TATGAATTC AAGATTTCAA
120

170	180	190	200	210	220	230	240	250	260	270
TTTTAAGAT	TCATGTTCTT	AAAAAATTGC	AGAAATTTCA	TCATTGGFTT	GGATTTTTCGT	GCTTTTTTTC	TAATATTTTAT	TTATTTTGTG	AATAAATTTG	TAAAATGTTT
280	290	300	310	320	330	340	350	360	370	380
TCCTCACAAA	TGAGAAGTGT	TGTCGAAAAA	TATAAANTTG	TGTCAAATTC	AATCAATTTT	AATATGAAAA	ATTTGAATGT	TCAAAGATGCA	TTCCGGTTTT	CTTCCACCCA
390	400	410	420	430	440	450	460	470	480	490
CATGGCCTGA	GTTTCTGAAA	TAAATTTGTC	ATGAATCGAC	TGAAAATAGA	TGTGTGTGTA	TAAATTTTAT	ATTGTAAAAA	TTTCAGAATC	TAAGCCTAAA	ATTTTAAGAA
500	510	520	-401	-391	-381	-371	-361	-351	-341	-331
TTCACTAGAA	TTTAAATGAT	GACAATTCGG	ATCGTTTAC	AACCCTTTTA	CGATGCATCG	ATTTTTTCGAT	TTTCAATGGA	CCATAACTCC	GGCATGCGTT	GACCGATTTT
CAATTTTAA	AGTTATTTT	TCTCCCCGAG	AGGAGACGTT	CCAAAATTTA	TAGCTAACGC	CAAAATTTCTT	TGGGTCAGTT	TCAANTGTTTA	CCTCAAACTT	GGGTAATTA
-321	-311	-301	-291	-281	-271	-261	-251	-241	-231	-221
ACCAACTACA	TCGGGGGGCT	TCCACACACTA	GTGGAAAACG	AGCGGGCCGAC	ACCAATCGAG	GCCCCGGTCCG	GCCACCCGAT	TTCCACACGG	TCCGGCCGCA	CACTCGTGGG
-211	-201	-191	-181	-171	-161	-151	-141	-131	-121	-111
TGAGTCTTCT	TCTACTGTTG	GGGGGACTGG	GAGAAATCGC	TCTTCTGCGT	TTCGATTTAT	TTCACTGATC	CCTAGTAGAG	TTAAAGGGG	AATGTAGAGG	TAGATGTGAT
-101	-91	-81	-71	-61	-51	-41	-31	-21	-11	-1
GCTTACGACC	ATATCACGTT	GAATGCACGC	CATCCCGTCC	GATCTGGCAA	GTTAANGAAC	GTTGAGTCCA	GTTAGTACTT	GGATCGGAGA	CGGCCCTGGGA	ATCCTGGATG
10	20	30	40	50	60	70	80	90	100	110



hybridization results. Only the 5S RNA coding region and 3' flanking sequences are directly conserved between the two repeats (Figure 3.6a). The 5S RNA coding sequence itself is perfectly conserved, while the 3' flanking 40 bp is slightly divergent. Further downstream, and upstream of the 5S RNA coding region, the two repeats are not significantly similar.

A second region of conserved sequence (approximately 250 bp in length) is revealed when Cb5S1 is compared to the complement of Cb5S0.7 (Figure 3.6a and 3.6b). The conservation is not as complete as noted above for the 5S RNA coding region, but is punctuated by blocks of divergent sequence.

Comparison of *C. elegans* and *C. briggsae* 5S DNA repeat sequences

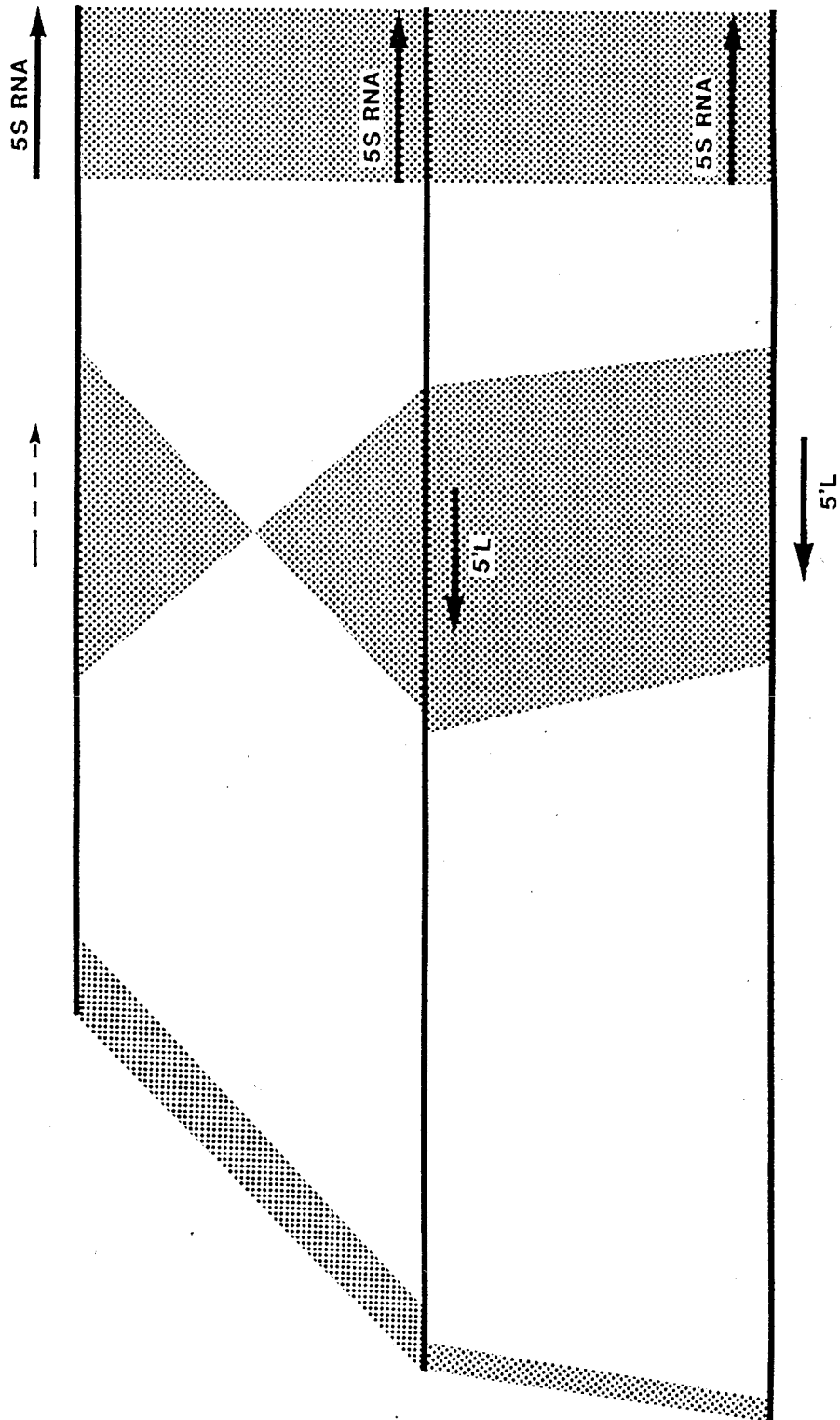
Comparison of *C. elegans* and *C. briggsae* 1 kb 5S DNA repeat sequences reveals two blocks of extended conserved sequence. One of these corresponds to, and is limited to the 5S RNA coding region and 15 bp of 3' flanking sequence. The other is located upstream in the spacer regions (-273 to -177 in Ce5S1 and -315 to -218 in Cb5S1, see figure 3.6c). This spacer sequence identity resembles that of the 5S RNA coding region; a core of 118 bp is nearly perfectly conserved while flanking sequences diverge rapidly on one side (within 20 bp upstream) and slowly on the other (within 60 bp downstream). The remainder of these repeats share little sequence identity.

Both of these conserved sequences are also found in the 0.7 kb *C. briggsae* 5S DNA repeat. The 5S RNA coding region is

Figure 3.6: Comparisons of C. elegans and C. briggsae
5S DNA repeat sequences.

- a) Line drawing showing the gross organization of the repeats and their conserved regions.

Each repeat (Cb5S0.7, top line, Cb5S1, middle line and Ce5S1, bottom line) is drawn to scale and the locations of the transcription units (5S RNA and 5'L) indicated by the arrows. Conserved regions are indicated by shading. Notice that the conserved region encompassing the 5'L coding region of Cb5S0.7 is inverted relative to the 5S RNA coding region.



(Figure 3.6 continued)

b) Computer generated alignments of the 5S RNA coding, 5' flanking and 3' flanking sequences of the three 5S DNA repeat families. The order of the repeats is:

- Cb5S0.7 (positions -20 to +187, Figure 5a)
- Cb5S1 (positions -31 to +186, Figure 5b)
- Ce5S1 (positions -26 to +195, Nelson and
Honda, 1985)

Only the Cb5S1 sequence is shown in its entirety. The 5S RNA coding region is delineated by the heavy arrows. Conserved positions are indicated by (*), non-conserved positions are indicated by the appropriate nucleotide and gaps are indicated by a space.

(Figure 3.6 continued)

- c) Computer generated alignments of the conserved spacer sequences.

The sequences are aligned to highlight the 5' L transcription unit which is shown in the 5' --> 3' orientation. Note that, as a result, the Cb5S1 and Ce5S1 sequences are shown in their complementary orientation relative to the 5S RNA coding sequences. The sequences follow the indicated order:

- Cb5S0.7 (positions -337 to -79, Figure 3a)
- Cb5S1 (positions -126 to -398, Figure 3b)
- Ce5S1 (positions -128 to -348, Nelson and
Honda, 1985)

Only the Cb5S1 sequence is shown in its entirety. The 5' L coding region is delineated by the heavy arrow. Conserved positions are indicated by (*), non-conserved positions are indicated by the appropriate nucleotide and gaps are indicated by a space.

perfectly conserved, while the spacer region discussed above is partially conserved, in the complementary orientation as noted in the previous section.

Are both *C. briggsae* 5S DNA repeat families functional?

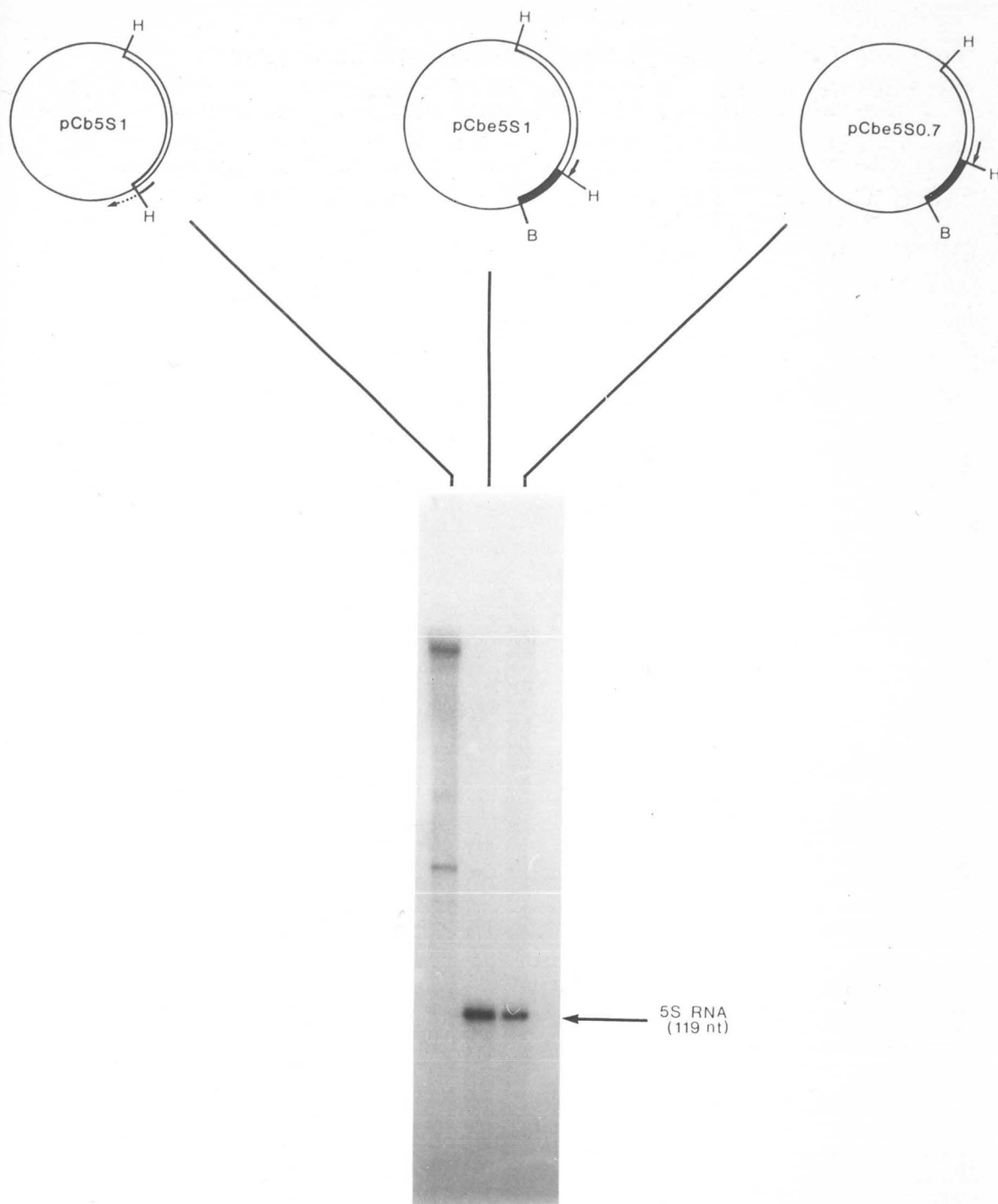
The identification of two related but distinct 5S DNA repeat families in the *C. briggsae* genome raises the possibility that one of these families may be partially or completely inactive in vivo.

The perfect conservation of the 5S RNA coding sequences of both repeat families makes it impossible to examine their expression in vivo. We have therefore examined their ability to program 5S RNA transcription in a *C. elegans* cell-free extract. Both Cb5S1 and Cb5S0.7 were cloned as Hind III fragments which results in the separation of 5S RNA coding sequences from the putative RNA pol III termination signals. The transcription of pCb5S1 and pCb5S0.7 results in the appearance of several large transcripts in vitro (Figure 3.7, lane 1). Rather than reisolate the *C. briggsae* genomic repeats, Cb5S1 and Cb5S0.7 were recloned adjacent to the RNA pol III termination signals carried by Ce5S1 to produce the templates pCbe5S1 and pCbe5S0.7 shown in Figure 3.7. These templates are efficiently transcribed in vitro to produce the 119 nt 5S RNA, indicating that both repeats can function as 5S RNA templates in vivo.

While this work was in progress, we learned of the results of Krause and Hirsh (1987) describing an abundant 100 nt

Figure 3.7: Transcription of C. briggsae 5S DNA in a C. elegans cell-free transcription system.

The structures of the templates pCb5S1, pCbe5S1 and pCbe5S0.7 are shown above the autoradiograph. An arrow indicates the position of the 5S RNA transcript programmed by pCbe5S1 and pCbe5S0.7.



transcript (5'L) in C. elegans which appears to donate its 5' 22 nt to the 5' termini of a variety of cellular mRNAs via a trans-splicing mechanism. Surprisingly, the sequence of this 100 nt 5'L transcript appears in the C. elegans 5S DNA repeat, 176 bp upstream of, and in the opposite orientation to, the 5S RNA coding sequence.

This 5'L coding region corresponds precisely to the highly conserved spacer region described above. The single nucleotide difference between the 5'L coding regions of Ce5S1 and Cb5S1 lies downstream of the 5' splice donor consensus sequence and is unlikely to affect transcription or the trans-splicing process. In contrast, the 5'L coding sequence of Cb5S0.7 has diverged considerably from its homologs. The 5' 22 nt is however perfectly conserved, as is most of the adjacent splice donor consensus sequence, suggesting that, if transcribed, the divergent 5'L transcript might function as a substrate in the trans-splicing process. Immediately adjacent to the splice donor site, the two repeats diverge extensively. We have used repeat-specific oligonucleotides complementary to this divergent region of the two putative 5'L transcripts to probe Northern blots of total C. briggsae RNA fractionated on denaturing polyacrylamide gels. The results (Figure 3.8) show that while the 1 kb repeat 5'L-specific probe detects an abundant 100 nt RNA analogous to that described in C. elegans, the 0.7 kb repeat 5'L-specific probe does not reveal a detectable transcript. Therefore, the divergence of 5'L coding sequences, and/or flanking sequences,

Figure 3.8: in vivo expression of the 5'L coding regions of the C. briggsae 5S DNA repeats.

- a) Origin of repeat specific oligonucleotide probes.
The putative 5'L transcripts derived from the C. briggsae 0.7 and 1 kb 5S DNA repeats are shown. Splice donor sites are underlined. The two oligonucleotide probes used to probe Northern blots are shown above the 5'L transcript sequences. Note that the oligonucleotides do not cross-hybridize under the conditions used.
- b) Northern blots of total C. briggsae RNA probed with repeat specific oligonucleotides.
10 ug of total C. briggsae RNA was fractionated on denaturing 10% polyacrylamide gels, transferred to nylon filters and hybridized to oligonucleotide probes, oligo 0.7-1 and oligo 1-1.

has resulted in the inactivation of the 5' L transcription unit of the 0.7 kb 5S DNA repeat.

DISCUSSION

The structure and organization of C. briggsae 5S RNA genes is typical of higher eukaryotes (Long and Dawid, 1980). C. briggsae 5S DNA is split into two discrete repeat families, approximately 65 copies of a 1 kb repeat and 20 copies of a 0.7 kb repeat. Each repeat family is homogeneous in size and structure, and is organized in an independent tandem cluster. There are few, if any, dispersed 5S DNA repeat sequences in the C. briggsae genome.

The identification of two repeat families encoding C. briggsae 5S RNA contrasts with the presence of a single 5S DNA repeat family previously characterized in C. elegans, a closely related nematode species. The 5S DNA repeats of both species appear to have diverged from a common ancestral repeat unit; all three share two blocks of highly conserved sequence. The relative orientation of these conserved sequences is the same for C. elegans and C. briggsae 1 kb repeats, suggesting that they most closely resemble the ancestral repeat. The complementary orientation of the conserved spacer sequences relative to the 5S RNA coding region in the 0.7 kb repeat indicates that this repeat may have diverged via a complex series of events including the inversion of spacer sequences. The organization of the C. briggsae 5S DNA repeat families into mutually exclusive tandem clusters may serve to maintain homogeneity within each repeat

family while allowing the families to diverge from one another.

The highly conserved nature of these 5S DNA repeat sequences attests to their functional importance. Both *C. briggsae* 5S DNA repeats contain perfectly conserved 5S RNA coding sequences and program the efficient transcription of 5S RNA in a *C. elegans* cell-free extract. While we cannot distinguish between 5S RNA transcripts derived from these two repeat families *in vivo*, the *in vitro* transcription results strongly suggest that both represent functional 5S RNA genes.

Surprisingly, the conserved spacer sequence also appears in the *C. elegans* cellular RNA pool, as an abundant 100 nt transcript (5'L) which is thought to participate in a trans-splicing process (Krause and Hirsh, 1987). Our results indicate that an analogous transcript is found in *C. briggsae*. Our Northern analysis using repeat-specific oligonucleotide probes complementary to putative 5'L transcripts suggests that the 1 kb repeat is transcribed *in vivo*, while the 0.7 kb repeat appears to be transcriptionally inactive.

The use of inter-species sequence comparisons has allowed us to identify two transcription units on the genomic repeats encoding 5S RNA of two closely related nematodes. The remainder of these repeats share limited sequence identity. We are now characterizing the template requirements for efficient 5S RNA transcription in the homologous cell-free extract.

REFERENCES

- Arnheim, N. (1983).
Concerted evolution of multigene families.
in: Evolution of Genes and Proteins. Edited by
M. Nei and R.K. Koehn. Sinauer Associates, Inc.,
Sunderland, MA. p. 38-61.
- Brenner, S. (1974)
The genetics of Caenorhabditis elegans.
Genetics 77 71-94.
- Butler, M.H., Wall, S.M., Luehrsen, K.R. Fox, G.E. and Hecht,
R.M. (1981).
Molecular relationships between closely related strains
and species of nematodes.
J. Mol. Evol. 18 18-23.
- Emmons, S.W., Klass, M.R. and Hirsh, D. (1979).
Analysis of the constancy of DNA sequences during
development and evolution of the nematode Caenorhabditis
elegans.
Proc. Natl. Acad. Sci. 76 1333-1337.
- Fedoroff, N.V. (1979).
On spacers.
Cell 16 697-710.
- Hattori, M. and Sakaki, Y. (1986).
Dideoxy sequencing method using denatured plasmid templates.
Anal. Biochem. 152 323-238.
- Henikoff, S. (1984).
Unidirectional digestion with exonuclease III creates
targeted breakpoints for DNA sequencing.
Gene 28 351-359.
- Hirsh, D., Emmons, S.W., Files, J.G. and Klass, M.R. (1979).
Stability of the C. elegans genome during development and
evolution.
in: Eukaryotic Gene Expression. Edited by R. Axel, T.
Maniatis and C.F. Fox. Academic Press, NY. p. 205-218.
- Honda, B.M., Devlin, R.H., Nelson, D.W. and Khosla, M. (1986).
Transcription of class III genes in cell-free extracts from
the nematode Caenorhabditis elegans.
Nuc. Acid. Res. 14 869-881.
- Krause, M. and Hirsh, D. (1987).
A trans-spliced leader sequence on actin mRNA in C. elegans.
Cell 49 753-761.

- Long, E.O. and Dawid, I.B. (1980).
Repeated genes in eukaryotes.
Ann. Rev. Biochem. 49 727-764.
- Maniatis, T., Fritsch, E.R. and Sambrook, J. (1982).
Molecular cloning, a laboratory manual.
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Nelson, D.W. and Honda, B.M. (1985)
Genes encoding 5S ribosomal RNA of the nematode
Caenorhabditis elegans.
Gene 38 245-251.
- Nelson, D.W. and Honda, B.M. (1986).
Genetic mapping of the 5S rRNA gene cluster of the nematode
Caenorhabditis elegans.
Can. J. Genet. Cytol. 28 545-553.
- Rigby, P.,W., Dieckmann, M., Rhodes, C. and Berg, P. (1977)
Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase.
J. Mol. Biol. 113 237-251.
- Samson, M-L. and Wegnez, M. (1984).
The 5S ribosomal genes in the Drosophila melanogaster
species subgroup. Nucleotide sequence of a 5S unit from
Drosophila simulans and Drosophila teissieri.
Nuc. Acid. Res. 12 1003-1014.
- Vieira, J. and Messing, J. (1982).
The pUC plasmids and M13mp7-derived system for insertion
mutagenesis and sequencing with synthetic universal primers.
Gene 19 259-268.

THE FOURTH PART

SEQUENCE REQUIREMENTS FOR EFFICIENT 5S RNA TRANSCRIPTION

IN A *C. elegans* CELL-FREE EXTRACT

INTRODUCTION

The transcription of 5S RNA in Xenopus laevis serves as a useful paradigm for our understanding of eukaryotic transcription mechanisms. The template requirements of 5S RNA transcription in Xenopus are limited to transcribed sequences and the termination signal immediately flanking them (Sakonju et.al., 1980; Bogenhagen et.al., 1980). An essential region, the internal control region (ICR), interacts directly with a 5S-specific transcription factor (TFIIIA) (Engelke et.al., 1980) to form a transient complex. This complex interacts with two general RNA polymerase III (RNA pol III) transcription factors (TFIIB and TFIIC) to form a stable transcription complex which remains in place through many rounds of transcription (Wormington et.al., 1981; Bogenhagen et.al., 1982; Lassar et.al., 1983).

In the Xenopus model system, altering the sequences flanking the 5S RNA coding region has little effect on the efficiency of 5S RNA transcription in vitro, although subtle effects on the formation of the stable transcription complex have been noted (Wormington et.al., 1981). In contrast, recent studies have shown that the transcription of 5S RNA in B. mori (Morton and Sprague, 1984), N. crassa (Selker et.al., 1986; Tyler, 1987) and D. melanogaster (Garcia et.al., 1987) requires both 5S RNA coding and 5' flanking sequences in homologous cell-free extracts. Comparison of 5S DNA repeat sequences from these species reveals a conserved sequence (TATA) centered approximately 25 bp upstream of the transcription initiation site (Morton and Sprague, 1984;

Rubacha et.al., 1984; Morzycka-Wroblewska et.al., 1985) which is required for transcription in vitro.

We have previously reported on the structure and organization of the 5S RNA gene families of two nematode species, C. elegans and C. briggsae (Nelson and Honda, 1985; Nelson and Honda, submitted). These repeats are efficiently transcribed in a homologous cell-free extract of C. elegans embryos (Honda et.al., 1986). In this paper we use two lines of evidence, cell-free transcription and evolutionary conservation, to identify the functionally important template sequences required for C. elegans 5S RNA transcription. Functional analysis of deletion/substitution derivatives of the C. elegans and C. briggsae 1 kb 5S DNA repeats suggest that a conserved 5' flanking sequence element is required for efficient 5S RNA transcription in vitro. Our results also indicate that 5S RNA coding sequences (corresponding to an internal control region) are required for 5' deletion/substitution templates to compete for extract components which are limiting for transcription.

MATERIALS AND METHODS

a) Template constructions (see also figures 4.1 and 4.4)

5' and 3' deletion/substitution (D/S) derivatives of pCe5S1 (Nelson and Honda, 1985) were constructed using the exo III-S1 strategy of Henikoff (1984). Briefly, pCe5S1 was digested with

pst I and Sal I (5'D/S templates) or Sma I and Sst I (3'D/S templates), and then treated with exo III for varying times. The products were trimmed with S1, repaired with Klenow DNA polymerase, recircularized and transformed into E. coli JM83.

pCe5S1MAXI was constructed by ligating the 563 bp Hind III fragment of pCe5S1 (pUC13 polylinker (Hind III to Bam HI) + 406 bp 5' flanking sequence + 120 bp 5S RNA coding sequence lacking the 3' transcription termination signal) into the Hind III site of a 5' deletion/substitution derivative of pCe5S1 which lacks all of these sequences and an additional 185 bp of 3' flanking sequences. As discussed further in the text, this template is efficiently transcribed in vitro, but the lack of the appropriate termination signal results in a longer transcript (approximately 165 nt) which terminates at an alternative termination signal at position +346.

pCbe5S1 and pCbe5S0.7 were constructed by recloning the 1 kb and 0.7 kb Hind III 5S DNA repeats of C. briggsae, carried by pCb5S1 and pCb5S0.7 (Nelson and Honda, submitted), adjacent to the 3' flanking sequences carried by pCe5S1. pCbe5S1 5' deletion/substitution (D/S) derivatives were obtained by first recloning the Eco RI fragment of pCbe5S1 containing 78 bp of 5' flanking sequence, 5S RNA coding sequences and 3' flanking sequences into the Eco RI site of pUC13 (pCbe5S1 5'D/S -78). Additional 5' deletions were obtained by exo III-S1 treatment of Pst I-Sal I linearized pCbe5S1 5'D/S -78 as described above. The single 5' deletion derivative of pCbe5S0.7 was obtained by

ligating Hind III linearized pCe5S1 and pCb5S0.7 5'D/S -38 (a 5' deletion derivative of pCb5S0.7 used for determining the 0.7 kb 5S DNA repeat sequence), and recloning the gel-purified 603 bp Eco RI fragment containing 38 bp of 5' flanking sequence, 120 bp 5S RNA coding sequence and 3' flanking sequence into the Eco RI site of pUC13.

Combined templates pCe5S1 5'D/S ___+MAXI and pCbe5S1 5'D/S___+MAXI were constructed by ligating the 795 bp Bam HI fragment of pCe5S1MAXI into the unique Bam HI site of pCe5S1 5'D/S___ and pCbe5S1 5'D/S___.

In each case, plasmid structure was determined by restriction analysis and confirmed by direct plasmid sequence analysis (Hattori and Sakaki, 1986).

Transcription reactions

C. elegans embryonic cell-free extracts were prepared essentially as previously described (Honda et.al., 1986). Transcription reactions were scaled down to 25 ul total volume. Incubations were at 25 °C for 1 hour and reactions were stopped by adding SDS to 1% and were then phenol-CHCl₃ extracted. Nucleic acids were ethanol precipitated, resuspended in formamide and loaded directly onto 10% sequencing gels. After electrophoresis, gels were dried and autoradiographed on Kodak XK-1 film. Transcription was quantitated by excising labeled transcript bands using the autoradiograph as a template and Cerenkov counting, or by scanning densitometry of the

autoradiogram.

Before using an extract for quantitative analysis, the [template] and [DNA] dependence of 5S RNA transcription was examined. In general, 5S RNA transcription was linear between 0 and 0.5 nM pCe5S1. Further increase in template concentration did not increase the amount of 5S RNA transcription. The addition of carrier DNA (as covalently closed circular (ccc) pUC13) up to a final [DNA] of 12 ug/ml doubled the efficiency of 5S RNA transcription from low [pCe5S1]. Higher [DNA] depress 5S RNA transcription slightly. Based on these results, we chose to hold [DNA] constant at 8 ug/ml (with ccc pUC13) and [templates] (both experimental and internal reference) at 0.1 nM.

For template competition assays (Wormington et.al., 1981), the competing template was preincubated with the extract in a reaction mix lacking NTPs for 30 minutes at 25 C before adding the reference template (pCe5S1MAXI) and NTPs.

RESULTS

Template requirements for 5S RNA transcription from pCe5S1

We previously described the structure and organization of the 5S RNA gene family of C. elegans (Nelson and Honda, 1985). The wild-type template used in these studies, pCe5S1, consists of a 979 bp C. elegans 5S DNA repeat (Bam HI fragment) cloned into the polylinker of pUC13. Two cellular transcripts are encoded by

the C. elegans 5S DNA repeat. The 119 bp 5S RNA coding sequence is positioned immediately adjacent to the centrally located Hind III site as shown in Figure 4.1. A second region located 176 bp upstream of the 5S RNA coding region is divergently transcribed to produce the 100 nt 5' leader transcript (5'L) described by Krause and Hirsh (1987).

We have also described a cell-free extract of C. elegans embryos which accurately and specifically transcribes 5S RNA using pCe5S1 as a template (Honda et.al., 1986). Note that the 5'L transcription unit is not transcribed under our conditions; either the appropriate transcription factors are not active in our extract, or 5'L is transcribed by RNA pol II which is inactivated by the level of α -amanitin present in the reaction.

In order to identify the template sequences required for efficient 5S RNA transcription in this extract, we constructed a series of 5' and 3' deletion/substitution (D/S) derivatives of pCe5S1 using the directional deletion strategy of Henikoff (1984). The strategy and extent of the deletions obtained is shown in Figure 4.1.

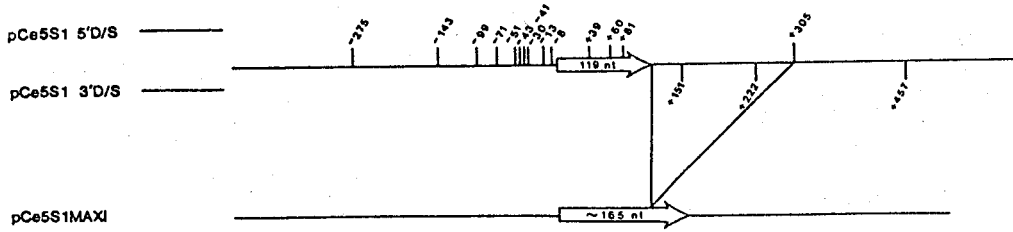
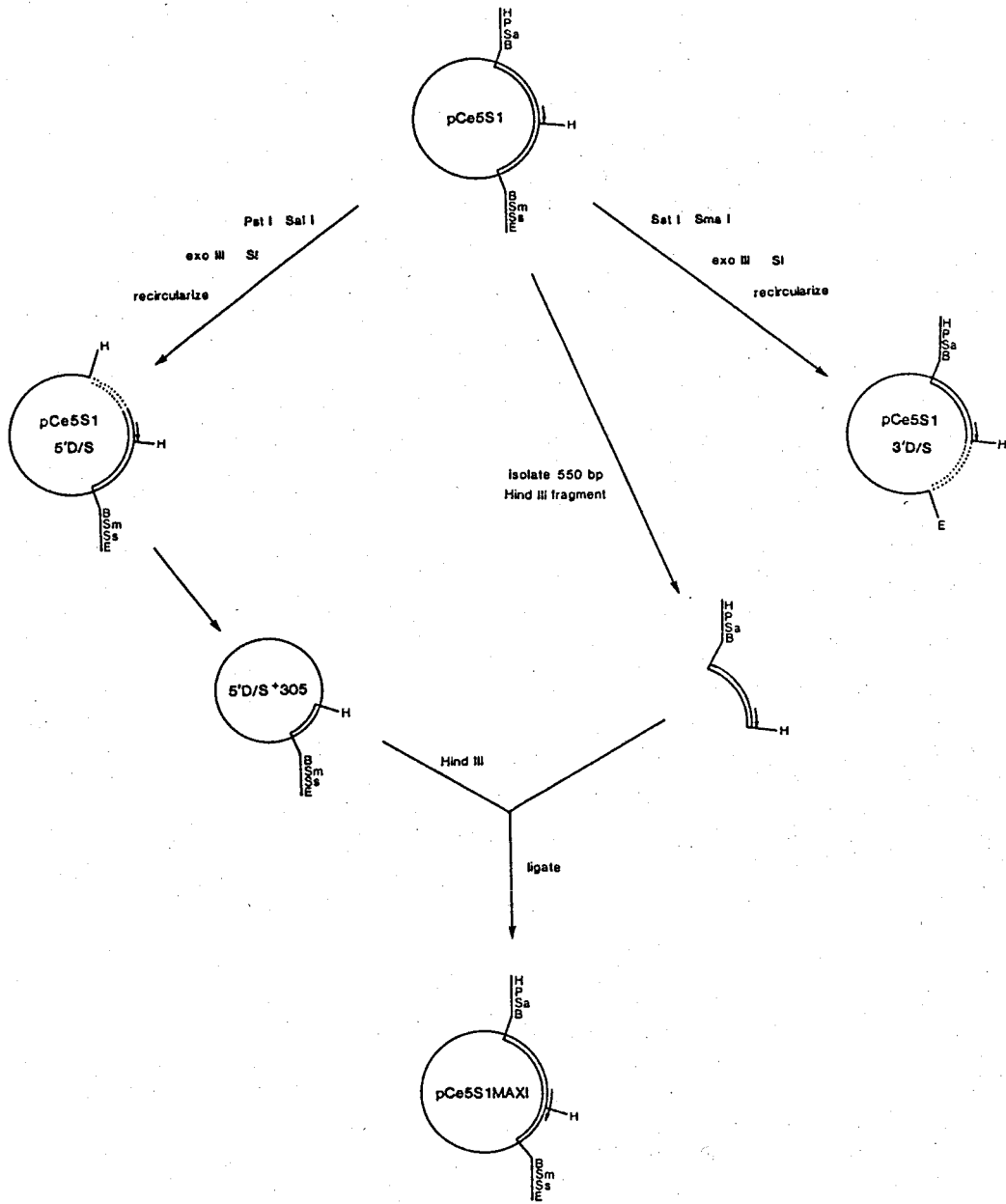
In order to quantitate 5S RNA transcription from each of these D/S templates, we constructed an internal reference template (pCe5SMAXI) in which 185 bp of 3' flanking sequence, including the RNA pol III termination signal, is deleted (see Figure 1). This template is efficiently transcribed in the C. elegans extract to produce a transcript doublet of approximately 165 nt, apparently terminating at an alternative T₄ signal in the

Figure 4.1: Construction and structure of pCe5S1 5' and 3'deletion/substitution templates, and the internal reference template, pCe5S1MAXI.

pCe5S1 (Nelson and Honda, 1985) was linearized with either Pst I and Sal I (5'D/S) or Sst I and Sma I (3'D/S), treated with exonuclease III and S1 nuclease, repaired with Klenow, recircularized and transformed into JM83. Individual isolates were characterized by restriction mapping and direct sequencing of the vector-insert junction.

pCe5S1MAXI was obtained by recloning the 550 bp Hind III fragment of pCe5S1 (pUC13 polylinker (Hind III to Bam HI plus 406 bp of 5' flanking sequence plus 5S RNA coding sequence, but lacking the 3' T₄ transcription termination signal) into the unique Hind III site of pCe5S1 5'D/S+305 which lacks all of these sequences and 185 bp of 3' flanking sequence. The internal deletion of 185 bp of 3' flanking sequence including the wild-type transcription termination signal results in the production of a longer transcript (approximately 165 bp) which appears to terminate at position +346.

The structures of pCe5S1 5'D/S and 3'D/S templates used in this work are shown in the line drawings.



substituted 3' flanking sequences. The efficiency of 5S RNA transcription from D/S templates was determined relative to pCe5SMAXI; total template concentration was maintained within the linear region of the [template] dependence curve (0.1 nM pCe5S1MAXI plus 0.1 nM D/S template). The results of these experiments are shown in Figure 4.2.

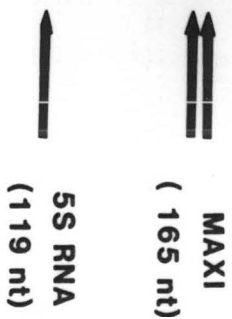
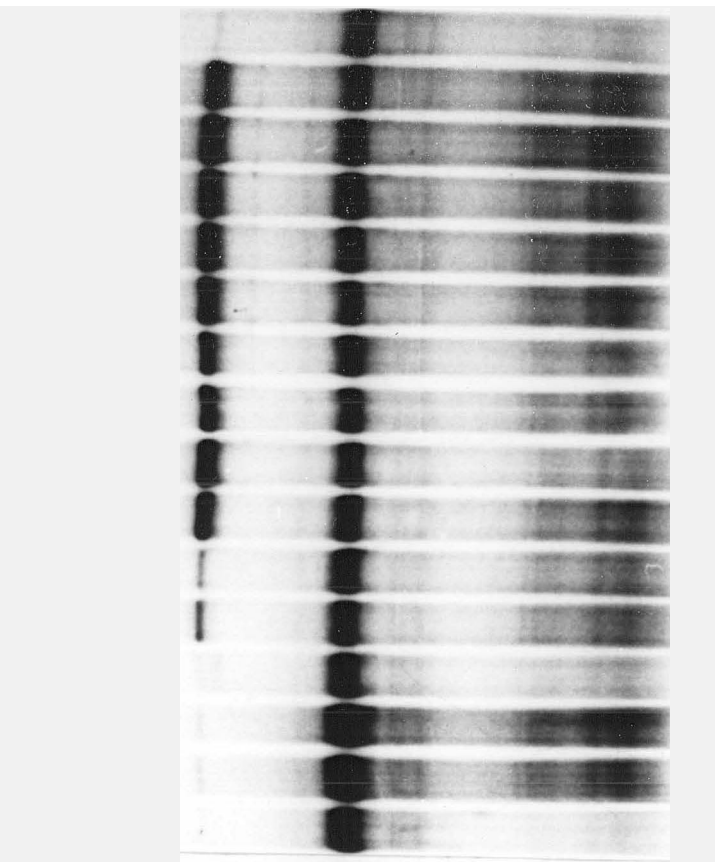
The deletion of 3' flanking sequences to within 30 bp of the 5S RNA coding sequence had no significant effect on 5S RNA transcription relative to the internal reference template. The deletion/substitution of 5' flanking sequences to within 30 bp of the 5S RNA coding region also has no dramatic effect on the efficiency of 5S RNA transcription. However, extending the 5' deletion/substitution to within 13 or 8 bp of the 5S RNA coding region results in a dramatic decrease in 5S RNA transcription to approximately 10% of the wild-type level. Deletion/substitutions extending into the 5S RNA coding region have additional effects on transcription. The deletion/substitution of the entire 5' flanking region and 39 bp of 5S RNA coding sequence does not appear to further reduce the efficiency of transcription, but the transcripts observed are heterogeneous in length. The two most extensive 5' D/S templates examined, affecting 60 and 81 bp of 5S RNA coding sequence, do not produce detectable transcripts.

Competition of pCe5S1 5'D/S templates for limiting extract components.

The transcription of eukaryotic 5S RNA is preceded by the

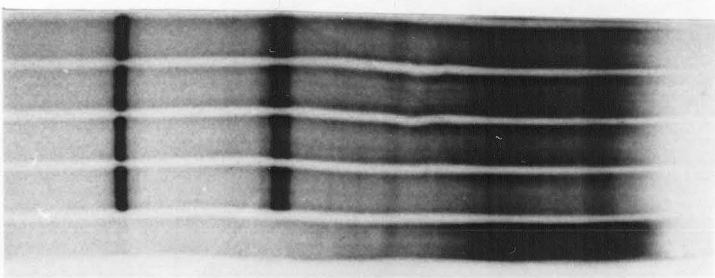
pCe5S1 5'D/S

pCe5S1 (wt)
5'D/S-275
5'D/S-143
5'D/S-99
5'D/S-71
5'D/S-51
5'D/S-43
5'D/S-41
5'D/S-30
5'D/S-13
5'D/S-8
5'D/S-39
5'D/S-60
5'D/S-81



pCe5S 3'D/S

3'D/S-151
3'D/S-222
3'D/S-457
pCe5S1 (wt)
pUC13



pCe5S1 DELETION TEMPLATES

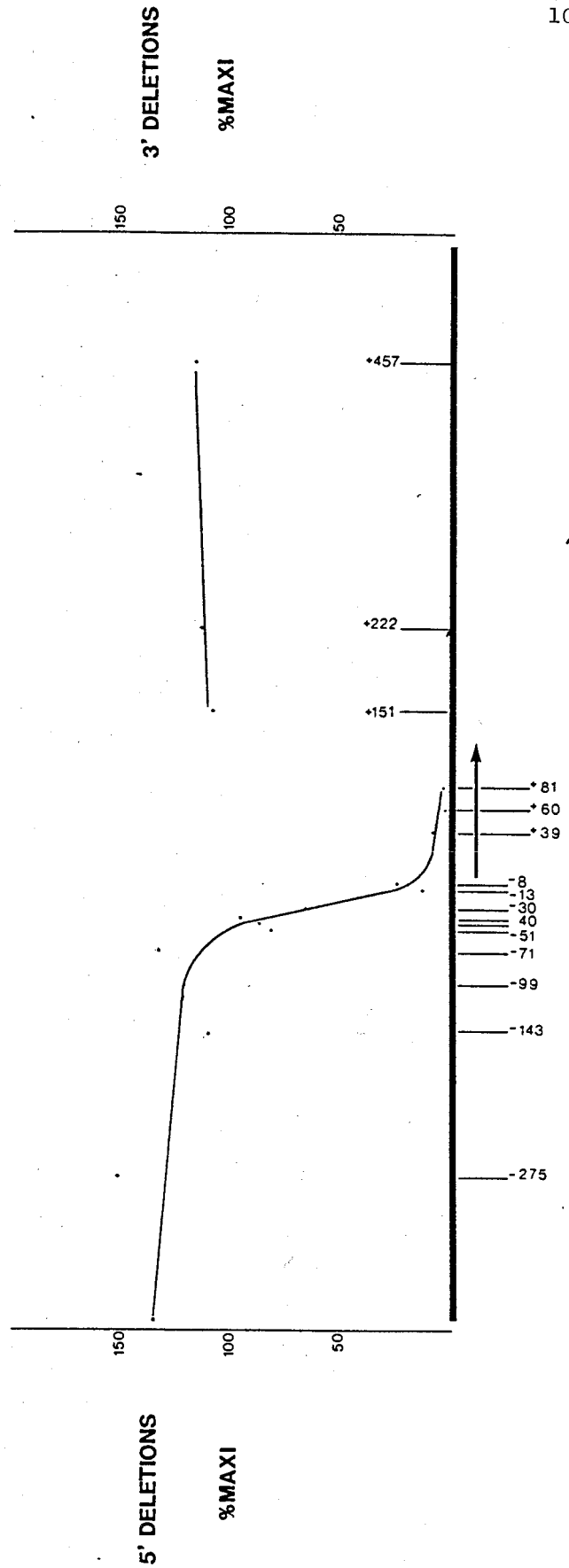
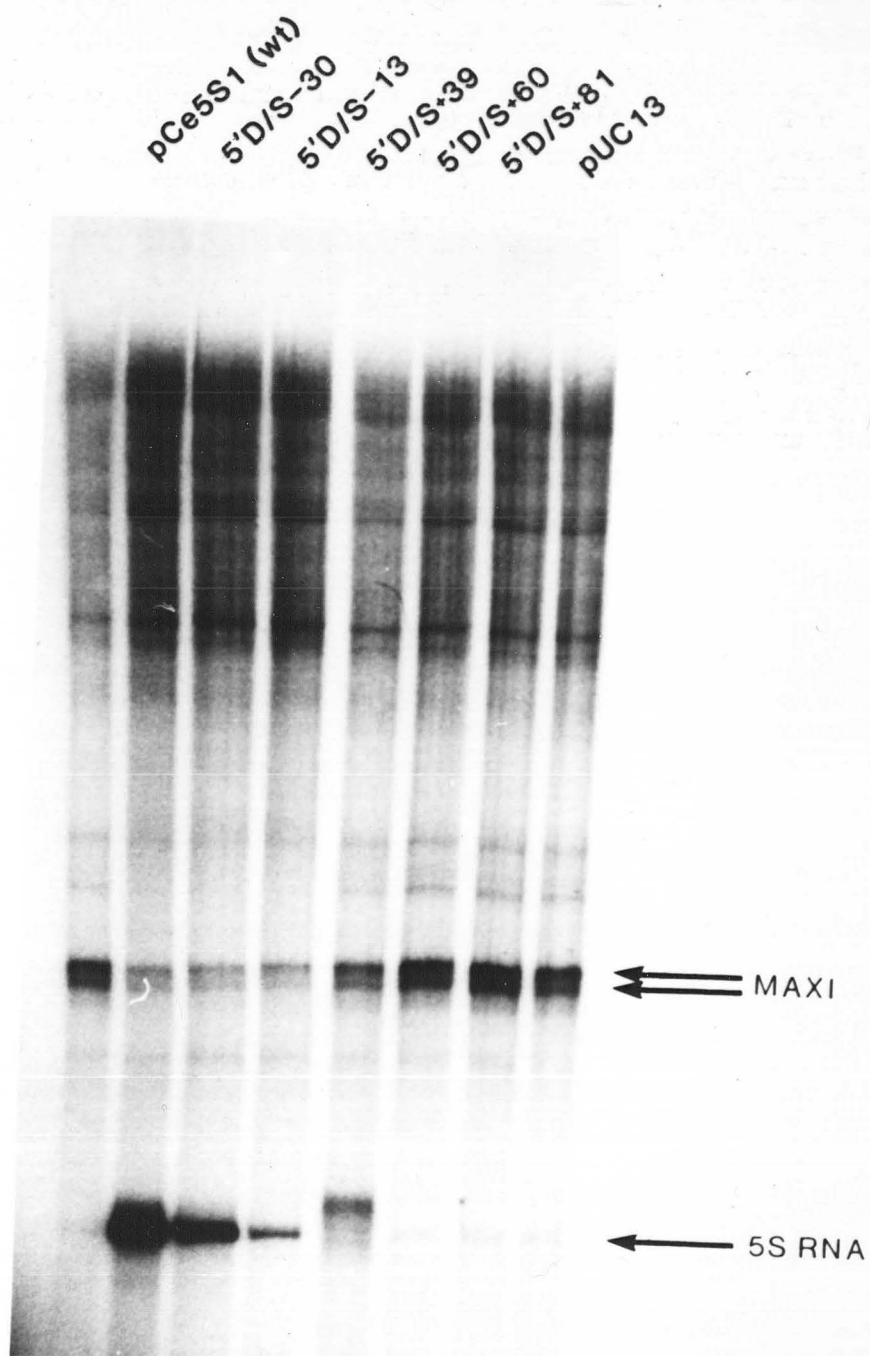


Figure 4.3: Competition of pCe5S1 5'D/S templates for limiting extract components.

pCe5S1 5'D/S templates (1 nM, 5X saturation), were preincubated with extract in the presence of ATP for 30 minutes at 20 °C, before addition of pCe5S1MAXI, and the remaining NTPs. The positions of the MAXI and 5S RNA transcripts are shown. The labels above each lane indicate the identity of the preincubated template.



formation of a stable transcription complex by the sequential binding of transcription factors TFIIIA, TFIIIC and TFIIIB to an internal control region located near the center of the gene (Lassar et.al., 1983). The stability of this transcription complex allows saturating amounts of a pre-incubated template to sequester limiting extract components thus reducing transcription of a subsequently added reference template (Wormington et.al., 1981).

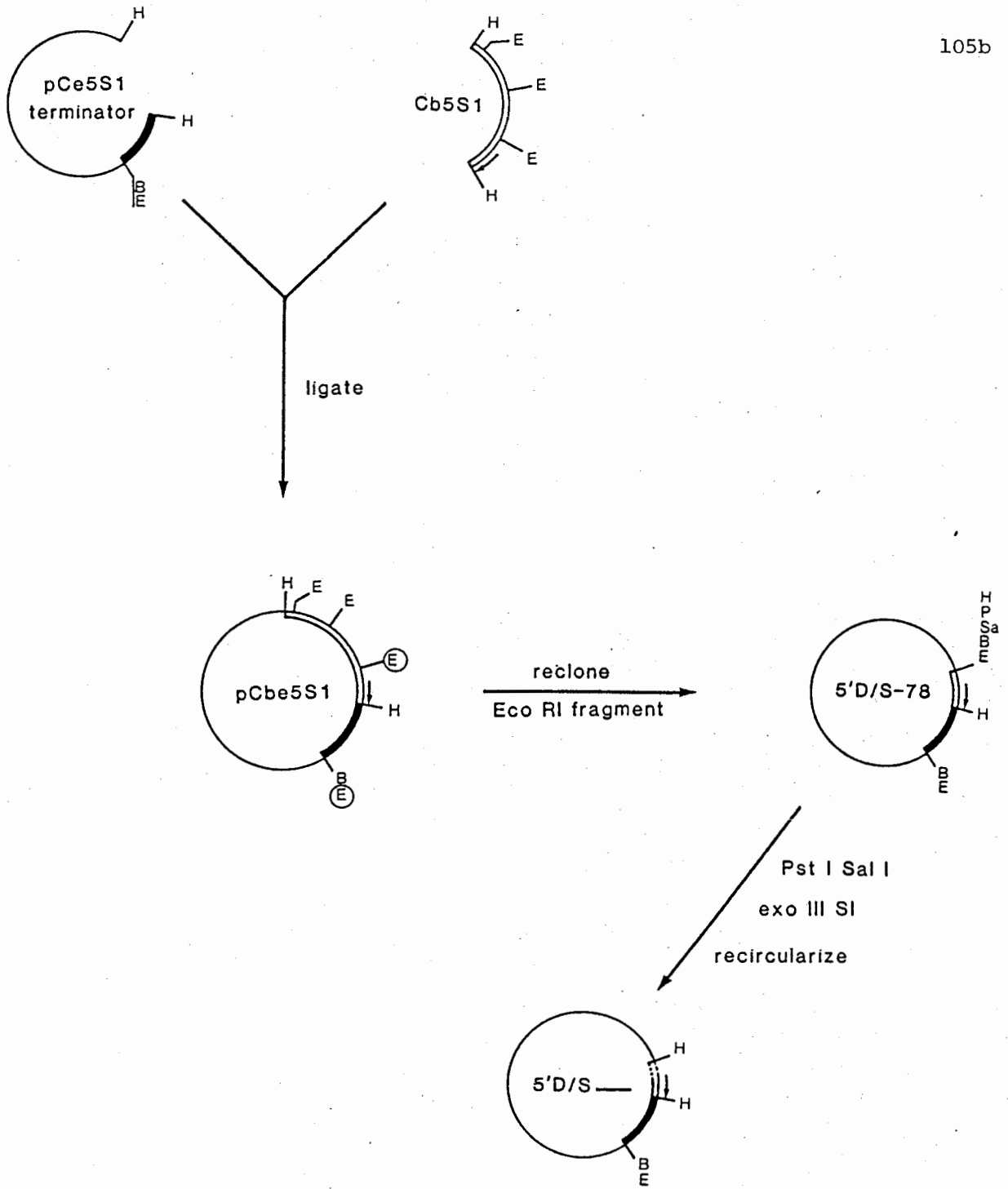
We have used a similar assay to examine the effect of 5' deletion/substitutions on the ability of the template to compete for limiting extract components (Figure 4.3). Under our conditions (pCe5S1 5'D/S template at five times saturation (1 nM) and preincubated with extract for 30 minutes), pCe5S1 and pCe5S1 5'D/S -30, -13 and -8 all reduce transcription from pCe5SMAXI. pCe5S1 5'D/S+39 also reduces pCe5S1MAXI transcription, but less efficiently than the others. pCe5S1 5'D/S+60 and +81, on the other hand, are not effective in reducing pCe5S1MAXI transcription.

5S RNA transcription from the C. briggsae 1 kb 5S DNA repeat requires a conserved 5' flanking sequence element

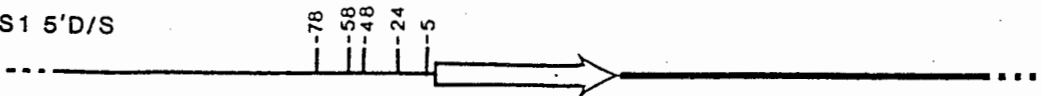
We have previously described the structure and organization of the two genomic repeat families encoding the 5S RNA of C. briggsae, a close relative of C. elegans' (Nelson and Honda, submitted). The 1 kb 5S DNA repeats of C. elegans and C. briggsae are clearly related, sharing perfectly conserved 5S RNA

Figure 4.4: Construction and structure of pCbe5S1 5'D/S templates.

The C. briggsae 1kb 5S DNA repeat (Hind III fragment) carried by pCb5S1 (Nelson and Honda, submitted) was recloned adjacent to the transcription termination signals carried by pCe5S1 (Hind III site). pCbe5S1 5'D/S-78 was obtained by recloning the 650 bp Eco RI fragment (circled sites in Figure) into pUC13. Additional 5'D/S templates were obtained by linearizing pCbe5S1 5'D/S-78 with Pst I and Sal I, treating with exonuclease III and S1 nuclease, recircularizing and transforming into JM83. The structures of the pCbe5S1 5'D/S templates used in this work are shown in the line drawing.



pCbe5S1 5'D/S



and 5' splice-leader RNA (5'L) coding sequences. For use as a transcription template, the C. briggsae 1 kb 5S DNA repeat (Hind III fragment) was recloned adjacent to the RNA pol III transcription termination signals of the C. elegans 5S DNA repeat (Figure 4.4). This template is efficiently transcribed in the C. elegans cell-free extract (see Figure 4.5).

The transcription results obtained for pCe5S1 5'D/S templates may reflect the participation of 5' flanking sequence elements in the transcription process, or may be the result of substituting wild-type sequences with interfering vector sequences. If 5' flanking sequences are indeed functionally important for the transcription process, they may be under selective pressure and exhibit inter-species conservation. Comparison of Ce5S1 and Cb5S1 sequences immediately upstream of the 5S RNA coding regions identifies a conserved sequence element, CACTTGTCATAG (-15 to -3 in Ce5S1) and CACTGATCCCTAG (-45 to -33 in Cb5S1), which may be required for efficient 5S RNA transcription as described above.

In order to determine whether this sequence is important for 5S RNA transcription from pCbe5S1, we constructed a set of 5' deletion derivatives of this template (see Figure 4.4) and examined their ability to program 5S RNA transcription. The results of this analysis are shown in Figure 4.5. The deletion/substitution of 5' flanking sequences immediately upstream of the conserved sequence do not significantly affect transcription efficiency, while 5' deletion/substitutions which

Figure 4.5: Transcription of pCbe5S1 5'D/S templates.

a) Autoradiograph of transcription results.

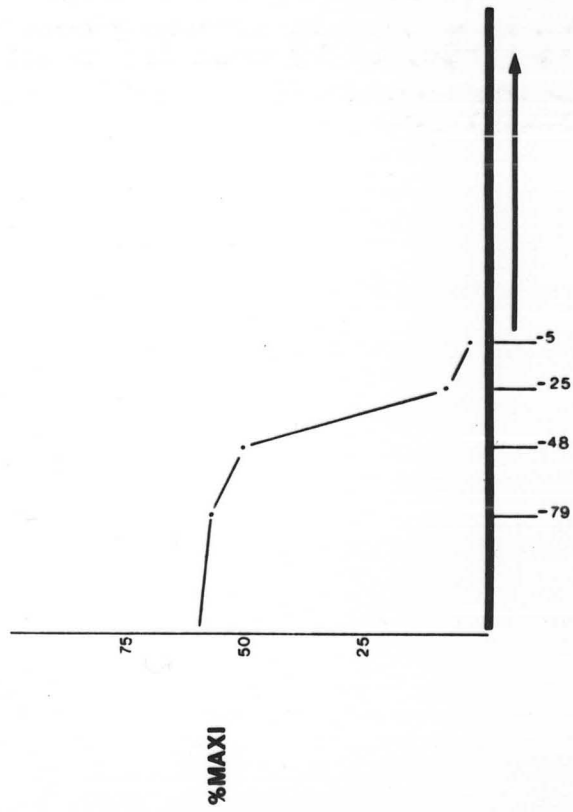
Transcription reactions all contained 0.1 nM pCe5S1MAXI as an internal reference template, 0.1 nM pCbe5S1 5'D/S template, and [DNA] held at 8 ug/ml with pUC13. The identity of each D/S template and the positions of the MAXI and 5S RNA transcripts are indicated.

b) Graphical representation of transcription results.

5S RNA transcription was quantitated relative to the internal reference MAXI transcript by Cerenkov counting excised bands, or by scanning densitometry of the autoradiograph.

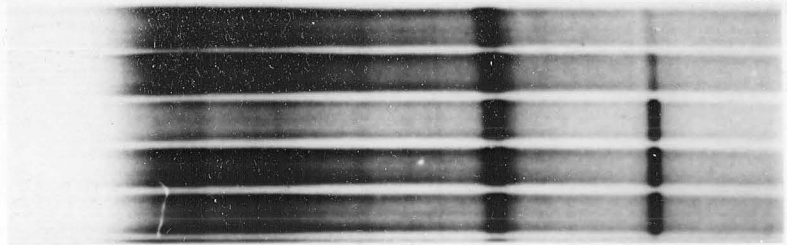
pCbe5S1 5'D/S

(a)



(b)

pCbe5S1
5'D/S-78
5'D/S-48
5'D/S-25
5'D/S-5



MAXI

5S RNA

107b

remove this conserved element decrease 5S RNA transcription to approximately 10% wild-type efficiency.

How does the 5' flanking sequence promote 5S RNA transcription?

5' deletion/substitutions affecting an upstream conserved sequence element decrease the efficiency of 5S RNA transcription from both *C. elegans* and *C. briggsae* 1 kb 5S DNA repeats to 10% of wild-type levels. Such upstream sequence elements might influence transcription efficiency in a variety of ways.

Eukaryotic enhancers are short sequence elements which influence transcription efficiency in a position and orientation independent fashion. The inclusion of such an enhancer element on a plasmid template might therefore be expected to activate transcription from a previously inactive transcription unit.

In order to examine the ability of this putative 5' regulatory element to activate efficient transcription of 5S RNA transcription from inefficient pCe5S1 5'D/S and pCbe5S1 5'D/S templates, we have cloned the MAXI-transcript encoding 800 bp Bam HI fragment adjacent to the D/S templates (Figure 4.6) and examined their transcription patterns. The efficiency of 5S RNA transcription from the D/S templates is not affected by the presence of the efficiently transcribed MAXI transcription unit on the same plasmid (Figure 4.6).

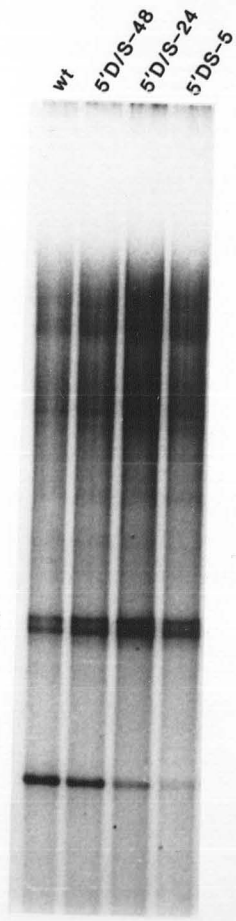
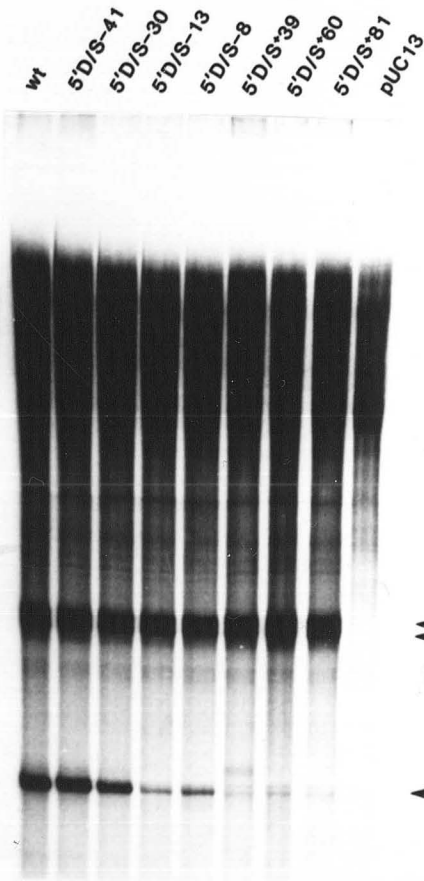
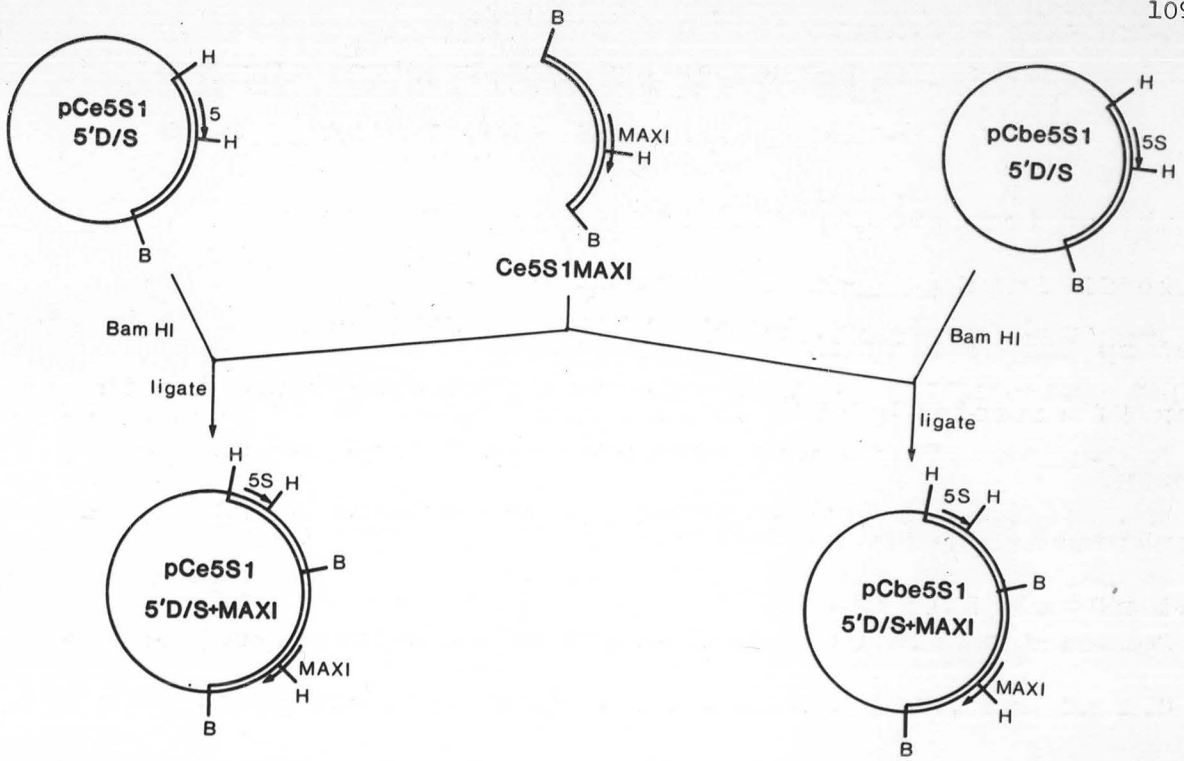
DISCUSSION

The results presented here identify the template sequences

Figure 4.6: Construction and transcription of pCe5S1
5'D/S____+MAXI and pCbe5S1 5'D/S____+MAXI
templates.

Combined templates, pCe5S1 5'D/S____+MAXI and
pCbe5S1 5'D/S____+MAXI, were constructed by recloning the
795 bp Bam HI fragment from pCe5S1MAXI into the unique
Bam HI sites of pCe5S1 5'D/S____ and pCbe5S1 5'D/S____ as
shown.

Combined templates were transcribed at a [template] of
0.1 nM and [DNA] was held at 8 ug/ml with ccc pUC13.



which influence 5S RNA transcription in a C. elegans cell-free extract.

A large portion of the wild-type template appears to be completely dispensable for efficient 5S RNA transcription. 5' deletion/substitutions affecting sequences more than 30 bp upstream of the C. elegans and 45 bp upstream of the C. briggsae 5S RNA coding region have no significant effect on 5S RNA transcription. Similarly, 3' deletion/substitutions affecting the entire C. elegans 3' flanking sequence are transcribed at approximately wild-type efficiency.

The putative RNA pol III termination signal sequence, T₄, immediately adjacent to the C. elegans 5S RNA coding region qualitatively affects transcription. Deletion/substitution of this sequence element results in the appearance of higher molecular weight transcripts which presumably arise by termination at analogous sites in the substituted sequence. This characteristic was exploited in the construction of pCe5S1MAXI for use as an internal reference template in quantitative transcription assays.

The template sequences immediately upstream of the 5S RNA coding region dramatically affect the efficiency of 5S RNA transcription. 5' deletion/substitutions affecting a conserved sequence element centered 7 bp upstream of the C. elegans 5S RNA coding sequence, or 38 bp upstream of the C. briggsae 5S RNA coding region, decrease the efficiency of 5S RNA transcription to approximately 10% wild-type levels. The conservation of this 12

bp sequence element and the dramatic decrease in 5S RNA transcription efficiency when it is altered, argues that this sequence is functionally important. A closely related sequence (AACTTCTCGCTAG, 9/13 matches) is also observed centered 25 bp upstream of the 5' end of the 18S coding sequence of the C. elegans large ribosomal DNA repeat (Ellis et.al., 1986). These observations suggest that this sequence element may play a role in the coordination of rRNA transcription by RNA polymerases I and III in C. elegans.

How might this sequence function to modulate RNA polymerase III transcription? It does not appear to be an essential feature of the template as 5S RNA is still transcribed, albeit at 10% the wild-type level, in its absence. It is unlikely to function as an enhancer element, as its presence on a plasmid does not restore 5S RNA transcription from an inefficient D/S template located approximately 1 kb away on the same plasmid.

We have previously shown that the C. briggsae 0.7 kb 5S DNA repeat is efficiently transcribed in the C. elegans cell-free extract (Nelson and Honda, submitted). This repeat does not contain 5' flanking sequences related to this sequence element, but is very rich in A and T residues. This feature resembles that of several other eukaryotic 5S RNA genes which require 5' flanking sequences for their efficient transcription in vitro (Morton and Sprague, 1984; Selker et.al., 1986, Tyler, 1987, Garcia et.al., 1987). In each case, the sequence TATA is found centered approximately 25 bp upstream of the 5S RNA transcription

initiation site.

These results indicate that the conserved sequence element required for efficient 5S RNA transcription may act to facilitate RNA polymerase III entry. Entry could be enhanced by a sequence-specific mechanism, or by other sequences which perhaps strand-separate easily in a manner reminiscent of the formation of the open initiation complex described in E. coli (Travers, 1987).

Finally, template sequences between +39 and +60 are absolutely required for 5S RNA transcription. 5' deletion/substitutions extending to position +60 of the 5S RNA coding sequence or beyond are not detectably transcribed. In addition, these templates do not appear to compete for limiting extract components. These results are consistent with a general requirement for internal promoter sequences in the formation of a stable RNA polymerase III transcription complex on eukaryotic 5S RNA genes.

REFERENCES

- Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
I. the 5' border of the region.
Cell 19 27-35.
- Bogenhagen, D.F., Wormington, W.M. and Brown, D.D. (1982).
Stable transcription complexes of *Xenopus* 5S RNA genes: a means to maintain the differentiated state.
Cell 28 413-421.
- Ellis, R.E., Sulston, J.E. and Coulson, A.R. (1986).
The rDNA of *C. elegans*: sequence and structure.
Nuc. Acid. Res. 14 2345-2364.
- Engelke, D.R., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1980).
Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes.
Cell 19 717-728.
- Garcia, A.D., O'Connell, A.M. and Sharp, S.J. (1987).
Formation of an active transcription complex in the *Drosophila melanogaster* 5S RNA gene is dependent on an upstream region.
Mol. Cell. Biol. 7 2046-2051.
- Hattori, M. and Sakaki, Y. (1986).
Dideoxy sequencing method using denatured plasmid templates.
Anal. Biochem. 152 232-238.
- Henikoff, S. (1984).
Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing.
Gene 28 351-359.
- Honda, B.M., Devlin, R.H., Nelson, D.W. and Khosla, M. (1986)
Transcription of class III genes in cell-free extracts from the nematode *Caenorhabditis elegans*.
Nuc. Acid. Res. 14 869-881.
- Krause, M. and Hirsh, D. (1987).
A trans-spliced leader sequence on actin mRNA in *C. elegans*.
Cell 49 753-761.
- Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983).
Transcription of class III genes: formation of preinitiation complexes.
Science 222 740-748.

- Morzycka-Wroblewska, E., Selker, E.U., Stevens, J.N. and Metzenberg, R.L. (1985).
Concerted evolution of dispersed Neurospora crassa 5S RNA genes:
Pattern of sequence conservation between allelic and nonallelic genes.
Mol. Cell. Biol. 5 46-51.
- Morton, D.G. and Sprague, K.U. (1984)
in vitro transcription of a silkworm 5S RNA gene requires an upstream signal.
Proc. Natl. Acad. Sci. 81 5519-5522.
- Nelson, D.W. and Honda, B.M. (1985).
Genes encoding 5S ribosomal RNA of the nematode Caenorhabditis elegans.
Gene 38 245-251.
- Rubacha, A., Sumner III, W., Richter, L and Beckingham, K. (1984).
Conserved 5' flank homologies in dipteran 5S RNA genes that would function on 'A' form DNA.
Nuc. Acid. Res. 12 8193-8207.
- Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
I. the 5' border of the region.
Cell 19 13-25.
- Selker, E.U., Morzycka-Wroblewska, E., Stevens, J.N. and Metzenberg, R.L. (1986).
An upstream signal is required for in vitro transcription of Neurospora 5S RNA genes.
Mol. Gen. Genet. 205 189-192.
- Travers, A.A. (1987).
Structure and function of E. coli promoter DNA.
CRC Crit. Rev. Biochem. 22 181-219.
- Tyler, B.M. (1987).
Transcription of Neurospora crassa 5S RNA genes requires a TATA box and three internal elements.
J. Mol. Biol. 196 801-811.
- Wormington, W.M., Bogenhagen, D.F. and Brown, D.D. (1981)
A quantitative assay for Xenopus 5S RNA gene transcription in vitro.
Cell 24 809-817.

GENERAL DISCUSSION

The object of this thesis has been to establish the application of cell-free extracts to the study of C. elegans transcription. In order to obtain a suitable template for the development of a transcriptionally active cell-free extract from C. elegans, I undertook the characterization of C. elegans 5S RNA genes. As eukaryotic 5S RNA genes also represent an attractive model system for the study of multigene family evolution, I extended my analysis to the 5S RNA genes of a closely related nematode species, C. briggsae. Finally, I used a cell-free extract of C. elegans developed in this lab to examine the template requirements of 5S RNA transcription in vitro.

For the purpose of this discussion, I have divided the results of these studies into two sections, the first concerning the evolutionary relationship between C. elegans and C. briggsae 5S RNA genes, and the second concerning the in vitro transcription of 5S RNA.

C. elegans and C. briggsae are morphologically indistinguishable but genetically distinct nematode species. Previous studies on the evolutionary relationship between C. elegans and C. briggsae have relied on using random C. elegans genomic sequences to probe the C. briggsae genome for homologous sequences (Hirsh et.al., 1979; Emmons et.al., 1979), or on electrophoretic mobility analysis of homologous enzymes of these two species (Butler et.al., 1981). The results of these studies suggest that these species diverged on the order of 10 to 100

million years ago, and that the genomes of these two species are more divergent than are their proteins. Recent work exploiting the conservation of C. elegans and C. briggsae genomic sequences as a tool for the identification of genomic transcription units indicates that only those sequences found in the cellular RNA pool are conserved; intron sequences diverge rapidly in comparison to protein coding sequences (T.S. Snutch, S. Prasad, M. Hechtl and D.L. Baillie, personal communication). This difference in sequence conservation presumably reflects selective pressures which act to maintain gene function in the face of random mutagenic events. Interspecies sequence conservation should therefore be applicable to the identification of any functionally important sequence, regardless of what that function might be (Batterham et.al., 1983; Bodmer and Ashburner, 1984; Henikoff, 1987).

The structure and genomic organization of C. elegans and C. briggsae 5S RNA genes is typical of higher eukaryotes (Long and Dawid, 1980). In C. elegans, 5S RNA is encoded on a 1 kb genomic repeat unit. The haploid genome contains approximately 110 homogeneous copies of this repeat unit, which are organized into a single tandem cluster. In addition to these tandemly organized 5S RNA genes, there appear to be a few copies of non-tandemly organized genomic sequences homologous to the 1 kb 5S DNA repeat. Two distinct genomic repeat units, 1 kb and 0.7 kb in size, appear to represent functional 5S RNA genes in C. briggsae. The haploid genome contains approximately 65 copies of the 1 kb

repeat and 20 copies of the 0.7 kb repeat. Each repeat family is organized in separate tandem clusters in the genome. There appear to be few if any copies of dispersed genomic sequences homologous to either the 1 kb or 0.7 kb 5S DNA repeats of C. briggsae.

The 5S DNA repeat families of C. elegans and C. briggsae can be divided into two classes on the basis of relative repeat sequence identity; the 1 kb repeats share two blocks of conserved sequence in the same relative orientation, while in the 0.7 kb repeat these sequences appear in the opposite orientation. The conservation of two discrete blocks of sequence between the C. elegans and C. briggsae 1 kb 5S DNA repeats is evidence for their divergence from a single ancestral repeat.

Surprisingly, each highly conserved region represents a discrete transcription unit, one encoding 5S RNA and the other an abundant 100 nt primary transcript (5'L) which participates in a trans-splicing process recently described in C. elegans (Krause and Hirsh, 1987). An analogous transcript has been identified in C. briggsae, and presumably serves the same function as it does in C. elegans. The divergent orientation of these two transcription units serves to divide the 1 kb repeat into four regions: the 5S RNA coding region (120 nt), the 3' intergenic region (580 bp and 505 bp), the 5'L coding region (100 bp) and the 5' intergenic region (176 bp and 217 bp). Only the 5S RNA and 5'L coding sequences are highly conserved. The 3' intergenic region in both repeats is AT rich, containing numerous runs of A

and T residues. The 5' intergenic region is GC rich with short runs of conserved sequence. This conservation of general sequence features rather than specific sequences may reflect a lack of rigid sequence-specificity in the initiation and termination of transcription from these two transcription units.

The 0.7 kb 5S DNA repeat of C. briggsae is clearly related to the 1 kb 5S DNA repeat class described above. In addition to the perfectly conserved 5S RNA coding region, the 5'L coding sequence is partially conserved, albeit in the complementary orientation. The change in relative orientation of these two conserved sequences suggests that the 0.7 kb repeat diverged from the 1 kb ancestral repeat via a series of events, one of which involved the inversion of a portion of the repeat. The 0.7 kb repeat appears to represent a functional 5S RNA gene, but does not produce a detectable 5'L transcript.

The genomic organization of C. elegans and C. briggsae 5S RNA genes into homogeneous repeat-specific tandem clusters may have several important functional consequences. This organization is thought to provide a substrate for the establishment and maintenance of repeat homogeneity and copy number by the mechanisms of unequal cross-over and gene conversion (Arnheim, 1983). The homogeneity of these repeat families provides evidence for an efficient correction mechanism in C. elegans and C. briggsae. The number of 5S RNA genes in both species is also similar, suggesting that 100 may be optimal for normal development. The inactivation of the 5'L

transcription unit on the C. briggsae 0.7 kb 5S DNA repeat suggests that the 5'L transcript may be overproduced in vivo, in agreement with the observation that the 5'L transcript is an abundant cellular RNA, although it only appears on approximately 10% of the poly-A+ mRNA population.

The organization of a multigene family into a large tandem array also has relevance to the coordinate regulation of their expression. Such tandem arrays may facilitate the formation of large chromatin domains in which a repetitive chromatin structure defines the transcriptional status of the multigene family as a whole.

In C. elegans, we were fortunate to find a strain-specific RFLD associated with the 5S DNA tandem cluster. The genetic mapping of this RFLD places this 5S DNA cluster between the visible markers unc-76 and dpy-21 on the right arm of linkage group V. This localization of the 5S RNA genes to a specific position on the genetic map may facilitate the use of classical genetic analysis to study 5S RNA gene organization and expression.

In situ hybridization of the C. elegans 5S DNA repeat to metaphase chromosomes identifies a single site of hybridization to a Hoescht 33258 dark band approximately two thirds of the way along an otherwise unidentified chromosome (Albertson, 1984). The binding of Hoescht 33258 to DNA is thought to be somewhat sequence specific, favoring AT rich sequences (Martin and Holmes, 1983; Harshman and Dervan, 1985). Given that such sequences are

not uncommon in the 5S DNA repeat, it is surprising that the site of 5S DNA hybridization binds the dye poorly. Perhaps the chromatin structure in this region is responsible for this poor dye binding phenotype.

We have not been able to define the genetic location of the two C. briggsae 5S DNA tandem clusters, or their positions relative to one another in the genome. The application of in situ hybridization techniques (Albertson, 1984) or the use of pulsed field gel electrophoresis (Schwartz and Cantor, 1984; Carle et.al., 1986) with appropriate restriction digests may be useful to resolve this level of organization.

The deletion/substitution analysis of the template sequences required for C. elegans transcription described here identify the regions of the C. elegans and C. briggsae 5S DNA repeats which affect 5S RNA transcription in a C. elegans cell-free extract.

5' flanking sequences appear to modulate the efficiency of 5S RNA transcription. The C. elegans and C. briggsae 1 kb 5S DNA repeats each contain a conserved 13 bp sequence element which appears to be required for efficient 5S RNA transcription. The position of this 13 bp element appears to be variable (-15 to -3 in Ce5S1 and -45 to -33 in Cb5S1) but its action is short-range; its presence on a plasmid does not activate transcription from a 5S RNA coding region 1 kb away. The deletion of this conserved sequence element does not affect the ability of a template to compete for limiting extract components.

The 0.7 kb 5S DNA repeat of C. briggsae is also efficiently transcribed in the C. elegans cell-free extract. This repeat does not contain a copy of the 13 bp conserved sequence element described above, but does have a 'TATA' element centered at position -25 relative to the initiation site. This sequence has also been found at this position upstream of the 5S RNA coding region in several other species which require intact 5' flanking sequences for 5S RNA transcription in vitro (Morton and Sprague, 1984; Selker et.al., 1986; Tyler, 1987; Garcia et.al., 1987).

The observation that the C. briggsae 0.7 kb 5S DNA repeat is efficiently transcribed in vitro, despite the lack of the conserved sequence element found upstream of the 5S RNA coding region in the C. elegans and C. briggsae 1 kb 5S DNA repeats, suggests that the modulation of 5S RNA transcription may not be rigidly sequence-specific. Instead, these sequences may modulate the initiation of transcription in a general way, perhaps analogous to the function of the prokaryotic -10 box in the transition of the closed to open initiation complex forms (Travers, 1987). The 12 bp conserved sequence may act as a general polymerase entry site, a function which an AT rich sequence can also perform by virtue of the ease with which it strand separates. It is worth noting that a closely related 12 bp sequence is found centered approximately 25 bp upstream of the 18S rRNA coding sequence of the large rDNA repeat of C. elegans (Ellis et.al., 1986). This observation suggests that rRNA transcription may be coordinately regulated through the action of

this conserved sequence element.

In addition to the modulation of transcription efficiency by 5' flanking sequence, the local sequence characteristics surrounding the transcription initiation site appear to influence the selection of the precise position at which initiation occurs (Sakonju et.al., 1980). This is reflected in the heterogeneity in transcripts derived from the deletion/substitution template pCe5S1 D/S +39.

Although not conclusive, the lack of detectable 5S RNA transcripts from deletion/substitution templates pCe5S1 D/S +60 and +81, and their inability to compete for limiting extract components suggests that 5S RNA transcription is directed by an internal control region analogous to that described in X. laevis (Sakonju et.al., 1980; Bogenhagen et.al., 1980).

Finally, the T₄ sequence located immediately adjacent to the 5S RNA coding sequence is responsible for the termination of 5S RNA transcription. The remainder of the 5S DNA repeat does not appear to affect 5S RNA transcription in vitro.

In general, the sequence requirements of C. elegans 5S RNA transcription described above correlate well with other well-characterized systems. X. laevis 5S RNA transcription has been shown to be insensitive to alteration of the 5' flanking sequences (Sakonju et.al., 1980). However, 5S RNA transcription in B. mori (Morton and Sprague, 1984), N. crassa (Selker et.al., 1986; Tyler, 1987) and D. melanogaster (Garcia et.al., 1987)

cell-free extracts is dependant on the presence of a 'TATA' sequence centered 25 bp upstream of the transcription initiation site. One possible explanation for this apparent contradiction is that X. laevis developmentally regulates 5S RNA transcription by activating a large oocyte-specific gene family (Korn, 1982), while the 5S RNA genes of B. mori (Morton and Sprague, 1982) and N. crassa (Selker et.al., 1981) are dispersed throughout the genome. D. melanogaster, like C. elegans contains a single 5S RNA gene family tandemly clustered at a single genomic site (Procunier and Tartoff, 1975; Procunier and Dunn, 1978). The developmental expression of these 5S RNA genes may be regulated by modulating the transcription of these genes as a whole rather than activating the transcription of a separate class of genes. This modulation may be effected through the 5' flanking sequences required for 5S RNA transcription in vitro.

The results presented herein define a model system for the biochemical dissection of RNA polymerase III transcription in C. elegans. Future research may proceed down many different avenues, a few of which deserve particular attention.

The deletion/substitution analysis of the template requirements of 5S RNA transcription is incomplete. The D/S templates used to identify the 5' flanking sequence which modulates 5S RNA transcription do not unequivocally rule out the inhibition of transcription by vector sequences (Peterson et.al., 1987). These templates should be reconstructed with random

sequences replacing the wild-type template sequences in order to demonstrate that approaching vector sequences are not 'poisoning' 5S RNA transcription. The fortuitous replacement by AT rich sequences might be expected to promote transcription back to wild-type levels as observed with the *C. briggsae* 0.7 kb 5S DNA template.

The conserved 12 bp 5' sequence element, if indeed responsible for modulating 5S RNA transcription, can be inactivated by altering the first 2 bp (ie: pCe5S1 5'D/S -13).. The use of linker-scanning or point mutations affecting this region may provide further information regarding the sequence specificity of this modulating function.

The ability of a preincubated template to sequester limiting extract components, thereby reducing transcription from a subsequently added reference template, reflects the stability of the 5S RNA transcription complex. This stability can be exploited in the characterization of sequence-specific template-protein interactions *in vitro*. The development of a 'gel mobility shift assay' (Fried and Crothers, 1981), for example, would facilitate the identification of the (as yet putative) internal control region and the protein-binding capacity of the 5' flanking sequence. This technique also allows the separation of free DNA from the stable complex, thereby simplifying the characterization of protein-DNA interaction by DNase footprinting analysis (Galas and Schmitz, 1978).

While the fractionation of crude extracts is a tedious

process, the exploitation of sequence-specific DNA-protein interactions by 'duplex oligonucleotide affinity chromatography' (Kadonaga and Tjian, 1986) and use of the rapid 'gel mobility shift assay' should simplify this task. The fractionation of the crude extract into its components would permit a more detailed analysis of their interaction and function in the formation of the stable transcription complex itself.

The isolation and characterization of the genomic sequences and cDNA encoding Xenopus TFIIIA (Ginsberg et.al., 1984; Tso et.al., 1986) has contributed to our understanding of the molecular nature of its interaction with the internal control region and has defined a new class of DNA-binding structures, the Zn²⁺-binding finger domain (Klug and Rhodes, 1987). Previous results (Honda et.al., 1986) show that the C. elegans 5S DNA repeat is not efficiently transcribed in a Xenopus cell-free extract (and vice versa), presumably because of multiple nucleotide changes in the essential internal control region (Honda et.al., 1986). Does the C. elegans analog of TFIIIA have a structure analogous to that of Xenopus? This question might be addressed using the heterologous cDNA to probe for homologous sequences in the C. elegans genome.

While these experiments provide valuable insight into the mechanisms of 5S RNA transcription, they do not accurately reflect the in vivo situation. Of particular interest is the poor in situ staining of 5S DNA by Hoescht 33258. This dye is known to bind specifically to DNA sequences containing four

adjacent A·T base pairs (Martin and Holmes, 1983; Harshman and Dervan, 1985). This sequence feature is common in the AT rich intergenic region of the C. elegans 5S DNA repeat. The poor staining by Hoescht 33258 might therefore reflect a unique chromatin organization over the 5S DNA tandem cluster. This problem may be examined by characterizing the nuclease sensitivity of 5S DNA in isolated nuclei.

Finally, the C. elegans 5S DNA repeat contains two divergent transcription units, both of which are actively transcribed in vivo. It is difficult to see how the transcription of these two genes, which are separated by only 175 bp of 5' intergenic sequence, could be completely independent of one another. In our hands, 5'L is not transcribed in the C. elegans cell-free extract. This might be a result of the inhibition of RNA polymerase II by the levels of α -amanitin in the reaction, or to loss of the appropriate transcription factors during extract preparation. Attempts should be made to develop a cell-free extract competent in the transcription of RNA polymerase II genes. Alternatively, the chromatin organization studies mentioned earlier may provide information regarding the conformation of both the 5S and 5'L transcription units and the 5' intergenic region separating them.

REFERENCES

- Albertson, D.G. (1984).
Localization of the ribosomal genes in Caenorhabditis elegans by in situ hybridization using biotin-labeled probes.
EMBO J. 3 1227-1234.
- Arnheim, N. (1983).
Concerted evolution of multigene families.
in: Evolution of Genes and Proteins. Edited by M. Nei and R.K. Koehn. Sinauer Associates, Inc, Sunderland, MA. p38-61.
- Batterham, P., Lovett, J.A., Starmer, W.T. and Sullivan, D.T. (1983).
Differential regulation of duplicate alcohol dehydrogenase genes in Drosophila mojavensis.
Dev. Biol. 96 346-354.
- Bodmer, M. and Ashburner, M. (1984).
Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila.
Nature, 309 425-429.
- Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980)
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
II. The 3' border of the region.
Cell 19 27-35.
- Butler, M.H., Wall, S.M., Luehrsen, K.R., Fox, G.E. and Hecht, R.M. (1981).
Molecular relationships between closely related strains and species of nematodes.
J. Mol. Evol. 18-23.
- Carle G.F., Frank M. and Olson, M.V. (1986).
Electrophoretic separations of large DNA molecules by periodic inversion of the electric field.
Science 232 65-68.
- Ellis, R.E., Sulston, J.E. and Coulson, A.R. (1986).
The rRNA of C. elegans: sequence and structure.
Nuc. Acid. Res. 14 2345-2364.
- Emmons, S.W., Klass, M.R. and Hirsh, D. (1979).
Analysis of the constancy of DNA sequences during development and evolution of the nematode Caenorhabditis elegans.
Proc. Natl. Acad. Sci. 76 1333-1337.

- Fried, M. and Crothers, D.M. (1981).
Equilibria and kinetics of Lac repressor-operator interactions by polyacrylamide gel electrophoresis.
Nuc. Acid. Res. 9 6505-6525.
- Galas, D.J. and Schmitz, A. (1978).
DNase footprinting: A simple method for the detection of protein-DNA binding specificity.
Nuc. Acid. Res. 5 3157-3170.
- Garcia, A.D., O'Connell, A.M. and Sharp, S.J. (1987).
Formation of an active transcription complex in the Drosophila melanogaster 5S RNA gene is dependent on an upstream region.
Mol. Cell. Biol. 7 2046-2051.
- Ginsberg, A.M., King, B.O. and Roeder, R.G. (1984)
Xenopus 5S gene transcription factor TFIIIA:
characterization of a cDNA clone and measurement of RNA levels throughout development.
Cell 39 479-489.
- Harshman, K.D. and Dervan, P.B. (1985).
Molecular recognition of B-DNA by Hoechst 33258.
Nuc. Acid. Res. 13 4825-4835.
- Henikoff, S. and Eghtedarzadeh, M.K. (1987).
Conserved arrangement of nested genes at the Drosophila GART locus.
Genetics 117 711-725.
- Hirsh, D., Emmons, S.W., Files, J.G. and Klass, M.R. (1979).
Stability of the C. elegans genome during development and evolution.
in: Eukaryotic Gene Regulation. Edited by R. Axel, T. Maniatis and C.F. Fox. Academic Press, NY. NY. p. 205-218.
- Honda, B.M., Devlin, R.H., Nelson, D.W. and Khosla, M. (1986).
Transcription of class III genes in cell-free extracts from the nematode Caenorhabditis elegans.
Nuc. Acid. Res. 14 869-881.
- Kadonaga, J.T. and Tjian, R. (1986).
Affinity purification of sequence-specific DNA binding proteins.
Proc. Natl. Acad. Sci. 83 5889-5893.
- Klug, A. and Rhodes, D. (1987)
'Zinc fingers': a novel protein motif for nucleic acid recognition.
TIBS 12 464-470.

- Korn, L.J. (1982).
Transcription of *Xenopus* 5S ribosomal RNA genes.
Nature 295 101-105.
- Krause, M. and Hirsh, D. (1987).
A trans-spliced leader sequence on actin mRNA in *C. elegans*.
Cell 49 753-761.
- Long E.O. and Dawid, I.B. (1980).
Repeated genes in eukaryotes.
Ann. Rev. Biochem. 49 727-264.
- Martin, R.F. and Holmes, N. (1983).
Use of an (125)I-labelled DNA ligand to probe DNA structure.
Nature 302 452-454.
- Morton, D. and Sprague, K.U. (1982).
Silkworm 5S RNA and alanine tRNA genes share highly conserved 5' flanking and coding sequences.
Mol. Cell. Biol. 2 1524-1531.
- Morton, D. and Sprague, K.U. (1984).
in vitro transcription of a silkworm 5S RNA gene requires an upstream signal.
Proc. Natl. Acad. Sci. 81 5519-5522.
- Peterson, D.O., Beifuss, K.K. and Morley, K.L. (1987).
Context-dependent gene expression:
Cis-acting negative effects of specific prokaryotic plasmid sequences on eukaryotic genes.
Mol. Cell. Biol. 4 1563-1567.
- Procunier, J.D. and Dunn, R.J. (1978).
Genetic and molecular organization of the 5S locus and mutants in *D. melanogaster*.
Cell 15 1087-1093.
- Procunier, J.D. and Tartoff, K.D. (1975).
Genetic analysis of the 5S RNA genes in *Drosophila melanogaster*.
Genetics 81 515-523.
- Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980).
A control region in the center of the 5S RNA gene directs specific initiation of transcription:
I. The 5' border of the region.
Cell 19 13-25.
- Schwartz, D.C. and Cantor, C.R. (1984).
Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.
Cell 37 67-75.

Selker, E.U., Yanofsky, C., Driftmier, K., Metzenberg, R.L., Alzner-DeWeerd, B. and RajBhandary, U.L. (1981).

Dispersed 5S RNA genes in N. crassa:
Structure, expression and evolution.
Cell 24 819-828.

Selker, E.U., Morzycka-Wroblewska, E., Stevens, J.N. and Metzenberg, R.L. (1986).

An upstream signal is required for in vitro transcription of Neurospora 5S RNA genes.
Mol. Gen. Genet. 205 189-192.

Travers, A.A. (1987).

Structure and function of E. coli promoter DNA.
CRC Crit. Rev. Biochem. 22 181-219.

Tso, J.Y., Van Den Berg, D.J. and Korn, L.J. (1986).

Structure of the gene for Xenopus transcription factor TFIIIA.
Nuc. Acid. Res. 14 2187-2200.

Tyler, B.M. (1987)

Transcription of Neurospora crassa 5S rRNA genes requires a TATA box and three internal elements.
J. Mol. Biol. 196 801-811.