## NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

## AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

# GENOMIC ORGANIZATION OF THE NUCLEAR RIBOSOMAL RNA GENES AND PRIMARY SEQUENCE ANALYSIS OF THE RIBOSOMAL SPACER DNA IN THE FLAGELLATE PROTOCTIST, *Crithidia fasciculata*

by

James Clark Collings

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

ISBN 0-315-71453-0

Canada

This work is dedicated to my paternal grandmother, **Beatrice Ann Collings,** who more than anyone I have known, has shaped my intellectual and emotional development. I thank her for instilling in me a good work ethic coupled with a degree of stubbornness. It has allowed me to complete this thesis in spite of having it interrupted by an M.D. degree and Internship.
My only wish is that she were alive to have participated in my graduation from medical school and to witness the completion of this thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

List of Figures

List of Tables

## Abstract

The cytoplasmic ribosome of Crithidia fasciculata, a trypanosomatid protoctist, possesses an unusual complement of rRNAs. Besides an unusually long small subunit (SSU) rRNA of 2206 nucleotides and a heat-labile "28S" rRNA comprised of two high molecular weight (c and d) species, there are four novel small rRNAs (Sp. e, f, g, and j), in addition to the conventional 5S (Sp. h) and 5.8S (Sp. i) rRNAs, which are also present. All together, nine separate rRNAs are present in the C. fasciculata cytoribosome. All except the 5S rRNA are encoded in an 11,373-bp rDNA cistron, which can be isolated as five PstI-rDNA fragments. The genes are arranged in a basically eukaryotic fashion (i.e., 18S-5.8S-28S): starting with the SSU rRNA gene and going 3', there is an Internal Transcribed Spacer (ITS1), followed by large subunit (LSU) rRNA coding and spacer regions in the order Sp. i - ITS2 - Sp. c - ITS3 - Sp. e - ITS4 - Sp. d - ITS5 - Sp. f - ITS6 - Sp. j - ITS7 - Sp. g, and finally the Intergenic Spacer (IGS). The rDNA is either an extra-chromosomal circle or genomically integrated and tandemly repeated. By primary and secondary structure comparisons, the novel small rRNAs can be shown to be part of the conventional LSU rRNA. The rRNAs are most likely transcribed from a single promoter in the IGS and produced by post-transcriptional processing. The IGS has a complex physical substructure, the most conspicuous feature of which is an array of 27, 19-nucleotide long, tandem repeats. There are 10 types of these repeats and when compared to one another they show differing degrees of similarity. Their role, if any, in transcription ini iation or termination, rDNA replication, or rDNA evolution remains to be elucidated. Flanking this array are three related palindromic sequences, Blocs A, B, and C. They are 40 nucleotides long and could potentionally form hairpin structures in the DNA (or in the RNA if they are transcribed). The rDNA exhibits a length heterogeneity that is stably inherited. There appear to be 6 or 7 size classes of rDNA cistrons differing from one another by a variable amount (200-300 bp). The region encompassing this heterogeneity has been localized to the IGS and is unstable in $rec^+$ strains of E. coli, giving rise to specific deletions. The rDNA seems to be devoid of methylation.

## List of Abbreviations and Symbols Used

| | |
|---|---|
| α | alpha |
| A | adenosine |
| ADP | adenosine diphosphate |
| AGE | agarose gel electrophoresis |
| β | beta |
| C | centigrade |
| CAT | chloramphenicol acetyltransferase |
| cfu | colony-forming units |
| CP | core promoter |
| cm | centimeter |
| CsCl | cesium chloride |
| Da | dalton |
| DEPC | diethylpyrocarbonate |
| Dr | tandem direct repeat |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| EtBr | ethidium bromide |
| ETS | external transcribed spacer |
| G | guanosine |
| GTP | guanosine triphosphate |
| γ | gamma |
| HMG | high-mobility-group |
| IGS | intergenic spacer |
| iRNA | salt-insoluble RNA |
| ITS | internal transcribed spacer |
| IVS | intervening spacer |
| kbp | kilobase pair |
| KCl | potassium chloride |
| kDa | kilodalton |
| LSU | ribosome large subunit |
| M | molar |
| $m^6Ade$ | 6-methyladenine |
| $m^5Cyt$ | 5-methylcytosine |
| mg | milligram |
| $Mg(OAc)_2$ | magnesium acetate |
| ml | milliliter |
| min | minute |
| $M_r$ | mass |
| MMS | methylmethanesulfonate |
| NaCl | sodium chloride |
| nDNA | nuclear DNA |

| | |
|---|---|
| $NH_4OAc$ | ammonium acetate |
| NMF | N-methyl formamide |
| nt | nucleotide |
| NTS | non-transcribed spacer |
| n | any number |
| ψ | pseudo |
| PAGE | polyacrylamide gel electrophoresis |
| PEG | polyethylene glycol |
| pre-rRNA | pre-ribosomal RNA |
| psi | pounds per square inch |
| pUC9 | a genetically engineered cloning plasmid |
| rDNA | ribosomal DNA |
| RNAP | RNA polymerase |
| rRNA | ribosomal RNA |
| S | Svedberg unit |
| sRNA | salt-soluble RNA |
| SDS | sodium dodecyl sulfate |
| SOC | 2% (w/v) Bacto Tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose |
| Sp. | species |
| SSC | standard sodium citrate |
| SSU | ribosomal small subunit |
| S100 | the supernatant following a timed centrifugation at 100,000xg |
| T | thymine |
| TEMED | tetraethylmethylethylenediamine |
| TIF-IB | transcription initiation factor IB |
| tRNA | transfer RNA |
| U | uridine |
| UCE | upstream control element |
| UPE | upstream promoter element |
| UV | ultraviolet |
| μg | microgram |
| μl | microliter |
| vol. | volume |
| V | volt |
| w | weight |
| x g | times gravity |

## Acknowledgements

I would like to acknowledge the support of my wife Brenda who has seen me through my two degrees at Dalhousie University. She has been a constant companion, never complaining when I did. I would also like to thank her for Ruth, our daughter, who was born during my internship.

I would like to thank my parents and family, especially my sister Louise, who spent many unselfish hours at the word-processor in a fine attempt to understand my handwriting. Thank you very much.

I feel that it would require a volume the size of this thesis to adequately express my gratitude to my supervisor, Dr. Mike Gray. I thank him for accepting me into his lab and providing help whenever it was needed. He supported me and continued to provide me with the encouragement and advice required to see my research and thesis through to the end. He remained principled and exacting to the final day.

Many people have come and gone (and some have even come back again) during my tenure at Dal. I would like to thank all those who have assisted me, whether in a small or large way, in my research work and thesis production. Space does not permit me to list all of these people but I would like to give special mention to the following: Murray Schnare-thank you for introducing me to the Graduate House, David Spencer, Paul and Pam Joyce, Pat Covello, Mike Coulthart, Spencer and Diana and Kim and Wan for the use of their couches in the dying days of thesis production.

A special thank you to two friends I made while a graduate student, David Jolliffe and Mike Power.

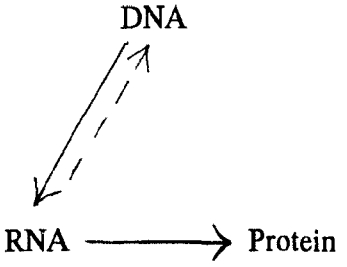
## SECTION A

## INTRODUCTION

### 1.) THE CENTRAL DOGMA OF MOLECULAR BIOLOGY WITH SPECIAL REFERENCE TO THE RIBOSOME

#### 1.i) "The Central Dogma": A Discussion

The concept of "The Central Dogma", which has become widely known in biology, was one of two articulated by F.H.C. Crick in an address he made to the Symposium of the Society for Experimental Biology in 1957. Crick was putting forward his thoughts on the directional flow of information that might exist in the cell. The "Central Dogma"[1], as stated by Crick, is "that once 'information' has passed into protein it cannot get out again". He continued: "The transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein or from protein to nucleic acid is impossible" (1). In 1970 (2) Crick outlined more precisely his thinking on the "Central Dogma" and summarized it in a now well-recognized figure.




Solid arrows show transfer of information that occurs in all cells and dashed arrows show special-case routes.

[1]Crick misunderstood the meaning of 'dogma' and, hence, misnamed this idea. For a discussion of this see p. 337 of H.F. Judson's "The Eighth Day of Creation" (207).

The idea then and the fact now, that information cannot be transferred back from protein to DNA, is of fundamental importance in evolution. It means that the Lamarkian Theory (3) of "The Inheritance of Acquired Traits" is impossible, and it brought to an end a long-held rival to Darwin's "Theory of Natural Selection" (4).

Even as early as 1957, Crick (1) and others suspected that the microsomal particles (now called ribosomes) observed in cells were the site of the transfer of information from nucleic acid to protein. The early data implicating these particles in such a role quickly gave way to more convincing evidence. Today a great deal is known about the structure and function of the ribosome (5-11,181), and all studies of these particles have borne out Crick's early suspicions of their pivotal role in the cell.

1.ii) The Ribosome: Its Place in the Central Dogma

The ribosome is now known to be a macromolecular complex involving ribonucleic acid and protein molecules. All ribosomes consist of two subunits, a small and a large. In a typical prokaryote (e.g., a eubacterium such as Escherichia coli) the small subunit (SSU) contains 21 proteins (totalling 350,000 Da) and a high molecular weight RNA (16S ; 500,000 Da), while the large subunit (LSU) contains 32 proteins (460,000 Da) and two RNA molecules [one high molecular weight (23S) and one low molecular weight (5S), with a combined mass of ca. 1,000,000 Da], giving a total ribosome mass of 2,310,000 Da (12).

The eukaryotic cytoplasmic ribosome is slightly larger than the prokaryotic (eubacterial) one, having a sedimentation coefficient of 80S rather than 70S. The eukaryotic SSU contains ca. 30 proteins (700,000 Da) and one

RNA molecule (18S; 700,000 Da), while the LSU has 45-50 proteins (1,000,000 Da) and one each of a high molecular weight RNA (28S; 1,700,000 Da) and 5S and 5.8S (combined mass 90,000 Da) RNAs. This gives a total ribosome mass of 4,190,000 Da (6).

The process of translating mRNA into protein is not accomplished by the SSU and LSU alone. Translation involves many other components, including a variety of factors (initiation, elongation, termination), aminoacylated tRNAs, and free nucleoside triphosphates (GTP), as well as the mRNA itself. These and other factors come together in a highly ordered manner to produce, accurately and efficiently, the encoded proteins (11).

It is in the hope of understanding this extremely complex, sophisticated, and centrally important phenomenon that investigators have undertaken a detailed examination of the ribosome and its components.

## 2.) RIBOSOMAL DNA (rDNA): WHAT IS IT AND WHERE IS IT FOUND?

In addition to studying the ribosomal components for biochemical reasons, investigators have also attempted to gain some insight into biological evolution and the processes of evolutionary molecular biology. The ribosome is very well-suited for this purpose because it is ubiquitous and evolutionarily well conserved in structure and function. All organisms, be they archaebacteria[2], eubacteria, or eukaryotes, possess ribosomes which carry out the same basic function, that of protein synthesis. The eukaryotic cell possesses, in addition to the cytoplasmic ribosome, two other sets of functionally and structurally distinct ribosomes, one set in mitochondria (organelles found in

[2]See page 7 for a description of these organisms.

almost all eukaryotes), the other in plastids (found in photosynthetic eukaryotes, i.e., plants and algae). Ribosomes thus provide the only easily accessible molecules which have been functionally and structurally conserved through evolutionary time and are found in all known organisms and organelles. The genomes of all organisms code for the rRNA components and even the eukaryotic organelles possess DNA which encodes all of the rRNA molecules used in translation in that organelle. The RNA components of the ribosome (ribosomal RNAs, or rRNAs) are the most easily studied and the most reliable molecular markers of evolutionary relatedness. These have been analyzed both at the RNA and the DNA level.

Figure 1 diagrammatically summarizes the structural organization of the so-called rDNA (ribosomal DNA) from the eukaryotic nucleus, eubacteria, and chloroplast.

The eubacteria provide a good starting point for a discussion of rRNA gene arrangement. In these organisms the 16S (SSU), 23S (LSU) and 5S rRNAs are transcribed in a 5' to 3' direction, giving rise to a single transcript which is processed to yield the individual mature rRNAs. These cistrons characteristically contain coding regions for two tRNAs as well (13).

In the eukaryotic nucleus the coding region for the 5'-end of the LSU rRNA is separated from the remaining portion of the LSU rRNA coding region by DNA sequence which does not appear in the mature rRNA. This 5'-end is commonly referred to as the 5.8S rRNA. As a rule the 5S rRNA in eukaryotic nuclear DNA does not constitute part of the rRNA coding unit. Normally it is physically unlinked and is invariably transcribed by a separate RNA polymerase (14).

rRNA Gene Arrangements

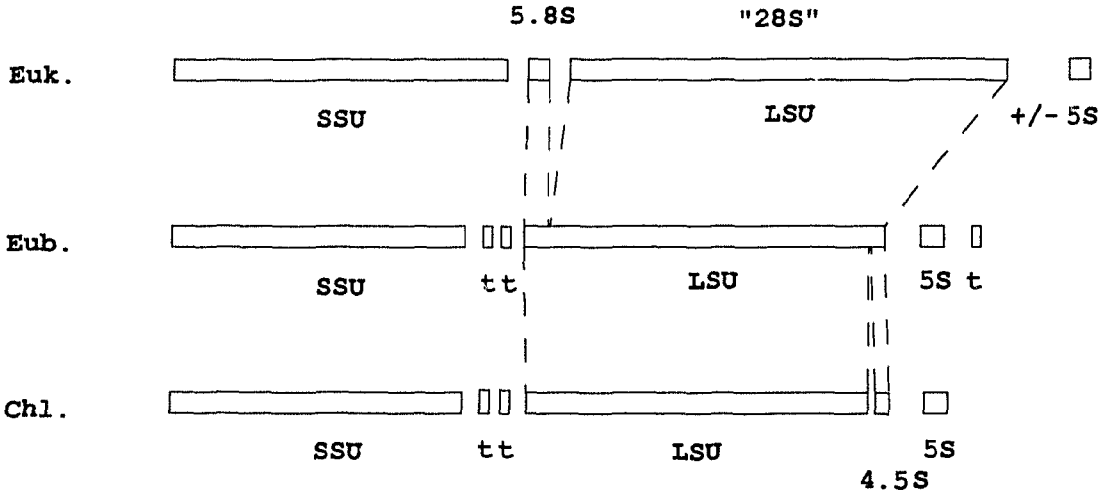



**Figure 1** Schematic presentation of typical rRNA coding arrangement from eukaryotic nucleus, eubacteria, and the chloroplast. Boxes represent the coding regions. **SSU** is the small subunit rRNA, **5.8S** is the 5.8S rRNA (the structural homologue of the 5'-end of the eubacterial LSU rRNA), **LSU** is the large subunit rRNA, **4.5S** is found only in chloroplasts and is homologous to the 3'-end of the eubacterial LSU rRNA, and the **5S** is a small rRNA found in the ribosomal large subunit and whose coding region is sometimes closely linked with the other rRNA coding regions in the eukaryotic nucleus. **t** is transfer RNA. **Euk.** is eukaryotic, **Eub.** is eubacterial, and **Chl.** is chloroplast. Dashed lines are to indicate regions of homology.

In Drosophila melanogaster, the 5.8S rRNA coding region is further interrupted by a small internal transcribed spacer, or ITS ( representing a sequence of DNA that is transcribed but is excised in the maturation process), resulting in the 5.8S rRNA being further subdivided into two rRNAs, 5.8Sa and 5.8Sb (15). D. melanogaster and Sciara coprophila also provide an example of an interrupted 28S rRNA coding region, with post-transcriptional processing giving rise to a bipartite 28S rRNA (16).

Plastid rRNA gene arrangement is very similar to that in eubacteria , reflecting the evolutionary origin of plastids from endosymbiotic eubacteria (specifically cyanobacteria) (182,183). Chloroplasts in Zea mays and other land plants have, in addition to the 5S rRNA, a 4.5S rRNA which is encoded immediately downstream of the "23S" rRNA gene. Sequence analysis has revealed that this rRNA is actually the 3'-end of the LSU rRNA, being separated from the remaining portion of the LSU rRNA gene by an ITS[3] which is subsequently excised (17). This theme is repeated in Chlamydomonas reinhardtii, but in the 5'-half of the chloroplast LSU rRNA, whose coding region is split into three parts, separated from one another by ITSs (18). C. reinhardtii also contains an intron within the 3'-half of the chloroplast LSU rRNA gene (18).

Chlamydomonas eugametos provides a dramatic example of an interrupted LSU rRNA coding region (184). The chloroplast LSU gene contains six group I introns and three internal transcribed spacers[3] resulting in a LSU rRNA consisting of four RNAs.

---

[3] Sequences which interrupt coding regions are termed intervening sequences (IVS). Introns and internal transcribed spacers are two types of IVS. Introns are distinguished by the fact that when they are removed in the process of RNA maturation, the flanking exons are ligated together, resulting in a gene product which is covalently continuous. In contrast, when internal transcibed spacers are removed during the post-transcriptional processing, the flanking RNA segments are not ligated, resulting in non-covalently associated products.

Ribosomal RNA gene arrangement in mitochondria is much more variable and species-specific (19,182,183).

The last major grouping of organisms is the archaebacteria. This recently recognized assemblage (20) is less well-studied but some details of the arrangement of archaebacterial rDNA are known. In most of the species studied, the genes follow the eubacterial convention rather closely (in euryarchaeotes, i.e., halobacteria and methanogens). The 5'-SSU rRNA gene-spacer-LSU rRNA gene-5S rRNA gene-3' configuration has been found to exist in most species studied (21), with some deviation in linkage seen in Thermoplasma acidophilum (a crenarchaeote). In this organism the SSU rRNA gene is separated by several kbp of DNA from the LSU and 5S rRNA genes (22).

## 3.) THE CLASSIFICATION OF LIFE: THE THINKING THEN AND NOW

Evolutionary thinking in recent years has been undergoing some fairly radical changes. With the introduction of molecular genetics in the early to mid 1970's, scientists have been able to make use of macromolecular sequences, both protein and nucleic acid, to establish phylogenetic relatedness for organisms and their cytoplasmic organelles (plastids and mitochondria). Recombinant-DNA technology allows the rapid isolation and sequence determination of any selectable gene. The genes coding for the ribonucleic acids found in the ribosome were among the first to be cloned and used in delineating evolutionary phylogenies (for a review, see 14).

These genes, particularly the 5S, 5.8S and SSU, possess characteristics which lent themselves to early investigation and phylogenetic

analysis. Their ubiquity and relative abundance (of both the gene and its product) allowed their ready isolation and manipulation, and the fact that they have the same function in all systems (structural-dynamic involvement in the mature ribosome) and provide quantifiable measurements, made these molecules welcomed evolutionary clocks (21,23-25).

Early workers made use of databases derived from the primary nucleotide sequences of 5S and 5.8S rRNAs and their genes (24,25). It quickly became apparent that the length of these RNAs (120 and 150-170 bp, respectively) were insufficient to allow comment on the relatedness of very divergent organisms. Efforts then turned to the primary sequence determination of the SSU rRNA, which possesses a mature sequence length of ca. 1600-2200 nucleotides. This molecule possesses all the characteristics desirable in a phylogenetic marker (i.e., ubiquity, constant function, and high information content). It was the study of SSU rRNA by G.E. Fox, C. Woese and their colleagues that allowed this group to propose a radically new interpretation of the evolutionary descent of the major groups of organisms. They proposed that all known organisms should be grouped into three "primary kingdoms" or primary lines of evolutionary descent (20). More recently Woese et al. have proposed a refinement to this classification (185), elevating "kingdom" to a higher taxonomic level termed "domain". These three domains are the Bacteria (which comprises the eubacteria), the Archaea (which comprises those organisms previously termed archaebacteria), and the Eucarya (comprising all of the organisms which contain a nucleus). They also proposed that the two major groups of Archaea be given kingdom status, and they named them Euryarchaeota (encompassing the methanogens and halophiles) and the Crenarchaeota (encompassing the extremely thermophilic archaebacteria).

Organisms have been classified according to various phylogenetic schemes using a variety of criteria. A recent and widely accepted classification is that of R.H. Whittaker (26). This system is a five-kingdom one which is based on placing an organism in a group with respect to its level of organization and mode of nutrition. The five kingdoms described are the Monera (prokaryotic cells), the Protista (unicellular eukaryotic cells), the Plantae (multicellular, walled, eukaryotic cells with photosynthetic plastids), the Fungi (primarily multi-nucleated eukaryotic cells with walled and often septate mycelial synctitium, not capable of photosynthesis), and the Animalia (multicellular organisms with wall-less eukaryotic cells lacking plastids and photosynthetic pigments). The three levels of organization upon which Whittaker based his classification are the prokaryotic, eukaryotic unicellular, and eukaryotic multicellular and multinuclear. It was his assumption that these levels were hierarchical in nature, with the Monera at the bottom, Protista in the middle and the multicellular and multinuclear organisms at the top. The three modes of nutrition (photosynthesis, absorption, and ingestion) were used further to divide organisms into major groups at each level. The Monera do not exhibit an ingestive mode of nutrition and the Protista contain within themselves all three modes, whereas the remaining three kingdoms are distinguished by their mode of nutrition.

A direct evolutionary flow from the Monera to the so-called "higher" organisms is implicit in Whittaker's arrangement: Whittaker has the Monera giving rise to the Protista and various groups of Protista giving rise to the Plantae, Fungi, and Animalia. This assumption is not unique to the Whittaker scheme but is a generally held tenet by most who have constructed phylogenetic trees (27,28). As mentioned previously, Fox and Woese challenged this notion by

determining phylogenetic relatedness using the primary sequences of SSU rRNA. They showed that all extant organisms fall into three phylogenetically coherent groups, these being the eukaryotes, the eubacteria, and the archaebacteria. The eukaryotic group should properly be called the eukaryotic nuclear (or nucleocytoplasmic) group because the plastid and the mitochondrion have been shown to have arisen as the result of the endosymbiotic acquisition of a cyanobacteria-like bacterium and a purple photosynthetic bacterium, respectively (29). The division of the prokaryotes (bacteria) into two distinct, anciently related groups was the revolutionary conclusion of the study of Fox and Woese. These workers argued that the microorganisms conventionally thought of as bacteria should indeed be grouped together: they termed these the eubacteria (to reflect their familiar nature). However, Fox and Woese identified a grouping of less familiar bacteria, which they termed the archaebacteria. It turned out that the archaebacteria possess formerly unrecognized common attributes, such as muramic acid-less cell walls, membranes containing abundant branched-chain ether-linked lipids, tRNAs devoid of the modified base ribothymidine, and distinctive RNA polymerase subunit structures (21).

The data of Fox and Woese did not allow them to assert that any one of their kingdoms was more closely related to either of the remaining two, hence, the "three primary kingdom" concept (30). A common identifiable ancestor could not be ascertained, so they named the assumed common ancestor the progenote. The name reflects their belief that the branching occurred very early in the evolution of life, possibly at the time the first genome(s) was being formed. Recently investigations have revealed that the eukaryotes (Eucarya) are more closely related to the archaebacteria (Archaea) than they are to the eubacteria (Bacteria) (186).

Due to the novelty of Fox and Woese's phylogenetic world, I will be using a more conventional classification of the biological world for my discussion of organismal rDNA structure and organization. This is done, of course, with the knowledge that it may soon be shown to be incorrect, and require modification.

The classification I will be using is one proposed by L. Margulis and K. Schwartz (31) in their book entitled Five Kingdoms. As the title might suggest, this classification strongly resembles Whittaker's scheme. Margulis and Schwartz have attempted to make the taxonomic level of phylum somewhat comparable in all major groups of organisms. They have retained Whittaker's definition for the Monera but recognize sixteen phyla to Whittaker's five. Their definition for the protist is one of exclusion: "they are neither animals (which develop from a blastula), plants (which develop from an embryo), fungi (which lack undulipodia (flagella) and develop from spores), nor prokaryote". They go on to say, "They comprise the eukaryotic microorganisms and their immediate descendants: all nucleated algae (including the seaweeds), undulipodiated (flagellated) water molds, the slime molds and slime nets, and the protozoa".

Because Margulis and Schwartz include multicellular, macroscopic organisms in their definition, they felt that a name change from protist (connoting single-celled) to protoctist (as defined by H.F. Copeland (27)) was required; hence, this kingdom is called the Protoctista.

In the Margulis-Schwartz scheme, the Fungi grouping loses several phyla, which instead are included among the protoctists; these include the water molds, slime molds, cell-net molds, and chytrids and oomycotes (undilipodia-bearing ones). The Kingdom Plantae loses all nucleated algae

including the seaweed to the Protoctista. The Kingdom Animalia remains largely intact.

## 4.) PROTOCTIST rDNA

### 4.i) Prologue

As previously mentioned, the gene cluster that codes for rRNA is referred to as the ribosomal DNA, or rDNA. In eukaryotes, the transcriptional unit found in these clusters is remarkably uniform in its organization. Traditionally, rDNA has been characterized as having a non-transcribed spacer (NTS) followed by an external transcribed spacer (ETS), then the SSU rRNA gene which in turn is followed by the LSU rRNA gene and then another ETS. The 5.8S rRNA gene is interposed between the SSU rRNA gene and the remainder of the LSU rRNA gene and flanked by relatively small and variably sized internal transcribed spacers (ITS) (14).

Although this unit of structure is very well conserved, some variation within this basic theme has been found. This includes different unit lengths (ranging in size from 7 kbp to 44 kbp; length heterogeneity can be found within an individual and is also an interspecies phenomenon) and different numbers of rDNA units encoding the rRNA complement found in the ribosome. The rDNA units, or cistrons, may be clustered together on a single nuclear chromosome, or they may be dispersed on several chromosomes; they may be found extra-chromosomally as linear molecules, circular molecules, or subchromosomal fragments (or in some cases, some combination thereof) (38). The greatest diversity in terms of rRNA complement and rDNA organization is found among the protoctists. This is both a reflection of the biological diversity of this group of organisms and a function of their antiquity, making this assemblage

a particularly interesting one to investigate. One incentive for doing so is the expectation that it will be possible to gain some insight into organismal relationships, and to infer what form the important rRNAs took in early evolutionary history. It is also hoped that some understanding of eukaryotic gene regulation can be gained by such comparative analyses.

What follows is a compilation of all protoctist rDNAs studied to date, with a description of: the size of the rDNA; its genomic arrangement; the number of rDNA units per haploid genome; whether there is any observable length heterogeneity; whether there are any discernible introns; whether there are any unusual spacers resulting in an altered complement of rRNAs; and whether the gene coding for the 5S rRNA is physically linked with the others.

Where appropriate, discussion of rDNAs from "higher" eukaryotes will be included for comparative purposes or because of some particular novelty.

4.ii) Amoeboid rDNA

Acanthamoeba castellanii is a free-living soil amoeba which can undergo a modest differentiation into a cyst. D'Alessio et al. (32) isolated the nuclear rDNA as two fragments in a bacteriophage vector. It was found that the entire repeat was 12 kbp in length, with no evidence of length heterogeneity or introns, but R-loop mapping did show that the LSU rRNA is composed of two coding regions separated by a gap of 200 bp. It was determined that the overall arrangement of the rDNA repeat is conventional, with the 5.8S rRNA gene positioned between the SSU and LSU rRNA genes. The 5S rRNA gene is not linked with the other rRNA genes. The rDNA units are arranged in a head-to-tail fashion and are tandemly repeated.

A second amoeba studied is Naegleria gruberi, an opportunistic pathogen of mammals which is normally a free-living soil organism. N. gruberi was discovered to have an extra-chromosomal, circular plasmid which contains the rRNA genes. This plasmid is 14 kbp long and is present in 3000 to 5000 copies per cell. Like Euglena gracilis but unlike Physarum and Dictyostelium there is no identifiable, chromosomally integrated rDNA. No IVS were detected in the rRNA coding regions either (192).

A third amoeboid-like organism which has had its rDNA studied is Entamoeba histolytica (193). This intestinal parasite lacks mitochondria and a Golgi apparatus and has a single nucleus. The nucleus harbours a large circular molecule (24.6 kbp) which contains two regions of ca. 5.2 kbp which are inverted relative to each other and separated by reiterated DNA. These inverted sequences code for the rRNA genes. No IVS were identified in the rRNA coding regions.

### 4.iii) Ciliate rDNA

The most extensively studied group of protoctists is the ciliates. To date the rDNA cistron has been investigated in five hypotrichous (cilia restricted to the ventral surface of the organism) species: Oxytricha fallax, Oxytricha nova, Stylonychia mytilus, Stylonychia pustulata, Euplotes aediculatus; and in the holotrichous (cilia arranged uniformly about the organism) ciliates: Paramecium tetraurelia, Glaucoma chattoni and several Tetrahymena species.

The ciliates are peculiar in that they contain two types of differentiated nuclei in the same cell. There is at least one germinal diploid micronucleus (there may be several but this is a species-specific feature) and a somatic macronucleus which is often polyploid. The micronucleus is responsible

for genetic continuity and is the nucleus involved in sexual exchange. The macronucleus is the transcriptionally active nucleus and, therefore, is responsible for the organism's phenotype. The rDNA has been characterized from the two types of nuclei in some species but only from the macronucleus in others. Such distinctions will be noted.

The Hypotrichous Ciliates: Stylonychia mytilus, Stylonychia pustulata, Oxytricha fallax, Oxytricha nova, and Euplotes aediculatus.

This group of hypotrichous ciliates has a macronuclear genome made up of "achromosomal" or "gene-sized" DNAs. These DNA pieces are generated (in post-conjugative cells) by fragmentation of polytene chromosomes, which are produced from the micronuclei in the process of macronuclear development, during which a great deal of DNA complexity is lost. Native macronuclear DNA shows a species-specific banding pattern when fractionated by electrophoresis in an agarose gel, but the fragments have a fairly restricted size range of 0.4 kbp to 20 kpb (34, 37). Although the macronuclear DNA tends to be heterogeneous in size, the rRNA genes are found associated with discrete-sized DNA fragments (33).

The rDNA of Stylonychia mytilus has been studied at the genomic level (35) whereas the rDNAs of the other hypotrichous ciliates have been isolated and studied using a recombinant plasmid vector (33,36,37). In all five organisms the rDNA was found to be 7.5 kbp in length and to exist naturally at that size in the macronucleus. Southern hybridization mapping studies of S. pustulata, O. fallax, O. nova, and E. aediculatus DNAs showed that the rDNA molecules are linear and that each contains only one set of rRNA genes. Each rDNA molecule was found to have a 1.3-kbp non-coding region 5' to the SSU rRNA gene, with the latter separated from the LSU rRNA gene by a small spacer

presumed to contain the 5.8S rRNA gene. No additional spacers or intervening sequences > 0.5-1.0 kbp were detected. As has been found in other organisms, the degree of sequence conservation, as judged by restriction site polymorphism, is higher in the coding regions than in the non-coding regions. Presumably S. mylitus rRNA gene arrangement is representative of that in other hypotrichous protists, but detailed mapping has yet to be carried out.

The Holotrichous Ciliates: Paramecium tetraurelia, Glaucoma chattoni, and Tetrahymena sp. In these ciliates, macronuclear development appears to take place by a gradual increase in chromatin content, with no polytenization of the chromosomes as is the case in the hypotrichous ciliates, and apparently there is not the massive loss of macronuclear DNA that is observed in the hypotrichs (38). In spite of this fact, one of these organisms (G. chattoni) has been shown to have discrete-sized, linear, subchromosomal DNA molecules, reminiscent of the situation in the hypotrichs (39).

The macronuclear rDNA of P. tetraurelia (40) exists as relatively small extra-chromosomal molecules. The rDNA is arranged as non-palindromic, head-to-tail repeats with an average repeat size of 8.3 kbp; it is found as a mixture of circular and linear forms. The linear molecules are thought to be randomly sheared circular molecules.

The number of rDNA repeats per molecule was found to be two to thirteen (the 5S rRNA gene is not linked to the other rRNA genes). Limited length heterogeneity was found in the spacer region but the latter was found to be more highly conserved than in other organisms. This length heterogeneity, although rather subtle in extent (some repeats were 500 bp larger than others), does allow speculation that the micronucleus contains at least two copies of the repeats, and possibly more.

In Tetrahymena thermophila (41), the best studied of the Tetrahymena species, it has been established that the micronucleus contains a single integrated copy of the rDNA repeats found in the macronucleus. The molecular form taken by the several hundred copies (per haploid genome) of macronuclear rDNA is perhaps the most unusual and best studied among the protoctists. It is a free, double-stranded extra-chromosomal molecule, either linear (90%) or circular (10%), and with a length of ca. 20 kbp. The linear molecules are not randomly sheared circular molecules as is the case in P. tetraurelia. The linear molecule consists of two identical 10 kbp sequences, each containing one rRNA transcriptional unit, arranged in a head-to-head (or inverted) fashion. In other words, it is a very large palindrome. This palindromic arrangement was found in all species and strains studied (42), with some variation in the overall length of the molecules (18-20 kbp). A significant difference in the rDNA molecules was seen in some strains of T. thermophilia and T. pigmentosa, where the presence of an intron was established in the LSU rRNA coding sequence (42). Cech and co-workers (43) demonstrated the autocatalytic or self-splicing nature of this intron and have worked out the protein-free mechanisms responsible for the elimination of the intron sequence from the pre-rRNA.

The closely related holotrichous ciliate, Glaucoma chattoni, has a macronuclear genomic organization reminiscent of that seen in the hypotrichous ciliates. Katzen et al. (39) in 1981 reported that the macronucleus of this organism possesses subchromosomal ("gene size") molecules ranging in size from 2.1-100 kbp and containing specific, discrete-sized fragments. The rRNAs hybridized to an abundant 9.3-kbp linear fragment which possessed terminal repeats having the sequence $C_4A_2$, as in Tetrahymena; such terminal repeats (telomeres) are similar to sequences found in the hypotrich Stylonychia (44).

4. iv) Slime Mold rDNA

The phylogenetic placement of the slime molds has long been controversial (26). I have chosen to abide by Margulis' definition of protoctist and, therefore, have included slime molds in this discussion of protoctist rDNA (Whittaker places them in the sub-kingdom Gymnomycota of the Kingdom Fungi). Dictyostelium discoideum is a cellular slime mold and Physarum polycephalum is an acellular slime mold.

Both organisms have been the subject of intense study because their simple developmental life cycles lend themselves to the investigation of fundamental biological problems, such as synchronous control of nuclear division (as in Physarum), and cAMP-mediated cellular communication (in Dictyostelium). The rRNA genes are found on a linear extra-chromosomal palindrome (45-48), an arrangement very similar to that in Tetrahymena. Each molecule possesses two transcription units arranged in a head-to-head fashion, with RNA synthesis being initiated near the centre and proceeding out to the termini. Each organism has about 150 copies of rDNA molecule per haploid genome. In Dictyostelium the 5S rRNA gene is found downstream of the LSU rRNA gene, but it is transcribed independently of the other rRNA genes. Overall rDNA length is ca. 44 kbp in D. discoideum and ca. 60 kbp in P. polycephalum. The P. polycephalum LSU rRNA gene contains within it two introns which are removed during maturation of the pre-rRNA to the final stable rRNAs.

4.v) Algal rDNA

In keeping with Margulis and Schwartz's definition of Protoctista, I have included the algae in the present discussion. Although rRNAs

have been analyzed in a number of algae, only in a few closely related ones have the rDNAs been studied in any detail. I will limit my presentation to those organisms in which something is known about the rDNA cistron, as opposed to the rRNAs only.

The Miller spread technique (49) was used by Berger and Schweiger (50-52) in the mid-1970's to study the rDNA from several green algae. This technique allows electron microscopic visualization of DNA isolated from nucleoli. A structure having the appearance of a Christmas tree can be seen, the branches of which correspond to newly synthesized pre-rRNA, the trunk being the rDNA, the apex the point of transcription initiation, and the base the position at which transcription termination is thought to occur. It was from such pictures that the terms "non-transcribed" and "transcribed" spacers arose. As can be imagined, not a great deal can be learned from such photographs but they did allow the authors to comment on the length of the non-transcribed and transcribed DNA, as well as the degree of cistron redundancy and cistron orientation.

The unicellular green algae studied by Berger and Schweiger included Acetabularia peniculus, A. dentata, A. ryukyuensis, A. major, A. dasycladus clavaeformis, Cymopolia van basseae, and Bataphora aerstedii. In general, all species exhibited head-to-tail tandemly repeated rDNA cistrons of uniform length. Acetabularia ryukyuensis has an rDNA length of 5.5 kbp with a spacer of 0.73 kbp. A. denta and A. peniculus have very similar rDNA configurations of 0.9-kbp and 2.9-kbp sized spacer and 7.38-kbp and 6.89-kbp sized transcribed regions, respectively. Although A. major has an overall rDNA length of 11.0 kbp, its transcribed region is smaller (5.8 kbp). C. van bosseae and D. clavaeformis have the shortest repeats visualized, being 1.2 kbp and 0.65 kbp respectively. The transcribed regions of A. major and B. aerstedii (5.9 kbp) more

closely resemble those of A. ryukyuensis than the others. B. oerstedii was the only species to exhibit no spacer region at all, although a minor number did possess a small (400 bp) non-transcribed spacer.

Marco and Rochaix (53) reported results of rDNA studies in the only other group of algae studied to date. Using recombinant-DNA technology and Southern hybridization they examined five species of the genus Chlamydomonas. C. reinhardtii has tandemly repeated rDNA having a homogenous length of 8.0 kbp. The same arrangement was seen in C. globosa and C. callosa. C. intermedia has an rDNA size of 14.8 kbp and C. eugametos has two rDNA size classes of 11.4 and 11.0 kbp. C. eugametos was the only protoctist alga to exhibit any length heterogeneity.

4.vi) Apicomplexan rDNA

The apicomplexans are a peculiar and medically important group of parasitic protoctists. They include the organisms responsible for human malaria; for this reason they have been studied extensively. In an attempt to elucidate the differential regulation of genes in the various blood and dipteran forms, investigators have cloned and studied the rDNA from several Plasmodium species, including P. lophurae (the avian malaria parasite) (54-56), P. berghei (the rodent malaria parasite) (57,58), and P. falciparum (the human malaria parasite) (59). Their rDNAs were found to be unique in several respects.

The rDNA of P. lophurae

The structural and organizational characteristics of the rDNA of the avian parasite include:

(1) a very small copy number of ca. eight repeats per haploid genome, which can be assigned to one of four classes as defined by restriction analysis; and

(2) an SSU rRNA gene that contains a 230-bp insert and an LSU rRNA gene that contains two insertions (one of 240 bp and a second of 110 bp). The rDNA units are not tandemly linked in the genome but are flanked by unique-sequence DNA.

The rDNA of P. berghei

As in the case in P. lophurae, P. berghei has a small copy number (four copies per haploid genome in this case) of rDNA which can be assigned to classes (two in this case) based upon restriction endonuclease analysis. The units are unlinked. The authors of this study were able to demonstrate that one of the two classes is preferentially transcribed in the blood stage forms. Restriction analysis revealed sequence heterogeneity within coding and non-coding regions between the units studied.

The rDNA of P. falciparum

P. falciparum has four copies of rDNA per haploid genome, divided into two classes. The LSU rRNA gene appears to have an intervening sequence.

4.vii) Flagellate rDNA

The rDNA has been studied in seven flagellate species from three genera. Included in this section of protoctist is the photosynthetic flagellate Euglena gracilis, an organism conventionally placed with the plants. The justification for placing this organism in the protoctists lies with the molecular biological investigation of its rDNA and rRNAs. This will be discussed in detail in the "Discussion" section.

This group of flagellates possesses the most unusual rDNAs and rRNA complements yet discovered. Interestingly this group includes the subgroup Kinetoplastida (those flagellates which have a single large mitochondrion ("kinetoplast") associated with the kinetophore). The Kinetoplastida contains a number of important human and animal pathogens including Leishmania donovani (responsible for visceral leishmaniasis), Trypanosoma cruzi (responsible for Chagas' disease) and Trypanosoma brucei (African sleeping sickness) (194).

Grouped with the Trypanosoma is Crithidia fasciculata, the subject of this thesis. A detailed discussion of its rDNA arrangement and the rRNA complement of these organisms will therefore be deferred to the "Results" and "Discussion" sections.

4.viii) Summary

Protoctist rDNA resembles the rDNA of so-called "higher" (multicellular) eukaryotes in the arrangement of SSU, 5.8S, and LSU rRNA genes, but it provides dramatic examples of aberrant rDNA genomic organization, low rDNA copy number, and varied rRNA complement. The rDNA can be found on nuclear chromosomes in a head-to-tail, tandemly repeated fashion, as is the

situation in most multicellular eukaryotes, or it can be on extra-chromosomal DNAs. These extra-chromosomal rDNAs can take the form of linear or circular molecules. The linear molecules may code for one or two rRNA cistrons; those containing two rRNA cistrons form a large palindrome with the cistrons oriented in a head-to-head manner. The circular molecules may be monomeric or oligomers of the rRNA cistron. The ciliate protoctists possess integrated rRNA cistrons in their micronucleus and rRNA cistrons (one per DNA fragment) on "gene-sized" or "achromosomal" DNA found in their macronucleus. These gene-sized copies are the result of chromosome amplification and fragmentation.

The slime molds, D. discoideum and P. polycephalum, possess extra-chromosomal rDNA very similar to that of the ciliate Tetrahymena, strengthening the case for including them with the more conventional protists.

The plasmodial organisms exhibit the lowest copy number of rDNA of any eukaryotes studied to date, and the first demonstration of differential developmental regulation of an rRNA cistron in a single organism.

The flagellates have an rDNA genomic organization not unlike that of multicellular eukaryotes, but have the most bizarre rRNA complement yet identified. As we will see the trypanosomes possess a "nicked" LSU rRNA resulting in two large rRNAs (a condition seen in other protoctists and in some animals) and four small rRNAs not including 5.8S or 5S rRNAs. The "genes" for these small rRNAs are separated from the large rRNA coding regions by transcribed spacers, and are produced by processing of the pre-rRNA.

## 5.) THE rDNA SPACER: PHYSICAL ORGANIZATION

### 5.i) Prologue

Historically the spacer DNA separating the end of the LSU rRNA gene in one repeat from the beginning of the SSU rRNA gene in the next repeat has been sub-divided into an external transcribed spacer (ETS) and a non-transcribed spacer (NTS). The ETS begins at the site of transcription initiation and proceeds to the start of the coding region for the SSU rRNA, leaving the remaining spacer DNA as the NTS (see Figure 2 for a description of a typical rDNA cistron). This terminology, particularly "non-transcribed spacer", was developed to describe what was seen in Miller spreads of transcriptionally active rDNA. Detailed restriction analysis, primarily sequence analysis, and transcriptional analysis have revealed that the NTS is a stretch of DNA with an unusual physical substructure, and is, in fact, transcriptionally active. For that reason, it has now been renamed the "intergenic spacer" (IGS).

The best studied rDNA spacers, those in Xenopus (60-76), Drosophila (77-82), and S. cerevisiae (83-86), are described in detail below. Also included is a discussion of mammalian and Tetrahymena (88,89) rDNA spacer structure.

### 5.ii) Xenopus

The dominant feature of the rDNA spacer in X. laevis, X. clivii, and X. borealis is that of repeating DNA sequences. In X. laevis there are four regions of repeated DNA designated 0, 1, 2, and 3. In addition to these stretches, there are at least two blocks of DNA (the "spacer promoters") that are homologous to the rDNA promoter located at the site of transcription initiation. The four DNA repeats (0,1,2,3) can exist up to several kbp away from the rDNA

promoter and the duplicated spacer promoter (between two and seven ) can be found among these DNA repeats. The spacer promoters are separated by the 2/3 repeats, and a variation in the number of spacer promoters and 2/3 repeats is responsible for the length heterogeneity observed in X. laevis rDNA. The spacer promoters are not identical to the rDNA promoter, but exhibit a very high sequence similarity to the promoter. They have been shown to promote RNAP I transcription, although they do not initiate complete rDNA transcription; instead, transcription from these promoters terminates prematurely ca. 215 bp upstream of the true pre-rRNA promoter, giving rise to short transcripts of the spacer DNA.

The repeated regions 2 and 3 are themselves a composite of homologous sub-repeats 60 and 81 bp in length. It seems that the 60-bp repeat is derived from the 81-bp repeat by a 21-bp deletion. The 60/81-bp repeat possesses a stretch of nucleotides that is homologous to the rDNA promoter, and it is thought that the 60/81-bp repeat has arisen from pre-rRNA promoter duplication and expansion. The functional significance of these sequences will be discussed below. An interesting observation is that the 60/81-bp repeats exhibit more inter-species than intra-species similarity with the pre-rRNA promoter.

X. clivii exhibits a spacer arrangement very much like that of X. laevis. Whereas X. laevis has from two to seven spacer promoters, X. clivii seems to have only one. This spacer promoter is separated from the pre-rRNA promoter by repeated DNA with a physical arrangement like that in X. laevis. X. clivii has two sets of sequences (3C/3C') similar in sequence and position, as the 2/3 repeats are in X. laevis. Indeed the 3C/3C' regions are composed of units 79-83 bp in length which are very similar to the 81-bp units in X. laevis. These 79- to 83-bp units are homologous to a segment of the pre-rRNA promoter, as is the case for the 81-bp unit in X. laevis. Presumably these regions are functionally

analogous. The 3C/3C' regions in X. clivii are separated by other repeated units whose significance is unknown.

The remaining repeated regions in X. laevis (0 and 1) show strong similarity with regions 1 and 2A, 2B, 2C, respectively, in X. clivii and have similar relative positions. Such similarity would suggest similar function, but functionality of these sequences remains to be demonstrated.

Sequence information for X. borealis spacer DNA is less complete than for the other two Xenopus species, but the spacer appears to have an arrangement similar to that found in X. laevis and X. clivii. The X. borealis spacer contains two spacer promoters, one termed the 5'-spacer promoter and the other the 3'-spacer promoter, separated by two 80-bp units. Sequence comparison between the 5'- and 3'-spacer promoters and the pre-rRNA promoter reveals that the 5' motif displays greater sequence similarity than does the 3' version: in fact, sufficient similarity to be potentially functional. The 3'-spacer promoter exhibits several deletions, one in a region that has been shown to be essential for function in X. laevis.

The two 82-bp units are very similar to the 81-bp unit of region 2/3 in X. laevis and the 79-83 bp unit of X. clivii. They exhibit the same promoter homology as is seen in the other Xenopus species. Moss (87) have proposed a generalized spacer structure for Xenopus sp. based on the data available from the three species studied. This generalized form is as follows: (ca. 30 bp)$_n$ (ca. 100 bp)$_n$ (Spacer Promoter/ ca. 80 bp$_n$)(Spacer Termination) (Pre-rRNA Promoter), where the ca. 80-bp units each are homologous to the -70 to -110 bp region of the true promoter (Pre-rRNA Promoter) and Spacer Termination refers to the spacer termination site.

Eukaryotic Nuclear rDNA Arrangements

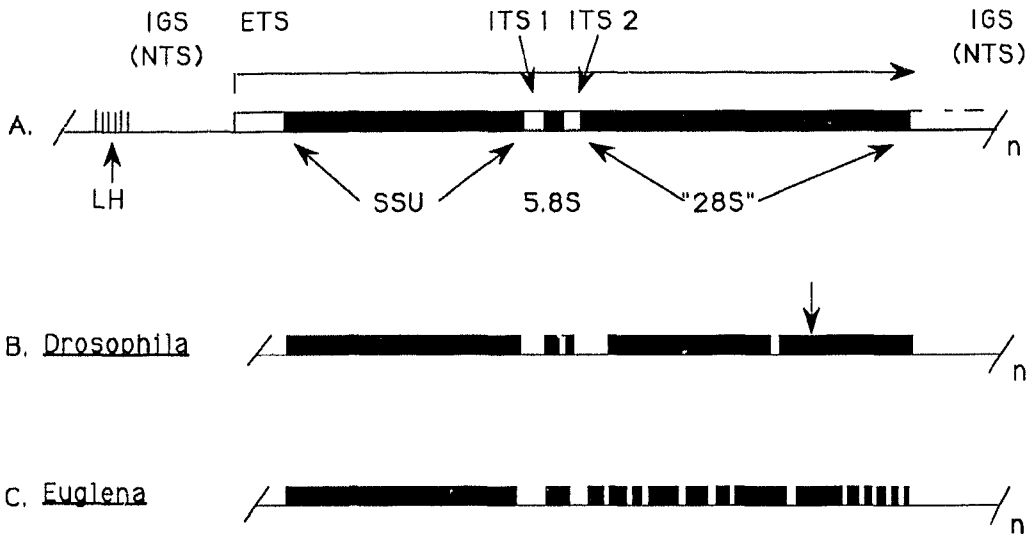



Figure 2. Schematic representation of a typical eukaryotic nuclear rDNA cistron and representation of the cistron organization in Drosophila and Euglena A. Boxed areas are transcribed regions Black boxes are rRNA coding regions and white boxes are transcribed regions which do not code for mature rRNAs Vertical line above the boxes represents the site of RNAP I transcription initiation and the horizontal line represents the direction of transcription **SSU** is the small subunit rRNA, **5.8S** is the 5 8S rRNA, and '28S" is the rest of the large subunit rRNA **ETS** is external transcribed spacer, **ITS** is internal transcribed spacer, and **IGS** is inter-genic spacer Vertical bars in the IGS are meant to represent the site of tandemly repeated sequences found in many organisms and responsible for observed length heterogeneity **(LH)** **n** is to indicate that the cistron is, frequently, directly and tandemly repeated on a single chromosome B. Schematic Drosophila rDNA cistron The arrow is to indicate the approximate location of an intron (16) C. Schematic Euglena rDNA cistron (171,172)

5.iii) Drosophila

Restriction and sequence analysis has revealed that the structure of the rDNA spacer in Drosophila melanogaster is comparable to that of Xenopus, in that it consists of three regions of repeated sequence. Immediately upstream of the pre-rRNA promoter are found 240 bp imperfect promoter duplications that are transcriptionally active in vivo and in vitro, the most upstream of which seems to dominate spacer transcription.

Further upstream of the 240-bp units are found 340-bp repeat units that are related to the pre-rRNA promoter. These ca. 340-bp repeats possess a direct repeat of the sequence found immediately downstream of the initiation site in the region 3 promoters. A short region of ca. 90 bp having no promoter homology is found upstream of the ca. 340-bp unit region and terminates just 3' to the end of the LSU rRNA coding region. Spacer transcription has been shown to be promoted by any of the 240-bp repeats, with the furthest upstream one dominating. These transcripts are thought to terminate in advance of the pre-rRNA promoter.

D. virilis and D. melanogoster share an almost identical organization, each having 240-bp repeats homologous to the pre-rRNA promoter, but with primary sequences that differ greatly.

5.iv) Mammals

As a result of the extreme size (ca. 30 kb) of mammalian rDNA spacers, little is known about their physical structure. The ETS does possess a variable number of 135 bp repeat units localized to the proximal 2 kbp of the spacer. These repeats do not exhibit any homology with the pre-rRNA promoter.

5.v) Yeast

No pre-rRNA promoter homologies have been found in the spacer DNA of S. cerevisae. The only resemblance to Xenopus or Drosophila spacers is the presence of short direct repeats scattered throughout the yeast spacer. Experiments using a chimeric rRNA gene to transform S. cerevisiae have revealed a region 3' to the LSU rRNA coding region that possesses the ability to enhance pre-rRNA transcription 20-30 fold. This region has been termed the major promoter element. It operates in an orientation-specific manner, and its action is strongly attenuated by placement of a copy of the presumed pre-rRNA sequence between it and the pre-rRNA promoter. This behaviour is reminiscent of what has been seen with Xenopus spacer DNA (see "Discussion").

5.vi) Tetrahymena

The rDNA in Tetrahymena is a highly amplified extra-chromosomal palindrome 21 kbp in length. It consists of two rDNA cistrons separated at the center of the molecule by unique sequence DNA. Higashinakagawa's group (202-205) has determined that the site of rDNA transcription initiation in both T. pyriformis and T. thermophila is at an A residue 647 bp upstream from the start of the coding region for the SSU rRNA. This conclusion was based on analysis of an RNA transcipt produced in a Tetrahymena in vitro transcription system. In $S_1$ nuclease mapping studies, Niles et al. (88,89) inferred a different site of rDNA transcription initiation in T. pyriformis than that determined by Higashinakagawa's group. Repeating the experiments with Niles' strain of T. pyriformis, Higashinakagawa et al. (204) determined that the site of transcription initiation in their strain of T. pyriformis was the same as that in T. thermophila. Recently, the complete rDNA sequence for T. thermophila was reported (206).

## 6.) THE rDNA SPACER: TRANSCRIPTION

### 6.i) Prologue

Eukaryotic nuclear-encoded rRNA genes are arranged in a very characteristic fashion. Moving from the point of transcription initiation to transcription termination, the RNAP I travels from a start site in the 5'-ETS through the SSU rRNA gene, a small internal transcribed spacer (ITS 1), the 5.8S rRNA gene, a second small internal transcribed spacer (ITS 2), and the rest of the LSU rRNA gene to a point 3' to the mature end of the LSU rRNA within the spacer DNA. The spacer DNA consists of the above-mentioned 5'- and 3'-ETS plus DNA that is not transcribed as part of the pre-rRNA (i.e., IGS). The IGS separates the tail of one transcriptional unit from the head of a second one. This head-to-tail arrangement of transcriptional units and spacer DNA can be repeated hundreds to thousands of times in a single chromosome. There are exceptions to this arrangement, most notably the linear extra-chromosomal rDNAs such as that found in Tetrahymena, where the transcriptional units and spacer DNA are arranged in a head-to-head arrangement instead of the usual head-to-tail.

### 6.ii) RNA Polymerases (RNAP)

Unlike the situation in prokaryotes, eukaryotic cells make use of three functionally different RNA polymerases. RNA polymerase II transcribes all protein-coding genes and RNAP III is responsible for transcribing 5S rRNA genes, tRNA genes, and genes for other small RNAs. RNA polymerase I (RNAP I) transcribes the rRNA genes.

6.iii) Transcription

While in the nucleus, the primary rRNA transcript produced by RNAP I undergoes a number of endonucleolytic cleavages to give rise to the mature SSU and LSU rRNA (the latter comprising the 28S and 5.8S rRNAs, the two together being the functional and structural equivalent of prokaryotic 23S rRNA). These rRNA species associate with newly synthesized ribosomal proteins to produce functional ribosomes.

The biosynthetic control of ribosome production is one of prime importance to the actively growing and dividing cell. It requires that all the ribosomal components be synthesized in a coordinated fashion. This control undoubtedly involves regulation at the level of RNAP I transcription. For this reason DNA sequences involved in RNAP I transcription initiation and termination have been studied in some detail. What follows is a description of what is known about these sequences in a number of different organisms.

6.iv) Transcription Initiation: The Mouse Model

Transcription initiation and termination studies are most often carried out using cloned rDNA sequences and in vitro cell-free systems. With this technique, Grummt's lab at the Institüt für Biochemie der Universität Wurzburg and Muramatsu's group in the Department of Biochemistry of the University of Tokyo (Japanese Foundation of Cancer Research) have worked out many of the details of transcription initiation in the mouse system.

The following insights have been provided by these studies (90-99):

(a) 5' deletion/competition experiments have shown that sequences upstream of -40 have a small but real effect on transcriptional

activity (+1 being the $S_1$ nuclease-mapped end of the putative precursor rRNA molecule);

(b) deletion of sequences between -40 and -35 results in a significant (up to 90%) decrease in transcriptional activity;

(c) removal of sequences between -39 and -13 results in complete elimination of transcriptional activity, and;

(d) the loss of sequences downstream of +11 leads to a substantial decrease in transcription.

To delineate more finely the 5' sequences involved in transcription initiation, these two groups carried out point mutational studies using cloned rDNA sequences (93,98). Grummt's group found that a G to A transition (conversion of a guanosine to adenosine) at position -15 resulted in a 20% increase in transcriptional activity while a G to A change (guanosine to adenosine transition) at -16 led to a 95% decrease in template transcription. Nucleotide changes were introduced only into the -35 to -14 region so they were unable to say anything about the role of nucleotides outside of this region. This was not the case with Muramatsu's experiments. This group was able to show that a G to A change at -7 completely abolishes transcriptional activity. Also, a change at -16 (G to A) resulted in a very similar outcome to that observed by Grummt's lab (i.e., significant decrease) and a G to A transition at -25 resulted in a 50% decrease in transcription.

It should also be mentioned that in competition experiments where the aforementioned (-7 to -16) mutated templates competed with wild-type ones, the -7 mutation competed effectively but the -16 mutation did not, suggesting a functional difference in the role of these two nucleotides.

6.v) RNAP I Promoters, Transcription Factors, and Species Specificity

In experiments reported in 1982 by Grummt's (100) laboratory, it was shown that in vitro transcription of cloned rDNA is a species-specific event. Using mouse rDNA template alone or in combination with rDNA templates from Physarum polycephalum, Drosophila melanogaster, or Xenopus laevis, these workers were able to show that only the mouse rDNA template could be utilized in a mouse cell-free transcriptional system.

To ensure that this result was not due to a peculiarity of the mouse transcription system, these workers designed additional experiments in which they reconstituted transcription systems from HeLa cells and Acanthamoeba castellanii. Through the use of mixing experiments they showed that the previously observed species specificity held true for the human and Acanthamoeba cases, as well. In addition, they were able to conclude that the specificity most likely did not reside with the RNAP I. It was shown that the purified RNAP I from Acanthamoeba would not transcribe the homologous rDNA in the absence of Acanthamoeba S100 extract, as was the case for mouse; but if purified Acanthamoeba RNAP I was added to the mouse system together with Acanthamoeba rDNA, there was an increase in nonselective transcription. The fact that supplementation of a nonhomologous mixture of transcription factor and template with homologous RNAP I results in an increase in nonspecific transcription indicates that transcription specificity resides with transcription factors that recognize promoter sequence(s) or polymerase or both. Miesfeld and Arnheim (101) and Mishima et al. (96) have reported the partial purification and characterization of factors required for RNAP I-dependent transcription of human and mouse rDNAs in vitro. Four fractions (designated A, B, C, and D) were

purified from the S100 extract on a phosphocellulose column, with fraction C containing the majority of the RNAP I activity (although some was also found in fraction D).

Various reconstitution experiments indicate that fraction D is an absolute requirement for correct and efficient transcription initiation by RNAP I on mouse and human rDNA. Fraction A enhances the transcriptional activity of fractions C and D, and fraction B (replaceable by poly (ADP-ribose) polymerase) very effectively suppresses random initiation between the two species. Fraction D is the only one that is not interchangeable. Although there is evidence that fraction D is responsible for the species specificity of the S100 extracts, this is not strictly so because it was found that mouse fraction D could support accurate transcription initiation using a rat template and human A, B, and C fractions. Human fraction D would not do the same.

Reconstitution/competition experiments using human fractions A-D and several different human rDNAs showed that some factor in fraction D was responsible for the formation of a very stable preinitiation complex with the 5' upstream rDNA sequence in a species-specific manner. The formation of this preinitiation complex is absolutely required for accurate and efficient transcription with RNAP I.

Clos et al. (95), using a four-step purification procedure, were able to partially purify a transcription initiation factor they designated TIF-IB. As deduced from its behaviour in prebinding competition experiments, TIF-IB appears to be analogous to fraction D described by Miesfeld and Arnheim (101) and Mishima et al. (96). TIF-IB formed stable preinitiation complexes as did fraction D. Using point-mutated and DNA-deleted rDNA templates, it was shown that TIF-IB (fraction D) binds to a region 5' to the transcription start site, between

-14 and -35, and that a G at position -16 (a position previously shown to be essential for in vitro cell-free transcription of rDNA) is essential for stable binding.

TIF-IB (fraction D) protects a stretch of rDNA bounded by nucleotides -7 and -21 in footprinting experiments. This protection was abolished if the rDNA contained the -16 G to A point mutation, emphasizing the importance of this nucleotide in the efficient stable binding of TIF-IB (fraction D). Initial size determination of TIF-IB by columns and SDS-PAGE indicates that it is an 80 kDa protein, but further purification may be necessary.

The 1981 report by Grummt's lab (90) presented experimental evidence that the species-specific nature of rDNA transcription is conserved in distantly related mammals and protozoans, although in the more closely related species of mouse and Xenopus, species specificity does not seem to hold up (187). Working with a protozoan system (the amoeba Acanthamoeba castellanii), Kownin et al. (102) constructed a number of rDNA deletion mutants with Bal31 and carried out similar experiments to those done with the mouse system. These workers showed that the region from -31 to +8 was essential for transcription and that the region from -47 to -32 would augment transcription.

Bateman et al. (103) have carried out protein binding and footprinting studies using Acanthamoeba RNAP I, rDNAs, and transcription initiation factors (TIFs). The TIFs protected a region extending maximally from -12 to -69, while an RNAP I plus TIF complex protected a region from -69 to +20 on wild-type rDNA templates. The TIFs from Acanthamoeba have biochemical and physical characteristics that differ from those determined for mouse. In prebinding studies, the Acanthamoeba TIFs could be competed out by a second wild-type rDNA template. This is not seen with the mouse TIF,

indicating that in Acanthamoeba, the TIF(s) must form a less stable complex with the rDNA than is the case for mouse. As well, the Acanthamoeba TIFs were found to have a $M_r$ of 250-350 kDa, whereas the mouse TIF-IB has a molecular mass of only 80 kDa.

6.vi) Transcription Promotion and Enhancement in Xenopus Spacer rDNA

The complete nucleotide sequence of Xenopus spacer rDNA has been determined (60). Putative limits to the Xenopus promoter were determined by studying mutant rDNA templates as had been done for the mouse and Acanthamoeba systems (63). The DNA sequences responsible for active and efficient transcription were found to reside between -145 and +6 bp (+1 being the immediate 5' nucleotide of the 40S pre-rRNA synthesized in vivo). This promoter region is imperfectly duplicated two to seven times in the spacer region. A BamHI site located within these promoter sequences led to their designation as "BamHI islands". The gene promoter and spacer promoters are separated from each other by regions of intermingled 60-bp and 81-bp repeating elements. The 60-bp and 81-bp elements are homologous in sequence except that they differ by a 21-bp insertion/deletion. It is also of interest that 60/81-bp repeats share a 42-bp sequence with the gene promoter.The functional significance of the spacer promoters and the 60/81-bp repeats will be discussed below.

To the left of the last spacer promoter is a region of ca. 1600 bp that shows no homology with the spacer and gene promoters or the 60/81-bp sequences but which contains two types of repetitive element, designated 0 and 1, interspersed with regions of unique sequence DNA.

Studies by Dunaway and Reeder (72) provided evidence of three stretches of Xenopus promoter DNA (I-III) that were protected from DNase I digestion when incubated in oocyte homogenate. Region I spans the site of transcription initiation (-10 to +15). Region II runs from -7 to -100 and includes the 42-bp (-72 to -119) sequence that shows homology with the 60/81-bp repeats; and region III is delineated by nucleotides -120 to -140. All three regions reside within a span of DNA that was shown by Moss (63) to be essential for transcription initiation in vitro.

Functional differences between the three regions came to light during DNA protection studies. Protein factors would bind to region I independent of other regions of rDNA; regions II and III would support binding independent of region I, but deletion of region III diminished binding to region II. Regions II and III are, therefore, somewhat interdependent. It seems that there are two functional domains or motifs within the DNA promoter region of Xenopus, one spanning the transcription initiation site (+15 to -10) and a second extending from -70 to -140.

The previously mentioned homology between the 60/81 bp repeats and the rDNA promoter suggested that these repeats might be involved in the binding of transcription factors. Biochemical experiments in vitro have shown that a promoter attached to a string of 60/81-bp repeats is dominant over one without the same repeats, and that these repeats alone will successfully compete with a promoter on a second rDNA template when incubated in an in vitro homogenate. All of the preceding evidence supports the notion that the 60/81-bp repeat actively binds a factor(s) that also binds to the rDNA promoter.

It came as a surprise, then, that the 60/81-bp repeats were not protected in footprinting studies under experimental conditions where binding to

the promoter was seen (72). It was argued that this result might be explained by the fact that in oocyte injection experiments, only supercoiled 60/81-bp repeats successfully competed with the promoter, whereas a linear template had been used in this study. In these oocyte experiments, enhancer-like behaviour of the 60/81-bp elements was seen. Both enhancers and the 60/81-bp repeat will function to stimulate the promotion of transcription when oriented in either direction relative to the promoter, and can act at a distance. Reeder's lab (73) further demonstrated that the 60/81-bp sequences will function even when inserted within the transcriptional unit whose promoter they enhance. Their insertion downstream of the promoter does not interfere with transcription. With the discovery in Xenopus of the spacer promoter sequences, it was felt that given their similarity with the true pre-rRNA promoter (BamHI islands), these particular repeats might be able to support transcription initiation. Moss (64) designed and carried out experiments which indeed demonstrated their ability to do so. He also found that the transcripts from these promoters terminate at a site 213 bp upstream of the pre-rRNA initiation site.

Competition experiments revealed that if the BamHI islands are deleted, such deleted rDNAs cannot compete with wild-type rDNA for transcription factors; therefore, it seems that these sequences act to produce in vitro what is in effect a "high density region" of bound RNAP I and any other transcription factors.

Moss's group (65) has also demonstrated that the 60/81-bp sequences will not enhance transcription without the presence of a spacer promoter. The removal of the spacer promoter reduces transcription from the pre-rRNA promoter by 90% of wild type. It seems, then, that both elements need to be present to obtain full enhancement of transcription by RNAP I.

The Xenopus promoter has been shown by deletion mapping to lie between bases -142 to +6 (63,69,104), and to contain two distinct domains. These two domains have been termed the Core Promoter (CP) and the Upstream Control Element (UCE). The core promoter is found in the region of +20 to -45 with respect to the site of initiation (+1). The UCE is found near -140, and is seen under stringent transcriptional conditions wherein it augments transcription 15-100 times above the basal level directed by the CP alone. Linker-scanning mutagenesis has better localized the CP to between -43 and +10 and the UCE to around -140 to -128 (105). The two promoter elements are very sensitive to the spacing between them.

Point-mutation analysis of the core promoter by Moss's group in 1989 (68) revealed several interesting features. Using a bisulfite mutagenesis technique, this group introduced 27 point mutations within a 78-bp region extending from -64 to +14 . These mutants were microinjected into oocytes for assay, with the experiments demonstrating that the 3' boundary of the core promoter was located at +4 and the 5' boundary was between -33 and -39. Interestingly the G residues at -7 and -16 were shown to be essential for promoter function as has been demonstrated for the mammalian core promoters (93).

Moss et al. (68) also found that the four regions, +2 to +4, -7, -12 to -20 and -28 to -33 interact with the proteins involved in transcription initiation. As well, single-point mutations between +3 and +5 could significantly reduce RNA stability, raising concerns that work done with deletion and linker-scanning mutants, which demonstrated reduced transcription, might in part be due to altered rRNA stability.

6.vii) Drosophila melanogaster Spacer rDNA: Physical Organization and Transcription Initiation

The physical organization of Drosophila spacer DNA is very reminiscent of that seen in Xenopus, in that the NTS is a complex array of different tandemly repeated DNA sequences. Starting at the 3' end of the LSU gene there is a 95-bp tandem repeat and then a 330-bp tandem repeat which is followed by a 240-bp repeat which terminates 140 bp upstream from the transcription initiation site. This general arrangement is found in all Drosophila species studied to date (D. mauritiana, D. simulans, D. virilis and D. hydei). Length heterogeneity of the rDNA repeats within and between species is due to fluctuations in the number of 240-bp repeats.

Using a recombinant plasmid containing the rRNA promoter attached to the bacterial chloramphenicol acetyltransferase ( CAT ) gene, and introducing this construct into Drosophila cells, Grimaldi, and DiNocera (114,115) showed that the DNA sequence between -180 and +34 is sufficient for accurate transcription. Kohorn and Rae (111,112) have demonstrated that a major component of the rRNA promoter is found between -43 and -27 and that a second component lies within the first four nucleotides of the 5'-ETS.

Interestingly, Miller et al. (113) have shown that transcription can be initiated from any of the repeated 240-bp sequences within the 'NTS', resulting in transcripts ranging in length from 240 to 1680 nt and differing from one another by 240 nt. In competition experiments the NTS was shown to confer a transcriptional advantage: the greater the amount of NTS DNA attached to the promoter, the more efficiently the promoter was transcribed. In other words, the NTS acted like an enhancer of transcription. Notably, when the 240-bp repeat is compared with the sequence between -19 bp and +25 bp at the site of

transcription initiation, 42 identical positions are found. Further work by these workers using rDNA co-transfected into cultured Drosophila cells revealed that the activity of the pre-rRNA promoter is directly correlated with the number of 240-bp repeats and that sequences upstream of these are also able to enhance transcription from this promoter. Unlike true enhancers these sequences are orientation-dependent, although they seem to be distance-independent.

6.viii) Human rDNA Transcription Initiation

In 1982 Financsek et al. (106) determined the site of RNAP I transcription initiation by $S_1$ nuclease protection experiments on human rDNA. Subsequently Tjian and coworkers worked out many of the details of the promoter structure and the proteins that interact with the DNA to initiate transcription (107-110).

Deletion mutational studies revealed that the sequence between -158 and +18 is necessary for efficient transcription initiation. This region was shown to be a composite of two functionally different DNA sequences: a central core of 40 bp that was required for transcription (-26 to +2) and flanking sequences that affect the efficiency of transcription in vitro (-158 to -26).

Further promoter deletion experiments in vivo and linker scanning mutation studies in vitro identified two separate blocks of sequence that play a role in directing efficient transcription of the human rDNA genes. The core promoter was located between +7 and -45 and is absolutely required for transcription, while the upstream control region (a region which modulates the level of transcription) was localized between -140 and -108.

As one aspect of these in vitro studies, whole cell HeLa extracts were fractionated to identify factors involved in transcription, thus identifying

two active fractions. One contained the endogenous RNAP I while the second contained a protein factor termed SL1 which was shown to confer RNAP I promoter selectivity. SL1 has been extensively purified and shown to be the factor responsible for species-specific recognition in whole-cell extracts, in spite of the fact that it could not be demonstrated that it interacted with the promoter DNA in DNase I footprinting experiments. What was also discovered was that a second factor, UBF1, recruits SL1 to the promoter and directs binding to a region encompassing sequences in the UCE. This binding was shown to modulate the efficiency of rRNA synthesis, a finding in keeping with the demonstration that this part of the promoter is not essential for transcription initiation, although its absence greatly reduces the efficiency of transcription in run-off experiments.

UBF1 recognizes sequences in the proximal portion of the UCE (termed site A) and protects the region between -75 and -115 from cleavage by DNase I. When SL1 and UBF1 are both present, the region between -75 and -165 is protected. This encompasses site A and a second site termed B. DNase I analysis of mutations induced in site A revealed that protection of site B requires sequences in both site A and site B. On the other hand mutations in site B interrupt SL1-dependent binding but do not alter UBF1 interaction with site A. If site A and site B are separated by insertion of base pairs, SL1 will not bind to site B but UBF1 will protect sequences in site A; thus, it seems that these two proteins require intimate interaction for binding and protection of the DNA in this area. Mutations introduced into the UCE that inhibit or prevent binding of UBF1 to site A or SL1 to site B will also diminish the UCE's ability to augment transcription from the core promoter.

UBF1 has been purified from the RNAP I pool fraction by DNA affinity chromatography. Two closely related polypeptides, 94 kDa and 97

kDa in size, were shown to possess the UBF1 DNA binding activity. The purified UBF1 was shown to bind to a region between -75 and -114 which is almost identical to that protected using the crude UBF1 preparation. In addition it was shown to act in the same way with SL1 and RNAP I in transcription initiation of the cruder extracts.

Tjian and coworkers have put forward a model of human rDNA transcription initiation taking into consideration all of their laboratory results. The first step is postulated to be the binding of UBF1 at both the UCE and core sites. Then SL1 interacts with bound UBF1 at both sites. The resulting UBF1-SL1 complexes are then recognized by RNAP I resulting in transcription initiation.

6.ix) Rat rDNA Transcription Initiation

Molecular biological studies of rat rDNA have elucidated the physical organization of its NTS DNA and the sites of transcription initiation, termination and enhancement (117-124). As has been found to be the case in the human and mouse systems, rat rDNA contains an RNAP I promoter which is comprised of a core promoter between -31 and +1 and an upstream promoter element located between -162 and -122. The core promoter is sufficient for accurate transcription of the 45S pre-rRNA, while the upstream promoter element (UPE) increases the efficiency of transcription.

In addition to the 45S pre-rRNA promoter there is a spacer promoter sequence located between -731 and -713. Thirteen nucleotides within this region are identical to nucleotides -18 to -6 of the 45S pre-rRNA promoter. This spacer promoter proved to be an order of magnitude less efficient than the 45S pre-rDNA promoter when the two were compared individually; but the 45S pre-rRNA promoter was fifty times more efficient when the two were part of the

same DNA in a cis-orientation (120). This stretch of DNA had previously been found to contain an enhancer (-1018 to -286). Within the region -620 to -417, this enhancer sequence has been shown to interact with high-mobility-group (HMG) - like DNA-binding proteins. These proteins also bind to rDNA in the 5'-ETS.

Protein-binding studies using the pre-rRNA and spacer promoters have shown that there are at least two transcriptional factors (SF1 and SL1) that interact with homologous regions both to initiate and enhance transcription. In rat, SL1 is needed for transcription by heterologous extracts in vitro.

DNase I footprinting experiments revealed that SF1 and SL1 interact with the UPE. SF1 protects a 40 bp region of the spacer promoter (-91 to -132) and SL1 protects a 25-bp region just upstream of the SF1 site (-165 to -187). The homologous sites were protected in the 45S pre-rRNA promoter.

## 6.x) Spacer rDNA and Transcription Termination

Transcription termination has become better understood but is a much more complex problem than was originally thought. Research carried out in the frog, mouse, yeast and Drosophila systems has unveiled a great deal about transcription termination. It appears that at least in the frog and mouse systems, termination is intimately involved in initiation and its enhancement (66,67,74-76,125-134).

### Xenopus rDNA

Initially, it had been presumed that rRNA transcription terminated at or very near the mature 3'-end of the LSU rRNA. However, in 1986 DeWinter and Moss (66) demonstrated that the actual site of termination is at

position -213 bp (+1 being the site of transcription initiation of the pre-rRNA) in the NTS. In the same year Labhart and Reeder (74) characterized three sites of rRNA 3'-end formation in Xenopus. The first, termed $T_1$, is a site of rapid rRNA processing and is located at the 3'-end of the LSU rRNA sequence. $T_2$, which is 235 bp further downstream of $T_1$, is a point at which the transcript is rapidly processed. The RNAP I continues to traverse the entire NTS until reaching position -213 where the polymerase is released. This site, $T_3$, is the one identified by DeWinter and Moss. It shows sequence similarity with $T_2$. Mutagenesis and deletion studies of this region have shown that a 7 bp sequence, which is conserved among species of Xenopus, is necessary for termination (76); interestingly, this same sequence has an enhancing effect on transcription promotion from the 45S pre-rRNA promoter (75). Mutations in the $T_3$ box which eliminate termination also eliminate this enhancement effect on initiation (75).

Mutations within the $T_2$ box (at the homologous 7 bp as in the $T_3$ box), result in the failure of rRNA processing at $T_2$. Curiously, mutations within the $T_3$ box also eliminate processing at $T_2$. The $T_2$ box will direct processing provided the $T_3$ box is downstream from it and in the correct orientation. In other words, the relationship between $T_2$ and $T_3$ is distance-independent but orientation-dependent.

A more detailed study of the $T_3$ box and its surrounding sequences and their involvement in transcription termination has been carried out using point mutational analysis and micro-injection techniques (76). These experiments demonstrated that the 7 bp sequence at the $T_3$ box was sufficient for correct and efficient termination, provided it was in the context of its usual flanking sequences. A G to A transition at -195 eliminated termination in this system, while mutations in flanking sequences affected termination as well. A G

to A mutation at -200 caused loss of termination if the assay conditions were less than optimal. A mutation at -216 resulted in some readthrough and caused a shift in the site of 3'-end formation. The above three mutations also caused a reduction in initiation. Other mutations actually stimulated transcription, particularly those in the GC -rich sequences flanking the $T_3$ box.

Mouse rDNA

As in the case of Xenopus early work indicated that transcription termination occurred very near the mature end of the LSU rRNA gene. (It was thought to occur 30 bp downstream of the 3'-end of the LSU rRNA.) Later work demonstrated that transcription actually terminates 565 bp downstream of the mature 3'-end of the LSU rRNA. This region of the rDNA possesses tandemly repeated sequences which contain within them a SalI site; hence, they are called SalI boxes. The DNA repeats contain a highly conserved 18 bp sequence that is actually made up of two elements, one 11 bp and the second 4 bp (both of which are perfectly conserved), and separated by 3 bp which are partially conserved. The remaining sequence is mostly stretches of pyrimidines.

Termination occurs upstream of the first SalI box, $T_1$. Competition experiments revealed that these sequences interact with some factor. Exonuclease III is precluded from digesting sequences in the 3' region of the SalI boxes. Mutations in the consensus sequence that reduce or eliminate transcription termination affect binding of the factor to the SalI box.

Like the situation in Xenopus a site of transcription termination has been identified just upstream of the transcription initiation site. This site contains a SalI box at position -171 to -154. The sequence of this site differs from the consensus sequence with changes at positions 16 and 18. This site, which has

been designated site $T_0$, interacts with the same factor(s) that binds at $T_1$, although it does so at a reduced efficiency (70%). In contrast to the Xenopus case, this site of termination in mouse rDNA seems to act on RNAP I which has initiated transcription from spacer promoters, as opposed to the 45S pre-rRNA promoter. Like the situation in Xenopus, this upstream termination site modulates transcription initiation from the pre-rRNA promoter.

Drosophilia rDNA

Tautz and Dover (133), using $S_1$ mapping and nuclear run-on assays, demonstrated that there was no identifiable site of RNAP I termination in the rDNA spacer and that transcription continued along the entire length of the spacer.

Yeast rDNA

Four putative sites for transcription termination have been identified in the yeast rRNA spacer. $T_1$ is a site between +15 and +50, with +1 the first nucleotide 3' to the end of the mature LSU rRNA. This site is actually three sites of rRNA processing (+20, +35, +50). $T_2$ is at position +210 and is thought to be the functional RNAP I terminator. $T_3A$ (+690) and $T_3B$ (+950) are thought to be "fail-safe" termination motifs.

Yeast is unusual in that the 5S rRNA gene is located in the rDNA spacer, thus dividing the spacer into two parts, $NTS_1$ and $NTS_2$. $T_1$, $T_2$, $T_3A$ and $T_3B$ are all located in $NTS_1$. No sequence similarity could be discerned around $T_1$, $T_2$, $T_3A$ and $T_3B$. It was also found that $NTS_2$ contains a terminator (Tp) 300 bp upstream from the rRNA promoter. It is less efficient than $T_2$ and does not seem to have any enhancing ability with respect to the pre-rRNA promoter.

The region which contains $T_2$ was also known to contain an enhancer activity. Mestel et al. (134), using deletion mutagenesis on a plasmid which contained a centromere and an rRNA minigene, were able to localize the enhancing ability to a 160 bp region. A short distance upstream (15 bp) from the 3'-end of the enhancing region was located 22 bp known to mediate RNAP I initiation in vitro. Deletion of the termination-active DNA region also resulted in reduction of 35S rRNA transcription.

6.xi) Summary

The efficiency and accuracy of RNAP I transcription initiation and termination depend upon circumscribed DNA sequences and trans-acting factors. In vitro transcriptional studies have mapped two important DNA motifs that together constitute the promoter. Sequences spanning the transcription start site (-45 to +10) have been shown to be the more important of the two regions. Within this region (the core promoter, CP) are located two highly conserved nucleotides, both of which (-16 and -7) are absolutely essential for transcription.

A protein factor, variously termed TIF-IB, fraction D, or SL1, is responsible for the species-specific nature of transcription initiation and requires the wild type nucleotide at position -16. The factor has been shown to protect this region from nuclease digestion in footprinting studies. This DNA-protein complex is a stable one and confers upon RNAP I the ability to carry out accurate and efficient initiation. In the presence of this transcription factor and RNAP I, the region of protection is extended through the initiation start site and into the 5' transcribed region.

The second rDNA motif linked to initiation spans the region from ca. -50 to -150 (the upstream control element, UCE). In the human rDNA

transcription model, the UCE is not essential for transcription initiation but modification of this region greatly reduces the efficiency of transcription. Experimentation has revealed that the UCE is bipartite in nature. Both parts of the UCE are bound by UBF1 and it is this protein which facilitates the binding of SL1 ( the protein which confers specificity to initiation) to the -115 to -165 region. A similar functional organization of the promoter has been determined for mouse, rat, Acanthamoeba, Xenopus, and Drosophila rDNA.

Enhancement of RNAP I transcription is known to occur with Xenopus rDNA. The elements responsible for this activity are 60/81-bp repeats of DNA found upstream of the gene promoter. These repeats are highly reiterated and have been found to be interspersed with two to seven spacer promoters. An internal conserved 42 bp of the gene promoter is found as part of the 60/81-bp repeat. This sequence, which is functionally important for transcription from the gene promoter, is thought to be responsible for binding a protein factor(s) and endowing the 60/81-bp repeat with enhancer-like properties. Without the spacer promoters (which have been shown to initiate transcription), the 60/81-repeats will not enhance transcription from the pre-rRNA promoter.

Drosophila exhibits a similar spacer organization to Xenopus (i.e., short repeat elements). One of these repeats (the '240') contains within it 42 bp of complete identity with sequences of the pre-rRNA core promoter. The 240-bp repeats can initiate transcription, resulting in NTS transcripts. They also "enhance" transcription from the true promoter, but in an orientation-dependent fashion.

Previously described transcription termination sites have been called into question by in vitro studies carried out in the mouse cell-free system. $S_1$ mapping studies using presumed legitimate primary pre-rRNAs placed the 3'

end of the primary transcript at or very near the 3' end of the mature LSU rRNA in a number of systems, including mouse. It is now known that the transcription complex traverses the entire NTS and terminates several hundred nucleotides upstream of the promoter at $T_0$. Rapid processing occurs at repeated SalI box sequences 565 bp downstream of the 3' end of the LSU rRNA ($T_2$). These sequences bind some factor(s) involved in termination. $T_0$ is also involved in initiation from the pre-rRNA promoter.

In Xenopus, transcription also extends through the NTS and terminates just upstream of the pre-rRNA promoter. Again, protein factors are involved and this site is implicated in promotion of transcription from the site of transcription initiation.

## 7.) RIBOSOMAL SPACER AND MECHANISMS OF GENE AND GENOME EVOLUTION

There is a long-standing enigma in molecular biology. It was proposed by Brown et al. in 1972 (135) and it concerns the now more widely recognized phenomenon of intra-species homogeneity versus inter-species heterogeneity of tandemly repeated gene families. In other words, there is an observed disparity between the "degree of sameness" between members of a repeated DNA family within the individual or within a population of individuals, and between members of the homologous family in a different species. To describe this phenomenon of intra-species homogeneity coupled with inter-species differences for sequences of a tandemly repeated gene family, the term "concerted evolution" has been coined.

The genes under study by Brown et al. (135) which led them to articulate this riddle were the rDNA repeats of Xenopus laevis and X. borealis.

These investigators discovered that the sequences of the repeat members were basically the same in one species, but when the sequences were compared to those in the second species, the NTS exhibited differences whereas the rRNA coding regions were virtually identical.

Although many mechanisms have been proposed to explain this puzzle, two are favoured; these are unequal cross-over and gene conversion (188). Gene conversion (or homology-dependent replacement) is a process that results in the replacement of one segment of DNA by another segment without reciprocal recombination of outside markers. Unequal cross-over is a process in which a region is first duplicated by recombination between misaligned stretches of DNA sequence which possess limited, but sufficient sequence similarity, and then amplified by recombination between improperly aligned duplicates.

Dover and colleagues (136) propose a third mechanism, transposition, that may account for the intra-species/inter-species heterogeneity problem. In what they call "molecular drive" -- "a process of fixation that arises from biases in the non-reciprocal spread of information between chromosomes"-- the authors argue that biased unequal cross-over is operating to maintain intra-chromosomal (single locus) homogeneity, and that transposition and gene conversion are responsible for the inter-chromosomal (or double/multiple loci) array homogeneity.

All these mechanisms can be seen to keep arrays of repeated genes homogeneous in structure within species as opposed to between species, but in the case of the rDNA NTS, questions remain regarding the significance of those regions within the NTS that exhibit greater heterogeneity than other regions.

Recently Hillis et al. (188) demonstrated concerted evolution of rDNA repeats in the unisexual, parthenogenetic hybrid, Heteronotia binoei. The observed homogeneity appears to have occurred as the result of biased gene conversion.

## 8.) DNA METHYLATION

### 8.i) Occurrence

#### (a) Modified Nucleosides

In addition to the four well-recognized bases found in DNA (adenine, cytosine, guanine and thymine), other modified bases have been identified, usually in trace amounts. These include 5-methylcytosine ($m^5Cyt$) and 6-methylaminopurine (6-methyladenine) ($m^6Ade$), as well as several more exotic but much less widely distributed modified bases (137). The occurrence of these modified residues in DNA has long been recognized (141) but their function has remained largely speculative.

#### (b) Maintenance and De Novo Methylation

It has been known for some time that modified bases in DNA owe their origin to a biochemical event occurring very shortly after replication, with hemi-methylated DNA serving as template for DNA methyltransferase (139). As a result, the DNA methylation pattern is conserved from generation to generation. Any alteration in the methylation pattern must come about as a result of loss of methyl groups, de novo addition of methyl groups, or inhibition of the post-replication addition of methyl groups. It is thought that the latter situation (which has been termed maintenance methylation) is largely responsible for changes in methylation patterns in eukaryotic nuclear DNA, but there have also been reports

of demethylation activity in mammalian nuclear extracts. De novo DNA methyltransferases which methylate previously unmethylated DNA or partially methylated DNA have been well studied in prokaryotes and are responsible for protecting specific host DNA sequences from endogenous restriction endonuclease activity. De novo methylation activity has been reported with eukaryotic nuclear extracts, but the full significance of this activity has not been determined (137).

(c) Distribution

In eukaryotic nuclear DNA, $m^5$Cyt is by far the commonest modified base, and it has been determined that it is predominantly located in 5'-CpG-3' sequences, with small amounts also being found in 5'-CpA-3', 5'-CpT-3', and 5'-CpC-3' dinucleotides. In land plants, $m^5$Cyt is located in all cytosine-containing dinucleotides, with all dinucleotides making up a subset of the sequence CpXpGp, X being any of the four normal nucleosides (139).

The degree of methylation is species-specific, with no obvious phylogenetic relationship. The extent of cytosine methylation ranges from 0.02% [$m^5$Cyt/(Cyt + $m^5$Cyt) x 100] in Tetrahymena, to 0.17% in mosquito, 5.4% in calf thymus, and 31% in wheat (138). The dinucleotide CpG is under-represented in most organisms, and indeed the greater the extent of $m^5$Cyt, the lower the percentage of the CpG nucleotide. It has been suggested that the increased mutability of $m^5$Cyt (via spontaneous deamination) provides a basis for this phenomenon. If indeed $m^5$Cyt is important in temporal control of gene expression, it is assumed that the CpG dinucleotide is unevenly distributed and that the $m^5$CpG dinucleotide is even more so.

(d) Regulation

Analysis of specific genes has pointed out some general trends in $m^5$Cyt occurrence and gene expression. In constitutive or "housekeeping" genes the same pattern of methylation is found in all tissues, whereas tissue-specific genes exhibit tissue-specific methylation patterns. Also in this vein, it has been shown that inactive genes are methylated and active genes are hypomethylated, or in some cases unmethylated (139).

The pattern of methylation is determined using methylation-sensitive and -insensitive restriction endonuclease isoschizomers. The two sites most often analyzed are CCGG (HpaII/RspI) and GCGC (HhaI), which constitute about 10% of the methylatable (CpG) sites (142). This leaves 90% of the methylatable sites unassayable, meaning that genes that are regarded as being fully methylated might be unmethylated at sites that are not analyzed, and vice versa.

Sites can be classified into two groups, one methylated to the same extent in all tissues and the other showing a tissue-specific pattern. The sites in either group may be fully methylated, partially methylated, or completely unmethylated.

Yisraeli and Szyf (141) have classified genes into five paradigms according to their methylation pattern. Paradigm I includes genes fully methylated in non-expressing tissues and unmethylated in active tissues. The second paradigm contains genes that exhibit some tissue-specific hypomethylation occurring exclusively in active tissues, while other hypomethylations are site-specific and occur in non-expressing tissues as well. In Paradigm IV at least some sites are undermethylated in sperm, as well as in all other tissues. Genes remaining fully methylated are placed in Paradigm IV.

Paradigm V contains genes in which tissue-specific hypomethylation is observed but no correlation with the state of activity of the gene is found. This last category (Paradigm V) contains the group of genes transcribed by RNA polymerase I. Ribosomal DNA in Xenopus (142) is fully methylated in sperm and oocyte and remains highly methylated in all somatic cell types except for two regions that contain the 60-bp tandemly repeated sequence in the 5' spacer region.

8. ii) Mechanisms of Action

If one assumes that methylation is causally associated with gene expression (as some evidence indicates) (137), the next question is, "How is it that methylation accomplishes this task?" Speculation about the "how question" has led to some suggested mechanisms of action. Experimental evidence lending credence to one or other of these mechanisms is available.

(a) Chromatin Structure

In prokaryotic DNA, methylation prevents cleavage of host DNA by the endogenous host restriction endonucleases (141). In eukaryotes, the nuclear DNA is complexed with a multitude of proteins, including histones and HMG proteins, in a highly ordered structure resulting in the condensed chromosome. The nature of the three-dimensional arrangement of the chromosome in the nucleus is not known in any detail, but it is thought to involve interactions of the chromosome with the nuclear protein skeletal architecture. DNA-protein interactions must be important, and have indeed been shown to be important with regard to the higher order structure of the chromosome based on the histone-containing nucleosome. Any changes in protein-binding ability of the DNA could result in alterations, allowing or causing gene activation.

DNase I, an endonuclease, preferentially digests actively transcribed genes in whole nuclei, suggesting a chromosomal alteration in spatial arrangement that renders the DNA more susceptible to approach by and interaction with this enzyme. Although no direct evidence exists to show a causal relationship between demethylation and chromosomal alteration (and hence, increased gene activity), there are some studies that suggest such a link (143).

(b) DNA Methylation and Structural Changes in DNA

The steric interaction of protein and DNA will depend to a large extent on the "proper" conformation of the protein and DNA molecules. DNA is known to assume a number of molecular conformations depending on the environmental conditions in which it is found. Three general forms of DNA helical conformation are known, designated A, B, and Z. Form B is the well-known Watson-Crick helix. The Z helix is assumed by poly (dG·dC): poly (dG·dC) in a high salt environment. This is a left-handed helix with a zig-zag phosphate backbone. If the dC is replaced with $m^5dC$, a transition from the B form to the Z form can occur near physiological salt concentrations, especially in the presence of di- and polyvalent ions. This Z DNA does not readily package nucleosomes. Methylated cytosine could cause dramatic molecular conformational changes in the cell (139).

(c) Site-Specific Methylation

Simple cytosine methylation at a select site or two in or around a gene could prevent the binding of necessary protein factors, and hence activation of the gene. A slight variation of this would be that the binding of a protein to $m^5Cyt$ would prevent the binding of other protein factors necessary for transcription. Indeed, Boyes and Bird (196) provide direct evidence that this method operates in cells to

inhibit transcription of genes. A methyl-CpG protein has been identified which is the presumed protein involved in this type of transcription inhibition.

### 8.iii) Summary

Inverse correlations between the degree of DNA methylation of certain genes and the extent to which these genes are expressed have now been established in many eukaryotic systems. The causality of this association remains to be determined for many of these. Indeed, there is evidence that is hard to rationalize in the context of a simple demethylation-active gene/methylation-inactive gene phenomenon, so a caveat is necessary. In those systems in which causality seems certain, the regulatory function appears to be exerted via specific DNA-protein interactions, and/or perhaps DNA molecular conformational changes.

## 9.) AIMS AND OBJECTIVES

There is a tendency in science to focus on one or a few "systems" in the course of investigation in the hope of elucidating "facts" or principles which may then be used to explain the workings in many other systems. This is as true in molecular biology as in any other branch of the sciences. A good example is the amazing degree of attention the eubacterium E. coli and the eukaryote Xenopus have received and are receiving in the field of molecular biology. This usually leads investigators to use what is known in a few organisms to explain a phenomenon described in another, perhaps not closely related, organism. As well, investigators will often be coloured or prejudiced in their thinking with regards to what should and should not be expected. It is for

this reason that it is important not to concentrate on one or a few organismal systems.

The Protoctists form a very large collection of organisms which are known for their phylogenetic diversity, evolutionary antiquity and ecological ubiquity and importance. In addition to being ecologically important, many human and animal pathogens/parasites are found in this group; these are responsible for the deaths of many millions of humans each year and serious morbidity in many more millions.

Considering their biological diversity, the protoctists have been little studied, but many of those organisms that have been investigated have revealed unexpected novelties and quirks in their manner of existence, method of metabolism, or genetic repertory.

Crithidia fasciculata is a member of the flagellated protoctists and is closely related to the medically/husbandry important group of trypanosomatids which includes those organisms responsible for Chagas' disease, leishmaniasis, and nagana disease. C. fasciculata is easily grown axenically in defined medium near room temperature (25-27$^{o}$C), and was readily available to us through Dr. F.B. St.C. Palmer of this department.

Early studies carried out by Dr. M.W. Gray on the ribosomes and rRNAs from the cytoplasm of C. fasciculata (149) revealed the existence of additional RNAs of a small size, which Dr. Gray was able to localize to the large subunit of the ribosome (144). Control experiments ruled out artifactual production of these RNAs during their isolation (144). This complement of rRNAs had never before been described in any organism, so the question of their significance, in terms of the structure/function and evolution of the ribosome, was raised.

At the time of initiating the research described in this thesis, not a great deal was known about trypanosomatid rRNAs and their rDNAs. In 1980, Cordingley and Turner (145) published a study in which they described the identification of five small rRNAs from Trypanosoma brucei. These RNAs ranged in size from 115 to 205 nt. Three of the small rRNAs were unique, the remaining two being the 5.8S and 5S rRNAs. Two of the unique RNAs were localized to the ribosome LSU with the location of the third being unknown.

Prior to 1980 it was known that T. brucei contained heat-labile LSU rRNA. When heat-treated, the LSU rRNA would dissociate into two smaller rRNA fragments of about equal size (ca. 2000 nt). This type of heat-labile LSU rRNA was known to exist in the closely related kinetoplastid flagellate, Crithidia oncopelti (146). Shortly after the T. brucei study, Castro et al. demonstrated the existence of a heat-labile LSU rRNA in T. cruzi (147).

The kinetoplastid protozoa appeared to be an extremely interesting group of organism with regard to their rRNA composition. In these three species, at least, the RNA complement of their ribosomes was the most unusual known at that time. To better understand the role of the small rRNAs in C. fasciculata ribosome structure, function and evolution, I therefore decided to clone the rDNA, map the rRNAs to the cloned rDNA and to the genome itself, and to determine the primary sequence of the rDNA.

## SECTION B

## MATERIALS AND METHODS

### 1.) ENZYMES

Restriction endonucleases were purchased from one of the following suppliers:

(a) Bethesda Research Laboratories (BRL),

(b) New England Biolabs (NEB), or

(c) P-L Pharmacia Biochemicals.

T4 polynucleotide kinase was obtained from NEB; T4 RNA ligase from P-L Pharmacia Biochemicals, and T4 DNA ligase from BRL.

### 2.) GROWTH OF C. fasciculata

C. fasciculata was kindly provided by Dr. F.B. St.C. Palmer of this department and cultured as described by Maclean and Amiro (148) with minor modifications. Cultures (one liter in 2.8 liter Fernbach flasks) were grown for three days in the light.

### 3.) HARVESTING OF C. fasciculata

Cells were harvested by one of two methods, depending upon the nucleic acid to be isolated. For DNA isolation, 250 ml of cells were collected and treated as described in Gray (144) and for RNA isolation one liter of cells was centrifuged (100 x g), washed three times in 0.85% NaCl, and resuspended in 0.05 M Tris-HCl (pH 8.0).

## 4.) ISOLATION OF TOTAL CELLULAR RNA

Cells resuspended in 0.05 M Tris-HCl (pH 8.0) were extracted by the addition of an equal vol. of phenol-cresol mix [phenol/m-cresol/8-hydroxyquinoline, 500/70/0.5, v/v/w, saturated with 1X ψTE (0.01 M Tris-HCl (pH 8.0), 0.1 mM EDTA)] and vigorously shaken for 20 min. Phases were separated by centrifugation. After two more re-extractions of the aqueous phase with the phenol-cresol mix, RNA was precipitated by the addition of two vol. of 95% ethanol. The precipitate was stored at -20°C.

## 5.) ISOLATION OF NUCLEAR DNA

Harvested cells were washed once in Solution A [0.005 M $Mg(OAc)_2$, 0.06 M KCl, 0.05 M Tris-HCl (pH 7.6)] and pelleted. The cell pellet was resuspended in Solution A, filtered through glass wool, and lysed by passage through a French press (750 psi, medium setting). Cell lysis was monitored by light microscopy. Following centrifugation at 15,000 x g for 15 min, the pellet ("nuclear" pellet) was washed once and adjusted to a vol. of 10 ml with Solution B [0.44 M sucrose, 50 mM Tris-HCl, 1 mM $Na_2EDTA$ (pH 8.0)] and one vol. of each of the following was added: a detergent mix [2% (w/v) tri-isopropylnaphthalene sulfonate, 12% (w/v) sodium p-aminosalicylate, 0.1 M NaCl, 0.02 M Tris-HCl (pH 7.4)] and phenol-cresol mix saturated with 1X ψTE. The phenol-detergent extract was gently mixed on ice and then centrifuged to resolve the phases. Sufficient solid NaCl was added to the aqueous phase to make it 0.5 M and extraction with the phenol-cresol mix was repeated until no material was visible at the interface. The extracted DNA was precipitated by the addition of two vol. of 95% ethanol and stored at -20°C.

## 6.) PREPARATION AND PURIFICATION OF END-LABELLED rRNAs

Crithidia fasciculata contains an unusual complement of rRNAs. The ribosome small subunit contains a single species which is ca. 2200 nt long (Sp. b). The ribosome large subunit contains nine rRNAs, one of which (Sp. a) is heat-labile and four of which are unique to C. fasciculata (Sp. e (212 nt); Sp. f (183 nt); Sp. g (136 nt); and Sp. j (73 nt)) (144). Sp. a melts when heated to give rise to the 5.8S rRNA (Sp. i; 172 nt) and two, nearly equal sized, rRNAs (Sp. c and Sp. d). Sp. a has an estimated combined size of 3880 nt.

Total cellular RNA was dissolved in a final concentration of ca. 5 mg/ml for selective salt fractionation. This was accomplished by adding solid NaCl until the final concentration was 3 M and placing the 5 mg/ml-RNA solution at 0°C and centrifuging the precipitated RNA for collection (149). The salt-insoluble fraction (iRNA) was used for isolation of Sp. b, c, d, e, f, i, and j. Sp. g and h were isolated from the salt-soluble fraction (sRNA). Electrophoresis in a 2.5% polyacrylamide gel or a 7 M urea-1% agarose (to resolve Sp. b, c, and d) or a 2.5-10% stacker polyacrylamide gel (to resolve Sp. e-j) was accomplished as described by Schnare and Gray (163). Extractions were performed using Solution C (0.5 M $NH_4OAc$, 0.01 M $Mg(OAc)_2$, 0.1% SDS, 0.0001 M EDTA) under conditions similar to those described by Maxam and Gilbert (150). The electrophoresis and extraction steps were repeated to ensure the purity of the rRNAs. The unfractionated iRNA or individual high molecular weight rRNAs were partially hydrolyzed [0.15 M $NH_4OH$, 90°C, 2 min] before being 5'-end-labelled (151) using synthetically prepared [ $\gamma$ -$^{32}$P]ATP (152) and commercial T4 polynucleotide kinase. The small rRNAs were 3'-end labelled with [5'-$^{32}$P] pCp and RNA ligase before being further purified by electrophoresis in a 10% denaturing polyacrylamide gel.

## 7.) DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES, FRACTIONATION IN AGAROSE GELS, AND TRANSFER TO NITROCELLULOSE FOR PROBING

DNA was hydrolyzed with several-fold excess of restriction endonucleases using conditions suggested by the suppliers. High molecular weight DNA samples (1-20 kbp) were subjected to an electric field of 1-2 V/cm in a 1% agarose gel contained in a horizontal flat bed apparatus. The separated fragments were stained with ethidium bromide, transilluminated with UV light, and photographed with an MP-4 Polaroid camera. Sizes of low molecular weight DNA fragments (<1 kbp) were determined by vertical electrophoresis (12.5 V/cm) in 5% polyacrylamide gels (20 cm x 20 cm x 0.3 cm). Lambda ($cI_{857}$) DNA digested with EcoRI, HindIII, or EcoR1 + HindIII was used to provide standards for molecular weight determinations of large DNA fragments, and pBR322 digested with AluI was used for small fragments. Denaturation, transfer of the DNA from the agarose gel to nitrocellulose, hybridization of end-labelled rRNAs to Southern blots (153), and autoradiography have been described (154).

## 8.) CLONING OF GENOMIC DNA AND ISOLATION OF rDNA RECOMBINANTS

PstI- or HindIII-digested pUC9 and nDNA, in a 1:4 mass ratio, were ligated with commercial T4 DNA ligase. Competent E. coli JM83 (156) cells were transformed with 10 ng of ligated DNA as described by Hanahan (157). Recombinant clones (colonies which were white as a result of the fact that the plasmid's β-galactosidase gene had been interrupted by insert DNA and therefore could not act on 5-bromo-4-chloro-3-indolyl-β-D-galactoside

in the medium to convert it to its blue product) were chosen for in situ colony hybridization (158) with 5'-end-labelled iRNA or 3'-end-labelled Sp. g+j. In situ colony hybridization was accomplished by floating the colony filters on 10 to 15 ml of the denaturation or renaturation solutions in Petri dishes, eliminating the need for the apparatus described by Grunstein and Hogness (158). The ethanol washes were performed using a Buchner funnel, and baked filters were prehybridized overnight in 50% (v/v) deionized formamide/2XSSC, 0.05 M sodium phosphate buffer (pH 7.0), 10-250 μg/ml sonicated salmon sperm DNA, 0.02% w/v bovine serum albumin-Ficoll, and 0.02% w/v polyvinylpyrrolidone (159). The probe was mixed with a small volume of prehybridization solution and added to the filter. Colonies seen to hybridize with iRNA or Sp. g+j were picked for further analysis. Mini-preparations of plasmid DNA (160) were digested with PstI or HindIII, electrophoresed in an agarose gel, transferred to nitrocellulose, and probed with labelled iRNA, Sp. g, or Sp. j. To obtain workable quantities of plasmid DNA, a large-scale plasmid preparation was carried out according to Schleif and Wensink (161), except that the chloramphenicol amplification step was eliminated.

## 9.) PRIMARY SEQUENCE DETERMINATION OF rDNA RECOMBINANTS

Primary sequence analysis of 5'-end labelled rDNA fragments was carried out using the chemical sequencing method (Maxam and Gilbert), as described by Spencer et al. (162). Five μg of DNA digested with the restriction endonuclease of interest was precipitated with ethanol, pelleted by centrifugation for 10 min., washed with 80% ethanol, dried under vacuum, and dissolved in 50 μl of Tris-HCl (pH 8.2). To this was added 0.2 units bacterial alkaline phosphatase and the mixture was then incubated at 65°C for 30 min. At

the end of the incubation, 0.1 vol. 3 M NaOAc was added and the solution was extracted with an equal vol. phenol-cresol three times before being precipitated by the addition of 200 µl of 95% ethanol at -70°C. The DNA was pelleted by centrifugation for 10 min, then redissolved in 50 µl TE (10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA) and precipitated again by the addition of 0.1 vol. of 3 M NaOAc and 3 vol. of 95% ethanol before being pelleted by centrifugation, washed with 80% ethanol, dried under vacuum, and dissolved in 37 µl of TE for 5'-end-labelling. To this was added 1 µl of 5 mM spermidine and the resulting 38 µl were used to dissolve the previously dried 50 µl [ $\gamma$-$^{32}$P]ATP. This was then heated at 65°C for 5 min and mixed by vortexing before 5 µl of 10X kinase buffer (500 mM Tris-HCl, 100 mM $MgCl_2$ (pH 7.6)) and 5 µl of 50 mM DTT were added and vortexed to mix. The liquid was spun down and 10 units of polynucleotide kinase were added to start the labelling reaction, which was carried out at 37°C for 1 hr. The DNA was precipitated by adding an equal vol. of 4 M $NH_4OAc$-1mM EDTA and 300 µl of 95% ethanol and placing at -70°C. The DNA was pelleted by centrifugation for 10 min, redissolved in 50 µl TE and precipitated in the same manner. It was redissolved in the same manner but precipitated by the addition of 0.1 vol. 3 M NaOAc and 3 vol. 95% ethanol. The pelleted DNA was washed with 80% ethanol, dried under vacuum, and redissolved in restriction endonuclease buffer for the secondary digestion, in preparation for agarose gel electrophoresis and fragment extraction by the hot phenol method.

5'-End labelled DNA which had been extracted from the agarose gel was precipitated in the usual way, dried, and redissolved in 20 µl of TE. It was then heated at 90°C for 30 sec, placed on ice and then distributed equally into four 1.5 ml Eppendorf tubes for chemical sequencing. These tubes were labelled G, A>G, C, and T.

To the G tube was added 200 µl of cacodylate buffer (50 mM Na cacodylate (pH 6.8)) which was mixed and spun briefly, after which 1 µl of MMS was added to initiate the reaction. The tube was vortexed immediately, spun briefly, and incubated at 90°C for 30 sec. To stop the reaction the tube was transferred to ice, followed by addition of 50 µl of G,A,T precipitation solution [1.0 M Tris-acetate (pH 7.5), 1.0 M β-mercaptoethanol, 1.5 M NaOAc and 0.1 mM EDTA] and 750 µl of 95% ethanol before being placed at -70°C.

To the A>G tube was added 200 µl of A>G buffer (50 mM NaOAc (pH 4.5)); the resulting mix was vortexed, spun briefly, and 0.5 µl DEPC was added to initiate the reaction. The solution was vortexed immediately, spun briefly, and placed at 90°C for 20 sec. It was then immediately placed on ice, after which 50 µl G precipitation solution and 750 µl 95% ethanol were added before the mix was placed at -70°C.

To the C tube was added 15 µl of 5 M CsCl, after which the solution was vortexed, spun briefly, and placed on ice. The tube was incubated at 25°C for between 5 and 6 min after the addition of 30 µl of hydrazine. At the end of the incubation period the tube was placed on ice, with the addition of 200 µl of U precipitation solution (0.3 M NaOAc) and 750 µl 95% ethanol at -70°C.

To tube T was added 200 µl of cacodylate buffer (50 mM Na cacodylate (pH 7.6)) which was mixed and spun briefly. Then 2 µl of fresh 10% $KMnO_4$ was added, mixed, spun briefly, and incubated at 90°C for 45 sec. The tube was then placed on ice and 50 µl G precipitation solution and 750 µl of 95% ethanol were added and the solution placed at -70°C.

The DNAs were pelleted by centrifugation for 10 min and the supernatants were discarded before the pellets were redissolved in 200 µl precipitation solution, followed by 600 µl of 95% ethanol at -70°C. The DNA was pelleted again (10 min centrifugation), the supernatant was removed, the tube

was washed well with 80% ethanol and the DNA was re-pelleted before being dried in the vacuum lyophilizer. The pelleted DNA was redissolved in 25 µl of $H_2O$ to which was added 25 µl of fresh 1:5 diluted piperidine solution. The tubes were vortexed lightly, spun briefly, and placed at 90°C for 30 min. They were then spun briefly and frozen at -70°C.

All the samples were lyophilized and then redissolved in 25 µl of $H_2O$, frozen, and lyophilized again. They were redissolved a final time in 25 µl of $H_2O$ and were frozen and lyophilized. Each sample was dissolved in 3 µl of NMF/urea loading buffer (600 µl N-methylformamide (deionized), 95 µl $H_2O$, 10 µl 1 M Tris-HCl (pH 7.6), 10 µl 100 mM EDTA (pH 7.0), 50 µl of a solution containing 2% bromophenol blue plus 2% xylene cyanol, 300 mg enzyme-grade urea), and vortexed twice for 10 sec. The samples were then transferred to 0.5 ml Eppendorf tubes and heated at 65°C for 2.5 min, before being stored at 4°C overnight.

A 6% sequencing gel (33 x 40 x 0.05 cm with at least 48 wells) was prepared from a mixture containing 5.7 g acrylamide, 0.3 g bis-acrylamide, and 42 g urea in a final volume of 100 ml TBE buffer [25 mM Tris-boric acid (pH 8.3), 0.5 mM EDTA]. The solution was filtered, degassed, and polymerized with 20 µl of TEMED and 0.65 ml of 10% ammonium persulfate, and the gel was then prerun for at least 3 hr before vortexed and heated (65°C) samples were loaded onto the gel with micropipettes. Electrophoresis was carried out at 1500 V.

## 10.) PREPARATION OF CLONALLY DERIVED nDNA FROM CRITHIDIA FASCICULATA

Mid-to-late log ($1x10^7$ cells/ml) phase cells, grown in medium as defined by Maclean and Amiro (148), were counted using a haemocytometer, diluted to a cfu of ca. 200 (assuming a plating efficiency of ca. 50%) and spread

on fresh heart-brain infusion agar plates supplemented with hemin (10 μg/ml). The plates were sealed with Parafilm, inverted, and incubated in a high-humidity chamber at 25-27°C for ca. seven days (until the colonies were 1-3 mm in diameter).

Agar plugs containing single colonies were used to inoculate 2 ml of heart-brain infusion medium (supplemented with 10 μg/ml hemin) in Falcon 10 ml snap cap tubes. They were incubated at 25-27°C on a New Brunswick Rollodrum at the highest setting. Turbidity was apparent after only three days and light microscopy revealed a high density of active cells (with both free-swimming and rosettes of cells being present). Half of each of the 2 ml cultures was transferred to sterile Eppendorf mini-tubes, the cells spun down, washed in 25 ml TE (pH 8.5), and resuspended in 200 μl of the same buffer for DNA extraction. One-tenth vol. of 10% SDS solution was added, along with 0.1 vol. of 3 M NaOAc, and the resulting solution was extracted with phenol and precipitated with 3 vol. of 95% ethanol. Contaminants were removed by precipitating the DNA from 0.55 M NaCl, 0.01 M Tris-HCl (pH 8.2), 0.0002 M $Na_2EDTA$ by addition of an equal vol. of 0.5 M NaCl, 20% polyethylene glycol 6000. The resulting pellet was redissolved in 10 mM Tris-HCl (pH 8.0), 0.1 mM $Na_2EDTA$, and precipitated with ethanol. The final purified DNA was used in Southern experiments.


**Figure 3.** Photograph of EtBr-stained 7 M urea-agarose gel containing cellular RNA and trans-illuminated with UV light. The loading wells are at the top of the photograph and the positive electrode would have been at the bottom of the page. The lower case letters at the left hand border of the gel refer to the RNA species; **a** is the undenatured LSU rRNA, **b** is the SSU rRNA, **c** is the 5'-half of the denatured LSU rRNA, and **d** is the 3'-half of the LSU rRNA. The bands at the bottom of the gel are the various small rRNAs plus tRNAs. Each lane contains 5 mg of total RNA loaded at a concentration of 10 mg/ml.

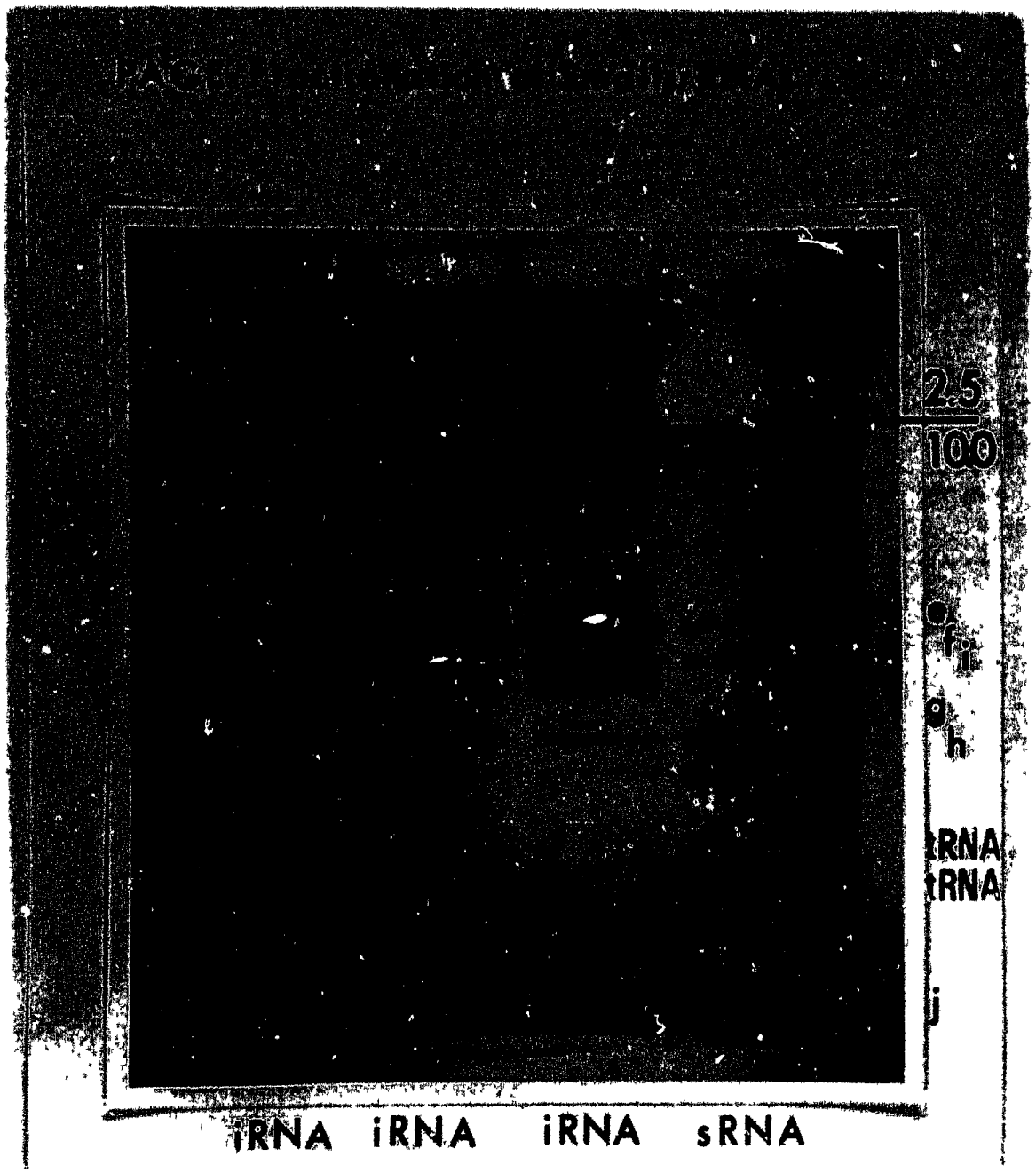



**Figure 4.** Photograph of 2.5%-10% stacking polyacrylamide gel shadowed with UV light. The three left hand wells contain 3 M NaCl-insoluble cellular RNA and the single right hand well contains 3 M NaCl-soluble cellular RNA. "2.5" refers to that portion of the gel which is 2.5% and "10.0" refers to that portion of the gel that is 10.0%. **iRNA** is 3 M NaCl-insoluble RNA and **sRNA** is that RNA which is soluble in 3 M-NaCl. **e-j** are the small rRNAs. **e** is Sp. e, **f** is Sp. f, **i** is Sp. i, **g** is Sp. g, **h** is Sp. h, and **j** is Sp. j. tRNA = transfer RNA.

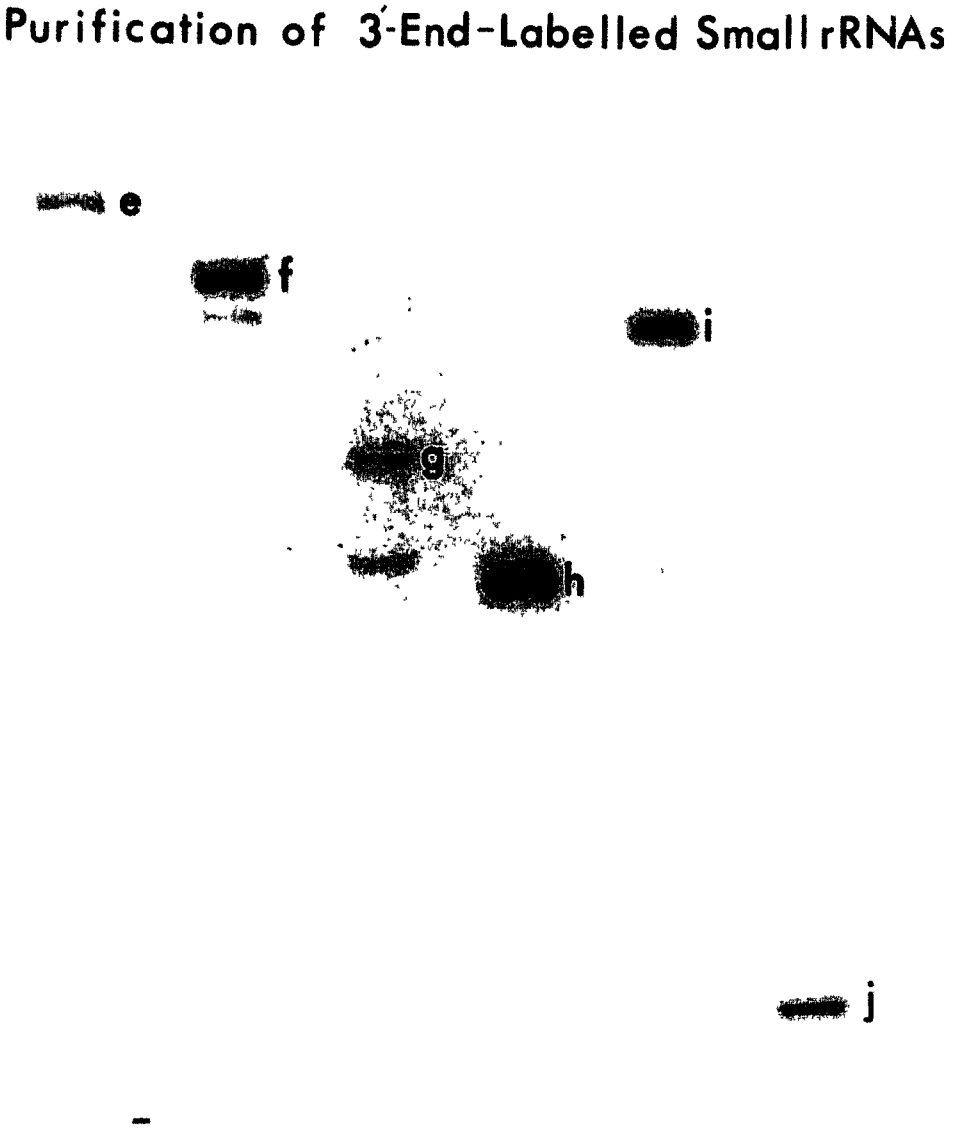



**Figure 5.** Autoradiogaph of 3'-[$^{32}$P]-end labelled Sp. e-j fractionated in a 10% polyacrylamide gel for purification purposes. **e** is Sp. e, **f** is Sp. f, **g** is Sp. g, **h** is Sp. h, **i** is Sp. i, and **j** is Sp. j.

## SECTION C

## RESULTS

### 1.) rRNA COMPONENTS OF C. FASCICULATA CYTOPLASMIC RIBOSOMES

In 1979 M. W. Gray identified an unusual complement of rRNAs from whole-cell extracts of C. fasciculata (149). In non-denaturing conditions he was able to isolate two high molecular weight RNA species. The largest ,Sp. a, was estimated to be ca. 3880 nt long and the shorter, Sp. b, was ca. 2460 nt in length. Upon heat denaturation Sp. a disappeared and three smaller RNAs appeared (Sp.c, d, and i). Sp. c and d were estimated to be ca. 2000 nt and 1720 nt long, respectively. Sp. i was a much smaller RNA, estimated to have a length of 176 nt. By analogy with the situation found in the closely related species, Crithidia oncopelti (146), Sp. a was considered to be the LSU rRNA and Sp. b the SSU rRNA. When the low molecular weight fraction of RNA was separated on a 2.4% polyacrylamide gel, four small RNAs were identified: Sp. e (238 nt), Sp. f (193 nt), Sp. g (135 nt), and Sp. h (120 nt). Sp. h was tentatively identified as the 5S rRNA due to the fact that it co-migrated with the wheat cytosol 5S rRNA. The identification of Sp. i as the 5.8S rRNA was based on its size and the fact that it associated with the LSU rRNA (Sp. a), a characteristic of other eukaryotic 5.8S rRNAs. The small RNAs were determined to be in near equimolar amounts to the large RNAs.

In a subsequent study (144), Gray was able to localize these RNAs to individual ribosomal subunits. RNA extracted from purified ribosomal subunits confirmed the existence of all of the RNAs previously identified plus an additional tRNA-sized RNA, Sp. j (73 nt). All the rRNAs were found to reside in the large subunit, except the unusually long Sp. b (2460 nt). Figures 3, 4 and 5 illustrate the polyacrylamide- and agarose gel-fractionated rRNAs.

At the time of initiating this work only the primary sequence for Sp. h (5S rRNA) was known (189). Shortly thereafter the primary sequences of Sp. e, f, g, i and j were determined by M.N.Schnare and D.F.Spencer in M.W.Gray's lab (163,166). In addition M.N.Schnare determined the primary sequence of the 3'-end of Sp. b (165) and the 3'- and 5'-ends of Sp. c and d, (unpublished information ).

Table 1 describes the various C. fasciculata RNA components and their subunit location.

This was the most unusual composition of cytoplasmic rRNAs discovered up to this time. It was with the hope of determining something about the origin and biochemistry of these rRNAs that the genomic mapping and rDNA cloning studies to be described here were undertaken.

## 2.) GENOMIC ORGANIZATION AND CLONING OF NUCLEAR rDNA

The exact nature of the nuclear chromatin in Crithidia fasciculata is unknown. No identifiable mitotic chromosomes have been seen in Crithidia or related organisms of the trypanosomatid group. Pulsed field gradient gel electrophoresis of the nuclear contents of Crithidia has recently revealed that, like other members of the same group studied in this manner, the chromosome complement consists of a defined number (>10 but <20) of chromosomes. Many of these are very small and have been called minichromosomes (190).

In an attempt to find a restriction endonuclease that would give suitably sized rDNA fragments for cloning purposes, nuclear DNA (nDNA) was digested with several different restriction endonucleases having a 6-bp recognition/cleavage site, and the resulting Southern blots were hybridized with the individual rRNAs, either 5'-end- or 3'-end-labelled.

TABLE 1

| RNA Species | Analog in Other Eukaryotic Ribosome | RNA Length (nt) *(estimate)*[1] | Gene Length (bp)[2] | Subunit Location | Positions in E. coli 23S rRNA[3] | Positions in Mouse 28S rRNA[4] | Gene Present in rDNA Cistron |
|---|---|---|---|---|---|---|---|
| **b** | 18S | *2460* | 2206 | Small | NA | NA | Yes |
| **i** | 5.8S | 176 | 171 | Large | 13-158 | 5.8S | Yes |
| **c** | 28S | *2000* | 1782 | Large | 168-1415 | 1-2302 | Yes |
| **e** | None | 238 | 212 | Large | 1420-1578 | 2325-2509 | Yes |
| **d** | 28S | *1720* | 1523 | Large | 1587-2625 | 2520-4222 | Yes |
| **f** | None | 193 | 183 | Large | 2630-2788 | 4226-4378 | Yes |
| **j** | None | 76 | 73 | Large | --- | (4379-4616) | Yes |
| **g** | None | 135 | 133 | Large | (2810-2904) | (4617-4712) | Yes |
| **h** | 5S | 120 | ND | Large | NA | NA | No |
| **Total** | | *7118* | | | | | |

[1] RNA lengths are estimates determined from polyacrylamide gel electrophoresis and using internal size standards.

[2] Gene length is the number of DNA bp which were determined to code for the individual rRNAs.

[3] See reference 199.

[4] See reference 200.

Figures 6, 7 and 8 show the results obtained when total genomic DNA was digested with SalI, EcoRI, and BamHI, respectively, and hybridized with the individual rRNAs. Each lane contains 1 μg of genomic DNA. A number of observations can be made from these experiments.

The resulting restriction pattern of C. fasciculata nDNA displays the characteristic complexity seen when nDNA from other eukaryotes is digested with a restriction endonuclease. Above the diffuse background, a distinct banding pattern can be discerned, with several DNA bands over-represented when compared to the smear of single-copy DNA. The pattern of such bands is unique to the particular restriction enzyme used.

In the SalI digest (Fig. 6, lane 8), the 5S rRNA (Sp. h) is seen to hybridize by itself to a very large DNA fragment ca. 30 kbp in size. The BamHI (Fig. 8, lane 8) digest also shows the 5S rRNA hybridizing by itself to give an intense band of ca. 25 kbp, as well as a second, less intense band of ca. 30 kbp. Interestingly, the 5S rRNA is seen to hybridize to three high molecular weight bands in the PstI digest (Figure 9, lane 9). The most intense band is ca. 11-12 kbp while the other two (much less intense) are ca. 20 kbp and 25 kbp in length. The 5S rRNA is, therefore, most likely physically unlinked with the remaining rRNAs. If the 5S rRNA coding region is organized physically like 5S rDNA in most other eukaryotes, these bands represent an island of DNA which contains multiple, tandemly repeated 5S rRNA genes. The multiple bands may indicate that these "islands" are located on more than one chromosome or they may represent sequence heterogeneity where a repeat has lost a PstI site, resulting in the production of a "dimer" of the 11- to 12-kbp repeat. A third possibility is that the large fragment represents a 5S rRNA repeat which flanks unique-sequence DNA.

All the rRNAs (except Sp. h, the 5S rRNA) hybridize to the same 11 to 12-kbp DNA fragment whether the DNA is digested with SalI or BamHI. The size of this DNA fragment is more than sufficient to accommodate the coding regions for the rRNAs (an estimated size of ca. 7100 nt, see Table 1 ).

The SSU rRNA (Sp. b) is also seen to hybridize in sub-molar amounts to a 6-to 7-kbp fragment in both SalI and BamHI digests. Two explanations for this have been considered. One is that these two enzymes liberate from the rDNA, a DNA fragment (of ca. 6-7 kbp) which has coding sequence for at least part of the SSU rRNA, leaving the remaining rDNA (ca. 11-12 kbp) intact (i.e., the SSU rRNA gene has internal SalI and BamHI sites). This would give a combined rDNA length of ca. 17-20 kbp. A second possibility is that a sub-population of SSU rRNA coding regions are contiguous with unique sequence DNA.

In the EcoRI-digested nDNA (Fig. 7) , the 5S rRNA is seen to hybridize to a single, tight band of rather large size (ca. 30 kbp; small arrow). This size is seen to be at the upper limit of DNA fragments visible in the stained agarose gel. All the remaining rRNAs hybridize strongly to a broad band of DNA fragments of between 8.5-9.5 kbp in size ( large arrow). Sp. b hybridizes lightly to the same broad band(s) but more intensely to a ca. 3000-bp-sized DNA fragment (curved arrow). The intensity of this band approximates that of the combined intensities of the larger DNA band(s) to which the remaining rRNAs (excluding Sp. h) hybridize. Interestingly the combined sizes of the Sp. b binding nDNA fragments are between 11.5-12.5 kbp; similar in size to the nDNA fragments hybridizing all of the rRNAs in the SalI and BamHI digestions (ignoring the weak band in the Sp. b lane). It seems that EcoRI has sites within the rDNA, so that EcoRI digestion separates a large part of the coding region of Sp. b from the remaining rDNA, which is left intact.

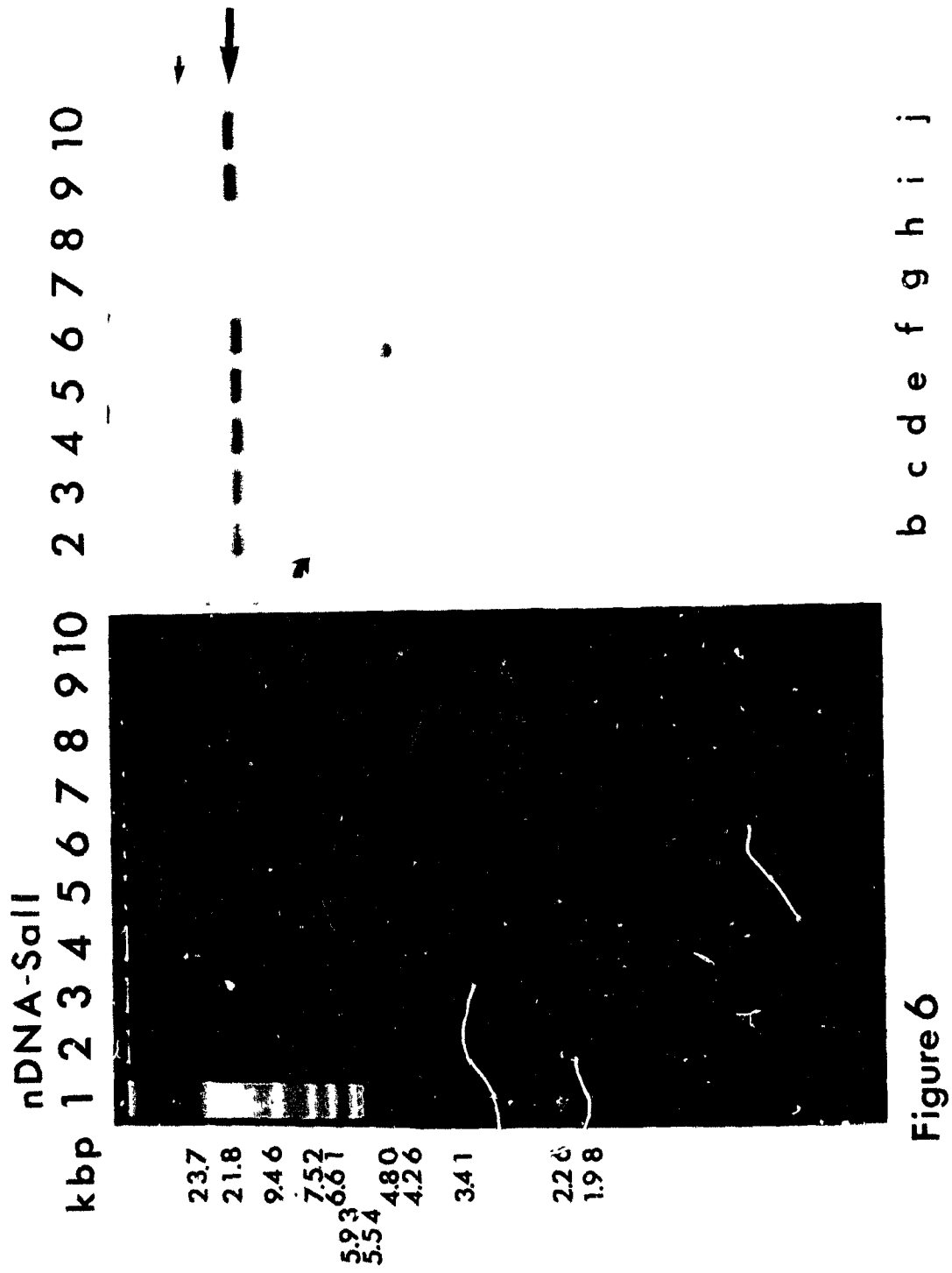



**Figure 6.** EtBr-stained 1% agarose gel and Southern blot autoradiograph of nDNA digested to completion with SalI and probed with $^{32}$P-labelled individual rRNAs. Lane 1 contains lambda DNA digested with HindIII and EcoRI; the sizes are in kbp. Lanes 2-10 contain nDNA digested with SalI. The right hand side of the figure contains the autoradiograph of the Southern blots probed with Sp. b-j. The large arrow points to the main band of rRNA hybridization and the small straight arrow indicates the ca. 30 kbp band that hybridizes to the 5S rRNA (Sp. h). The small curved arrow points to the 6-7 kbp weakly hybridizing band in lane 2. b-j indicate which rRNA species was used to probe that particular lane (i.e., b is Sp. b).

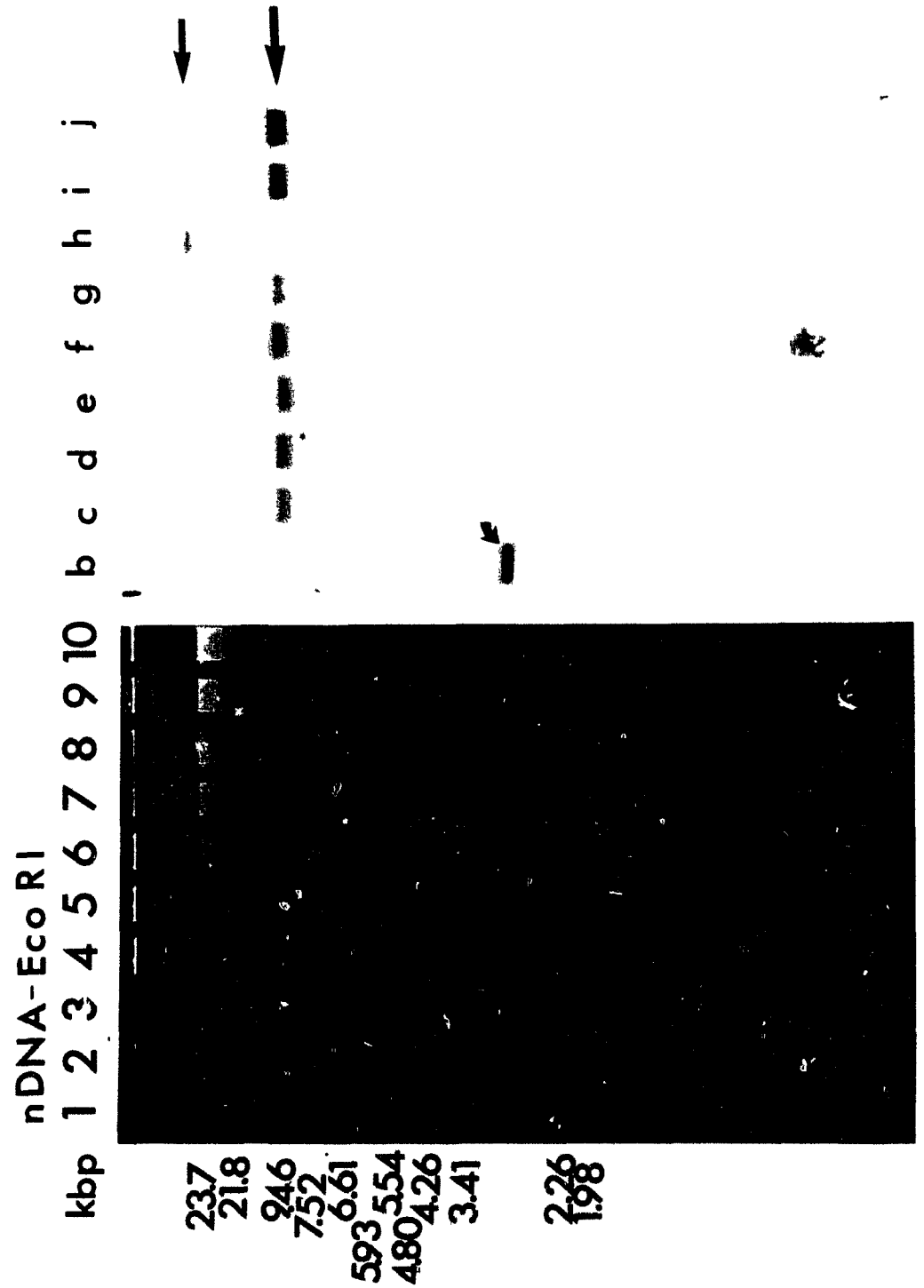



Figure 7

**Figure 7.** EtBr-stained 1% agarose gel and Southern blot autoradiograph of nDNA digested to completion with EcoRI and probed with $^{32}$P-labelled rRNAs. Lane 1 contains lambda DNA digested with EcoRI and HindIII; sizes are in kbp. Lanes 2-10 contain nDNA digested to completion with EcoRI. The right hand portion of the figure contains the Southern blots of the gel probed with the individual rRNA species (Sp. b-j). The large arrow points to the wide band of ca. 8.5-9.5 kbp to which most of the rRNAs hybridize. The small arrow points to the ca. 30 kbp band to which the 5S rRNA (Sp. h) hybridizes. The curved arrow points to the tight 3.0 kbp band to which only Sp. b (SSU) hybridizes.

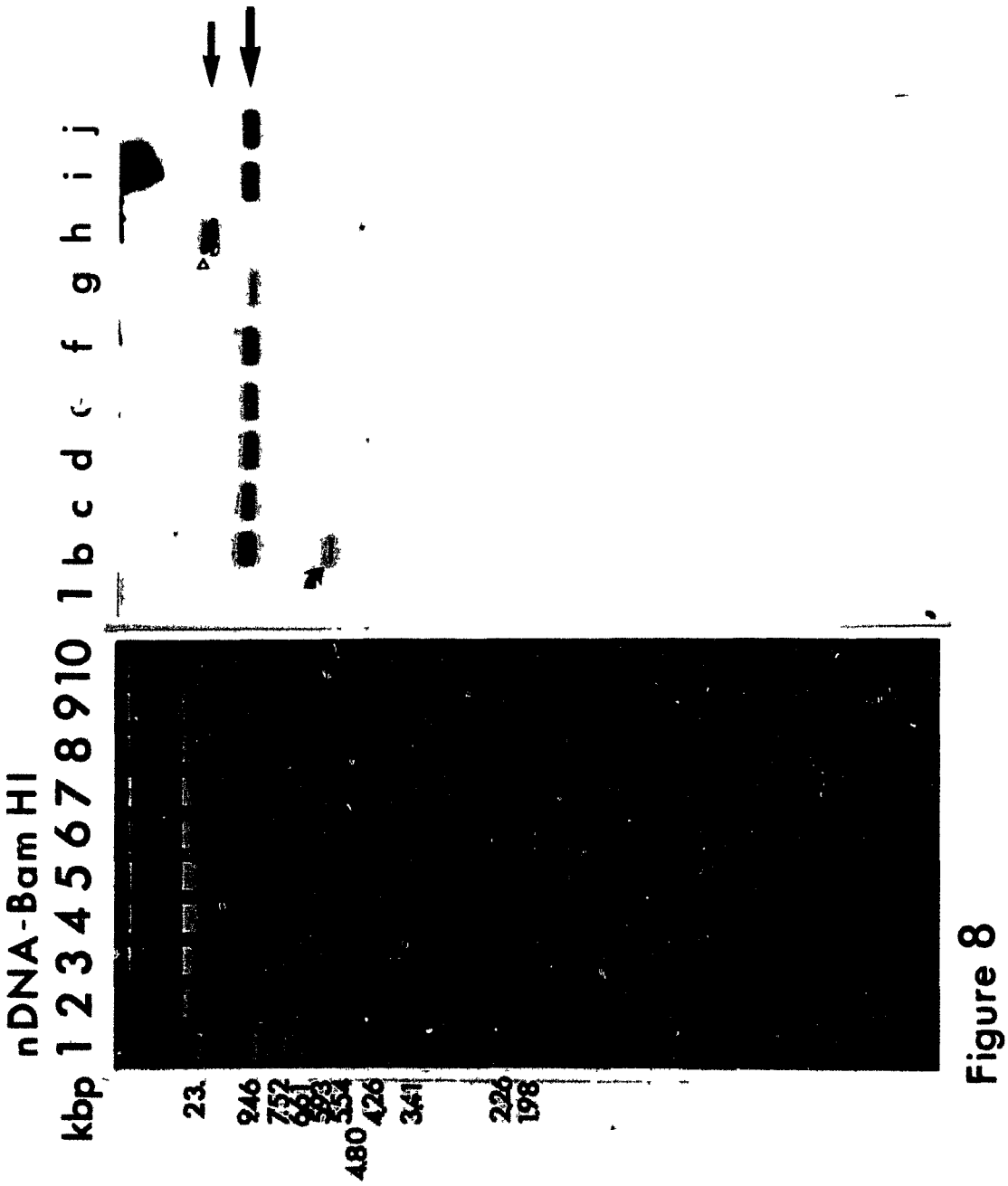



Figure 8

**Figure 8.** EtBr-stained 1% agarose gel and Southern blot autoradiograph of nDNA digested with BamHI to completion and probed with $^{32}$P-labelled individual rRNAs. Lane 1 contains lambda DNA digested with EcoRI and HindIII; sizes are in kbp. Lanes 2-10 contain the digested nDNA. The right side of the figure contains the autoradiograph of the Southern blots probed with Sp. b-j. The large arrow points to the 11-12 kbp band to which all the rRNAs hybridize except the 5S rRNA (Sp. h). The small arrow points to the ca. 25 kbp band to which Sp. h hybridizes strongly and the open triangle indicates the ca. 30 kbp band to which Sp. h hybridizes weakly. The small curved arrow indicates the 6-7 kbp band to which Sp. b hybridizes weakly.

Considerable effort was expended to clone most or all of the rDNA repeat as the SalI or BamHI fragments, but these attempts proved unsuccessful. Therefore, it was decided to focus on a restriction endonuclease that would generate rDNA fragments of smaller size. One such enzyme is PstI. Figure 9 shows the results obtained when the PstI-digested genomic DNA was probed with the individual end-labelled rRNAs.

The SSU rRNA was seen to hybridize by itself to an ca. 3.8-kbp fragment. Sp. c (identified as the 5'-end of the "28S" rRNA by primary sequence analysis), Sp. e, and the 5.8S rRNA (Sp. i) hybridized to an ca. 3.2-kbp PstI fragment. Sp. d (identified as the 3'-half of the "28S" rRNA by primary sequence analysis) and Sp. f (one of the novel small rRNAs) hybridized to fragments of the same size (ca. 2.0 kbp).

The remaining two small novel rRNAs (Sp. g and Sp. j) were observed to hybridize to a ladder of DNA fragments ranging in size from ca. 2.0 kbp to ca. 3.0 kbp. The intensity of hybridization to these bands implies that each is in sub-stoichiometric amount relative to the 3.8 kbp, 3.1 kbp, or 2.0 kbp fragments. Cumulatively, the intensities of the fragments hybridizing to Sp. g or j seem to approximate that of any of the other rRNA-hybridizing fragments.

If the sizes of the three main rDNA fragments (3.8, 3.2 and 2.0 kbp) are added to that of the smallest rDNA fragment in the hybridization ladder for Sp. g and Sp. j in one case, and to that of the largest rDNA fragment in the second case, the result is a combined rDNA size of between 11 and 12 kbp. These sizes are in good agreement with those deduced in the SalI or BamHI digests, implying that these three major rDNA fragments plus any one of the rDNA fragments hybridizing to Sp. g or Sp. j constitute the entire rDNA repeat present in the nucleus of this organism.

**Figure 9.**EtBr-stained 1% agarose gel and Southern blot autoradiograph of nDNA digested to completion with PstI and probed with $^{32}$P-labelled individual rRNAs (Sp.b-j). Lane 1 contains lambda DNA digested with HindIII and EcoRI; sizes are in kbp. Lanes 2-10 contain nDNA digested with PstI. The right hand side of the figure contains the autoradiograph of the Southern blots probed with the individual rRNA species (Sp. b-j). Arrow 1 points to the strongest hybridizing band, ca. 11-12 kbp in length, to which Sp. h hybridizes. Arrow 2 points to the 3.95 kbp band which hybridizes Sp. b; arrow 3 points to the 3.2 kbp band which selectively hybridizes Sp. c, e, and i; and arrow 4 points to the 2.0 kbp band which hybridizes Sp. d and f. The open triangles point to the 25 and 30 kbp bands to which Sp. h hybridizes. The four closed triangles point to the bands to which Sp. g and j selectively hybridize.

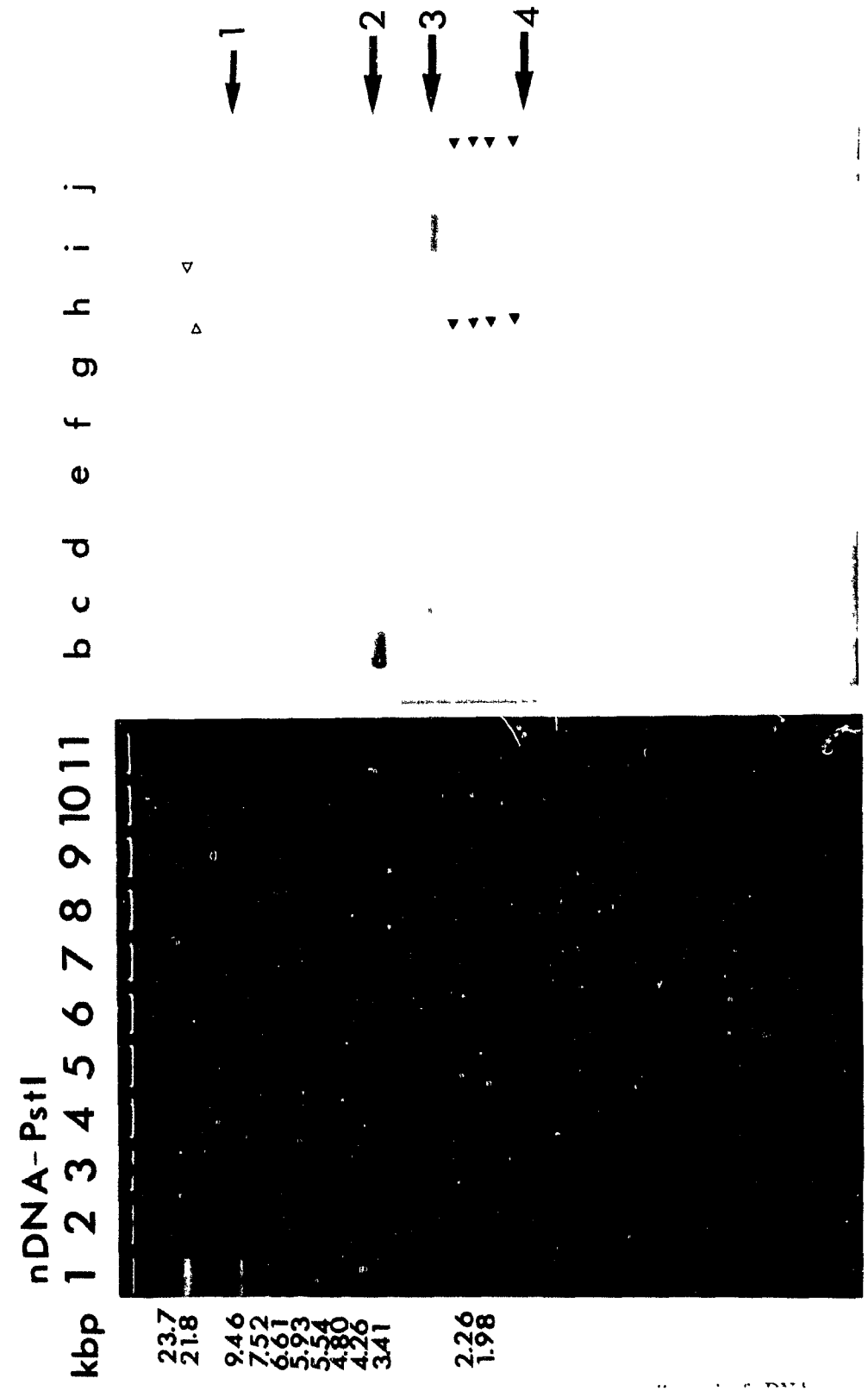

nDNA-PstI
kbp 1 2 3 4 5 6 7 8 9 10 11
23.7
21.8
9.46
7.52
6.61
5.93
5.54
4.80
4.26
3.41
2.26
1.98
b c d e f g h i j
1
2
3
4

Figure 9

Due to the fact that PstI digested the rDNA into more conveniently sized rDNA fragments, it was decided to "shotgun"-clone the PstI-digested nDNA into pUC9 and to screen for rDNA-pUC9 recombinants.

## 3.) CHARACTERIZATION OF pUC9-rDNA PLASMIDS

pUC9-rDNA recombinants were constructed and isolated by ligating the PstI-digested nDNA (in a 4:1 mass ratio, i.e. 1 μg:0.25 μg) with PstI-digested pUC9 in the presence of T4 DNA ligase in a final volume of 25 μl. Ten μl (0.5 μg of total DNA) was used to transform 210 μl of competent JM83 cells. To this was added 800 μl of SOC liquid medium and the cells were incubated at 37° C for 1 hr. A 1/10 volume (100 μl) of the medium containing the transformed cells was used to inoculate each agar plate (a total of 10). Approximately 1000 white (recombinant) colonies were found per plate (with an equal number of non-recombinant blue colonies). Of these ca. 65 large and well-separated white colonies were selected from each plate for in situ colony hybridization with $^{32}$P-end-labelled iRNA or Sp. g or Sp. j. A total of 1274 white colonies were screened in this manner. A total of 33 colonies hybridized to the labelled rRNA. Of those, 8 were found to hybridize to labelled rRNA in Southern blots of PstI-digested mini-preparations of their plasmid DNAs. Of these eight, 3 were of the 3.9 kbp size, 1 was of the 3.2 kbp size, 3 were of the 2.0 kbp size, and 1 was of the 2.0-3.0 kbp size. In this way, four distinct recombinant plasmid clones, designated pCf1-pCf4, were chosen for further detailed analysis.

Figure 10 the shows four PstI plasmid DNAs digested with PstI and fractionated in a 1% agarose slab gel by electrophoresis. pCf1 contains an ca. 3.8 kbp rDNA insert which hybridizes to the SSU rRNA (Sp. b) only; pCf2 contains an ca. 3.2 kbp rDNA insert which hybridizes to Sp. c, Sp. i, and Sp. e; and pCf3 contains an ca. 2.0 kbp rDNA which hybridizes to Sp. d and Sp. f (data

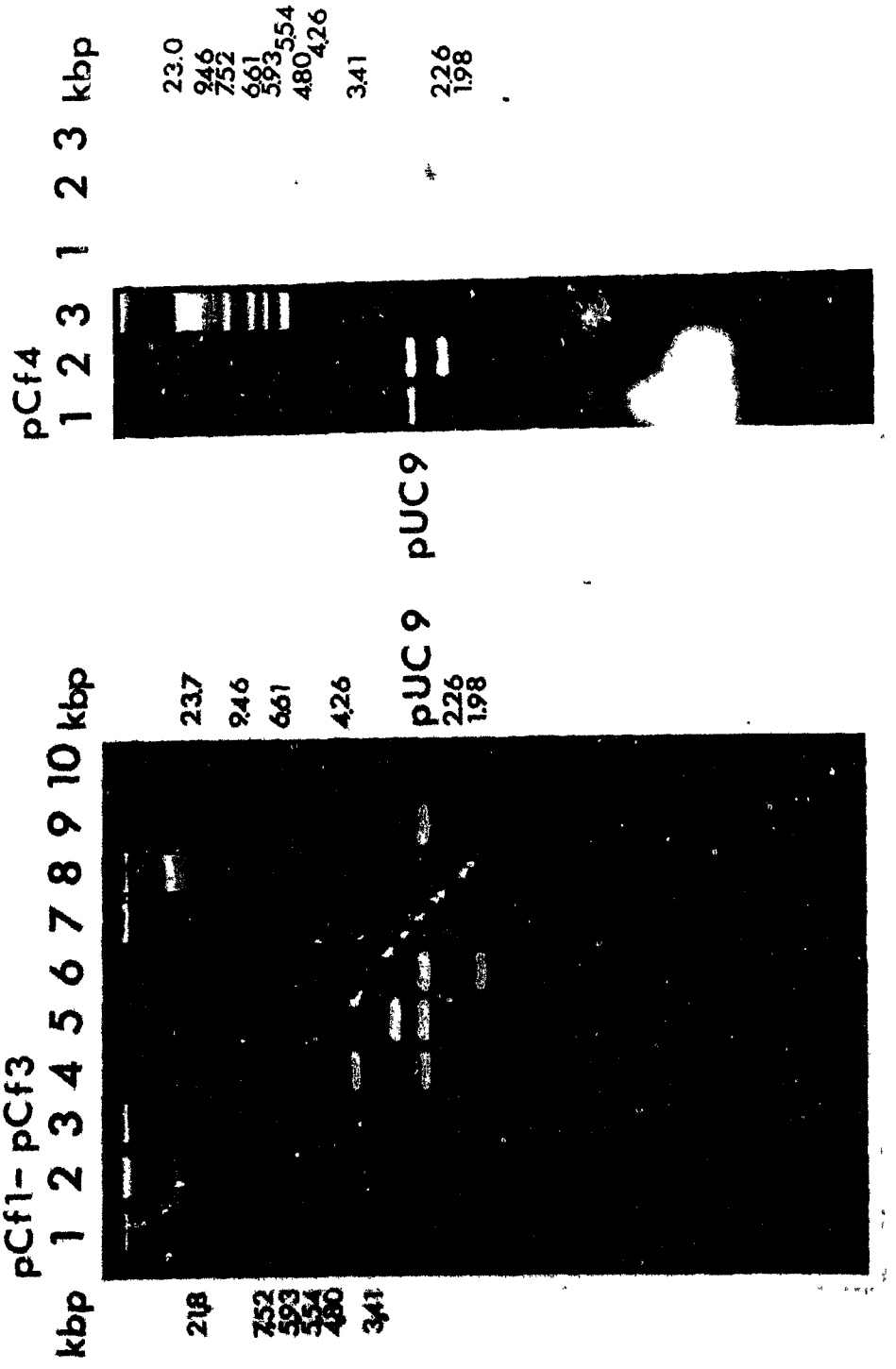



**Figure 10.** Composite figure. Left hand portion is a photograph of an EtBr-stained 1% agarose gel. Lane 1 contains lambda DNA digested with EcoRI. Lane 10 contains lambda DNA digested with HindIII. Sizes are in kbp. Lanes 2 and 8 contain undigested nDNA. Lanes 3 and 7 contain nDNA digested with PstI. Lane 4 contains pCf1 digested with PstI, lane 5 contains pCf2 digested with PstI, and lane 6 contains pCf3 digested with PstI. Lane 9 contains pUC9 digested with PstI (linear pUC9). The right hand portion of the figure contains a photograph of a 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. g and j. Lane 3 contains lambda DNA digested with HindIII and EcoRI.

not shown). The plasmid pCf4 contains an rDNA of ca. 2.2 kbp in length which hybridizes to Sp. j and Sp. g.

After having localized the individual rRNAs to one of the four rDNA clones, a more detailed restriction enzyme map of each was undertaken in order to obtain a better understanding of the positions of the coding regions for the rRNAs and to gain a better overall appreciation of the organization of the rDNA..

## 4.) MAPPING OF INDIVIDUAL rRNAs TO RECOMBINANT rDNA CLONES

The mapping of the coding regions for the rRNAs to their rDNA clones was facilitated by sequence information obtained by Dr. M.N. Schnare about the individual rRNAs (166). Primary sequence was available for the 5'-and 3'-ends of the large rRNAs (Sp. b, Sp. c, and Sp. d), while the complete sequence of each of the small rRNAs was known. Restriction endonuclease sites identified by RNA sequence analysis of some of the rRNAs proved invaluable in determining the location and orientation of the rRNA coding regions in the cloned rDNA.

The pCf insert rDNA (Cf) was released by digestion with PstI in all mapping experiments. Mapping enzymes were selected which either did not have a recognition site in pUC9 or had a site(s) located in the poly-linker, thereby keeping pUC9 essentially intact and so simplifying analysis of the fragment patterns.

### 4.i) pCf1

Clone pCf1 contains a rDNA insert of ca. 3.95 kbp (Cf1). Primary sequence analysis of the 3'-end of the SSU rRNA identified a PstI site 15

nt from the end (165). This information, coupled with the estimated size of the SSU rRNA (2460 nt) and the mapping information in Figure 11 , made it possible to construct a preliminary map of this rDNA. BglII is seen to split Cf1 (the pCf1 rDNA insert) into two unequally sized fragments of ca. 2200 bp and 1800 bp (lane 2), with Sp. b hybridizing to the ca. 2200-bp fragment only. When Cf1 is doubly digested with BglII plus AvaI (lane 10), AvaI is seen to split the ca. 2200-bp Pst-Bgl fragment into two fragments of different size, ca. 1050-bp and 1150-bp in length. When Cf1 is digested with AvaI alone, only the ca. 1050-bp fragment is produced (data not shown), indicating that the AvaI site is ca. 1050-bp from the PstI end of the ca. 2200-bp BglII-PstI fragment. Sp. b is seen to hybridize to both the ca. 1150-bp and 1050-bp fragments in the Bgl-Ava double digest but not to the ca. 1800-bp fragment (which is present in the BglII digest). EcoRI is seen to cut Cf1 at least twice (lane 3) to give rise to three fragments of ca 400 bp, 650 bp, and 2800 bp in length. Sp. b is seen to hybridize to the large 2800-bp fragment only. HindIII splits Cf1 into two unequal fragments of ca. 750 bp and ca. 3200 bp in length (lane 4). Sp. b is seen to hybridize to both fragments, although only weakly to the 750-bp one.

The EcoRI plus HindIII double digest of Cf1 (lane 8) reveals that the two small fragments (ca. 650 and 400 bp) in lane 3 (EcoRI) are retained in this digest but the large ca. 2800-bp fragment has been reduced by ca. 350 bp to an ca. 2450-bp sized fragment, indicating that the HindIII site is located at one end of Cf1. Sp. b is seen to hybridize to the ca. 2450-bp fragment as well as the two smallest fragments (ca. 350 and 400 bp). This suggests that the EcoRI ca. 650-bp fragment does not contain any Sp. b coding sequence. Due to the fact that the small HindIII fragment is lost in the HindIII-EcoRI double digest, the EcoRI site which generates the ca. 400-bp EcoRI-PstI fragment must be next to the HindIII site which generated the HindIII-PstI ca. 750-bp fragment. This means

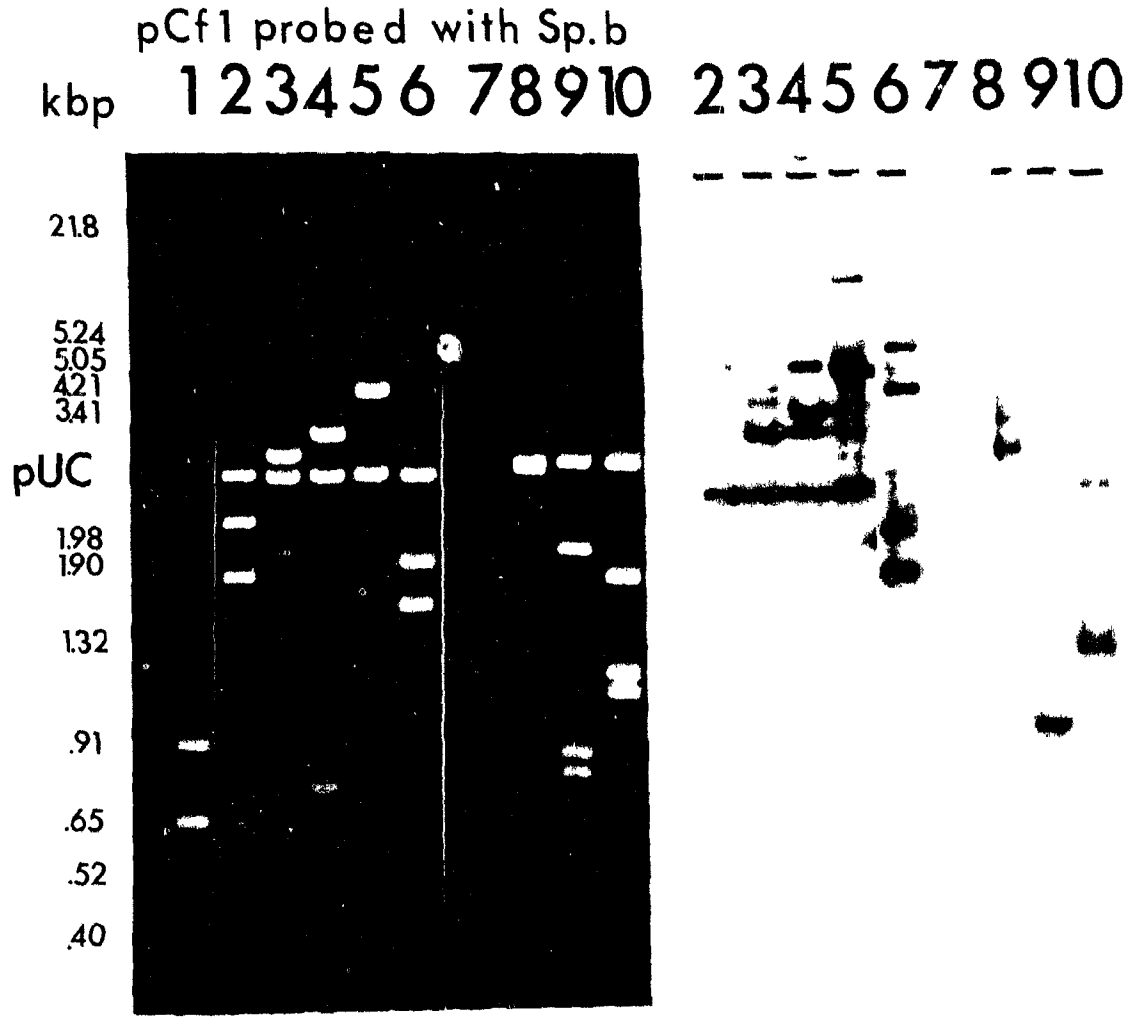
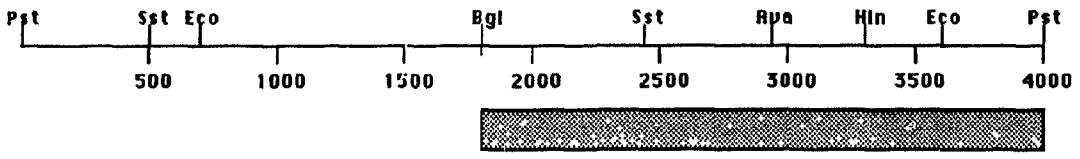



**Figure 11.** Photograph of an EtBr-stained 1% agarose gel and the autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. b. Lanes 1 and 7 contain lambda DNA digested with EcoRI and HindIII; sizes are indicated in kbp. All the remaining lanes contain pCf1 DNA digested with PstI and at least one other restriction endonuclease. The added enzyme(s) is (are): lane 2, BglII; lane 3, EcoRI; lane 4, HindIII; lane 5, only PstI; lane 6, SstII; lane 8, EcoRI and HindIII; lane 9, SstII and HindIII; and lane 10, BglII and AvaI.

Restriction Endonuclease and Gene Map of Cf1




Sp. b

Ava = AvaI
Bgl = BglII
Eco = EcoRI
Hin = HindIII
Pst = PstI
Sst = SstII

Figure 12. Vertical lines above the horizontal indicate the sites of restriction endonuclease cleavage and the vertical lines below the horizontal indicate the length in bp. The shaded box is the approximate location of the rRNA gene.

that the ca. 650-bp EcoRI fragment which does not hybridize to Sp. b contains a PstI site at one end and is therefore located at the opposite end of Cf1 relative to the other EcoRI site (and the HindIII site).

SstII digests Cf1 into at least three fragments (lane 6), indicating that it must have two sites in this DNA. Sp. b is seen to hybridize to the two large fragments (ca. 1900 and 1600 bp) but not to the small (ca. 500 bp) fragment. In the SstII plus HindIII double digest (lane 9) it can be seen that the ca. 1600-bp fragment present in the SstII digest is lost and replaced by two fragments of ca. 750 bp and 850 bp in length. The 750-bp fragment is seen in the HindIII digest. Sp. b is seen to hybridize to the ca. 1900-bp and 950-bp fragments and perhaps weakly to the ca. 750-bp fragment.

These data indicate that the Sp. b (SSU rRNA) coding region (ca. 2400 nt) is primarily located on one half of this rDNA fragment, the 3' end of which is marked by the PstI site adjacent to the HindIII site. The coding region contains within it (going 3' to 5') EcoRI, HindIII, AvaI, SstII, and BglII sites at distances indicated in the map (see Figure 12). In the remaining half of this rDNA are located two other sites, one EcoRI and one SstII. No sites for BamHI, EcoRV, SalI, SstI, or XhoI were identified in these mapping experiments.

4.ii) pCf2

Three rRNAs, namely Sp. c (known from primary sequence comparison to be the 5' half of a conventional 28S rRNA; unpublished data), i (the 5.8S rRNA), and e (one of the novel small rRNAs), selectively hybridized to pCf2, which contains an ca. 3.2-kbp rDNA insert (Cf2). Restriction mapping studies and Southern blotting analysis shown in Figures 13, 14 and 15 (Sp. c, i, and e, respectively) allowed the construction of a restriction map and rough placement of the coding regions for these rRNAs. Sizes of Sp. i and Sp. e are

known through primary sequence analysis. The size of Sp. c is an estimate (see Table 1).

AvaI digestion of Cf2 results in two DNA fragments ca. 2000 and 1200 bp long. Sp. c (Fig. 13, lane 1) hybridizes to both fragments and Sp. i binds to the ca. 2000-bp fragment only (Fig. 14, lane 1). BglII digestion of Cf2 results in two fragments as well, with a site ca. 650 bp from one of the PstI ends. Sp. c hybridizes to the large fragment but not to the smaller (Fig. 13, lane 2). Sp. i binds that same fragment (Fig. 14, lane 9) but Sp. e hybridizes to the small ca. 650-bp fragment (Fig. 15, lane2), indicating that this RNA coding region is located within 650 bp of one of the PstI ends. HindIII digestion of Cf2 produces two fragments ca. 660 bp and 2550 bp (Fig. 14, lane 7; Fig. 13, lane 4) in length. Sp. c hybridizes to the 2550-bp fragment selectively (Fig. 13, lane 4), as does Sp. i (Fig. 14, lane 7). NruI digestion produces two fragments, one 400 bp and the other ca. 2880 bp long. Sp. c selectively hybridizes to the 2880-bp fragment (Fig. 13, lane 5), as does Sp. e and i (Fig. 15, lane 4 and Fig. 14, lane 6). The 400-bp PstI-NruI fragment seems to be devoid of rRNA coding sequences (Fig. 15, lane 4). EcoRV digestion produces two fragments (ca. 1100 and 2100 bp), indicating that there is one recognition site. Sp. c hybridizes with the ca. 2100-bp fragment only (Fig. 13, lane 3), Sp. i hybridizes the ca. 1100-bp fragment, and Sp. e binds to the 2100-bp fragment.

When Cf2 is digested with HindIII plus NruI (Fig. 14, lane 3; Fig. 13, lane 7), it can be seen that the small fragment in the HindIII digestion (ca. 660 bp) and the small fragment in the NruI digestion (Fig. 14 and 15, lane 4) (ca. 400 bp) are retained. This means that these two fragments are at the two ends of this rDNA insert. Sp. i hybridizes to the large DNA fragment in this digest (Fig. 14, lane 3), as does Sp. c (Fig. 13, lane 7).

When Cf2 is doubly digested with NruI plus EcoRV, it can be seen (Fig. 14, lane 2; Fig. 13, lane 8) that the large DNA fragment in the EcoRV digestion (ca. 2100 bp) is retained (Fig. 13, lane 8), as is the small DNA fragment in the NruI digest (ca. 400 bp) (Fig. 14, lane 6). The smaller EcoRV DNA fragment (ca. 1100 bp) is lost, but an ca. 600-bp fragment appears (Fig. 14, lane 2). Sp. i binds to the 600-bp fragment exclusively (Fig. 14, lane 2), while Sp. c hybridizes to the ca. 2100-bp fragment exclusively (Fig. 13, lane 8). This means that the coding region for Sp. i is at least 400 bp from one of the PstI ends on the ca. 1100-bp EcoRV-PstI fragment to which Sp. c does not hybridize.

SphI has a single site in Cf2 (Fig. 14, lane 4) resulting in ca. 500-bp and 2700-bp fragments. Sp. i is seen to hybridize with the smaller of the two (Fig. 14, lane 4) exclusively. It was known from primary sequence analysis that the coding region for Sp. i contains within it a recognition site for SphI. This helps to place the coding region for this rRNA ca. 500 bp from one of the PstI ends.

When Cf2 is doubly digested with BglII plus EcoRV (Fig. 15, lane 6), it can be seen that the small BglII-PstI (ca. 650 bp) fragment (Fig. 15, lane 2) and the small EcoRV-PstI fragment (Fig. 14, lane 8) are retained, indicating that they represent the two ends of Cf2. Sp. e hybridizes to the ca. 650-bp fragment. When Cf2 is doubly digested with BglII plus NruI (Fig. 15, lane 7), again it can be seen that the smallest fragment in the BglII-PstI digestion (ca. 650 bp) (Fig. 15, lane 2) is retained, as is the smallest fragment in the NruI-PstI digest (ca. 400 bp) (Fig. 15, lane 4). Again Sp. e hybridizes to the ca. 650-bp fragment exclusively. Taken together this information establishes that the coding region for Sp. i is ca. 500 bp from one end of Cf2 and that the coding region for Sp. e is within 650 bp of the opposite end of Cf2. Sp. c as well as Sp. e hybridizes to the ca. 2200-bp EcoRV-PstI fragment. The deduced map would look as follows:

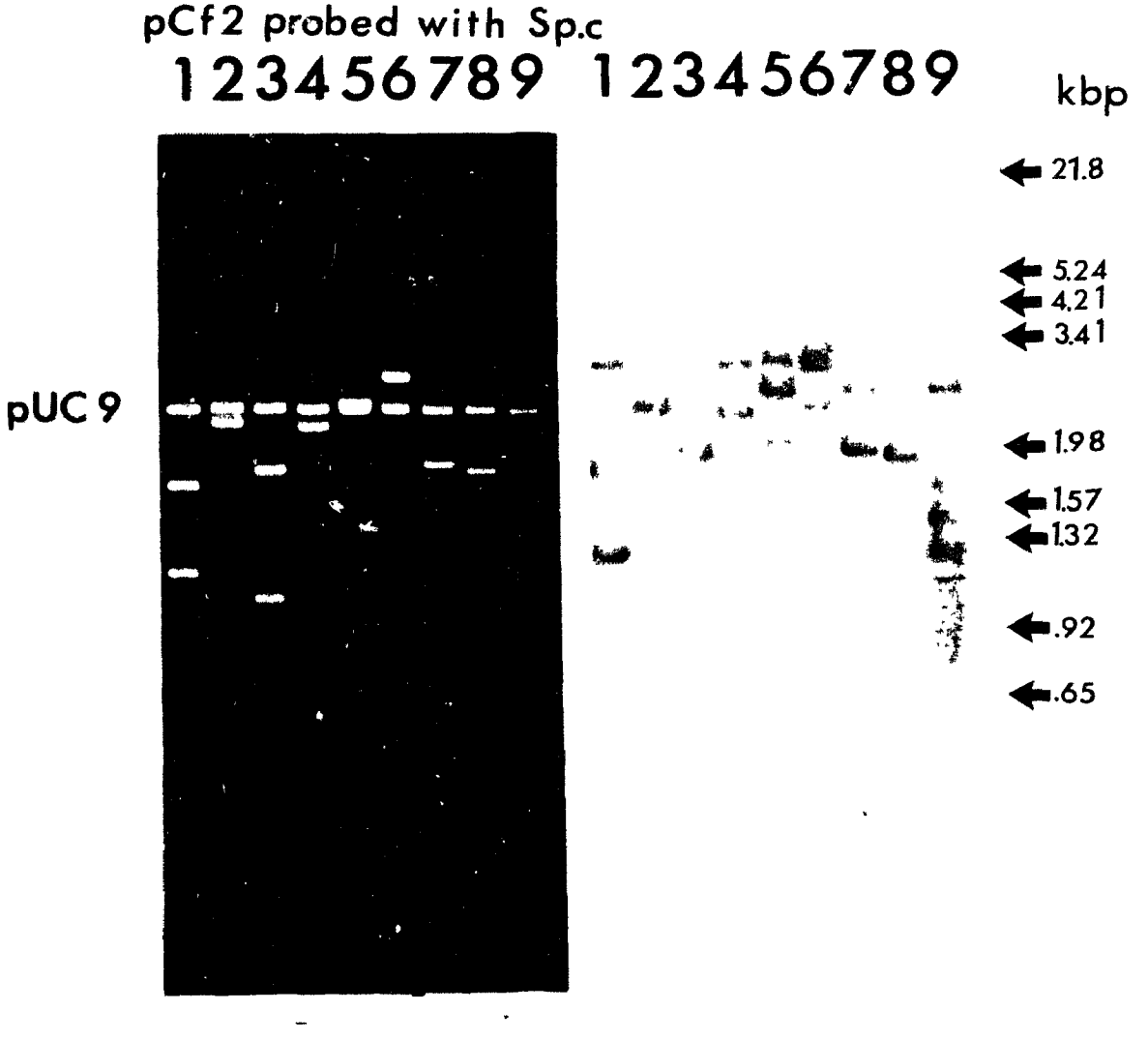



**Figure 13.** Photograph of an EtBr-stained 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. c. Lanes 1-9 contain pCf2 DNA digested with PstI and at least one other restriction endonuclease. Sizes are indicated in kbp at the right hand boundary of the autoradiograph. The added enzyme(s) is (are): lane 1, AvaI; lane 2, BglII; lane 3, EcoRV; lane 4, HindIII; lane 5, NruI; lane 6, PstI alone; lane 7, HindIII and NruI; lane 8, NruI and EcoRV; and lane 9, NruI and AvaI.

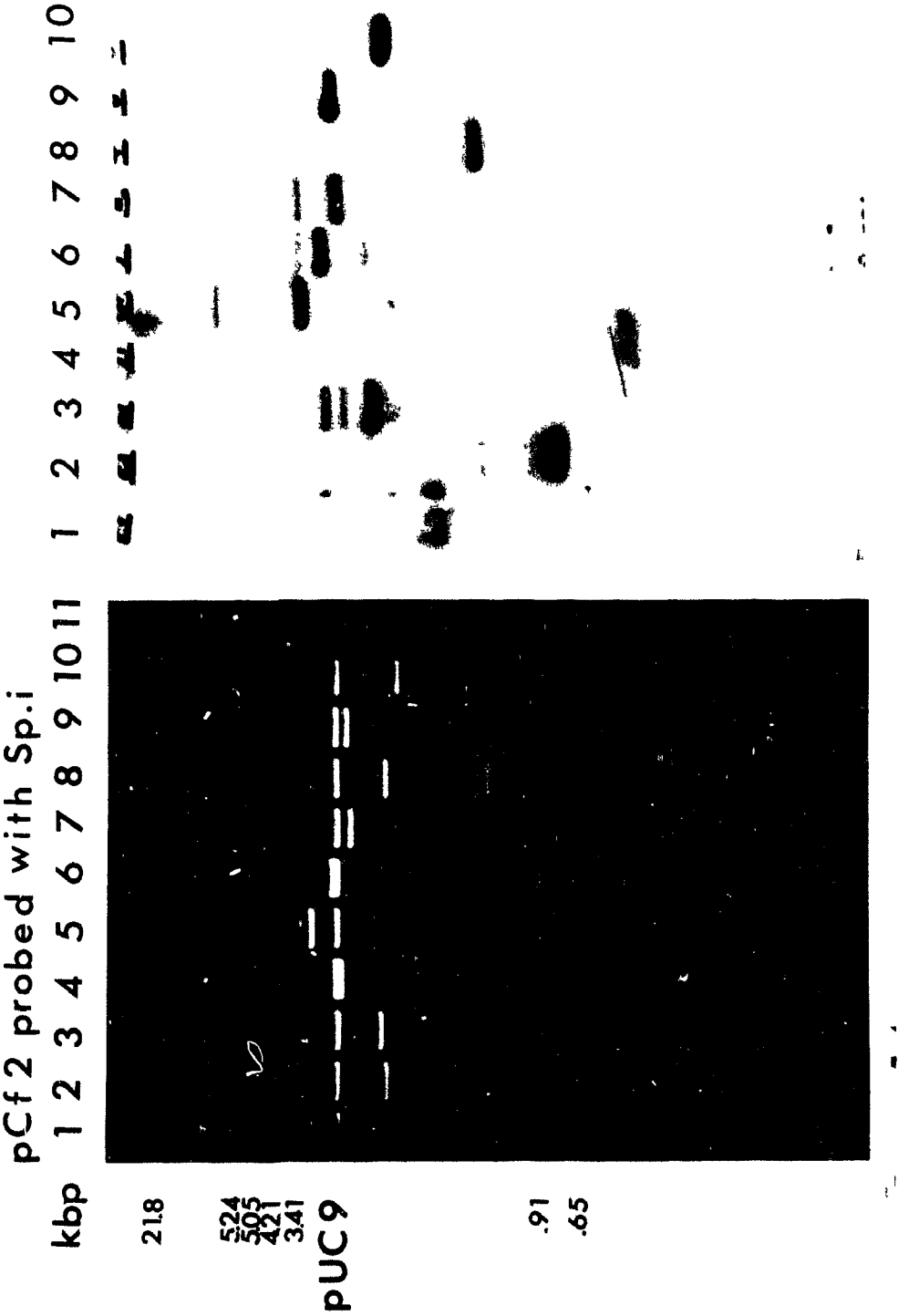



**Figure 14.** Photograph of an EtBr-stained 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. i. Lane 11 contains lambda DNA digested with EcoRI plus pBR322 digested with AluI; the marker sizes are indicated in kbp. Lanes 1-10 contain pCf2 DNA digested with PstI and at least one other restriction endonuclease. Enzyme(s) is(are): lane 1, NruI and AvaI; lane 2, NruI and EcoRV; lane 3, HindIII and NruI; lane 4, SphI; lane 5, PstI only; lane 6, NruI; lane 7, HindIII; lane 8, EcoRV; lane 9, BglII and lane 10, AvaI.

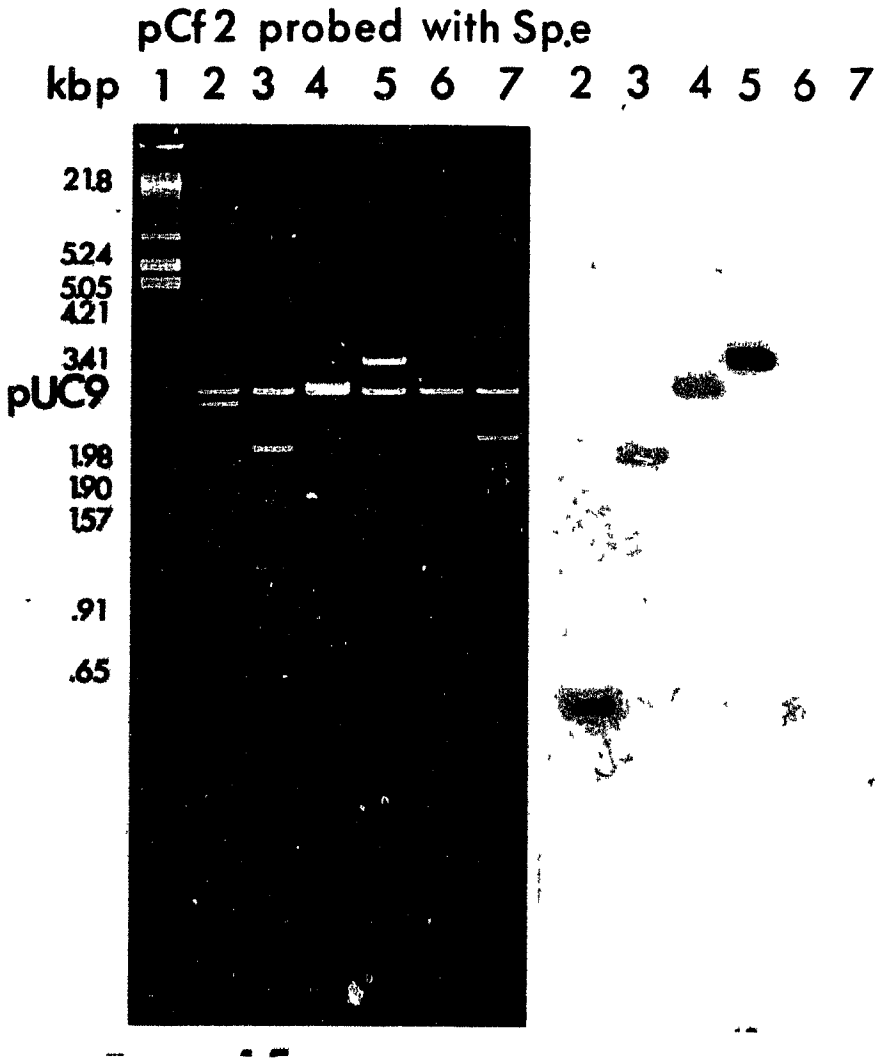
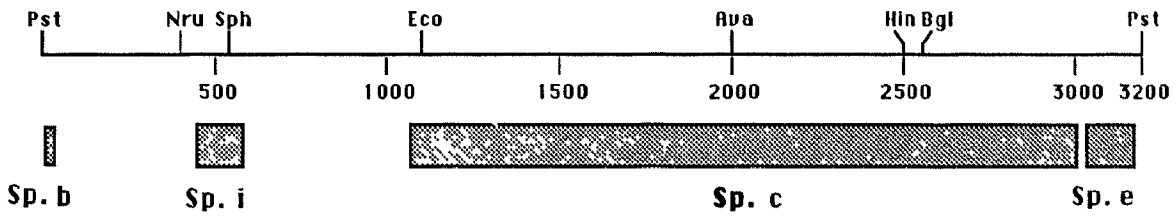



**Figure 15.** Photograph of an EtBr-stained 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. e. Lane 1 contains the marker DNA with sizes indicated in kbp. All the remaining lanes contain pCf2 digested with PstI and at least one other restriction endonuclease: lane 2, BglII; lane 3, EcoRV; lane 4, NruI; lane 5, PstI only; lane 6, BglII and EcoRV; and lane 7, BglII and NruI.

Restriction Endonuclease and Gene Map of Cf2




Ava = AvaI
Bgl = BglII
Eco = EcoRV
Hin = HindIII
Nru = NruI
Pst = PstI
Sph = SphI

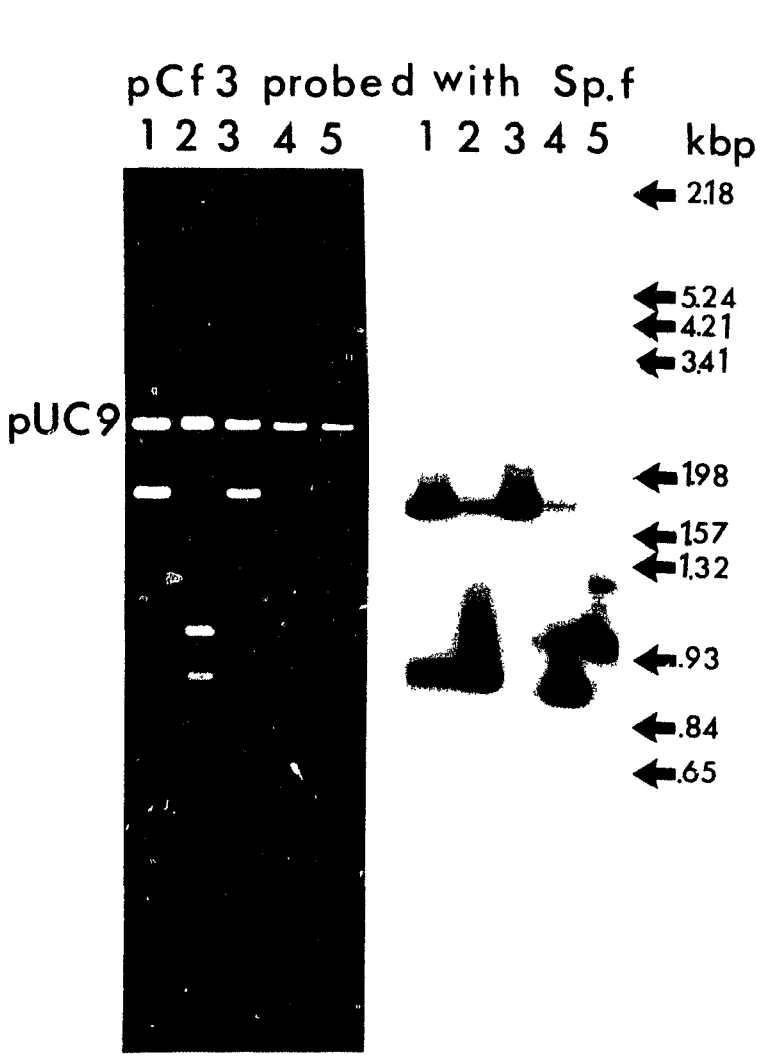
**Figure 16.** Vertical lines above the horizontal indicate sites of restriction endonuclease cleavage and the vertical lines below the horizontal indicate length in bp. The shaded boxes mark the locations of the coding regions for the rRNAs.

Starting at one PstI end and moving ca. 400 bp along its length one would encounter an NruI site. Next would follow the SphI site at ca. 500 bp, marking the coding region for Sp. i. An EcoRV site would be found at bp 1100 and this approximates the start site for the coding region for Sp. c. An AvaI site is found at bp 2000 and a HindIII site is located at bp 2500. The BglII site must follow it at bp 2550. The coding region for Sp. e is between the HindIII and BglII sites and the PstI end. The possibility of small IVS in the coding sequences (for the large rRNA in particular) cannot be ruled out completely, although they would have to be rather small. Cf2 did not contain any detectable sites for BamHI, EcoRI, SalI, SstI, SstII, or XhoI. Figure 16 depicts this map.

4.iii) pCf3

Localization of Sp. d and Sp. f on the rDNA insert of pCf3 was facilitated by the fact that primary sequence analysis of the 5'-end of Sp. d identified a PstI recognition site 8 bp from that end. The insert rDNA (Cf3) is ca. 2.0 kbp long. The length of Sp. f is known to be 183 bp and Sp. d is estimated to be ca. 1700 bp long.

Figure 17a and b shows the results obtained when the insert rDNA was digested with HindIII or SstI and hybridized with end-labelled Sp. d or Sp. f. Cf3 is cut once by HindIII and SstI. HindIII digestion gives rise to ca. 1200-bp- and 800-bp-sized fragments. Sp. d is seen to hybridize strongly with the 1200-bp fragment and weakly to the 800-bp one (Fig. 17b, lane 1). Sp f hybridizes with the ca. 800-bp fragment only (Fig. 17b, lane 4). SstI digestion results in ca. 1100-bp- and 900-bp-sized fragments (Fig. 17b, lane 3). Again Sp. d (Fig. 17b, lane 3) hybridizes to the larger fragment more strongly than to the smaller one. As well, Sp. f hybridizes to the smaller fragment exclusively (Fig. 17a, lane 2). It seems that most of the coding region of Sp. d is confined to the



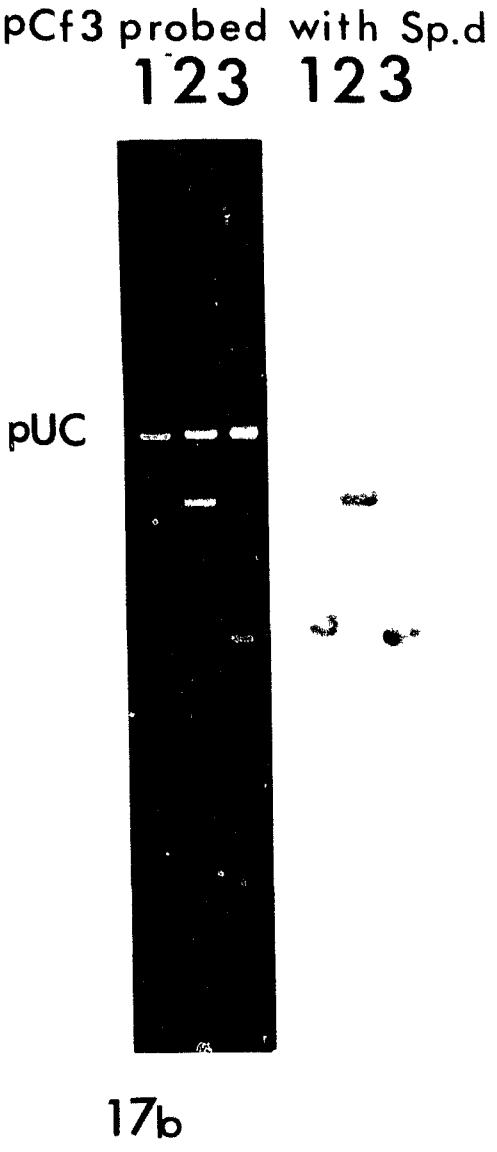
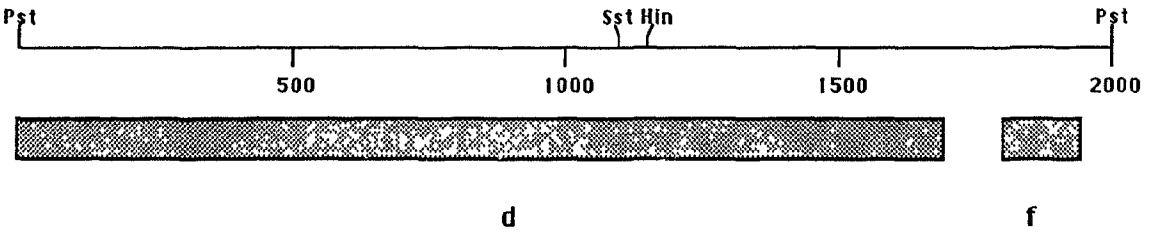




**Figure 17.** Photograph of EtBr-stained 1% agarose gels with accompanying Southern blot autoradiographs. Sp. f and d are the labelled probes in Fig. 17a and 17b, respectively. Size markers, in kbp, are indicated by arrows. pCf3 was digested with PstI and with at least one other restriction endonuclease. In 17a, lane 1 contains pCf3 digested with PstI only, lane 2 contains pCf3 digested with PstI and SstI, in lane 3 PstI is the only enzyme used, in lane 4 the additional enzyme is HindIII, and in lane 5 the additional enzyme is AvaI. In lane 1 of Fig. 17b, the DNA was digested with PstI and HindIII, lane 2 contains DNA digested with PstI only, and in lane 3 the added enzyme is SstI.

**Restriction Endonuclease and Gene Map of Cf3**




**Pst = PstI**

**Sst = SstI**

**Hin = HindIII**

**Figure 18.** Vertical lines above the horizontal indicate the sites of restriction endonuclease cleavage and those below indicate length in bp. Shaded boxes represent approximate locations of the rRNA coding regions.

large fragment side of Cf3 and that the coding region of Sp. f is at the opposite side of the insert DNA to Sp. d. Given the fact that Sp. d has a PstI site 8 nt from the 5'-coding end (unpublished data) and is ca. 1700 nt in length, the map of the coding regions would appear as depicted in Figure 18 . It is not possible, at this degree of mapping resolution, to determine if the coding region for Sp. f may be imbedded within a sequence of DNA which interrupts the coding region of Sp. d (i.e., Sp. d may be encoded in discontinuous stretches of DNA and Sp. f could be encoded in the intervening DNA). Cf3 does not contain any detectable site for BglII, BamHI, EcoRI, EcoRV, SalI, SstII, or XhoI.

4.iv) pCf4

pCf4 has a 2.2-kbp rDNA insert (Cf4) which hybridizes to Sp. g and Sp. j only. These rRNAs are 133 and 73 nt long, respectively. Figures 19 and 20 display the results obtained after single or double digestion of the insert rDNA followed by Southern blot hybridization with $^{32}$P-labelled Sp. g or Sp. j. Primary sequence analysis of Sp. g showed that it contained within its coding sequence single recognition sites for XhoI and SphI. This allowed precise placement of the coding region for this rRNA on the map.

BamHI cuts Cf4 into two unequal sized fragments of ca. 600 and 1600 bp in length. Sp. j and g hybridize exclusively to the ca. 600-bp fragment (Fig. 19, lane 1 and Fig. 20, lane 9, respectively). XhoI digestion of Cf4 generates ca. 500-bp- and 1700-bp-sized fragments, with Sp. j hybridizing with the 500 bp (Fig. 19, lane 5) and Sp. g with the 1700-bp fragment (Fig. 20, lane 2). SalI divides Cf4 into 1000-bp and 1200-bp fragments, with both Sp. j and g hybridizing to the 1000-bp fragment (Fig. 19, lane 3 and Fig. 20, lane 5, respectively). SstI divides Cf4 into two fragments, also, ca. 1300 and 900 bp in length. Sp. j and g hybridize to the 900-bp fragment only (Fig. 19, lane 4 and Fig.

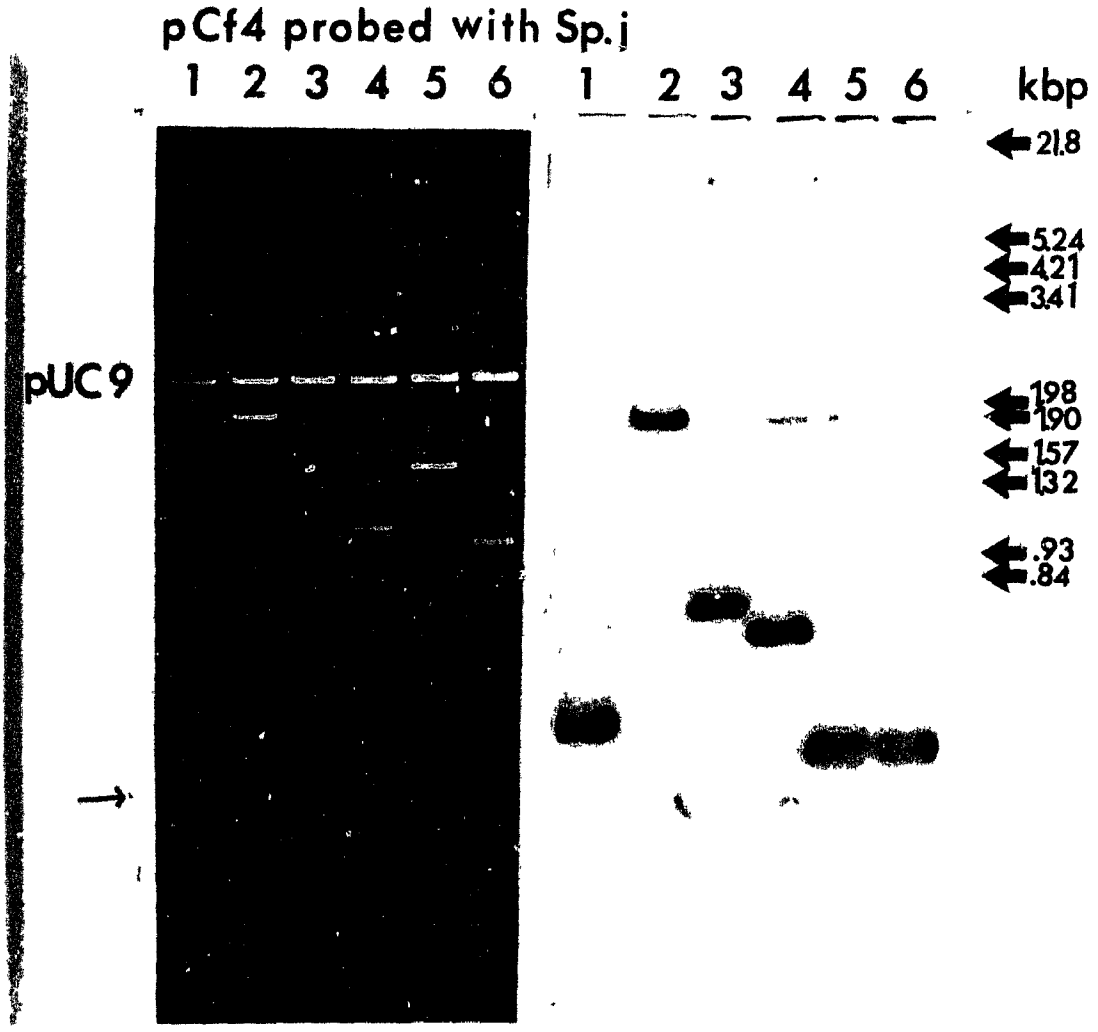



**Figure 19.** Photograph of EtBr-stained 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. j. DNA marker sizes (kbp) are indicated by broad arrows. All lanes contain pCf4 digested with PstI and at least one other restriction endonuclease: lane 1, BamHI; lane 2, PstI only; lane 3, SalI; lane 4, SstII; lane 5, XhoI; and lane 6, XhoI and SalI. The thin arrow along the left hand border of the photograph points out the ca. 200 bp contaminating PstI fragment seen in all digestions. This PstI fragment was co-cloned with the 2.2-kbp PstI rDNA fragment and most likely represents contaminating non-rDNA.

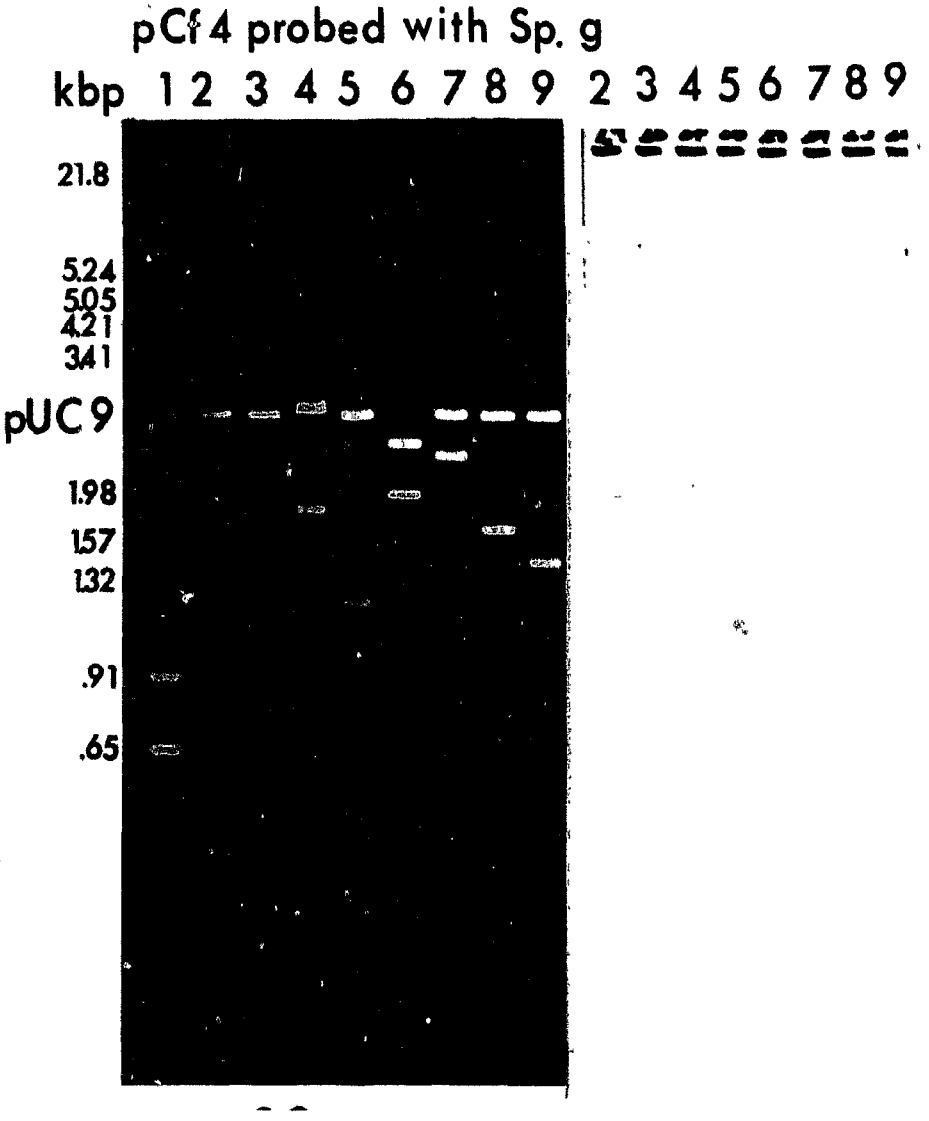
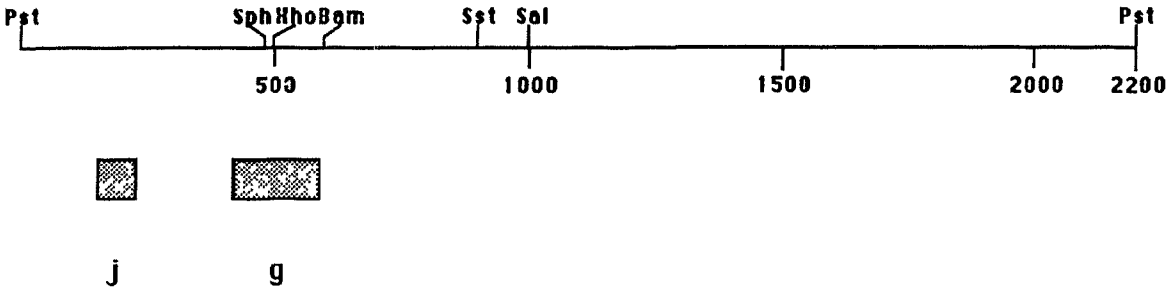



**Figure 20.** Photograph of an EtBr-stained 1% agarose gel and an autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. g. Lane 1 contains lambda DNA digested with EcoRI plus pBR322 digested with AluI; sizes are in kbp. Lanes 2-9 contain pCf4 DNA digested with PstI and at least one other restriction endonuclease: lane 2, XhoI; lane 3, SstII; lane 4, SphI; lane 5, SalI; lane 6, PvuII; lane 7, PstI only; lane 8, BamHI and lane 9, AvaI.

**Restriction Endonuclease and Gene Map of Cf4**




**Bam = BamHI**
**Pst = PstI**
**Sal = SalI**
**Sph = SphI**
**Sst = SstII**
**Xho = XhoI**

**Figure 21. Vertical lines above the horizontal indicate the sites of restriction endonuclease cleavage and those below indicate length in bp. Shaded boxes represent rRNA coding regions.**

20, lane 3, respectively). When Cf4 is digested with XhoI plus SalI, the small ca. 500-bp XhoI-PstI fragment is retained as is the SalI-PstI 1200-bp fragment (Fig. 19, lane 6). Sp. j hybridizes with the 500-bp fragment, as expected. The SalI-PstI 1000-bp fragment is lost and is replaced by an ca. 550-bp one. When Cf4 is digested with SphI, two fragments are produced, indicating that only the one predicted recognition site is present (i.e., in the coding region for Sp. g). Sp. g hybridizes with the ca. 1700-bp fragment.

In summary, Sp. g can be placed by the position of the SphI and XhoI sites, which are ca. 500 bp from one of the PstI ends. Both Sp. j and g hybridize to the small 600-bp BamHI-PstI fragment, placing the coding region for Sp. j within this 600-bp region. Indeed Sp. j is within the XhoI-PstI 500-bp fragment placing it between the coding region for Sp. g and the PstI end. Beyond the BamHI site at ca. 900 bp is the SstI site and a SalI site can be found at ca. 1000 bp. Cf4 does not contain any detectable site for BglII, EcoRI, EcoRV, HindIII, or SstI. Figure 21 depicts the restriction map of Cf4 with the coding regions for Sp. j and g indicated.

## 5.) OVERALL ORGANIZATION OF rDNA CLONES

Given that the complete rDNA repeat can be liberated from the nuclear genome with BamHI or SalI as an 11-to 12-kbp size fragment, it seemed reasonable that the four PstI rDNA inserts would together constitute the entire rDNA repeat of C. fasciculata (Cf1-Cf4; total size = 11,300 bp). The relative arrangement of Cf1 through Cf3 in this repeat is rather straightforward. Given that a PstI site is located at the 3'-end of the 18S rRNA and 5'-end of Sp. d, the only possible arrangement (assuming that Crithidia conforms to the conventional eukaryotic nuclear rDNA arrangement) is Cf1 followed by Cf2 which is directly

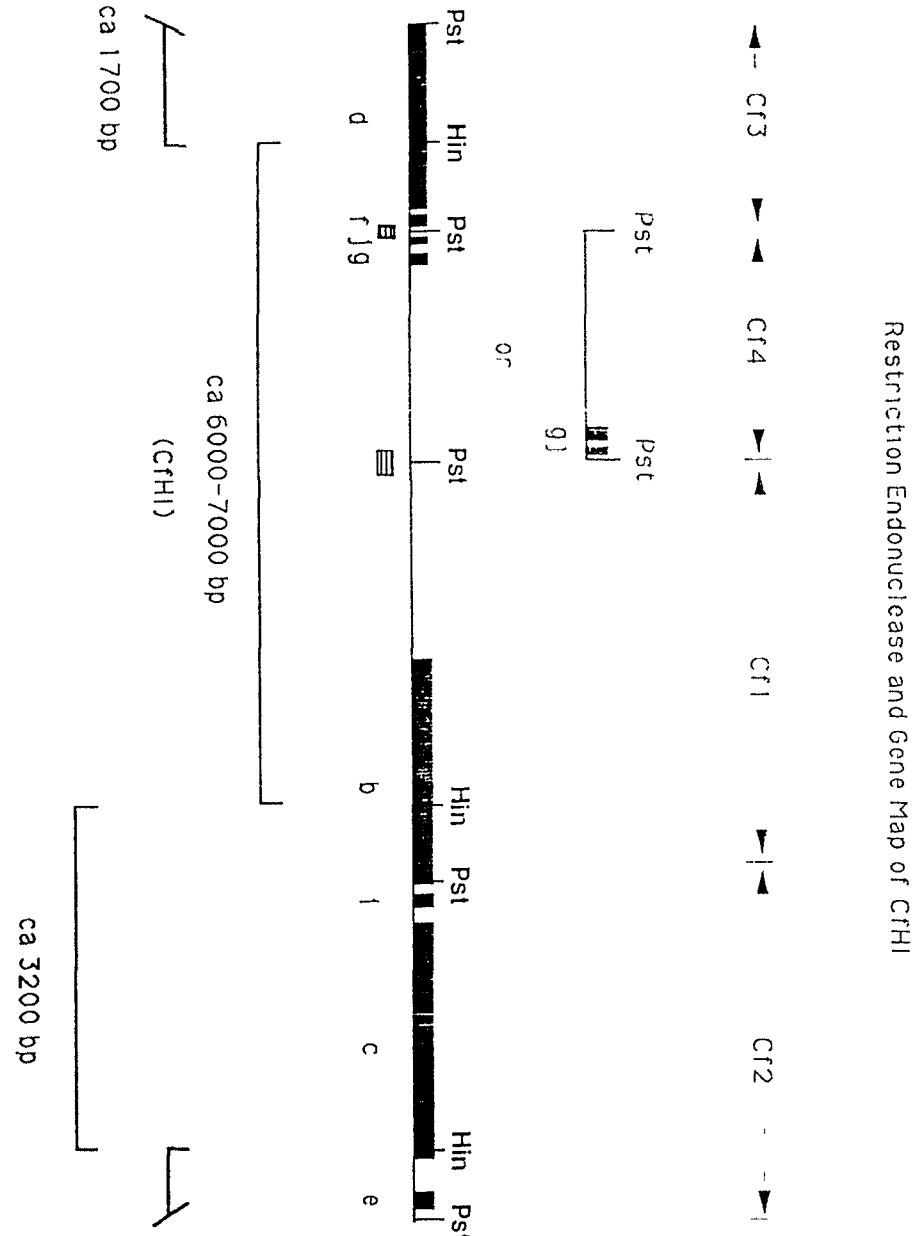



**Figure 22.** Schematic diagram of the deduced rDNA restriction and gene map. rRNA coding regions are depicted by black boxes and the boundary areas between Cf3 and Cf4, as well as between Cf4 and Cf1, are indicated by the hatched boxes. The bracketed Cf4 is the alternative orientaton to that depicted in the rDNA map below it. The horizontal lines below the rDNA map indicate the predicted HindIII rDNA fragments. The one of interest for cloning purposes is that labelled CfH1. Pst is PstI and Hin is HindIII. The coding regions are labelled with the corresponding letter.

followed by Cf3. Coding regions are oriented in the 5' to 3' direction going from left to right in Figure 22 .

At this point, the only uncertainty was the orientation and placement of Cf4. It was known that Cf4 was derived from a region of the genomic rDNA that was very heterogeneous in size (see Figure 9). Heterogeneity of this sort is known to exist in other eukaryotes and has been localized to the IGS. It seemed reasonable, then, to place Cf4 after Cf3. Indeed if the rDNA in Crithidia is tandemly repeated as in most other eukaryotes (or circular as in some eukaryotes), placement of Cf4 after (i.e., at the 3'-end of) Cf3 would then put Cf4 before (i.e., at the 5'-end of) Cf1. However, from the data accumulated, it was not possible to determine if the coding regions for Sp. g and Sp. j were proximal or distal to Sp. f (see Figure 22 ).

To resolve this uncertainty, and to ensure that any additional PstI rDNA fragment which might exist between Cf3 and Cf4 or Cf4 and Cf1 in the native DNA had not been missed, it was decided to clone rDNA fragments which overlap the junctions between Cf3-Cf4 and Cf4-Cf1. The restriction endonuclease HindIII proved to be an appropriate choice for carrying out these experiments.

## 6.) CLONING AND IDENTIFICATION OF HindIII rDNA CLONES

The restriction endonuclease HindIII was chosen as a good candidate enzyme to provide an rDNA clone which would possess the pCf3-pCf4 and pCf4-pCf1 boundaries because the mapping studies of Cf4 demonstrated the lack of a HindIII restriction site (see Figure 21).

At this stage, the mapping data suggested that a HindIII rDNA fragment should extend from the HindIII site in Cf3 through Cf4 to the next

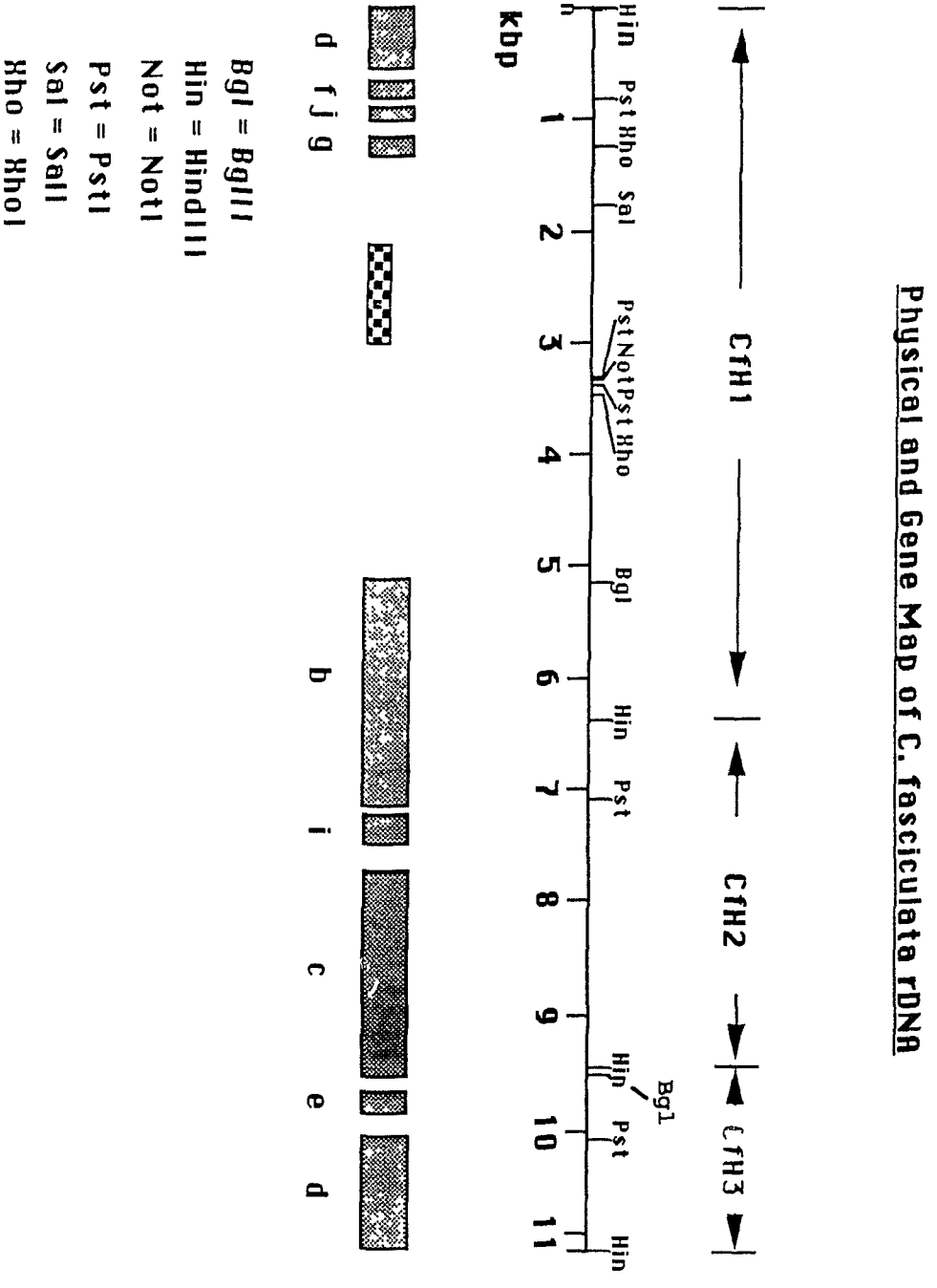



**Figure 23.** Schematic line diagram of the restriction and gene map of Crithidia rDNA. The restriction sites are indicated by the vertical lines above the horizontal and the length in kbp below. The rRNA coding regions are represented by shaded boxes and labelled with the corresponding letter. The checkered box indicates that area of the IGS that exhibits length heterogeneity.

HindIII site in the coding region of the SSU rRNA in Cf1 (see Figure 23). If such a fragment could be isolated intact, mapping and primary sequence analysis of the DNA would allow determination of the pCf4 orientation and the presence or absence of any additional PstI rDNA fragments which may have been missed in the initial cloning experiments. Due to the fact that this presumed HindIII rDNA fragment would possess the region in common with pCf4 (indeed it should possess the entire pCf4 rDNA fragment) which is responsible for the length heterogeneity, it seemed it should be possible to detect evidence of such heterogeneity on a Southern blot of HindIII-digested nDNA, using Sp. g or Sp. j or iRNA as a probe. Figure 24 shows the results of such an experiment. The iRNA was seen to hybridize to fragments ca. 1700 bp and 3200 bp in size, as well as a ladder of sub-stoichiometric fragments ranging in size from ca. 6000 to 8000 bp. From the restriction map of the rDNA constructed from pCf1 through pCf4, it can be determined that a genomic blot should produce an ca. 1700-bp fragment, an ca. 3200-bp size fragment, and a ladder of rDNA fragments ranging in size from 6000 bp to 7000 bp. The observed fragments (Fig. 24) are in good agreement with those predicted.

Cloning of HindIII rDNA fragments was carried out as in the case of PstI rDNA fragments. Colonies which hybridized with radioactively labelled rRNA in the in situ colony hybridization experiments were selected for plasmid extraction, restriction enzyme digestion, agarose gel electrophoresis, and Southern hybridization. A total of ten HindIII rDNA clones were selected for more detailed restriction enzyme analysis. Among the 10 rDNA recombinants selected, four contained inserts corresponding in size to the ca. 3200-bp size HindIII fragment, two contained inserts corresponding to the ca. 1700-bp size HindIII fragment, and one HindIII rDNA clone contained an rDNA fragment in the size range of the hybridization ladder (between 6000 and 7000 bp). Double

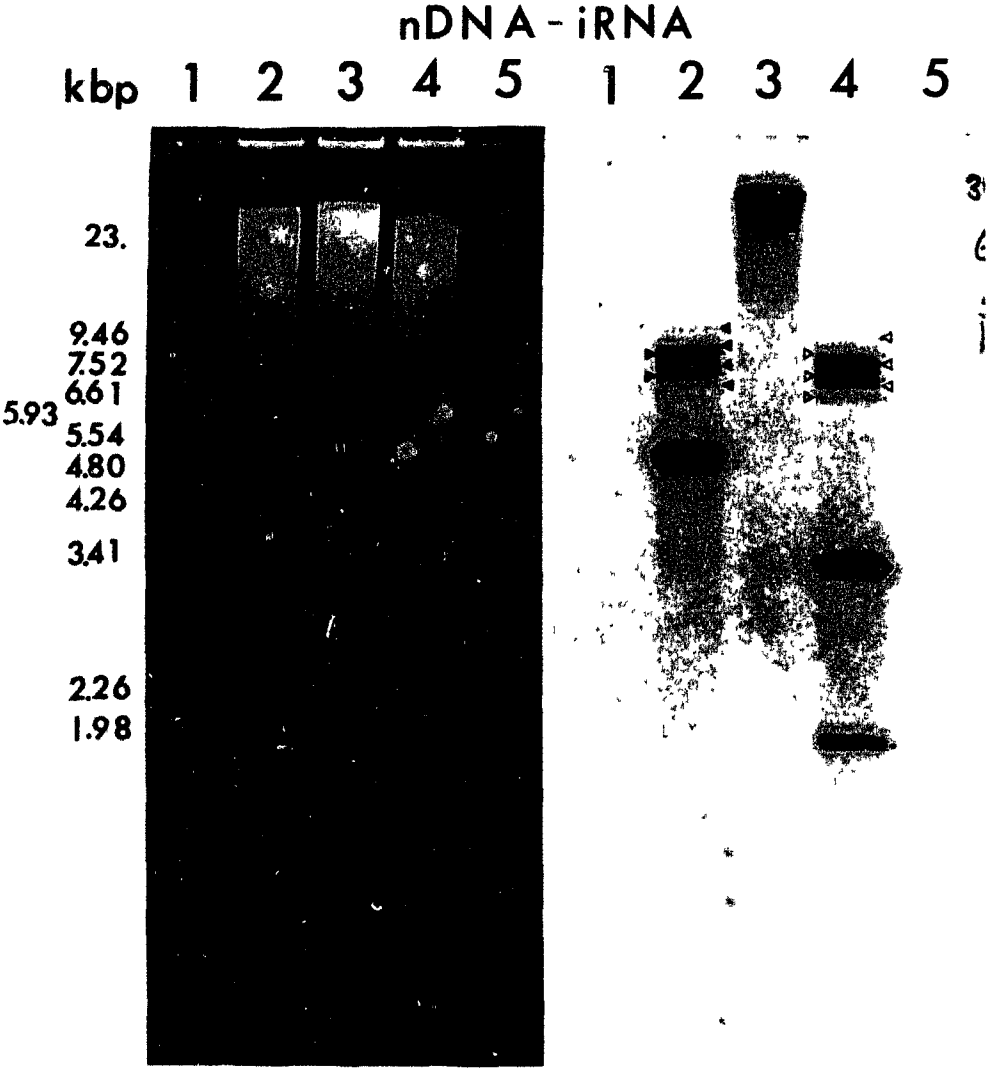



**Figure 24.** Photograph of EtBr-stained 1% agarose gel and autoradiograph of its Southern blot probed with $^{32}$P-labelled iRNA. Lanes 1 and 5 contain lambda DNA digested with EcoRI and HindIII, respectively. Marker sizes are indicated in kbp. Lane 2 contains nDNA digested with BglII, lane 3 nDNA digested with KpnI, and lane 4 nDNA digested with HindIII.

digestions of representative clones of each size class with HindIII plus PstI yielded the fragments of the expected size, confirming their identity. Figure 25 contains this information.

Lanes 1, 2, 3, 5, 6, 8, 9, 11, 12 and 14 contain the 10 different HindIII rDNA recombinants digested to completion with HindIII. Lanes 1, 3, 5 and 6 contain a plasmid with an rDNA insert of the same size (ca. 3200 bp). Lane 4 contains the same plasmid as in lane 5 but doubly digested with HindIII plus PstI. The insert contains a single PstI site as demonstrated by the two fragments present (ca. 2500 and 700 bp). This is the predicted outcome of digesting CfH1 with PstI (see Figure 22). Lane 2 has an rDNA recombinant which has undergone a mutation resulting in an alteration in the size of pUC9 (now ca. 3500 bp). Its insert DNA is ca. 1200 bp in length.

Lanes 8 and 14 contain plasmid DNA which has the same size insert DNA of ca. 1800 bp in length. This is the expected size of the smallest rDNA HindIII fragment (see Figure 22). When it is digested with PstI plus HindIII (lanes 7 and 13), the result is two fragments of ca. 650 and 1150 bp in length in each case. These are the sizes predicted from the restriction map (Figure 22).

Lane 11 contains the largest plasmid. Its insert is ca. 6900 bp in length and is expected to be the HindIII rDNA fragment that contains the homologue of Cf4 plus the Cf3-4 and Cf4-1 boundaries. Given this, I would expect fragments of ca. 850 bp (HindIII-PstI), ca. 3200 bp (HindIII-PstI), and 2000 to 3000 bp (PstI-PstI) when this insert is digested with these two enzymes. The 2000-to 3000-bp fragment would be the homologue of Cf4. The results of such a digestion can be seen in lane 10. The ca. 850-and 3200-bp fragments are present, as is an ca. 2800-bp one. This strongly suggests that this is the anticipated clone and that the 2800-bp fragment is the homologue of Cf4, albeit 600 bp

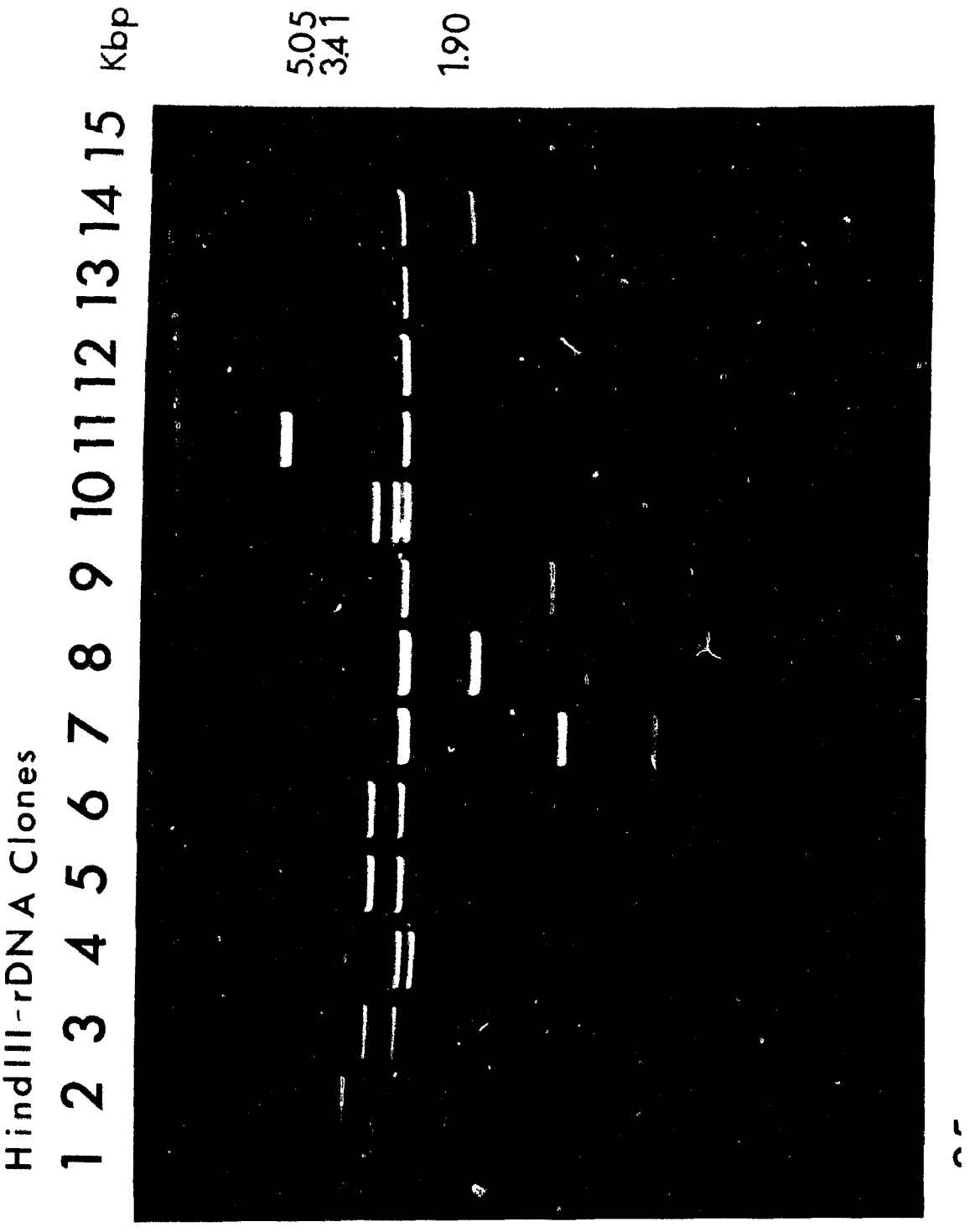



**Figure 25.** EtBr-stained 1% agarose gel containing in lanes 1-14, HindIII pUC9-rDNA clones digested with either HindIII or HindIII plus PstI. Lane 11 contains CfH1 digested with HindIII alone; while in lane 10 CfH1 is digested with HindIII and PstI. Lanes 1, 3, 5 and 6 contain clones with the same sized HindIII-rDNA insert (3.2 kbp). Lanes 8 and 14 contain the same HindIII-rDNA insert of a different class (1.7 kbp). Lane 2 contains a mutated pUC9-rDNA clone resulting in an alteration in the size of the pUC9 insert liberated by HindIII digestion. Lane 4 contains the clone in lane 5 digested with PstI and HindIII. Lane 7 contains the lane 8 clone digested with PstI and HindIII. Lane 15 shows lambda DNA markers, whose sizes are indicated in kbp.

longer. Lanes 9 and 12 illustrate two different recombinant plasmids with unanticipated insert size DNAs (ca. 1200 and 300 bp respectively).

The 6.9-kbp HindIII rDNA clone was used for more detailed restriction mapping, in order to confirm that this clone was indeed the desired one, and for sequencing purposes. Primary sequence analysis was carried out to determine the exact nature of the Cf3-Cf4 boundary and the orientation of Cf4 relative to its neighbors, and to allow a better delineation of the Cf4-Cf1 boundary region. The DNA sequence information obtained from pCfH1 is presented elsewhere in this thesis.

Figure 26 contains the relevant restriction mapping experiments. Lane 2 contains pCfH1 digested with HindIII only. The insert is ca. 6900 bp long. When doubly digested with BamHI and HindIII (lane 4), it can be seen that the two fragments of expected size (ca. 1450 and 5500 bp) are produced. SalI (lane 5) cuts the insert DNA once only, as expected, resulting in two fragments, one ca. 1850 and one ca. 5000 bp long. Lane 7 demonstrates that the insert DNA contains one NotI site near the middle of the DNA, resulting in two fragments of ca. 3300 and 3600 bp in length upon NotI digestion. An unexpected finding is that XhoI (lane 3) cuts the insert twice and not once. The small (ca. 1350 bp) fragment which is predicted is observed but the other expected fragment of ca. 5450 bp is absent, and instead replaced by two fragments of ca. 2400 and 3150 bp. This extra XhoI site either represents a sequence difference between this DNA and the corresponding DNA already cloned as Cf1-4, or this XhoI site was missed in the mapping of the Cf1-4 DNAs. It is possible that if the second XhoI site in this insert DNA was ca. 2400 bp downstream of the confirmed XhoI site in the Cf4 homologue, it would map <100 bp from the PstI site which marks the boundary between Cf4 and Cf1. Thus this site in Cf1 could well have been missed. If the site is positioned ca. 3150 bp downstream of the known XhoI site

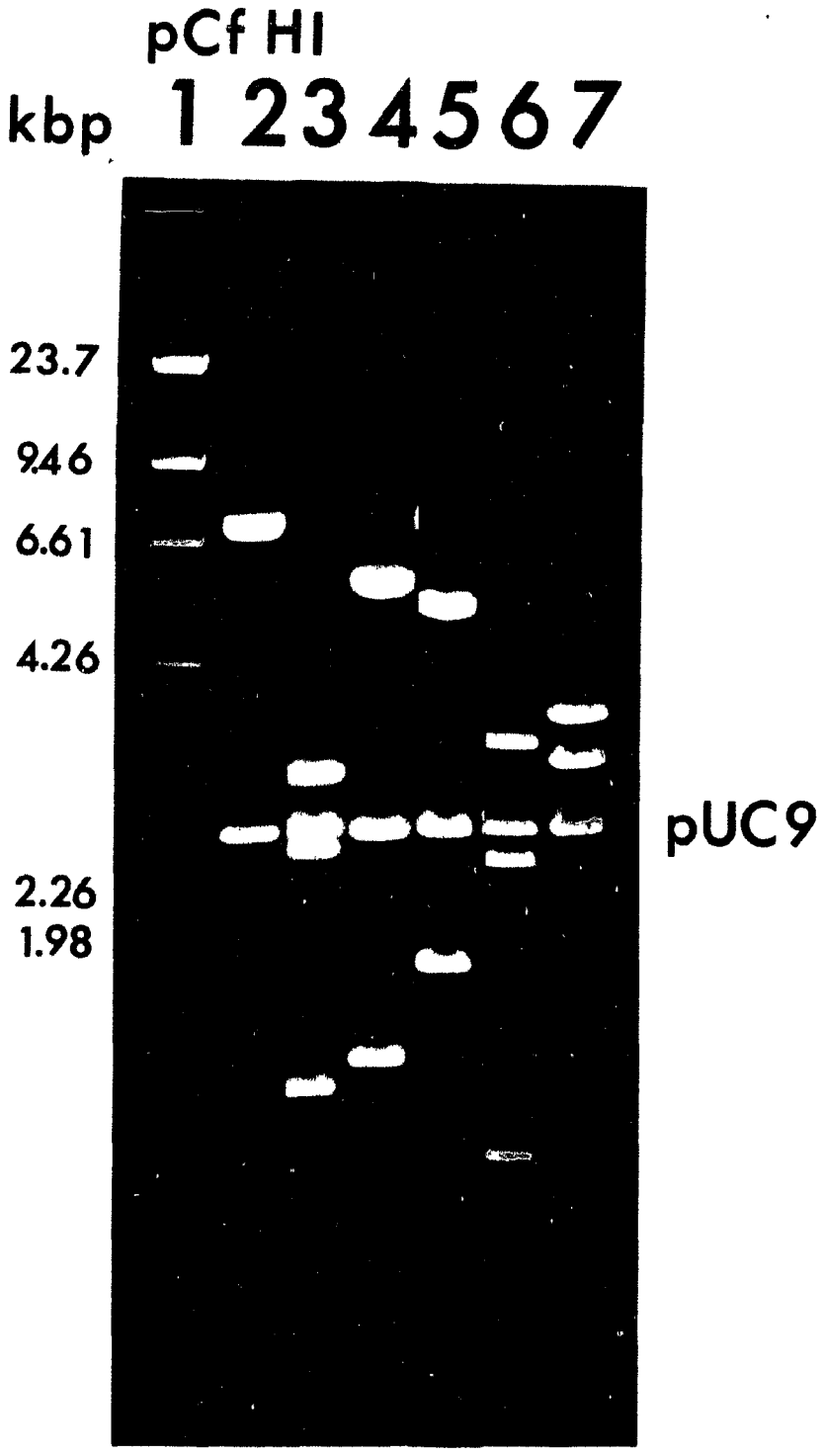
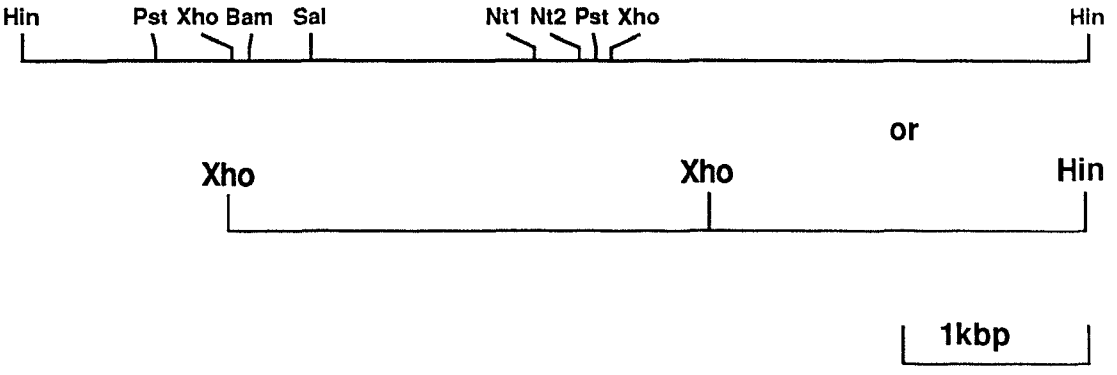



**Figure 26.** EtBr-stained 1% agarose gel. Lane 1 contains lambda DNA digested with HindIII; marker sizes are indicated in kbp. Lanes 2-7 contain CfH1 DNA digested with HindIII and one other restriction endonuclease: lane 2, HindIII alone; lane 3, XhoI; lane 4, BamHI; lane 5, SalI; lane 6, NcoI and lane 7, NotI.

Restriction Map of CfH1




Nt1 or Nt2 = NotI

Hin = HindIII

Pst = PstI

Xho = XhoI

Bam = BamHI

Sal = SalI

**Figure 27. Schematic representation of two alternative restriction endonuclease maps for CfH1. Vertical line refers to the position of restriction endonuclease cleavage sites.**

in Cf4, it would map to a site well within Cf1 and should therefore have been discovered during characterization of the fragment (see Figure 27 for a visual depiction of these possible arrangements). If this is the case, then it is necessary to postulate a sequence difference between Cf1 and CfH1. Figure 27 depicts the determined restriction map of CfH1, the 6.9-kbp HindIII rDNA insert. The pCfH1 clone contains within it a Cf4 homolog which is ca. 600 bp longer in length. The DNA fragments homologous to Cf1 and Cf3 are the expected size.

## 7.) rDNA LENGTH HETEROGENEITY STUDIES

As previously mentioned,the length heterogeneity associated with a particular region of the rDNA repeat was noted when genomic DNA was digested with PstI and probed with either Sp. g or Sp. j, or when the nDNA was digested with HindIII and probed with 5'-end-labelled iRNA. With PstI-digested genomic DNA, species g and j each hybridizes to a ladder of rDNA fragments ranging in size from ca. 2.0 kbp to ca. 3.0 kbp (Fig. 28). The individual rDNA fragments differ from one another by about 200-300 bp, although the size difference is not a constant. There is a great deal of variability in the degree to which any individual rDNA fragment in the ladder hybridizes with the rRNA probe. This indicates that the rDNA fragments are not equimolar, but are present in the genome in varying amounts.

Figure 24 illustrates the results obtained when HindIII-digested rDNA and BglII-digested nDNA were probed with 5'-end-labelled iRNA. It can be seen that both enzymes generate an rRNA hybridization pattern not unlike that seen with PstI-digested rDNA. In both cases, there is a very similar ladder of rDNA fragments which also closely resembles that observed with PstI-digested nDNA. Indeed, the banding pattern (both band intensities and band separations) of the ladder of fragments appear identical with all three enzymes.

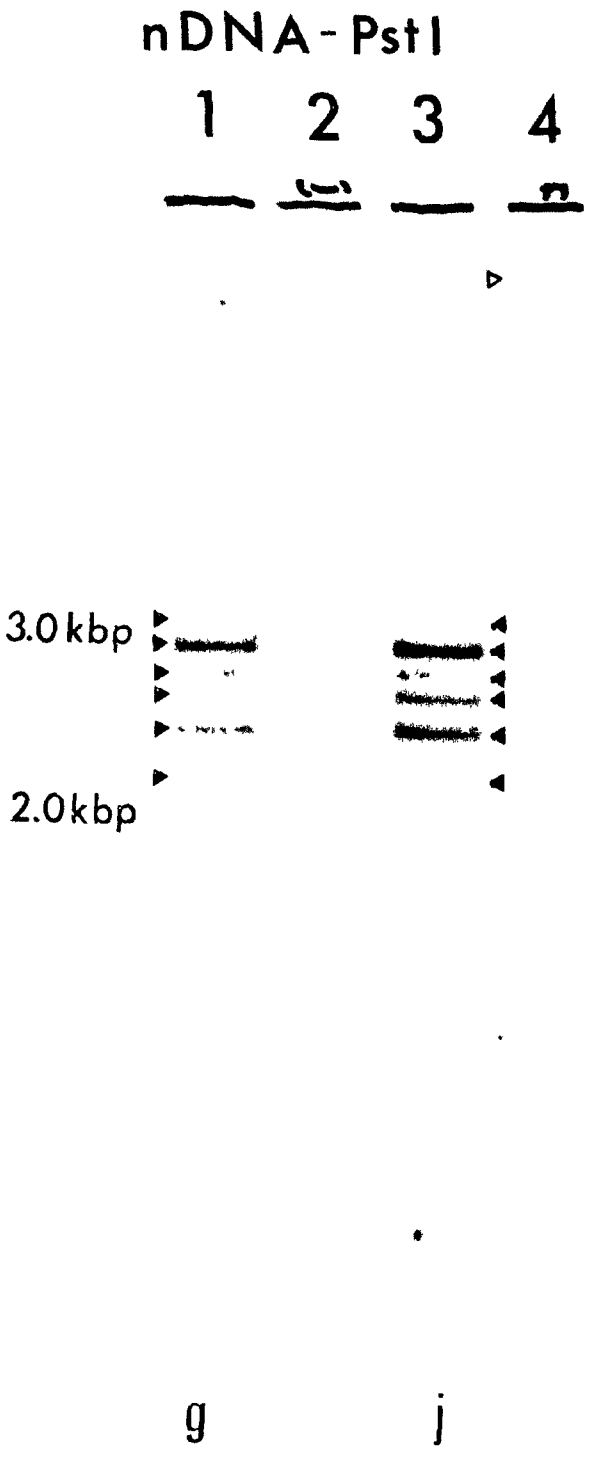



**Figure 28.** Autoradiograph of a Southern blot probed with $^{32}$P-labelled Sp. g or Sp. j. Lanes 1 and 3 contain nDNA digested to completion with PstI and probed with either Sp.g (g) or Sp. j (j). Lane 4 contains undigested nDNA probed with Sp. j (j). The closed triangles mark the rDNA fragments that hybridize either rRNA probe. Size markers are indicated (kbp). Lane 2 does not contain any DNA.

It was demonstrated in the mapping studies of pCf1-pCf4 and pCfH1 that the region of rDNA that is responsible for the length heterogeneity maps to a region 3' to the coding regions for Sp. g and Sp. j and 5' to the SSU rRNA gene, i.e., to the IGS. Indeed this sort of length heterogeneity has been noted in a number of other eukaryotic species, most notably Xenopus sp., mouse, and Drosophila . It is known in these organisms that the length differences of various rDNA repeats is due to a varying number of direct repeats in the IGS. In the above-mentioned species, these directly repeated sequences are homologous to the transcription initiation promoter, and indeed are thought to be capable of promoting and/or enhancing initiation.

To determine if this length heterogeneity was a population phenomenon (inter-individual) or if it was present in a single cell (intra-individual), C. fasciculata was plated at low density on agar plates and allowed to grow into single colonies. The plating densities were adjusted to ensure that only individual organisms would give rise to colonies, which were picked to be grown up in small liquid cultures (2 ml). One half of the cells was used to extract nDNA for analysis.

Crithidia fasciculata is not known to have a sexual cycle (164). Coupled with the fact that single colonies were grown in small liquid cultures, and hence the number of cell divisions would be small, this means that any DNA recombination due to sexual exchange of DNA should be negligible.

The clonally derived DNA was digested with one of several restriction endonucleases and probed with one of the end-labelled rRNAs. The results are shown in Figure 29. 3'-End-labelled Sp. j was used as a hybridization probe to five PstI-digested, clonally-derived preparations of nDNA. The two right-hand lanes contain PstI-digested nDNA derived from cells grown in large liquid culture and seeded with a multi-cell inoculum. In all five cases of clonally-

derived nDNA, the hybridization pattern with Sp. j was the same as with the non-clonal nDNA. In Figure 29 , a seventh rDNA fragment of about 4.0 kbp in size is seen to hybridize with Sp. j. Although this fragment was not detected in previous hybridization experiments, it can be seen in the nonclonal as well as clonal rDNA. Undoubtedly, its appearance in this particular Southern blot is attributable to a longer exposure time and/or the use of a higher-specific-activity rRNA probe. The pattern of band intensities is constant in all five clonal rDNAs; given the degree of sensitivity of this technique, it appears then that this pattern of rDNA length heterogeneity is found in each individual and is rather stable. This is demonstrated by the fact that lane 7 contains nDNA extracted from a large liquid culture seven years prior to that in lane 8. During these eight years the organisms were maintained in semi-solid agar by serial inoculations.

Figure 30 demonstrates the pattern obtained when fifteen clonally derived rDNAs were digested with PstI and probed with end-labelled Sp. f. It can be seen that in all cases, the probe hybridizes to an ca. 2.0-kbp fragment, the Cf3 homologue. In lane 11 it is also seen to hybridize to an additional, sub-stoichiometric fragment ca. 4.9 kbp in length. It thus appears that one of the fifteen clonally-derived nDNAs has undergone some kind of rearrangement involving rDNA in the vicinity of the coding region for Sp. f.

The results obtained when the clonally derived nDNAs were digested with BglII are almost identical to those obtained when large-scale preparations of nDNA were digested with BglII and probed with end-labelled Sp. g and Sp. f. As can be seen (Figures 31 and 32), the pattern of rDNA fragments and the intensity of hybridization of the end-labelled rRNAs to the individual rDNA fragments are identical, within the limitations of this technique.

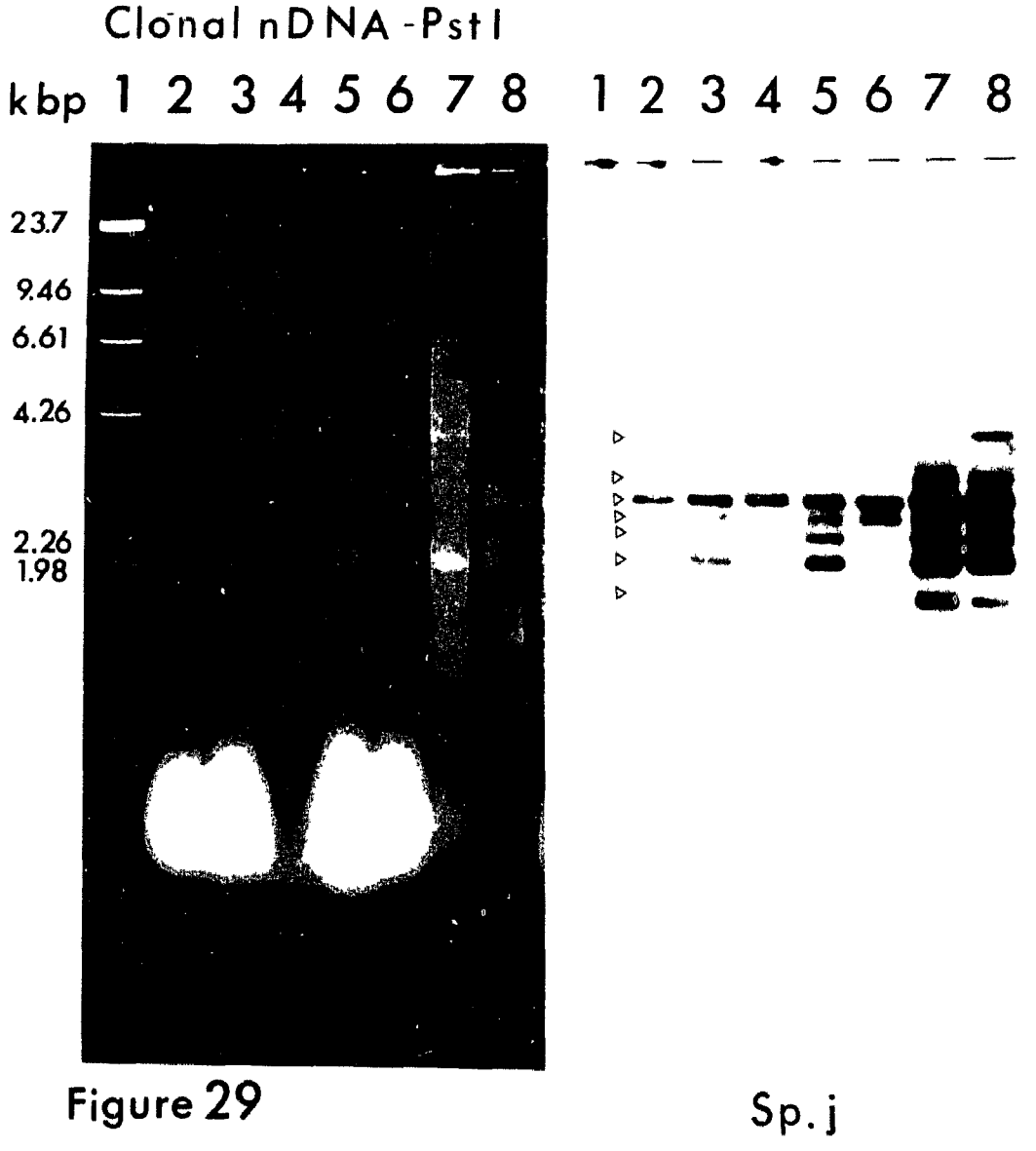



**Figure 29.** EtBr-stained 1% agarose gel and autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. j. Lane 1 contains lambda DNA digested with EcoRI; marker sizes are in kbp. Lanes 2-6 contain clonally derived Crithidia nDNA digested to completion with PstI. Lane 7 contains nDNA extracted from a large culture of cells in 1979 and digested to completion with PstI. Lane 8 contains nDNA extracted from a large culture of cells in 1986 and digested with PstI. The open triangles mark the ladder of PstI rDNA fragments which hybridize Sp. j. Note should be made of the consistency of the pattern between the different DNA samples.

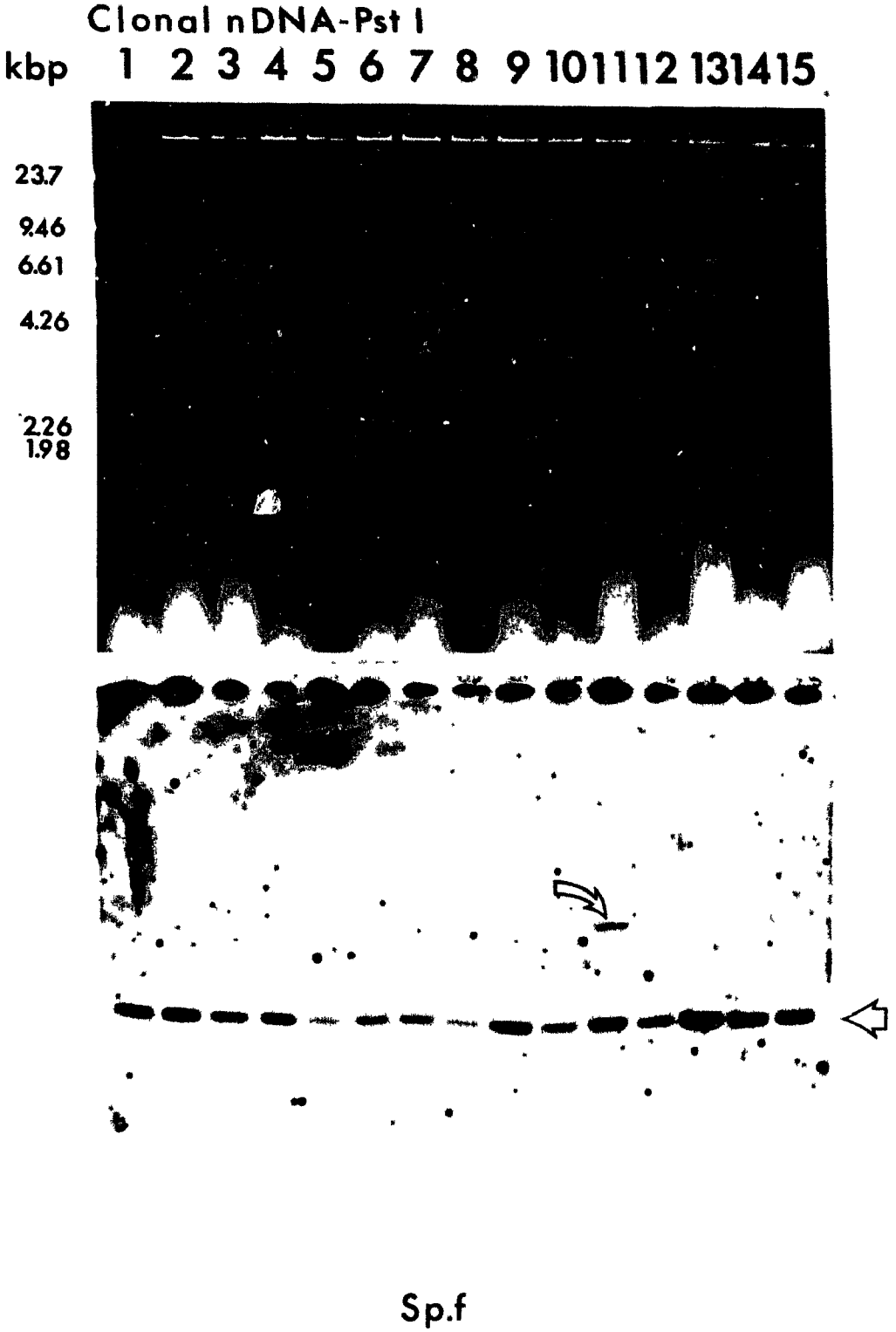



**Figure 30.** EtBr-stained 1% agarose gel and its autoradiograph probed with $^{32}$P-labelled Sp. f. DNA marker sizes are indicated in kbp at the left hand margin of the agarose gel. Lanes 1-15 contain different clonally derived Crithidia nDNA samples digested to completion with PstI. The straight arrow points to the 2.0 kbp Cf3 homologue which hybridizes Sp. f in each sample. The curved arrow points to the additional hybridizing band in lane 11.

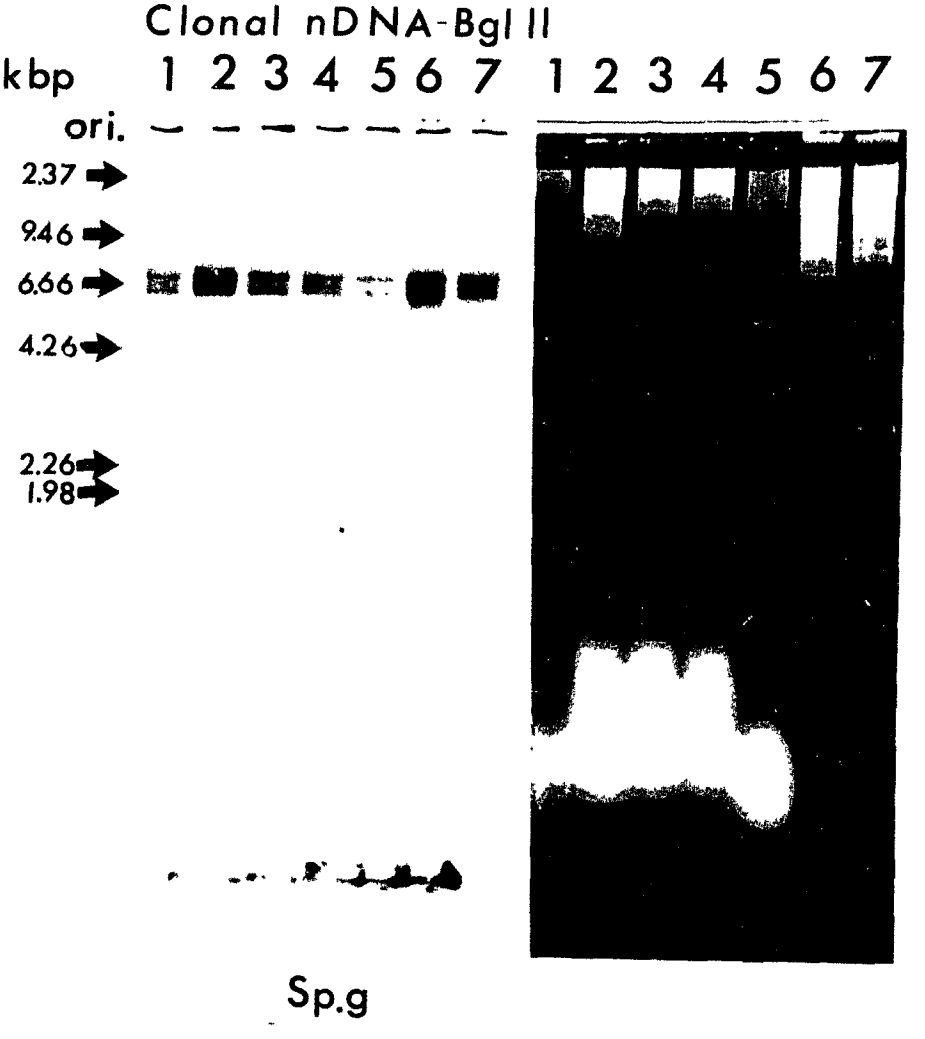



**Figure 31.** EtBr-stained 1% agarose gel and autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. g. Lanes 1-5 contain clonally derived nDNA digested to completion with BglII. Lane 6 contains poly-clonal nDNA purified in 1979 and digested with BglII. Lane 7 contains poly-clonal nDNA purified in 1986 and digested with BglII. DNA markers, in kbp, are indicated at the left hand border of the autoradiograph.

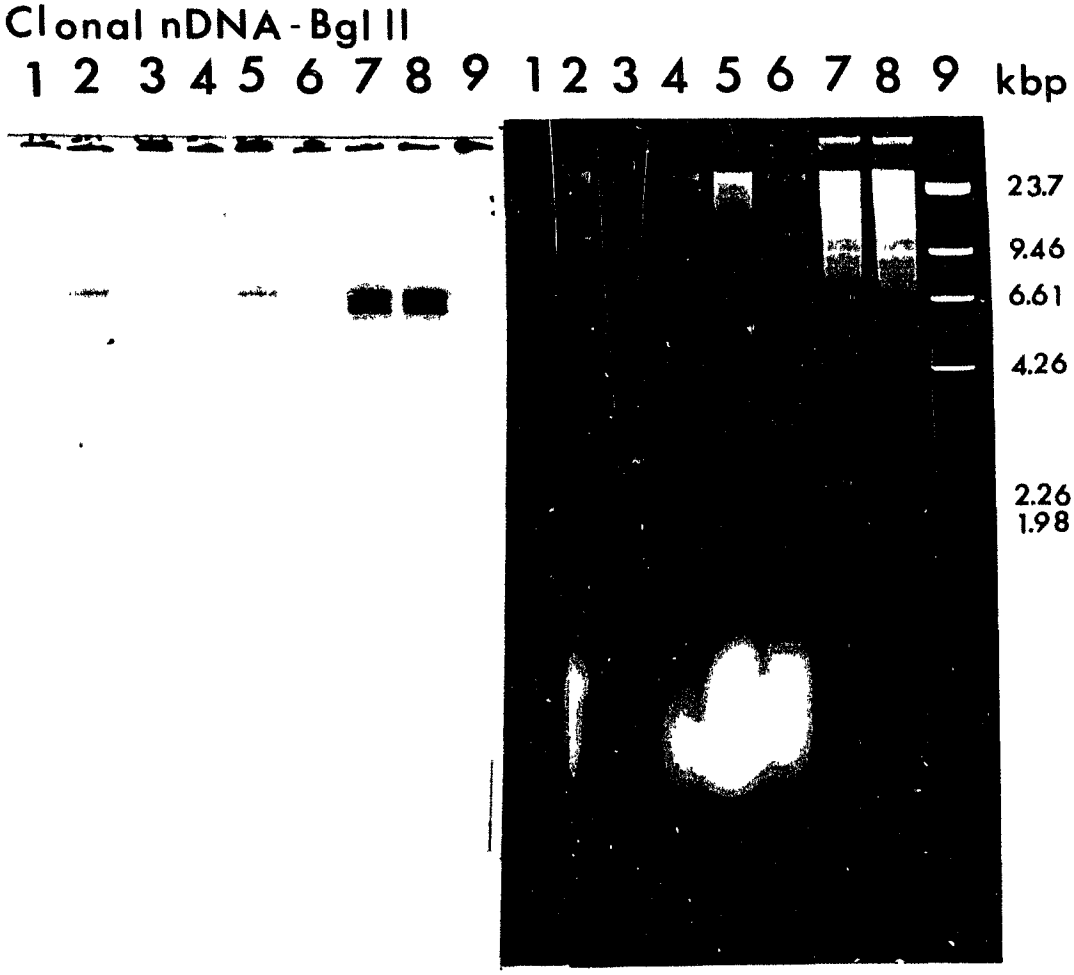



Figure 32

**Figure 32.** EtBr-stained 1% agarose gel and autoradiograph of its Southern blot probed with $^{32}$P-labelled Sp. f. Lanes 1-6 contain six different clonally derived Crithidia nDNAs digested to completion with BglII. Lane 7 contains poly-clonal nDNA extracted in 1979 and lane 8 contains poly-clonal nDNA extracted in 1986, both digested to completion with BglII. Lane 9 contains lambda DNA digested with EcoRI.

## 8.) DNA METHYLATION STUDIES

Methylation of DNA has been shown to occur in the nucleus of eukaryotes. Most often it takes the form of a methyl group attached at the 5 position of cytidine when cytidine is adjacent to guanosine i.e., 5' $m^5$CG 3'. Methylation of adenosine residues (to give $N^6$-methyladenine, $m^6$Ade) also occurs in nDNA.

Methylation of DNA sequences has been implicated in the control of transcription (see the Introduction for a detailed discussion of DNA methylation). To determine the extent of DNA methylation in C. fasciculata rDNA, nDNA was digested with the isoschizomers HpaII and MspI, which recognize the DNA sequence 5' CCGG 3', but have different sensitivities to cytidine methylation, and MboI and Sau3A, which recognize the DNA sequence 5' GATC 3', and have different sensitivities to adenine methylation.

Nuclear DNA digested with HpaII or MspI and Sau3A or MboI was size-fractionated by electrophoresis in an agarose gel, transferred to a nitrocellulose membrane, and probed with 5'-end-labelled iRNA. The results obtained are shown in Figure 33. To ensure complete digestion the DNA was incubated under conditions recommended by the manufacturer using 5 to 10-fold excess restriction endonuclease. It can be seen that the patterns obtained with each pair of isoschizomer are the same. This means in the case of HpaII that the rDNA contains at least five CCGG sites (five sites instead of four sites becouse the total size of the hybridizing rDNA fragments do not add up to the total size of the rDNA cistron) in which the C residues are unmethylated. The results obtained in the Sau3A/MboI experiment indicate that there are at least four GATC sequences in the rDNA that do not contain any appreciable amount of $m^6$Ade. In the Sau3A lane a fourth very faint band, ca. 2100 bp, can be discerned. This may indicate that a small amount of $m^6$Ade is present in the rDNA.

Figure 33

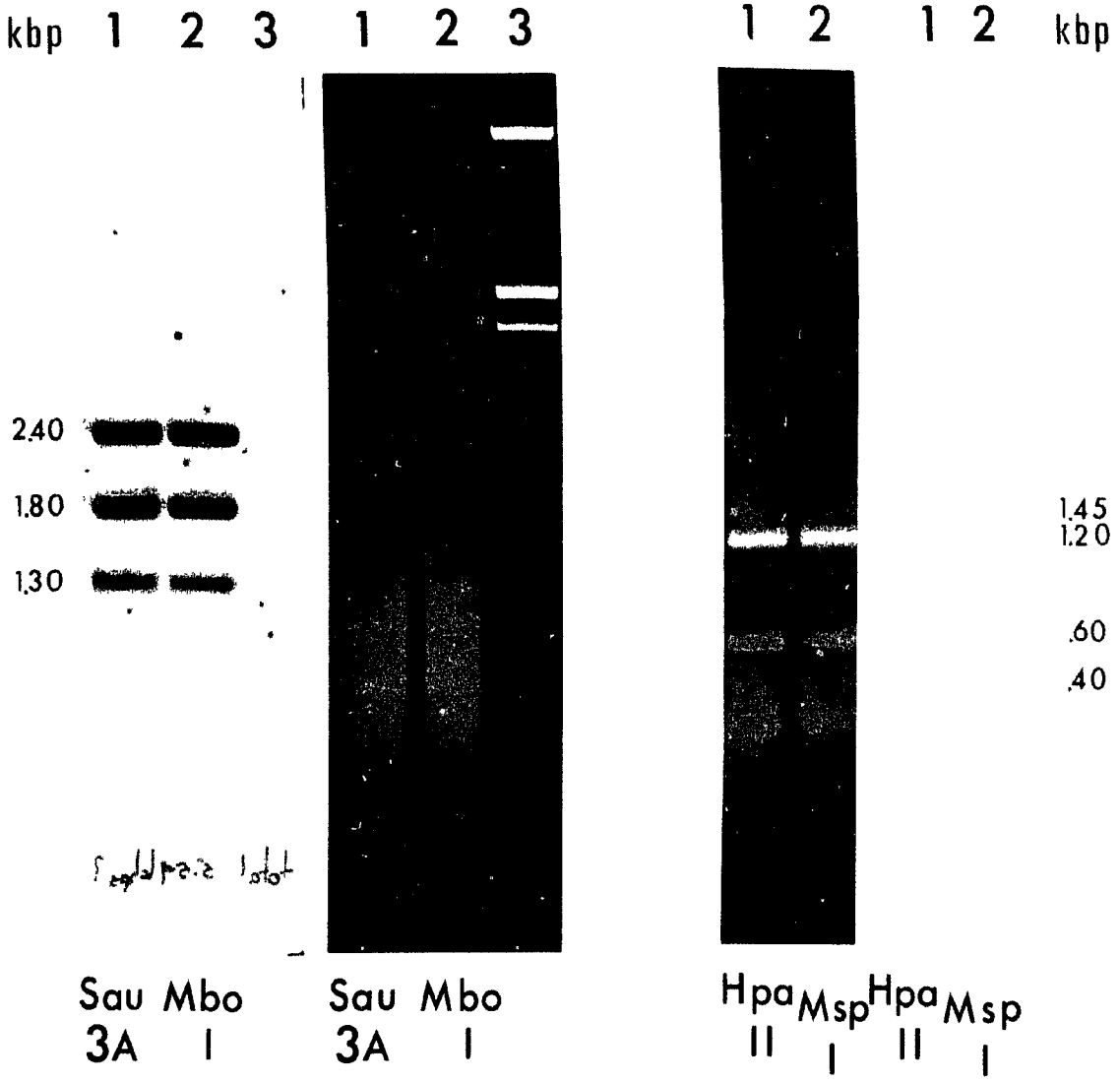



**Figure 33.** Composite figure of two EtBr-stained 1% agarose gels and the autoradiographs of their Southern blots probed with $^{32}$P-labelled iRNA. The left hand lanes 1 and 2 contain Crithidia nDNA digested to completion with either Sau3A or MboI. When probed with iRNA three main bands are seen to hybridize. A fourth faint band may be seen in lane 1, perhaps indicating a minimal amount of rDNA methylation. Lane 3 contains lambda DNA digested with HindIII for size markers. The right hand lanes 1 and 2 contain Crithidia nDNA digested with either HpaII or MspI. The labelled iRNA is seen to hybridize to four rDNA fragments in both lanes 1 and 2. No differences in the band patterns can be discerned, suggesting that little if any methylation is present at the assayable sites.

## 9.) PRIMARY SEQUENCE DETERMINATION OF C. FASCICULATA rDNA

Clones pCf1, pCf2, pCf3, pCf4 and pCfH1 were used to determine the primary sequence of the complete rDNA unit. pCfH1 was used to obtain sequence at the Cf3-CF4 and Cf4-Cf1 boundaries (Fig. 34). To expedite analysis of the coding regions for the SSU and LSU rRNAs, the 18S rRNA coding region in Cf1 was determined by Dr. M. N. Schnare while the sequences of Cf2 and Cf3 were determined by Dr. D. F. Spencer.

In 1986 Schnare et al. (191) reported the gene sequence of the Crithidia fasciculata SSU rRNA, as determined from Cf1. As previously mentioned, the 3' end of this rRNA was found to contain a recognition site for PstI 13 nt from the end. Mapping studies localized the coding region to one half of Cf1 with the 5' end of the rRNA in the vicinity of the BglII site (see Figure 12). The exact position of the 5' end was determined by comparing the RNA primary sequence with that of the DNA. This positioned the mature 5' end exactly at the BglII site. The SSU rRNA gene has a G plus C content of 49.7% and is 2206 nt long. At the time, this was the longest SSU rRNA known [334 nt longer then the next longest, that of D. discoideum (1872 nt)] and consistent in size with that determined by acrylamide gel electrophoresis (ca. 2400 nt). The primary sequence was analyzed and found to fit a phylogenetically conserved secondary structure model with the extra length being accommodated within discrete variable domains that are present in eukaryotic SSU rRNAs (191). These domains (V1, V2, V4-V9) are dispersed throughout the length of the molecule and are separated by regions of highly conserved primary or secondary structure. These variable regions correspond closely in position to regions of pronounced structural variation identified in a comparison of eubacterial and organellar SSU rRNA sequences (162). V3 is highly conserved among eukaryotes but is different

from the same region in eubacteria. Most of the extra length in the Crithidia SSU rRNA can be accounted for by extra nucleotides in V7 (ca. 200) and in V8 (ca. 100). Pairwise comparisons of the primary sequences of non-variable (or universal) regions of various SSU rRNAs revealed that Crithidia represented the earliest branching within the eukaryotic lineage, earlier than either D. discoideum or T. thermophila (191).

Primary sequence analysis of that portion of the cloned rDNA that contained the coding regions for Sp. c-j (excluding Sp. h) was published in 1987 (178). Dr. D.F.Spencer determined the DNA sequence of Cf2 and Cf3 while I determined the sequence of Cf4, the overlap between Cf3-Cf4 and Cf4-Cf1 in CfH1 and the left hand portion of Cf1 to the start of the coding for the SSU rRNA (see Figure 34).

As noted previously, Sp. i, one of the seven RNAs found in the large subunit of the Crithidia cytoplasmic ribosomes, proved to be the 5.8S rRNA (163). The primary sequences of the remaining small rRNAs had been determined and were already published by the time the DNA sequencing work was undertaken (166). Primary sequence analysis of these four small rRNAs could not unequivocally identify them as part of the more conventional LSU rRNA; therefore to better understand the relationship of the small rRNAs and the 5.8S rRNA to the larger fragments of the LSU rRNA (i.e., Sp.c and d), the regions encoding these RNAs were sequenced.

Sp. c-j (excluding Sp. h) are encoded in their entirety and without interruption in the cloned rDNA (see Figure 37 ). Their coding regions are arranged as follows: Sp.b/ITS1/Sp.i/ITS2/Sp.c (28S)/ITS3/Sp.e/ITS4/Sp.d (28S) /ITS5/Sp.f/ITS6/Sp.j/ITS7/Sp.g/IGS.

Thirteen nucleotides of the 3'-end of Sp. b are encoded at the start of Cf2, with ITS1 being 366 bp long. Sp. i exhibits some length

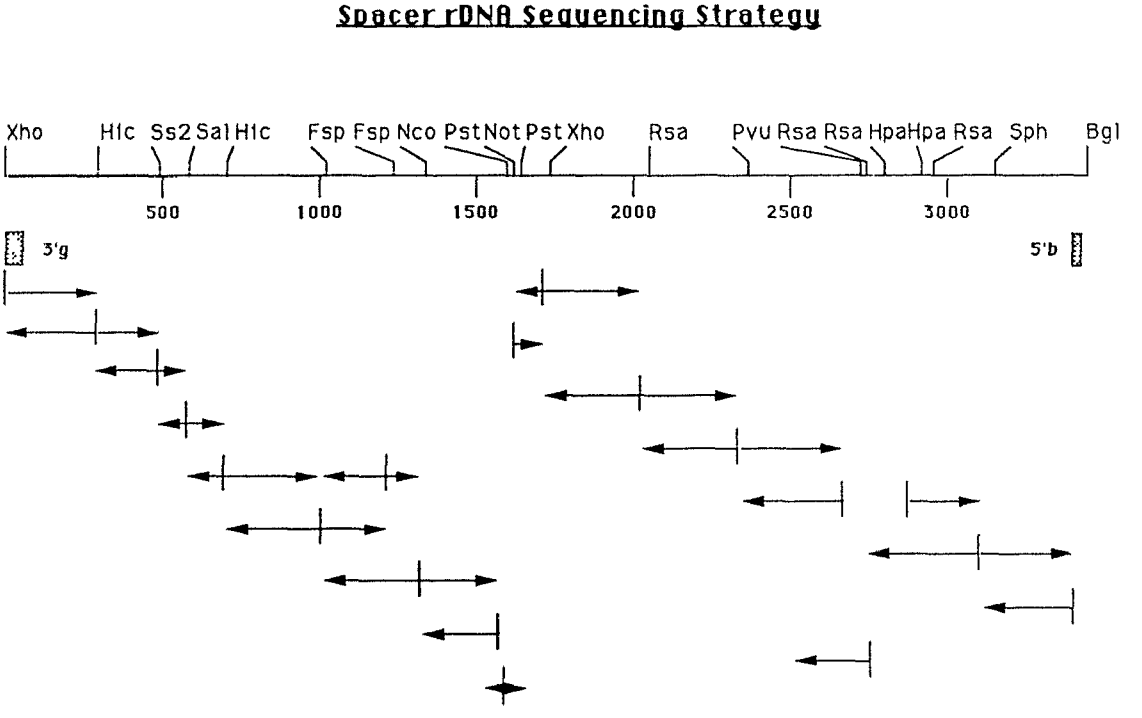



**Figure 34** Schematic diagram depicting the sequencing strategy employed to determine the primary sequence of the rDNA IGS. Vertical lines above the long horizontal indicate the sites of restriction endonuclease cleavage and the ones below indicate the length in bp. Shaded boxes indicate the coding regions for 3'-end of Sp. g (g) and the 5'-terminal region of Sp. b (b). Below the shaded boxes the vertical lines represent restriction endonuclease sites from which chemical sequencing was carried out. The arrowhead indicates the direction of sequencing and marks the extent of sequence obtained from that particular site. Note should be made of an approximately 250 nt stretch of DNA between bp 2500 and 3000 which was determined for one strand only ( between the right hand HpaI site and the third RsaI site from the BglII end). Note should also be made of the fact that the 55 nt block of sequence centered at the NotI site and bounded by the flanking PstI sites was determined from one strand only. The remaining sequence was determined for both strands. Abbreviations are: BglII=Bgl; FspII=Fsp; HincII=Hic; HpaII=Hpa; PstI=Pst; PvuII=Pvu; NcoI=Nco; NotI=Not; RsaI=Rsa; SalI=Sal; SphI=Sph; SstII=Ss2; XhoI=Xho.

heterogeneity at the 3'-end (some of the molecules have an extra C) but all of its nucleotides are encoded in the rDNA. Following the 5.8S rRNA gene is ITS2, 416 bp long. The 3' and 5' ends of Sp. c were mapped using the determined RNA primary sequences (M.N. Schnare). Sp. c is followed by ITS3 which is 124 bp long. The coding region of Sp. e is found between Sp. c and d; ITS4 follows Sp. e and is very short (31 bp). Primary sequence analysis of the 5'-end of Sp. d identified a PstI site 5 nt from the end, and this was confirmed by DNA sequence analysis. The 3'-end of Sp. d exhibits length heterogeneity (some molecules contain an additional C, which is encoded in the DNA). ITS5 is 78 bp long and this is followed by the coding region for Sp. f. ITS6, which is 262 bp long and contains the boundary between Cf3 and Cf4, separates Sp. f and Sp. j. Sp. j exhibits 5'-end heterogeneity with some molecules containing an additional U. Sp. j and g are separated by ITS7 (247 bp long). Sp. g is actually a population of molecules, some of which contain an additional A at the 5'-end. The longest of these possesses three 3' A residues not found in the DNA, and which must therefore be added post-transcriptionally.

From primary and secondary structural analysis, as well as from the relative positions of their coding regions in the rDNA, Sp.i, c, d, and f were easily identified as parts of the LSU rRNA complex of other eukaryotes and E. coli (180). Sp. i is the 5.8S rRNA, Sp. c is equivalent to the 5'-half of mouse 28S rRNA and Sp. d represents most of the 3'-half of the LSU rRNA. Sp. d lacks ca. 500 nt 3'-terminal which are present in mouse LSU rRNA, including the highly conserved $\alpha$ -sarcin cleavage site (197). This functionally important domain is found in Sp. f and therefore allows the localization of this rRNA to the corresponding region of conventional LSU rRNA.

Sp. e, which is found between the coding regions for Sp. c and d, maps to a region in eukaryotic 28S rRNA that has limited primary sequence

conservation but where potential secondary structure is maintained. Based on the fact that Sp. e contains 13 nt which are identical to a sequence in helix 43 of known eukaryotic 28S rRNAs and that Sp. e can assume a secondary structure like that in the region of helix 41-43 in other eukaryotic 28S rRNAs, it was concluded that it did indeed correspond to this region of the LSU rRNA.

Sp. g is encoded last in the linear array of Crithidia rRNA genes. This should place it at the extreme 3' end of a typical eukaryotic 28S rRNA. This region of the LSU rRNA is known to be poorly conserved in primary and secondary structure. Primary sequence analysis did demonstrate some sequence similarity (77%) between residues 98-119 of Sp. g and the terminal 21 residues of E. coli LSU rRNA. Based on this and the fact that Sp.g appears to be oligo-adenylated post-transcriptionally (as occurs with animal mitochondrial LSU rRNA (198)), this rRNA was considered to be homologous to the extreme 3'-terminal region of the 28S rRNA.

The last remaining rRNA, Sp. j, maps between Sp. f and g. In eukaryotic 28S rRNA, this is a highly variable region known as D12 (178). The only primary sequence similarity uncovered is the 9 nt stretch, residues 41-49, which are also present in the Xenopus LSU rRNA, but not in any other LSU rRNA examined. Sp. j was tentatively assigned to this location.

A search for processing signals in the rDNA (i.e., conserved or repeated primary sequences or potential secondary structures) revealed only a few features of interest. Coding/spacer boundaries of 3'-i/ITS2; 3'-d/ITS5; 3'-f/ITS6 and ITS7/5'-g are located within A+C-rich stretches. As well, A+C stretches occur just upstream of ITS3/5'-e and downstream of 3'-b/ITS1 and 3'-e/ITS4. The only conserved primary sequence is a near perfect 10 bp direct repeat found at the 3' and 5' ends of Sp. e; indeed, the ends of this RNA map to exactly the same site in each repeat.

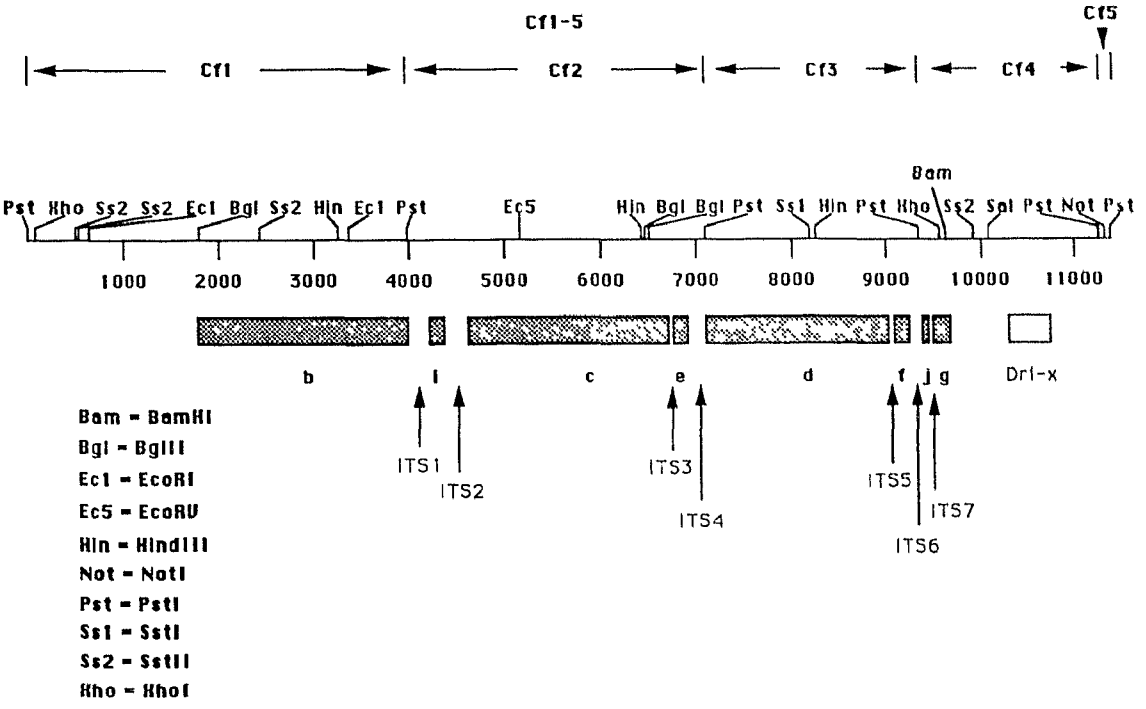



**Figure 35.** Schematic diagram of the restriction site and gene map of the entire Crithidia rDNA. Vertical lines above the horizontal line indicate the sites of restriction endonuclease cleavage and the ones below indicate the distance in bp. The darkly shaded boxes are the coding regions for the rRNA genes. They are labelled with their respective letter (i.e., **b** is Sp. b). The lightly shaded box is the site of the direct repeat, tandem array.

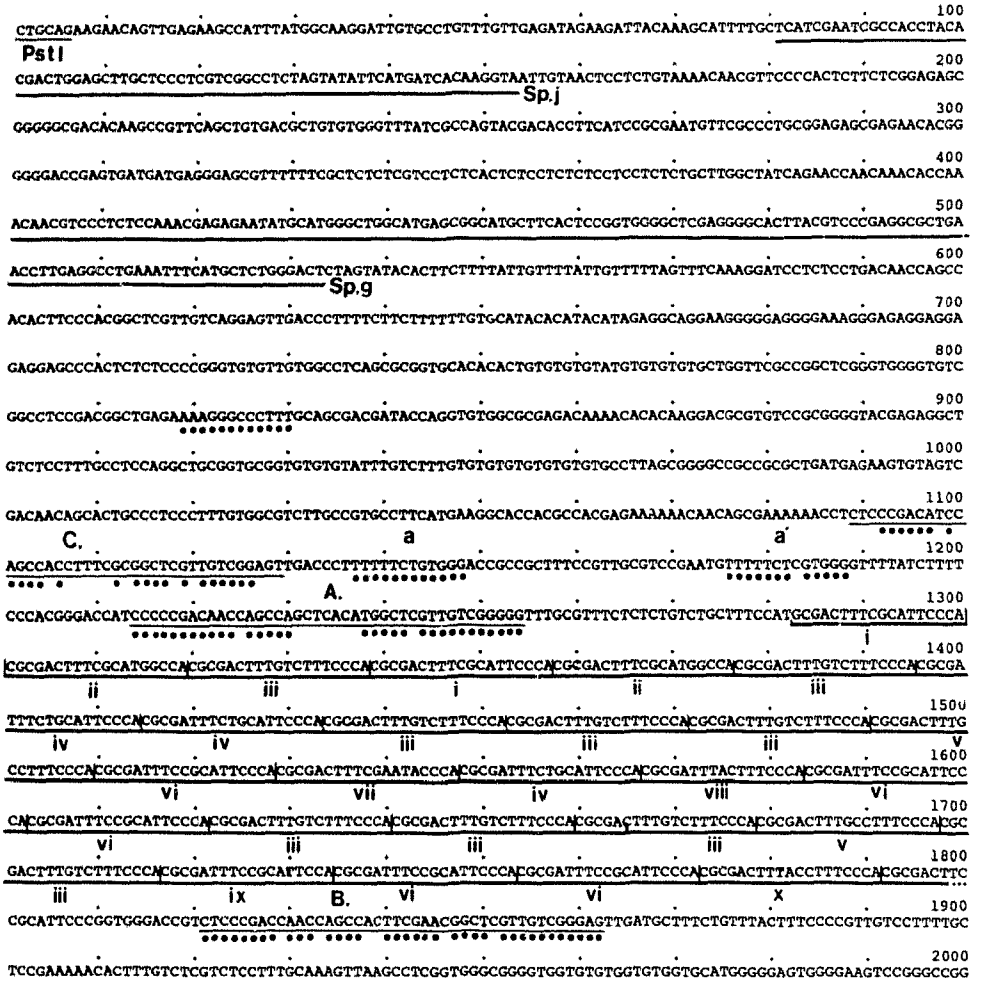



**Figure 36a.** Primary sequence of the 5'-half of the rDNA IGS. The sequence is presented as the coding strand and begins at the left hand PstI site of Cf4 and terminates at nt 2000 of Cf4. The PstI site is indicated and the coding regions for Sp. j and g are underlined. The three terminal As of Sp. g are not encoded by the DNA. The sequence starting at nt 819 and ending at 830 (delineated by the dots) is a short, perfect palindrome. **Blocs A, B,** and **C** are underlined by a thin line. These are three related sequences (see text for details) which are capable of forming hairpin-like structures in the DNA (or in the RNA if they are transcribed). The bases involved in base-pairing are marked with dots. Sequences **a** and **a'** are separated from one another by 26 nt. **a'** differs from **a** by a single nt insertion (a C). The sequence between nt 1282 and 1799 is composed of short (19 nt) tandemly repeated sequences (27) which can be assigned to one of 10 related sequences (**Dr i-x**). See text for a detailed analysis of **Dr** repeats.

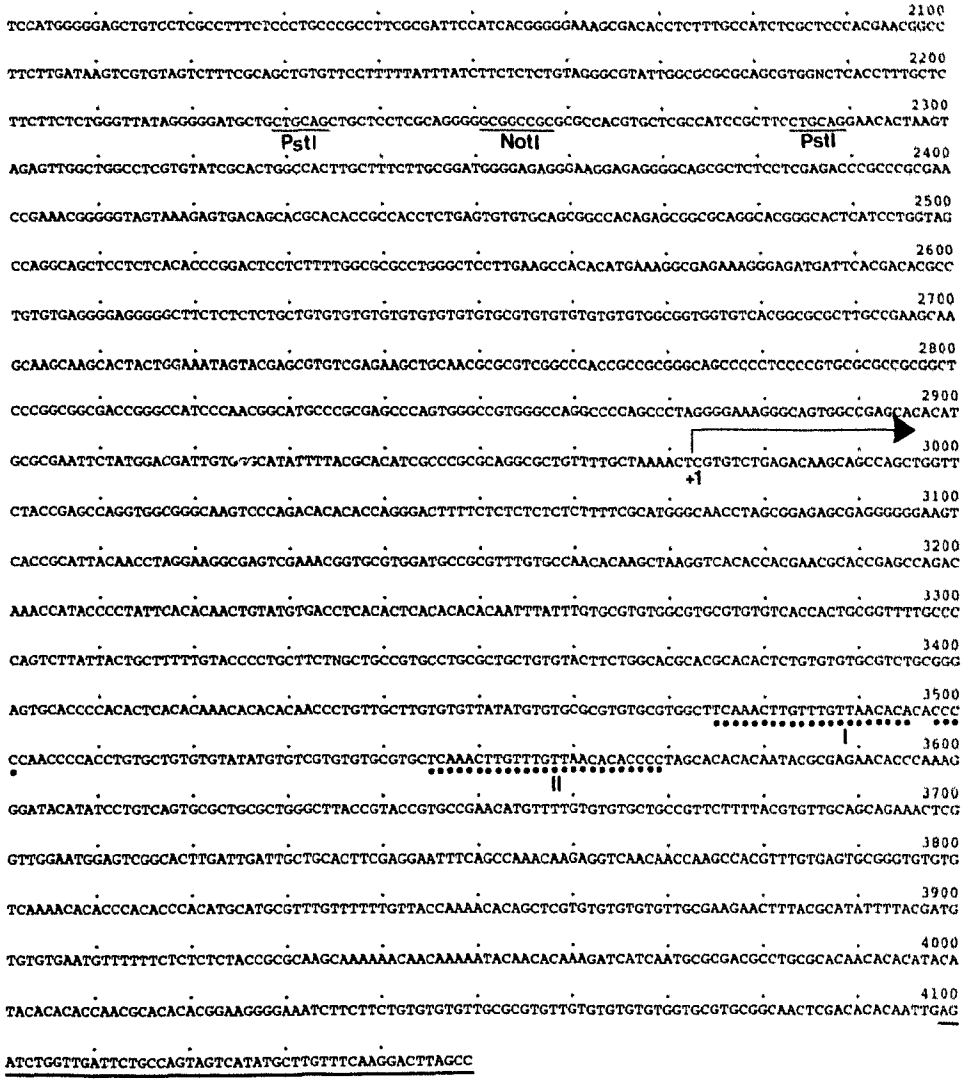



**Figure 36b.** Primary sequence of the 3'-half of the rDNA IGS. This sequence begins at nt 2001 and ends at nt 51 of the coding region for Sp. b. The sequence between the PstI site at nt 2229 and 2283 (Cf5) was determined from CfH1 and only from one strand. The sequence from 2284 to 4149 was determined from Cf1. The +1 and arrow at nt 2973 indicates the site of transcription initiation as determined by Grondal et al. (195). **I** and **II** are a pair of related sequences which differ by two nt only (a CA insertion at nt 3497 in **I**). The underlined sequence starting at nucleotide 4099 is the start of the coding region for Sp. b (SSU rRNA).

| DNA | A | C | G | T | A + T | C + G | AA | AC | AG | AT | CA | CC | CG | CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rDNA | 23 | 25 | 27 | 25 | 48 | 52 | 6.9 | 6.1 | 5.5 | 4.7 | 6.7 | 5.8 | 6.3 | 6.0 |
| SPACER | 19 | 29 | 26 | 25 | 44 | 56 | 4.2 | 7.0 | 4.4 | 3.3 | 7.5 | 7.4 | 7.4 | 7.0 |
| ITS1 | 25 | 21 | 26 | 28 | 53 | 47 | 9.6 | 5.8 | 3.3 | 6.6 | 6.0 | 3.3 | 5.8 | 5.8 |
| ITS2 | 24 | 21 | 24 | 31 | 55 | 45 | 7.2 | 6.3 | 2.7 | 7.5 | 6.5 | 2.9 | 6.5 | 5.1 |
| ITS3 | 32 | 26 | 19 | 23 | 56 | 44 | 11 | 11 | 7.3 | 3.3 | 8.9 | 2.4 | 5.7 | 8.9 |
| ITS4 | 65 | 16 | 7 | 13 | 77 | 23 | 43 | 17 | 0 | 6.7 | 17 | 0 | 0 | 0 |
| ITS5 | 32 | 13 | 17 | 38 | 71 | 29 | 17 | 3.9 | 2.6 | 9.1 | 2.6 | 0 | 3.9 | 6.5 |
| ITS6 | 25 | 18 | 28 | 28 | 53 | 47 | 6.9 | 4.2 | 9.6 | 4.6 | 6.9 | 2.7 | 1.5 | 6.9 |
| ITS7 | 21 | 31 | 24 | 24 | 45 | 55 | 5.3 | 7.3 | 4.9 | 2.8 | 6.1 | 6.9 | 7.3 | 11 |
| Sp.b | 25 | 23 | 27 | 25 | 50 | 50 | 7.4 | 5.5 | 6.1 | 5.8 | 7.0 | 5.4 | 5.4 | 5.0 |
| Sp.c | 28 | 21 | 29 | 22 | 50 | 50 | 9.9 | 5.3 | 7.4 | 5.3 | 5.7 | 5.0 | 5.4 | 4.7 |
| Sp.d | 24 | 24 | 28 | 24 | 48 | 52 | 7.1 | 5.9 | 6.4 | 4.4 | 5.6 | 5.3 | 6.9 | 6.1 |
| Sp.e | 22 | 25 | 24 | 29 | 51 | 49 | 9.0 | 4.7 | 3.8 | 4.7 | 6.6 | 6.2 | 4.3 | 8.1 |
| Sp.f | 22 | 27 | 31 | 20 | 42 | 58 | 6.0 | 3.3 | 8.8 | 4.4 | 5.5 | 9.9 | 6.0 | 4.9 |
| Sp.g | 21 | 28 | 30 | 21 | 42 | 58 | 5.3 | 6.1 | 4.5 | 5.3 | 6.1 | 6.1 | 6.1 | 9.8 |
| Sp.i | 27 | 22 | 24 | 27 | 54 | 46 | 8.8 | 4.1 | 5.3 | 8.8 | 6.5 | 3.5 | 7.1 | 4.7 |
| Sp.j | 23 | 31 | 19 | 26 | 49 | 51 | 2.8 | 6.9 | 4.2 | 8.3 | 8.3 | 6.9 | 6.9 | 9.7 |
| All Coding | 25 | 23 | 28 | 24 | 49 | 51 | 8.0 | 5.4 | 6.4 | 5.3 | 6.2 | 5.4 | 5.8 | 5.4 |
| All Spacers | 21 | 27 | 26 | 26 | 47 | 53 | 5.5 | 6.9 | 4.4 | 4.0 | 7.3 | 6.3 | 6.8 | 6.9 |

| DNA | GA | GC | GG | GT | TA | TC | TG | TT | LENGTH bp |
|---|---|---|---|---|---|---|---|---|---|
| rDNA | 6.0 | 7.0 | 6.9 | 7.0 | 3.6 | 5.9 | 8.1 | 7.5 | **11373** |
| Spacer | 5.0 | 8.3 | 6.3 | 6.8 | 2.3 | 6.5 | 8.2 | 8.3 | **4149** |
| ITS1 | 3.6 | 6.3 | 6.3 | 9.9 | 6.0 | 5.2 | 11 | 6.0 | **366** |
| ITS2 | 2.2 | 8.0 | 3.9 | 10 | 7.7 | 3.9 | 11 | 8.7 | **416** |
| ITS3 | 7.3 | 3.3 | 4.9 | 3.3 | 4.9 | 8.9 | 0.8 | 8.1 | **124** |
| ITS4 | 3.3 | 0 | 3.3 | 0 | 3.3 | 0 | 3.3 | 3.3 | **31** |
| ITS5 | 1.3 | 2.6 | 1.3 | 12 | 10 | 6.5 | 9.1 | 12 | **78** |
| ITS6 | 8.4 | 5.4 | 7.7 | 6.9 | 2.7 | 6.1 | 9.6 | 10 | **262** |
| ITS7 | 6.9 | 5.7 | 6.5 | 5.3 | 2.0 | 11 | 5.7 | 5.7 | **247** |
| Sp.b | 6.8 | 5.9 | 7.7 | 6.5 | 3.6 | 6.0 | 7.7 | 8.2 | **2206** |
| Sp.c | 8.0 | 6.2 | 7.7 | 6.9 | 4.4 | 4.3 | 8.3 | 5.6 | **1783** |
| Sp.d | 6.6 | 6.8 | 7.8 | 7.1 | 4.5 | 5.9 | 7.2 | 6.4 | **1523** |
| Sp.e | 3.8 | 7.1 | 7.6 | 5.7 | 2.8 | 7.1 | 8.5 | 10 | **212** |
| Sp.f | 8.2 | 7.1 | 8.8 | 7.1 | 2.7 | 6.6 | 7.1 | 3.3 | **183** |
| Sp.g | 7.6 | 9.1 | 11 | 2.3 | 1.5 | 6.8 | 8.3 | 3.8 | **133** |
| Sp.i | 7.6 | 6.5 | 3.5 | 6.5 | 3.5 | 8.2 | 8.2 | 7.1 | **171** |
| Sp.j | 5.6 | 5.6 | 4.2 | 4.2 | 6.9 | 12 | 4.2 | 2.8 | **73** |
| All Coding | 7.1 | 6.4 | 7.7 | 6.6 | 4.0 | 5.7 | 7.7 | 6.8 | **6284** |
| All Spacer | 4.9 | 7.7 | 6.1 | 7.1 | 3.1 | 6.4 | 8.4 | 8.2 | **5673** |

**TABLE 2** **NUCLEOTIDE FREQUENCIES**

**Legend:** All numbers are percentages, except the length which is expressed as base pairs. rDNA is the entire 11,373 nt of the rDNA cistron and spacer is the sequence depicted in Fig. 36a and b, (i.e., the IGS). ITS is the internal transcribed spacer. All Coding is all of the coding regions considered as a single sequence and All Spacer is all the spacer sequences considered as one sequence (i.e. ITSs and IGS).

Figure 35 depicts the restriction map of the inserts of pCf1, pCf2, pCf3, and pCf4. These have been aligned to show the organization of an intact rDNA unit. The restriction maps of CfH1, CfH2, and CfH3 are depicted in Fig. 23. The sequence responsible for the longer homologue of Cf4 in CfH1 is also indicated. The sequencing strategy of a composite spacer region is shown in Figure 34.

Figure 45a-g presents the entire rDNA sequence beginning at the left-hand PstI site of the pCf1 insert running past the NotI site in the small Cf5 PstI rDNA sequence (identified while sequencing CfH1) to the same PstI site from which it started.

Figure 36a and 36b shows the DNA sequence of the entire IGS region of C. fasciculata rDNA.

## 10.) ANALYSIS OF THE rDNA PRIMARY SEQUENCE

The rDNA repeat assembled from the analyzed cloned fragments (Fig. 45a-g) is 11,373 bp in length and has a G+C content of 51.7%. Figure 36a and 36b show the primary sequence of the rDNA spacer, which is 4,153 bp long and 55.6% G+C.

Table 2 contains a tabulation of the complete nucleotide base analysis. Single nucleotide frequency analysis reveals that the rDNA in toto contains 23.2% A compared to 19.0% A in the spacer rDNA. The G and T contents are almost identical, being (respectively) 26.8% and 25.1% in the rDNA in toto and 26.3% and 25.3% in the spacer rDNA. The reduction in A content in the spacer rDNA is compensated by an increase in the C content from 24.9% to 29.3%. This loss of A seems to be most marked in the A-A, A-T, and T-A dinucleotides, while the additional C is found most frequently in C-C and G-C dinucleotides.

Further visual and computer analysis reveals stretches of DNA which exhibit a marked preponderance of the dinucleotides AG, CT, TG or TT. These sequences comprise at least 80% of these particular regions. Some of the most impressive runs of dinucleotide-rich sequences include:

(1) Positions 281 to 324, 663 to 706, 1061 to 1085, 2347 to 2384, 2403 to 2428, 2562 to 2587, 3083 to 3113, and 3934 to 3967, all of which are AG-rich.

(2) Positions 327 to 375, 707 to 720, 902 to 915, 1129 to 1145, 1150 to 1166, 1174 to 1185, 1190 to 1205, 1210 to 1218, 1255 to 1299, 1325 to 1356, 1382 to 1398, 1400 to 1413, 1420 to 1432, 1439 to 1527, 1553 to 1565,1572 to 1601, 1608 to 1620, 1627 to 1809, 1867 to 1903, 1912 to 1929, 2013 to 2044, 2068 to 2105, 2130 to 2160, 2509 to 2545, 3049 to 3069, 3297 to 3338, and 3925 to 3933 ( TG-rich).

(3) Positions 545 to 573, 635 to 650, 1136 to 1145, 2139 to 2149, and 3835 to 3847 (TT- rich).

(4) Positions 229 to 246, 721 to 733, 754 to 802, 917 to 963, 1943 to 1987, 2111 to 2120, 2631 to 2678, 3262 to 3285, 3387 to 3408, 3440 to 3479, 3516 to 3549, and 4041 to 4084 ( GT-rich).

The significance of these stretches, if any, remains to be determined.

## 11.) SUBSTRUCTURE OF THE rDNA 'SPACER'

The 'spacer' DNA is defined for the purposes of the present discussion as that DNA which comprises the rDNA in Cf4, Cf5 and the first 1876 bp of Cf1. The sequence is illustrated as these PstI fragments would appear in the

native rDNA (Fig. 36a and 36b). Position one is the first nucleotide of the PstI recognition site at the 5'-end of Cf4 while position 4149 corresponds to bp 51 of the SSU rRNA gene. The spacer DNA also contains the coding regions for Sp. j and g.

Positions 819-830 constitute a perfect 12-nt palindromic sequence, (AAAGGGCCCTTT), which could potentially form a tight hairpin in the DNA, or in an RNA transcript.

Position 1137 is the start of a 12 bp sequence

TTTTTCTGTGGG (a)

which is directly and imperfectly repeated 28 bp directly downstream

TTTTTCTCGTGGG (a').

Position 1214 marks the start of a stretch of DNA that could potentially form a perfectly matched 17 bp hairpin with a 7-nt loop structure (Bloc A, Fig. 38). Interestingly, these 41 bp are imperfectly repeated starting at bp 1821 (B) and bp 1089 (C) (Fig. 36a). Compared to Bloc A, Bloc B is 42 bp in length due to an insertion of a C residue between the eighth and ninth bp of the stem. Its overall identity with Bloc A is 80%, (91% if the 7-nt loop is excluded). Bloc C is 41 bp long and has an overall identity with Bloc A of 78% (or 85% if the loop nucleotides in Bloc A are not considered). Bloc C has a large number of hairpin mismatches, which make the whole structure less impressive than in the case of A and B (Figure 38). Blocs B and C show 78% sequence identity, increasing to 90% if the loop nucleotides and the additional C residue in Bloc B are excluded (see Figures 37 and 38).

These palindromic sequences flank a portion of the rDNA that is made up of short, tandem, direct repeats. This region can be nicely demonstrated using a matrix comparison computer program which is designed to identify such sequences. Using a minimum match parameter of 15 or 20 bp and

DIRECT REPEATS A, B, AND C

a)

1089 1129
C. CTCCCGACATCCAGCCACCTTTCGCGGCTCGTTGTCGGAGT

1214 1254
A. CCCCCGACAACCAGCCAGCTCACATGGCTCGTTGTCGGGGG

1821 1862
B. CTCCCGACCAACCAGCCACTTCGAACGGCTCGTTGTCGGGAG

b)

A. CCCCCGACAACCAGCCAGCTCACATGGCTCGTTGTCGGGGG
B. CTCCCGAC AACCAGCCACTTCGAACGGCTCGTTGTCGGGAG
C

80% Overall Identity 91% Loopless Identity

A. CCCCCGACAACCAGCCAGCTCACATGGCTCGTTGTCGGGGG
C. CTCCCGACATCCAGCCACCTTTCGCGGCTCGTTGTCGGAGT

78% Overall Identity 85% Loopless Identity

C. CTCCCGACATCCAGCCACCTTTCGCGGCTCGTTGTCGGAGT
B. CTCCCGAC AACCAGCCACTTCGAACGGCTCGTTGTCGGGAG
C

78% Overall Identity 90% Loopless Identity

**Figure 37.** Analysis of blocs A, B, and C. **a.)** Bold numbers refer to the nt position in **Fig. 36a**. The underlined nt encompass the loop. **b)** Pairwise comparisons of the three blocs.

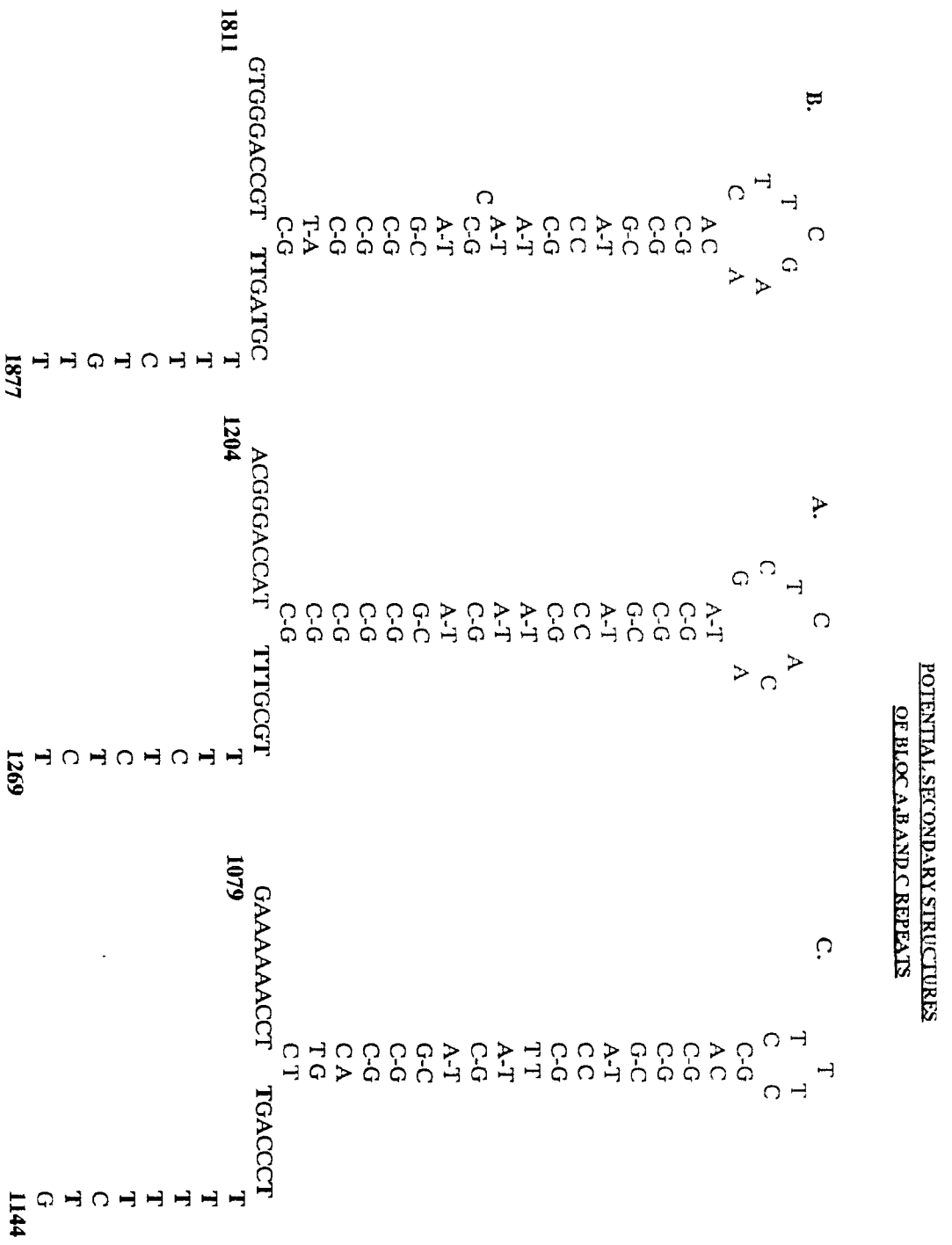



**Figure 38.** Potential secondary structures of blocs A, B, and C. Base-pairing is indicated by - . The bold Ts highlight U-rich sequence in any RNA transcribed from this region. Bold numbers indicate the start and finish position of the blocs in Fig. 36a.

minimum identity of 90%, it was possible to generate the matrices shown in Figures 39 and 40, respectively. Figure 40 describes the entire rDNA spacer sequence while Figure 39 focuses on the region between bp 1000 and 2000. In Figure 40, the "GT" denotes a region in the matrix that corresponds to sequences rich in simple repeats of the dinucleotide GT.

Figure 39 displays the matrix of the region in the spacer between bp 1000 and 2000 which contains tandem direct repeats. The gaps in the matrix indicate that these are not all perfect repeats of each other. When the individual sequences are analyzed, a number of points can be made:

(1) There is a total of 27 direct repeats and one incomplete one. They have been numbered i (i*) through x. There is one each of vii, viii, ix, and x; two each of i, ii, and v; three of iv; five of vi; and nine of iii. Repeats i(i*), ii and iii are themselves repeated tandemly in that order, to make up the first six direct repeats. The remaining repeats following these can be seen in Figure 36a. Repeat i* is identical to i but is missing the initial C. All the repeats (except i*) share the first five and the last three nt in common. The differences among them are the result of base insertions, deletions, transitions and/or transversions.

(2) Pairwise comparisons of all 10 different Dr was carried out. Direct repeats of the same length were simply paired by matching the first base in each one and counting the mismatches. A percentage was calculated for each pairwise comparison. The two shorter Dr (Dr viii (by 2) and Dr ix (by 1)) were expanded to 19 bp by introducing gaps in their sequences. These gaps were introduced to allow maximum percent identity.

(3) Direct repeat i (Dr i) exhibits an 89% identity with Dr ii, iv, v, vi and viii. It is least like Dr viii (74%). The remaining direct repeats (Dr iii,ix, and xi) exhibit 84% identity. Dr i has the highest aggregate percentage at 85.6 and is found twice in the array.

# Figure 39

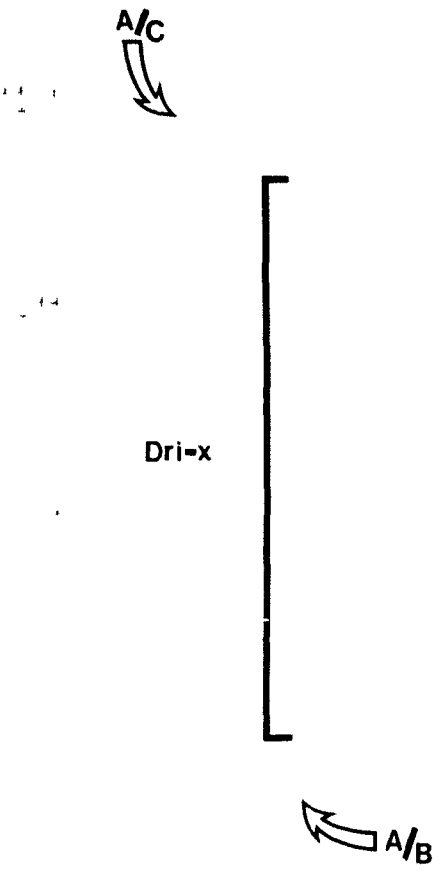



**Figure 39.** Dot matrix analysis of the IGS between nt 1000 and 2000. The analysis was performed using the Matrix Homology Comparison function of the Microgenie$^{R}$ Sequence Analysis Program, with parameters of minmatch of 15 and minpercent 90. The program was run on a Best Elite 386-33/0 computer. That portion of the plot bracketed and labelled Dri-x is the region of the IGS which possesses the direct tandem repeats, Dri-x. A/C refers to the sequence blocs A and C. A/B refers to the sequence blocs A and B. The numbers along the axis of the plot refer to the nt positions in Fig. 36a.

Figure 40

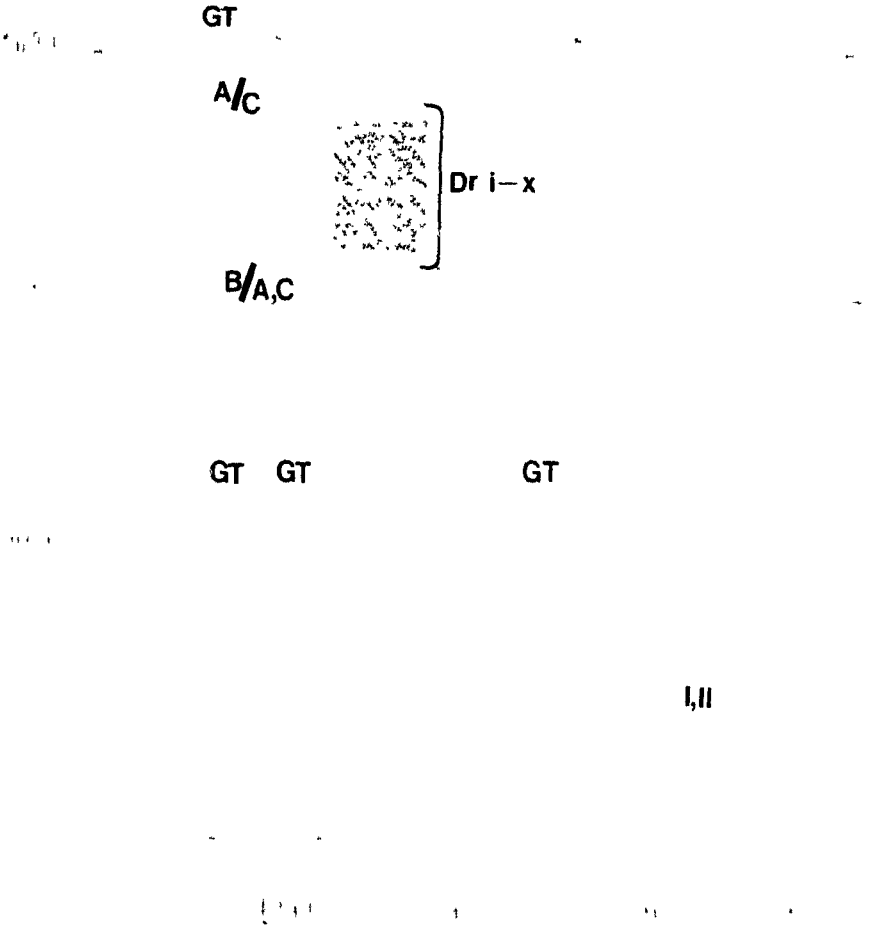



**Figure 40.** Dot matrix analysis of the IGS. The analysis was performed using the Matrix Homology Comparison function of the Microgenie$^{R}$ Sequence Analysis Program with parameters of minmatch 20 and minpercent 90. The program was run on a Best Elite 386-33/0 computer. That portion of the plot bracketed and labelled Dr i-x refers to the sequence of the IGS which possesses the tandem direct repeats (Dri-x). A, B, and C refer to the bloc A, B, and C sequences. GT refers to regions of the sequence which possess GT redundant sequence. I and II refer to the repeated sequences I and II located at nt 3475 and 3545, respectively. The numbers along the axis of the plot refer to the nt positions as in Fig. 36a and b.

(4) Dr ii is most like Dr i (84%) and least like viii (68%). Its aggregate percentage is 77.7, which places it seventh overall. It is found twice.

(5) Dr iii is the most common repeat in the array; it is found nine times. It is most like Dr v (95%) and is least like Dr ix (63%). It exhibits an aggregate percentage of 79, which places it sixth overall. Dr i,ii, and iii are themselves tandemly repeated to make up the first six repeats of the array.

(6) Dr iv is most like Dr vi (95%). It also exhibits a high homology with Dr i and ix (89%). It is least like Dr ii, iii, v, vii, and x, with a 74% identity. It has an aggregate percentage of 80.8, placing it fourth overall. This repeat is found three times in the array.

(7) Dr v is found twice in the array and is most like Dr iii and x (95%). It is least like Dr ix (68%). Its aggregate percentage (81.3) places it third.

(8) Dr vi exhibits marked identity with Dr iv and ix (95%). It is least like Dr iii, v and x (74%). Its aggregate percentage (82) places it second in the ranking. This repeat is found five times.

(9) Dr vii is found once and is least like all the other direct repeats (76.6%) in spite of being full length. It is most like Dr i (89%) and least like Dr viii (68%).

(10) Dr viii exhibits poor identity with the other repeats as demonstrated by its aggregate percentage of 77.6 which places it eighth overall. It is most like Dr x (89%) and least like Dr ii (68%). It is found once.

(11) Dr ix is found only once as well. It is least like Dr iii (63%) and most like Dr vi (95%). Its aggregate percentage (77.1) places it ninth in the ranking.

(12) Dr x is most like Dr v (95%) and least like Dr ix (68%) and is found only once. It is ranked fifth.

(13) When the averages of the pairwise percent identities are compared, it can be said that the repeat most like the other nine repeats is Dr i, exhibiting a 85.6% overall homology. In descending order of overall repeat similarity are: Dr vi

(82%), Dr v (81.3%), Dr iv (80.8%), Dr x (80.1%), Dr iii (79%), Dr ii (77.7%), Dr viii (77.6%), Dr ix (77.1%) and Dr vii (76.6%).

(14) The position of one direct repeat relative to another seems not to correlate with the degree of relatedness of one to the other. Although Dr i is most like Dr ii, Dr ii or Dr i is not most like Dr iii. Indeed Dr iii is most like Dr v. Dr v is most like Dr x even though it is four direct repeats from it. A consensus sequence can be derived from a comparison of the twelve different direct repeats (Figure 42). There are no sequences homologous to these repeats elsewhere in the spacer DNA or within rRNA coding or ITS sequences, including the site of pre-rDNA transcription initiation.

Once outside the repeated DNA region and past the palindromic repeat sequence B, the rDNA becomes unique sequence again. This sequence has islands of purine-, pyrimidine- and dinucleotide-rich elements, but for the most part it seems to approximate random sequence.

Position 2229 is the start of Cf5 and position 2284 is its 3'-end. Cf5 contains within it a NotI (2251) site, which is unique in the entire rDNA repeat.

The direct-repeat dot-matrix analysis identified two sequences in the vicinity of bp 3500 that show greater than 90% identity with one another. These sequences have been labelled I and II; they are unrelated to Blocs A, B, C or a/a' or the i-xii direct repeats. These motifs are 27 and 25 nt in length, respectively. Repeat I has an additional CA at position 3492 compared to II; otherwise the two repeats are identical, and both contain a HpaI site (Figure 43).

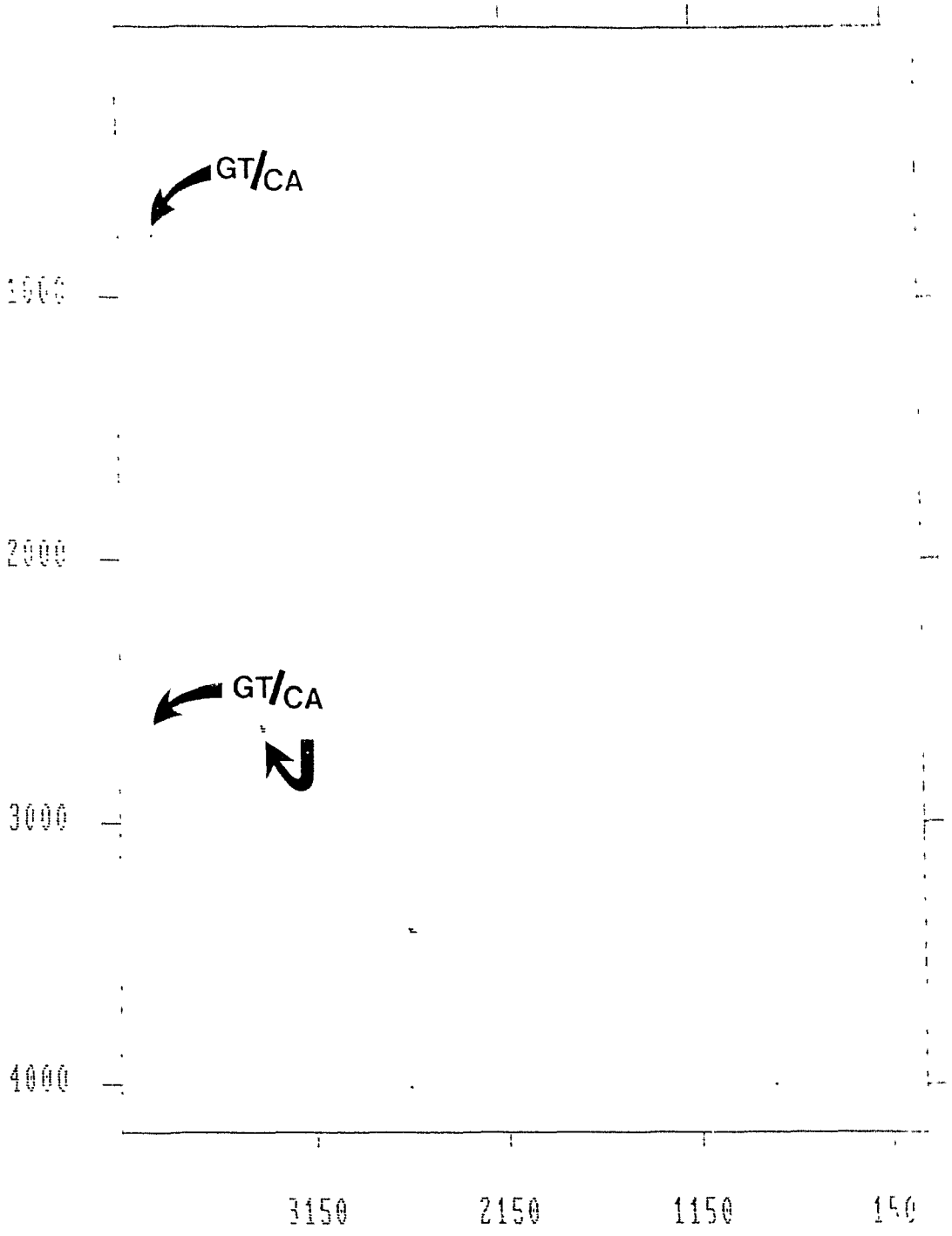



**Figure 41.** Dot matrix analysis of the IGS. The analysis was performed using the Matrix Comparison for Inverted Repeats function of the Microgenie$^{R}$ Sequence Analysis Program with parameters of minmatch 20 and minpercent 90. The program was run on a Best Elite 386-33/0 computer. GT/CA indicate sites in the IGS which plot out because they consist of the simple redundant sequences GT and CA.

**TABLE 3**

| Dr | i | ii | iii | iv | v | vi | vii | viii | ix | x | Aggregrate % | Ranking |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i | 89 | 84 | 89 | 89 | 89 | 89 | 74 | 84 | 84 | | 85.6 | 1 |
| ii | | 74 | 74 | 79 | 79 | 84 | 68 | 79 | 74 | | 77.7 | 7 |
| iii | | | 74 | 95 | 74 | 74 | 84 | 63 | 89 | | 79.0 | 6 |
| iv | | | | 74 | 95 | 74 | 79 | 89 | 74 | | 80.8 | 4 |
| v | | | | | 74 | 74 | 84 | 68 | 95 | | 81.3 | 3 |
| vi | | | | | | 79 | 79 | 95 | 74 | | 82.0 | 2 |
| vii | | | | | | | 68 | 74 | 74 | | 76.6 | 10 |
| viii | | | | | | | | 74 | 89 | | 77.6 | 8 |
| ix | | | | | | | | | 68 | | 77.1 | 9 |
| x | | | | | | | | | | | 80.1 | 5 |

**IGS DIRECT TANDEM REPEATS (Dr)**

| | | No. |
|---|---|---|
| Dri | CGCGACTTTCGCATTCCCA | 2 |
| Drii | CGCGACTTTCGCATGGCCA | 2 |
| Driii | CGCGACTTTCGCATGGCCA | 9 |
| Driv | CGCGATTTCTGCATTCCCA | 3 |
| Drv | CGCGACTTTGCCTTTCCCA | 2 |
| Drvi | CGCGATTTCCGCATTCCCA | 5 |
| Drvii | CGCGACTTTCGAATACCCA | 1 |
| Drviii | CGCGA TTTAC TTTCCCA | 1 |
| Drix | CGCGATTTCCGCATT CCA | 1 |
| Drx | CGCGACTTTACCTTTCCCA | 1 |
| | | 27 |
| Consensus | CGCGA$^{C}/_{T}$TT$^{T}/_{C}{}^{C}/_{Pu}{}^{G}/_{Py}$C$^{T}/_{A}$T$^{T}/_{Pu}$C/GCCA | |

**Figure 42.** Comparison of the IGS direct tandem repeats (Dri-x). Dri* is the first repeat in the array and differs from Dri by having a T as the first nt instead of a C. The consensus was constructed to highlight invariant nts or positions which have a predisposition for two bases only. Pu means any purine and Py means any pyrimidine.

**DIRECT REPEATS I/II AND a/a'**

a)

**I**

**3475** **3501**
TCAAACTTGTTTGTTAACACACACCCC

**II**

**3545** **3569**
TCAAACTTGTTTGTTAACACA CCCC

TCAAACTTGTTTGTTAACACACACCCC
TCAAACTTGTTTGTTAACACA* *CCCC

b)

**a**

**1137** **1148**
TTTTTCTGTGGG

**a'**

**1176** **1188**
TTTTTCTCGTGGG

TTTTTCT *GTGGG
TTTTTCTCGTGGG

**Figure 43.** Direct repeats I/II and a/a'. Bold numbers refer to nt positions as in **Fig 36a** and b. * indicates gaps inserted to allow a maximum alignment.

## 12.) THE SITE OF TRANSCRIPTION INITIATION IN THE C. FASCICULATA rDNA REPEAT.

Grondal et al. (195) have determined the putative site of transcription initiation in Crithidia fasciculata rDNA using $S_1$ nuclease protection and nuclear run-on experiments. Comparison of their results with those presented here establishes that the position of transcription initiation corresponds to position 2973 in Fig. 36b.

Figure 44 compares the sequences around the transcription initiation site, as determined by Grondal et al. and by me. Differences between the two sequences are apparent and are indicated in the figure. These differences are located in the upstream IGS, as well as in the region corresponding to the transcribed 5'-end of the pre-rRNA. Most of the changes are base substitutions but there is a single-nt insertion (a G at position 2876 in my sequence) and a single-nt deletion at position 3020.

Given the extent of the differences (10/420), the most likely explanation would seem to be that the sequence of Grondal et al. corresponds to a different rDNA repeat than the one characterized here. Indeed, I encountered minor sequence differences in the homologous regions of Cf4-Cf1 and CfH1. In a total of 246 bp compared, I observed 4 transitions, 3 tranversions and two, 3-bp insertions/deletions between the two sequences.

**SITES OF RNAP I TRANSCRIPTION INITIATION**

| | -50 | -40 | -30 | -20 | -16 | -7 | +1 | +10 |
|---|---|---|---|---|---|---|---|---|
| S.pombe | CAACCACAAATGTTTCTATAAAATCGAGGAAAATAGGTCCAAGGAACTATGTAAAAGGAG | | | | | | | |
| K.lactis | CTAATAGAAAGAAGANCATCGGGTGAGAAAAATAGGCCGGGAGGTACTTCATGCGAAAGC | | | | | | | |
| S.cerevisiae | GGTTTAGTCATGGAGTACAAGTGTGAGGAAAAGTAGTTGGGAGGTACTTCATGCGAAAGC | | | | | | | |
| S.rosei | GACAAGGTGACAATGTTTATAGTTCAATATTTTTGGGCAGAAGGAACTTCATGCGAAAGC | | | | | | | |
| H.winaei | TATTCACAATTGTTTAAGGCTAAAACACATTTCAGTGCCAAAGGAACAGTATGCGAAGGC | | | | | | | |
| N.crassa | CTCAGACCGTCGGGCCGCACCGTCCCCAGACTCGAGGTCGGAGGGTATACAAGAAGGAGC | | | | | | | |
| D.discoideum | CTAGGTGT ACCATTAGCCATTGAGCTCGGTGACTACCCCAAATACATATACAAGAAGAG | | | | | | | |
| P.polycephalum | TAATCCCGCGCACCAGCTATGCTTCTTAAAAAGAAACCCAAGATACATATAGGGGGGGTA | | | | | | | |
| T.pyriformis | TTCTGGCAAAAAAAAAAAAAAAAGTATCAGGGGGGTAAAAATGCATATTTAAGAAGGGGA | | | | | | | |
| A.castellanii | GGAGCACTTTTCTGGCACCTAAACTGGTCGGACCGTCCGAAAGTATATATAAAGGGACGG | | | | | | | |
| T.brucei | ATATATCTTATATAGGAAAGATTAAGCAGTAAAAGTAGCGCTTACGGCGTACGGAGCAGG | | | | | | | |
| C.fasciculata | TGGGCATATTTTACGCACATCGCCCGCGCAGGCGCTGTTTTGCTAAAACTCGTGTCTGAG | | | | | | | |
| D.melanogaster | ATGGAATTGAAAATACCCGCTTTGAGGACAGCGGGTTCAAAAACTACTATAGGTAGGCAG | | | | | | | |
| X.laevis | CGCACGACGCCTCCATGCTACGCTTTTTTGGCATGTGCGGGCAGGAAGGTAGGGGAAGAC | | | | | | | |
| Human | GAGGTATATCTTTCGCTCCGAGTCGGCAATTTTGGGCCGCCGGGTTATATGCTGACACGC | | | | | | | |

**Figure 44.** A listing of 60 nt around the site of RNAP I transcription initiation for various organisms. The heavy bold line centered on **-7** is a shared sequence between the fungal organisms. The light underline highlights sequence shared by the fungal species only. The bold nt are ones which have been shown to be important in transcription initiation studies and are more widely conserved.

**Figure 45a-g.** The complete primary sequence of Crithidia rDNA. This is a composite sequence of that determined by the author and by M.N. Schnare and D.F. Spencer (180,194). The sequence is depicted as the coding strand and begins at the PstI site which marks the left hand end of Cf1. The genes coding for the rRNAs are underlined and identified by their designations (i.e., Sp. b, etc) along the right hand margin. The **+1** and dashed arrow refers to the $S_1$ nuclease-determined start of transcription initiation (195). I and II refer to the repeated sequences I and II. All the PstI sites marking the boundaries of Cf1-5 are noted. Numbers refer to the nt position with 1 being the first nucleotide of the initial PstI site and 11373 the last nucleotide of the rDNA repeat. The underlined sequence ending at nt 1000 is that commented on in the legend of Fig. 36a. Blocs A, B, and C are also underlined. The array of tandem direct repeats is also underlined and designated in the right hand border as Dr i-x.

FIGURE 45a. C. fasciculata rDNA

CTGCAGGAACACTAAGTAGAGTTGGCTGGCCTCGTGTATCGCACTGGCCACTTGCTTTCTTGCGGATGGGGAGAGGGAAG
PstI
GAGAGGGGCAGCGCTCTCCTCGAGACCCGCCCGCGAACCGAAACGGGGGTAGTAAAGAGTGACAGCACGCACACCGCCAC

CTCTGAGTGTGTGCAGCGGCCACAGAGCGGCGCAGGCACGGGCACTCATCCTGGTAGCCAGGCAGCTCCTCTCACACCCG

GACTCCTCTTTTGGCGCGCCTGGGCTCCTTGAAGCCACACATGAAAGGCGAGAAAGGGAGATGATTCACGACACGCCTGT
400
GTGAGGGGAGGGGGCTTCTCTCTCTGCTGTGTGTGTGTGTGTGTGTGTGCGTGTGTGTGTGTGGCGGTGGTGTCACGG

CGCGCTTGCCGAAGCAAGCAAGCAAGCACTACTGGAAATAGTACGAGCGTGTCGAGAAGCTGCAACGCGCGTCGGCCCAC

CGCCGCGGGCAGCCCCCTCCCCGTGCGCGCCGCGGCTCCCGGCGGCGACCGGGCCATCCCAACGGCATGCCCGCGAGCCC

AGTGGGCCGTGGGCCAGGCCCCAGCCCTAGGGGAAAGGGCAGTGGCCGAGCACACATGCGCGAATTCTATGGACGATTGT
+1---------->
GGGCATATTTTACGCACATCGCCCGCGCAGGCGCTGTTTTGCTAAAACTCGTGTCTGAGACAAGCAGCCAGCTGGTTCTA
800
CCGAGCCAGGTGGCGGGCAAGTCCCAGACACACACCAGGGACTTTTCTCTCTCTCTCTTTTCGCATGGGCAACCTAGCGG

AGAGCGAGGGGGGAAGTCACCGCATTACAACCTAGGAAGGCGAGTCGAAACGGTGCGTGGATGCCGCGTTTGTGCCAACA

CAAGCTAAGGTCACACCACGAACGCACCGAGCCAGACAAACCATACCCCTATTCACACAACTGTATGTGACCTCACACTC

ACACACACAATTTATTTGTGCGTGTGGCGTGCGTGTGTCACCACTGCGGTTTTGCCCCAGTCTTATTACTGCTTTTTGTA

CCCCTGCTTCTNGCTGCCGTGCCTGCGCTGCTGTGTACTTCTGGCACGCACGCACACTCTGTGTGTGCGTCTGCGGGAGT
1200
GCACCCCACACTCACACAAACACACACAACCCTGTTGCTTGTGTGTTATATGTGTGCGCGTGTGCGTGGCTTCAAACTTG
I
TTTGTTAACACACACCCCCAACCCCACCTGTGCTGTGTGTATATGTGTCGTGTGTGCGTGCTCAAACTTGTTTGTTAACA
II
CACCCCTAGCACACACAATACGCGAGAACACCCAAAGGGATACATATCCTGTCAGTGCGCTGCGCTGGGCTTACCGTACC

GTGCCGAACATGTTTTGTGTGTGCTGCCGTTCTTTTACGTGTTGCAGCAGAAACTCGGTTGGAATGGAGTCGGCACTTGA

TTGATTGCTGCACTTCGAGGAATTTCAGCCAAACAAGAGGTCAACAACCAAGCCACGTTTGTGAGTGCGGGTGTGTGTCA

**Figure 45b**

1600

AAACACACCCACACCCACATGCATGCGTTTGTTTTTTGTTACCAAAACACAGCTCGTGTGTGTGTTGCGAAGAACTTT

ACGCATATTTTACGATGTGTGTGAATGTTTTTTCTCTCTCTACCGCGCAAGCAAAAAACAACAAAAATACAACACAAAGA

TCATCAATGCGCGACGCCTGCGCACAACACACATACATACACACACCAACGCACACACGGAAGGGGAAATCTTCTTCTGT

GTGTGTTGCGCGTGTTGTGTGTGTGTGGTGCGTGCGGCAACTCGACACACAATTGAGATCTGGTTGATTCTGCCAGTAGT

CATATGCTTGTTTCAAGGACTTAGCCATGCATGCCTCAGAATCACTGCATTTGCAGGAATCTGCGCATGGCTCATTACAT

2000

CAGACGTAATCTGCCGCAAAAATCTTGCGGTTTCCGCAAAATTGGATAACTTGGCGAAACGCCAAGCTAATACATGAACC

AACCGGGTGTTCTCCATCCCAGACAGTGGGCAACCATTGTCGTGAGACGCCCAGCGAATGAATGACAGTAAAACCAATGC

CTTCACTGGCAGTAACACCCAGCAGTGTTGACTCAATTCATTCCGTGCGAAAGCCGGCTTGTTCCGGCGTCTTTTGACGA

ACAACTGCCCTATCAGCTGGTGATGGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTC

CGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAAACAAAACG

2400

ATGAGGCAGCGAAAAGAAATAGAGTTGTCAGTCCATCTGGATTGTCATTTCAATGGGGGATATTTAAACCCATCCAATAT

CGAGTAACAATTGGAGGACAAGTCTGGTGCCAGCACCCGCGGTAATTCCAGCTCCAAAAGCGTATATTAATGCTGTTGCT

GTTAAAGGGTTCGTAGTTGAACTGTGGGCTGTGTAGGTTTGTTCCTGGTCGTCCCGTCCATGTCGGATTTGGTGGCCCAG

GCCCTTGCAGCCCGTGAACATTCAAAGAAACAAGAAACACGGGAGTGGTTCCTTTCCTGATTTACGCATGTCATGCATGC

CAGGGGGCGTCCGTGATTTTTTACTGTGACTAAAGAAGTGTGACTAAAGCAGTCATCCGACTTGAATTAGAAAGCATGGG

2800

ATAACAAAGGAGCAGCCTCTAGGCTACCGTTTCGGCTTTTGTTGGTTTTAAAGGTCTATTGGAGATTATGGAGCTGTCG

ACCAGTGCTTTCACATCGTACTTTCGTGCGGTGTGTGGTGCCTTTGGAGGGGTTTAGTGCGTTCGGCTCGGGCTTCGGTC

CGTCCGGGCGTAACGCCTCTTCAACTCACGGCCTCTAGGAATGAAGGAGGGTAGTTCGGGGGAGAACGTACTGGGGCGTC

AGAGGTGAAATTCTTAGACCGCACCAAGACGAACTACAGCGAAGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCA

AAGTGTGGAGATCGAAGATGATTAGAGACCATTGTAGTCCACACTGCAAACGATGACACCCATGAATTGGGGATCTTTTG

Sp.b

**Figure 45c.**

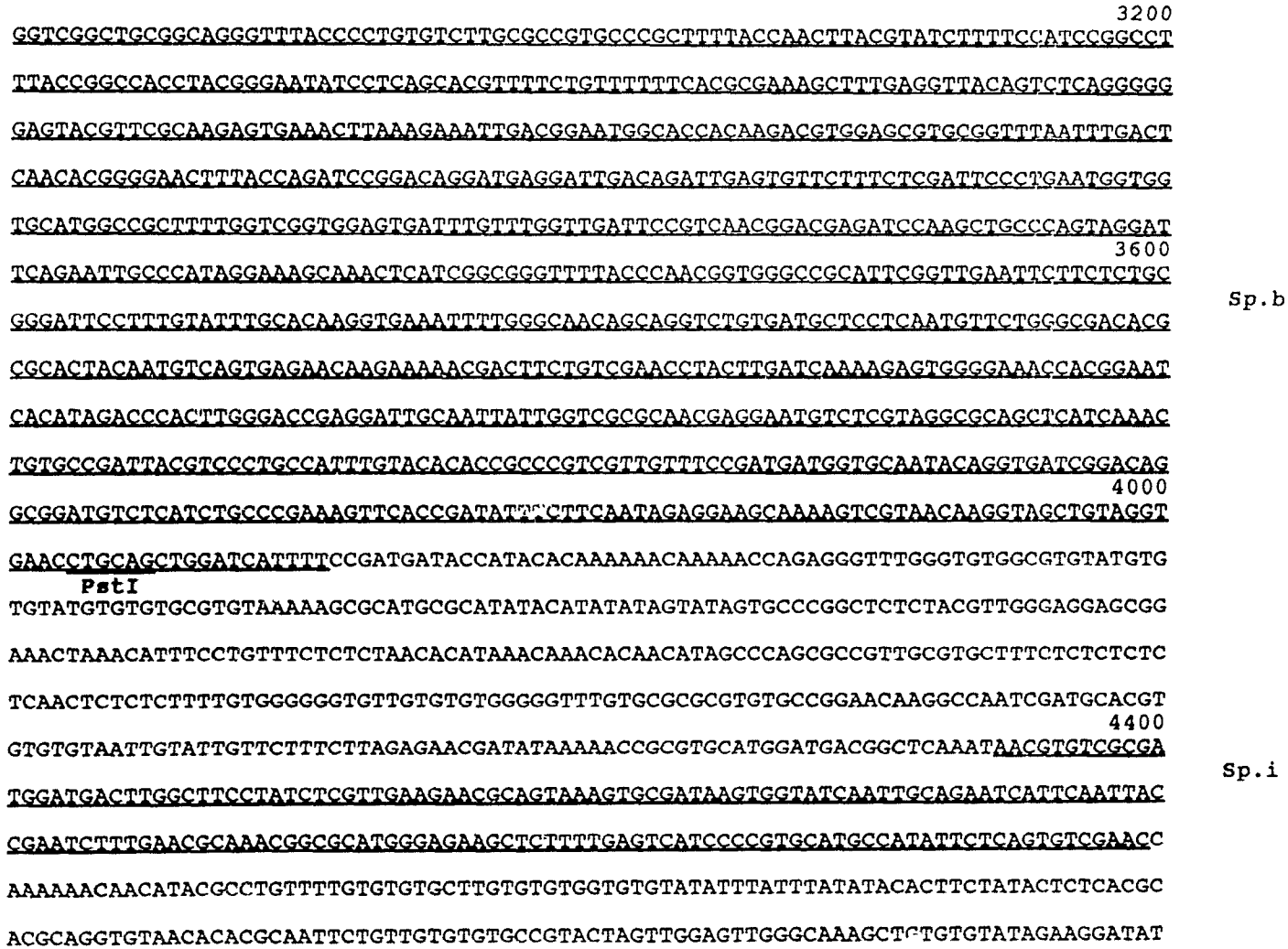

3200
GGTCGGCTGCGGCAGGGTTTACCCCTGTGTCTTGCGCCGTGCCCGCTTTTACCAACTTACGTATCTTTTCCATCCGGCCT
TTACCGGCCACCTACGGGAATATCCTCAGCACGTTTTCTGTTTTTTCACGCGAAAGCTTTGAGGTTACAGTCTCAGGGGG
GAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAATGGCACCACAAGACGTGGAGCGTGCGGTTTAATTTGACT
CAACACGGGGAACTTTACCAGATCCGGACAGGATGAGGATTGACAGATTGAGTGTTCTTTCTCGATTCCCTGAATGGTGG
TGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGACGAGATCCAAGCTGCCCAGTAGGAT
3600
TCAGAATTGCCCATAGGAAAGCAAACTCATCGGCGGGTTTTACCCAACGGTGGGCCGCATTCGGTTGAATTCTTCTCTGC
Sp.b
GGGATTCCTTTGTATTTGCACAAGGTGAAATTTTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCGACACG
CGCACTACAATGTCAGTGAGAACAAGAAAAACGACTTCTGTCGAACCTACTTGATCAAAAGAGTGGGGAAACCACGGAAT
CACATAGACCCACTTGGGACCGAGGATTGCAATTATTGGTCGCGCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAAC
TGTGCCGATTACGTCCCTGCCATTTGTACACACCGCCCGTCGTTGTTTCCGATGATGGTGCAATACAGGTGATCGGACAG
4000
GCGGATGTCTCATCTGCCCGAAAGTTCACCGATATTTCTTCAATAGAGGAAGCAAAAGTCGTAACAAGGTAGCTGTAGGT
GAACCTGCAGCTGGATCATTTTCCGATGATACCATACACAAAAAACAAAAACCAGAGGGTTTGGGTGTGGCGTGTATGTG
PstI
TGTATGTGTGTGCGTGTAAAAAGCGCATGCGCATATACATATATAGTATAGTGCCCGGCTCTCTACGTTGGGAGGAGCGG
AAACTAAACATTTCCTGTTTCTCTCTAACACATAAACAAACACAACATAGCCCAGCGCCGTTGCGTGCTTTCTCTCTCTC
TCAACTCTCTCTTTTGTGGGGGGTGTTGTGTGTGGGGGTTTGTGCGCGCGTGTGCCGGAACAAGGCCAATCGATGCACGT
4400
GTGTGTAATTGTATTGTTCTTTCTTAGAGAACGATATAAAAACCGCGTGCATGGATGACGGCTCAAATAACGTGTCGCGA
Sp.i
TGGATGACTTGGCTTCCTATCTCGTTGAAGAACGCAGTAAAGTGCGATAAGTGGTATCAATTGCAGAATCATTCAATTAC
CGAATCTTTGAACGCAAACGGCGCATGGGAGAAGCTCTTTTGAGTCATCCCCGTGCATGCCATATTCTCAGTGTCGAACC
AAAAAACAACATACGCCTGTTTTGTGTGTGCTTGTGTGTGGTGTGTATATTTATTTATATACACTTCTATACTCTCACGC
ACGCAGGTGTAACACACGCAATTCTGTTGTGTGTGCCGTACTAGTTGGAGTTGGGCAAAGCTCTGTGTATAGAAGGATAT

**Figure 45d.**

4800
GTTCATGTATGCAAATACGTGGCCATATTTGTATATATACTGTGGCGCAGTTCTCTCTGGCGATAGCGAAGCGCGCGTGT

GTATCCGCATTTTTTGTGTGTGGTACATACGCGCGCATTCGGGATAATTAACAAAAAACCAAACGAGAACACAACAAAAT

TGCATGACACCTCGCGGTGTGGTGCAACGCGTTGCGCGTTAGGCACTTTTTGCTTAATCTATCTATGGCGCCTCGTGCGC

CTCCCCCTTTTGTTCACAGACCTGAGTGTGGCAGGACTACCCGCCGAACTTAAGCATATTACTCAGCGGAGGAAAAGAAA

ACAACCGTGATTCTCTCAGTGAGCGGCGAGCGAAGAGGGAAAGAACTCGTTGCCGAATCGGGTCTTACAAGGCCTTGAGT
5200
TGTGGCAATGATATCATTTCGTGGTTGGTGCAGGAATGGCGAAATTCAAGCGCAAAGCAAACCCGTTGCTGAATACAACC

CTTCATGTGAGTATTGAGCCAAAGAAGGTGTTAGCCCATTGAGCCATAAACCTGAGCGCCTCATGAACTGTATTAAGAGA

GTAGCACTGTTTGGGAATGCAGTGTCAAGTTGGCAGGTATTTGTCTGCTAAAGTTAAATACAGAGTAGGAAGACCGATAG

CGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGAAATAGAACCTGAAGTCGTGACAACGACAA

TGGAAGTACCTCCATTTCGTTTTTGCCAGAACCAGCCCTACTGCATTATCTCGTCGGACGTCAAACTCCGGCGGGCAACG
5600
AAGTGCAAGAACAAACATTGGGTGTTGGTGAAATGGGCGGAGTATATGGGCTGTGCACGGCGCAAGCCGCGAGCACGGTT

TCTTCTCTGTGATTCTCCATGCTGGCGCAGAAAATGGGGTGCCCCCACCCGTCTTGAAACACGGACCAAGGAGTCAAACA

GACGCGCAAGGGAGAAGATATGGTTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCAAAATGCATCTGTGGTGTGCAC

ACGTCAAATCCGTTTGATTTGCCCAACACCGACCGGCCCTCCGTGGGTTGCGTTGGAGTGTGCCTGTTTGGACCCGAAAG

GTGGTGAACTATGCCTGAACAAGGTGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCT
6000
CGTATGATTTGGGTATCGGAGCGAAAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGC

TGGTGCTAGTAGAAGTATTGTGCGGTAAAGCAAATGATTAGAGGACTTGGTGTTCTTAGAATATCGACCTATTCTCAAAC

TTTAAATGTGCAAACAAACCGTACCTTAGCCAACTGCAAGGTCGGAAAGAAAAGAATTAGAGGCACCAAGTGGGCCTTCT

TCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATATGTGGCGTCAAAACTTATGGGCTCATTCTCGATACG

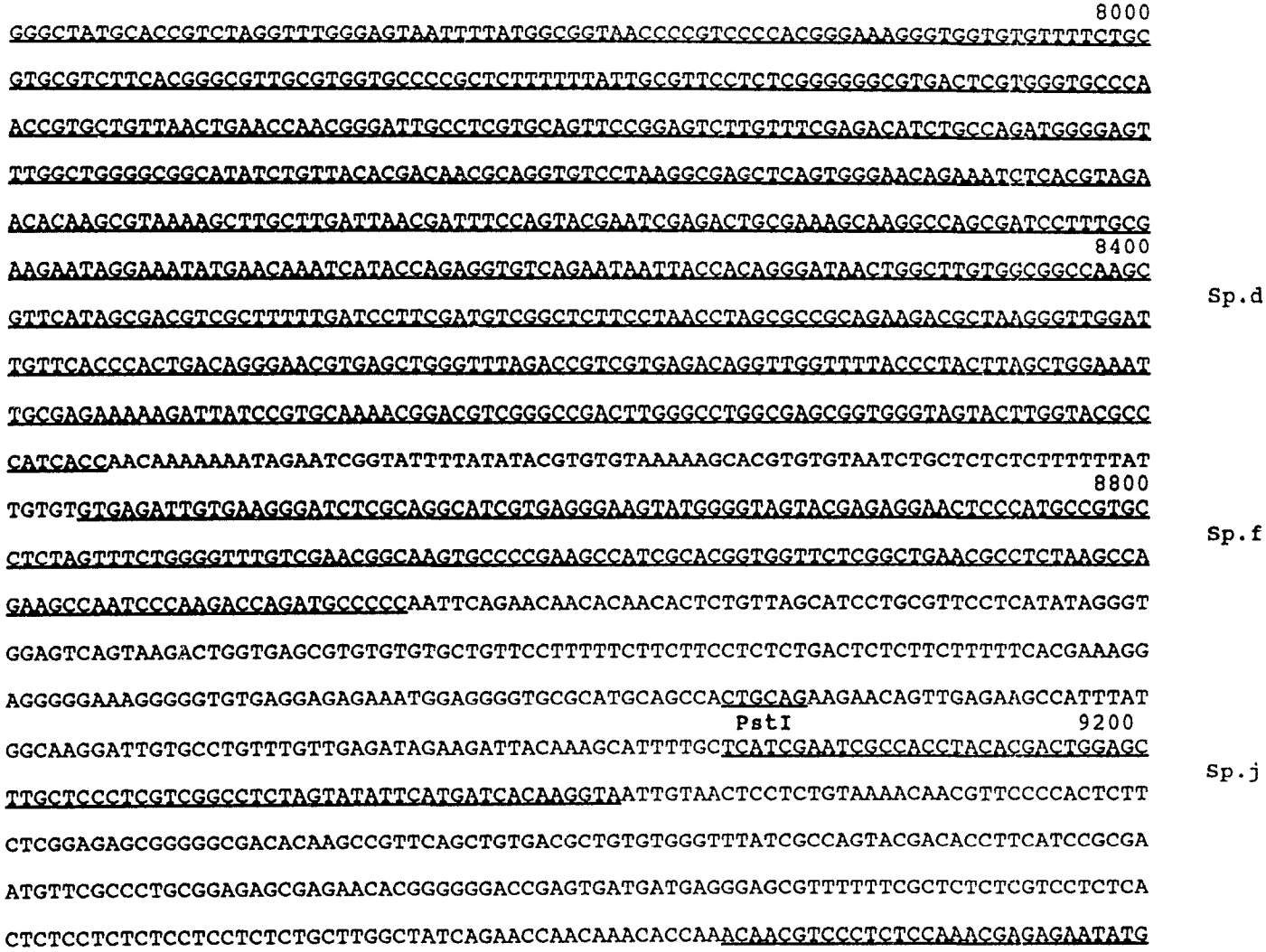
TGAAAAGGGTGTTGGTGGATATGGACAGTTGGACGGTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACTCA

Sp.c

Figure 45e.


6400
CCAACCGAATCCACCGGCCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCCATTATTCCTTTGGTAAGGGCG
GAAAAACTTGGAGAAAGTGTAGCCCTTCCGATGTAGATGTGGCCTGGAGGTCAGGACGAAGCTTATGGCGTGAGCCTAAG
ATGGACCGGCCTCTAGTGCAGATCTTGGTTGGCGTAGCAAAGATCTAACGGAGATATACTCAACATGCAACGTTGGATAC
Sp.c
TGGAGCGGGGAAGGATTTCGTGCCAACGGCACTCGTACACGAGTTGTTCGGAAACTGAGCACAACGTTATATCGTTTTGT
TAGGAAAGTGAAGGTCTGTCGGCGGTGGCTGTCTTCGGGCGGTTATCGACTGGCCATAACTGAAAAGGGGCAACAGAGAA
6800
CCTGGGATTATATTCCAAAAAGAAATTGCATGTGGGAAAAACACACACGAACGTACATTTTCTCTCCTCACACTCTCTCT
TTTCTTGTCCTCGCTCTTTACCGAGAGAGGGCAGGAGGAGGAAGAGCGCACACGGAGTTAAAAAATATACAAACACGTAA
TTAGTGGAAATGCGAAACACTTGCCAGGTGACAAATCAATCCTCCCACGGTGAGCTTTCTTTTCACCATAATCCACATCT
Sp.e
CCGGCTTTGCTGGGCTTGGGCCTTTTTACTTCTCGCGTTGTTCGGAGCGGGGGCCCAAGATTGAAAAATGCAGCTCTCCC
TACGTACTGTCATTGTTGTGAGTTCTGCGCATTAAAGCAAAAACCTGGGGTGTTATTGGAAAACACACAAAAAAACAAAC
7200
AAATCCCAACTGCAGACCGTACTCATCACCGCATCAGGTCCCCAAGCATAGAATGCTCTGGTACATAGAGAAAGGTACAC
PstI
TCAGGGAAGTCGGCAAAATAGATGCGTAAGTTCGCAAGAAGCATTGGCTCTGAGGGCAAAGTCAGAGAAAACCGGAGGGC
CAGTGTCCTCTCACGTGCAGCCGTTCTTCAGATGTTGTTTTGGCGCTCGCCGCCTTAGCACGCTGTTAGCGACTGTATCG
TTGTCTGTGAAGGCTACCTTCCCGAGGCGTTGTAACCCAGCGTCTCACTCTGGCACGGGCCGGTCTTGGCGGGTTCACGC
TCGTCTTGTGCCGCGTCCCACAAATTGCCAACTCAGAACTGCTTACGGCGGGGAATCCAACTGTATAATTAAAACATAGG
7600
TTTGTGATGCATCCAAGTGGTGTATGTCGCAAACTGATTTCTGCCCAGTGCTCTGAATGTCAACGTGACGAGATTCACCG
Sp.d
ACGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCTTCCAATTAGAGACGCGCATG
AATGGATTAATGAGATTCCCTCTGTCCCGAGTTACTATCTAGCGAAACGACAGTCAAGGGAACGGACTTGAAGGGCTAAG
CGGGGAAAGAAGACCCTGTTGAGTTTGACTCCAGTCTGGCTCTGTACGGCGACATCTGAGGTGTAGTATAGGTGGAAGCG
CGAGCACAAATGAAATACCACCACTCGGAACGTTGCTTTACTTATCGAATGAAGAGACCAATGGGTTTGGCGTAGTCTTC

Figure 45f.




Sp.d

Sp.f

Sp.j

**Figure 45g.**

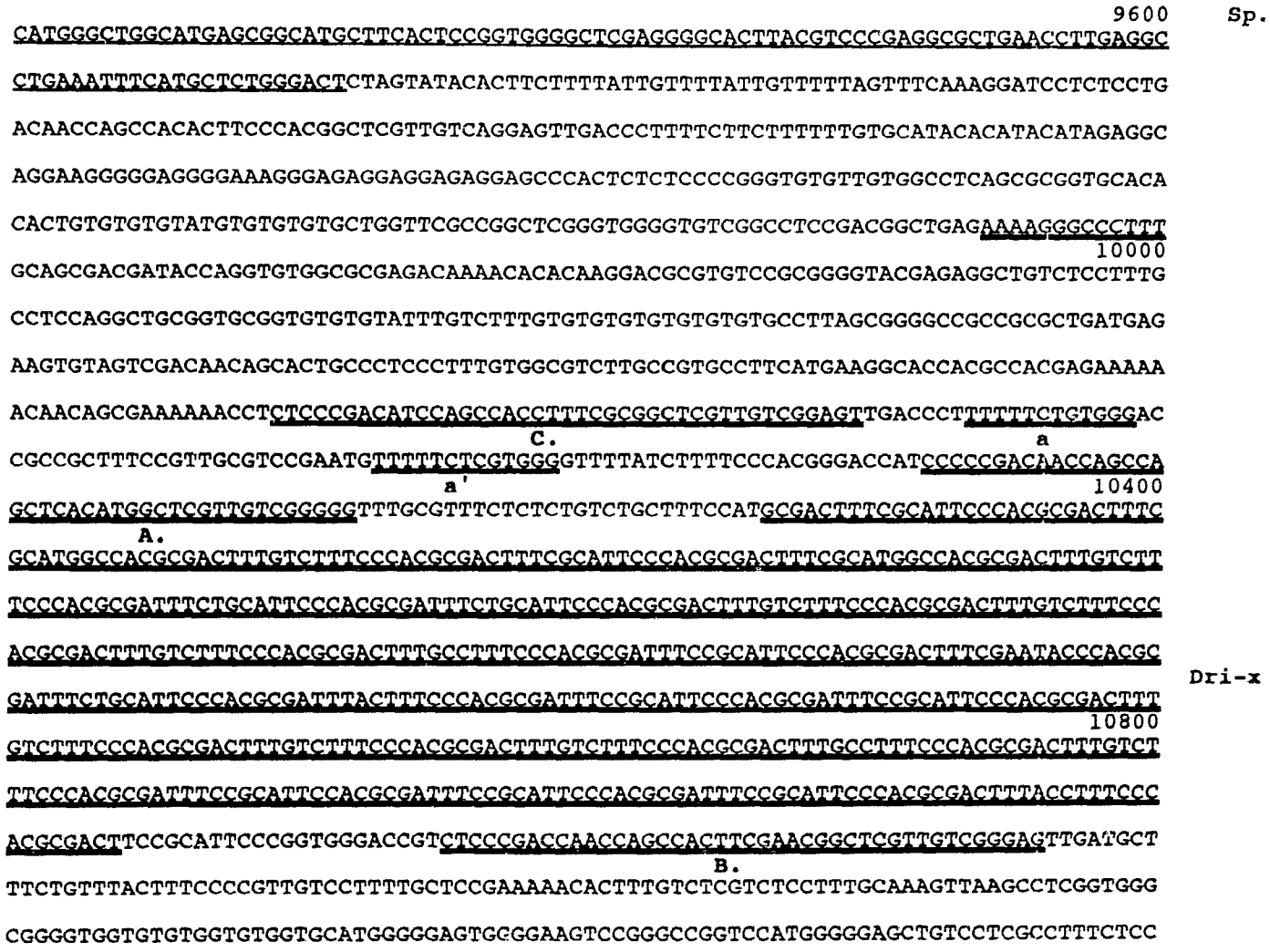
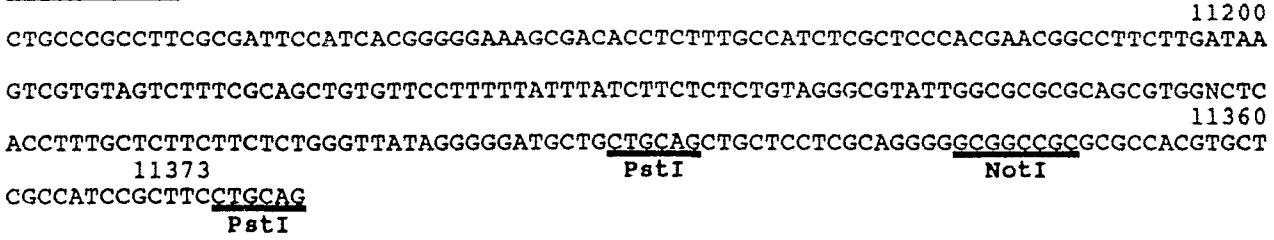


Figure 45h.


11200
CTGCCCGCCTTCGCGATTCCATCACGGGGGAAAGCGACACCTCTTTGCCATCTCGCTCCCACGAACGGCCTTCTTGATAA
GTCGTGTAGTCTTTCGCAGCTGTGTTCCTTTTTATTTATCTTCTCTCTGTAGGGCGTATTGGCGCGCGCAGCGTGGNCTC
11360
ACCTTTGCTCTTCTTCTCTGGGTTATAGGGGGATGCTGCTGCAGCTGCTCCTCGCAGGGGGCGGCCGCGCGCCACGTGCT
11373
PstI
NotI
CGCCATCCGCTTCCTGCAG
PstI

## SECTION D

# DISCUSSION

## 1. QUESTIONS TO BE ANSWERED

The work in this thesis was undertaken to investigate and establish the genomic organization of the nucleus-encoded rRNAs of the protoctist, Crithidia fasciculata. Prior to initiating my research M. W. Gray had determined that the cytoplasmic ribosome of Crithidia possessed the most unusual collection of rRNA molecules discovered up to that time. As has been outlined in detail in this thesis, the ribosomal SSU contains an unusually long RNA while the LSU contains a highly fragmented RNA consisting of six separate pieces, in addition to the 5.8S and 5S rRNAs.

These initial observations raised a number of questions:

1. How are these rRNAs encoded in the nDNA: are their genes integrated into the nDNA, or are they found on extra-chromosomal elements?

2. Are the rRNA genes grouped together at a single locus, or are they scattered throughout the genome at many loci?

3. If the rRNA genes are located at one locus, how are they arranged relative to one another?

4. Are the rRNAs processed from a single transcript (i.e., transcribed from a single promoter), or are they derived from multiple transcripts (i.e., transcribed from separate promoters)?

5. If the rRNA genes map together and are transcribed from a single promoter, how is the pre-rRNA processed to yield the mature products?

6. What is the evolutionary significance of the multiple rRNAs and their genic arrangement?

7. What can the primary sequence analysis of the rDNA tell us about rRNA and rDNA function and evolution?

## 2. GENOMIC ORGANIZATION OF rRNA GENES: COMPARISONS WITH OTHER SYSTEMS

Initial steps to determine the genomic organization of the rDNA involved restriction endonuclease digestion of the nDNA and Southern blot analysis with individually $^{32}$P-labelled rRNAs. These experiments revealed that the rDNA could be isolated as an 11- to 12-kbp fragment as demonstrated by the fact that all the rRNAs except 5S (Sp.h) hybridized to an 11- to 12-kbp fragment in BamHI or SalI digest.

Many attempts were made to shotgun-clone the presumptive rDNA repeat as a single SalI fragment but this approach proved unsuccessful. Concern that these failures were due to the large size of the desired insert prompted attempts to clone the rDNA in more manageable-sized fragments. For this reason PstI was chosen as the next cloning enzyme. PstI conveniently digests Crithidia rDNA into four identifiable rDNA fragments. Sp. b hybridized by itself to a 3.9-kbp fragment (Cf1), Sp. c, i, and e to a 3.2-kbp fragment (Cf2), Sp. d and f to a 2.0-kbp fragment (Cf3), and Sp. g and j to a ladder of rDNA fragments ranging in size from 2.0 to 3.0 kbp (Cf4; 2.2 kbp). The combined sizes of these fragments approximated the size of the presumptive rDNA repeat (11-12 kbp).

Shotgun cloning of PstI-digested nDNA proved successful, and the four desired pUC9-rDNA recombinants (pCf1-pCf4) were identified and isolated for gene mapping and sequencing. Restriction mapping, Southern blot experiments, and rRNA primary sequence information allowed unambiguous arrangement of the rRNA coding regions and relative orientations of Cf1, Cf2,

and Cf3 (as they would appear in the native rDNA). An overlapping HindIII rDNA clone (CfH1) was required to orient Cf4 relative to Cf1, Cf2, and Cf3, and to provide the primary sequence data of the Cf3-Cf4 and Cf4-Cf1 boundaries. It was during the sequencing of the Cf4-Cf1 boundary that a fifth PstI rDNA fragment (Cf5; 50 bp) was discovered.

With completion of the mapping and sequencing studies, comparison with other eukaryotic nuclear rRNA gene arrangements was possible, although the location of the rDNA (chromosomal or extra-chromosomal) has not yet been definitively established. The permutated arrangement of Crithidia rDNA (i.e. IGS-SSU-LSU-IGS) could arise as a result of its being chromosomally integrated and tandemly repeated, or its being contained in an extra-chromosomal circular molecule. Although undigested nDNA fractionated on 1% agarose gels did not reveal any 11- to 12-kbp bands, the possibility of multimeric (i.e., 22- to 24-kbp dimers) rDNA plasmid-like species could not be ruled out, as they could very well have remained hidden in the broad band of nDNA.

Several techniques could help determine if Crithidia rDNA is in the form of extra-chromosomal molecules. Preferential enrichment for circular DNA could be accomplished by separation from the chromosomal DNA using an alkali treatment of cell lysates followed by EtBr/CsCl equilibrium centrifugation of the fractionated DNA. Electron microscopy of nuclear contents could also reveal plasmid-like DNAs. Also, partial digestion with SalI or BamHI could be carried out to generate ladders (monomer, dimer, etc.) that could be separated by pulsed field gradient gel electrophoresis. If ithe rDNA were in the form of a circular monomer, one would expect to see only a single band rather than a ladder (unless the rDNA was a circular dimer, in which case one would see two bands).

As is becoming increasingly apparent, extra-chromosomal rDNA, particularly in the Protoctista, may be a much more widespread

phenomenon than originally thought. Whereas in most "higher" eukaryotes, the rDNA is chromosomally integrated and directly repeated (although extra-chromosomal rDNA circles have been reported in Xenopus and yeast), several protoctists possess extra-chromosomal rDNA exclusively, or chromosomally integrated rDNA with extra-chromosomal amplification.

The closely related kinetoplastids Leishmania donovanii, Trypanosoma brucei, and T. cruzi appear to have tandemly repeated, chromosomally integrated rDNA (167,145,170). Euglena gracilis, an organism which appears to be as ancient as Crithidia, has no integrated rDNA but instead has an 11.5 kbp extra-chromosomal circle (172). Euglena gracilis is of particular interest because of its recently described, multiply split LSU rRNA, not unlike that in Crithidia. This finding and its implications are discussed in detail later in this section.

Other protoctists possessing extra-chromosomal rDNA include:

1. Naegleria gruberi (an organism conventionally placed with the amoeba), whose rDNA is exclusively in the form of a 14 kbp extra-chromosomal circle (192);

2. Entamoebe histolytica (an organism which lacks a Golgi apparatus and mitochondria), whose rDNA appears to be found exclusively as a large (26.6 kbp) palindromic circular molecule (193);

3. the slime molds Dictyostelium discoideum (47) and Physarum polycephalum(45), whose rDNAs are large (44 and 60 kbp, respectively) palindromic molecules, as is the case in,

4. the ciliate Tetrahymena thermophila (20 kbp) (42); however the related ciliate ,

5. Paramecium tetraurelia, possesses a genomically integrated rDNA with extra-chromosomal circular molecules (40).

## 3. rRNA GENE ARRANGEMENT AND rDNA EVOLUTION

In Crithidia fasciculata, the rRNA gene arrangement is conventional in spite of the fact that the ribosome LSU contains seven RNA components compared to the usual two. Starting at the 5' end of the SSU rRNA gene and going in the 3' direction, the rRNA cistron is organized like the nuclear rDNA in a typical eukaryote, with the 5.8S rRNA gene located between the SSU (18S) rRNA and "28S" rRNA genes.

The SSU rRNA gene sequence, as determined by Dr. M. N. Schnare (191), revealed that the RNA it encodes is unusually long; indeed the longest yet determined at 2206 nt (consistent with the size (2400 nt) estimated from polyacrylamide gel electrophoresis), and that the coding sequence is uninterrupted by any IVS. The extra nucleotides could be accounted for by the increased lengths in two variable regions (V7 and V8) as defined by comparative SSU rRNA primary and secondary structure analysis. Pairwise comparisons of this SSU rRNA gene sequence with those of other eukaryotic nuclear SSU rRNAs revealed that Crithidia represents one of the earliest branchings within the eukaryotic line of descent, earlier than the branching that gave rise to the ciliates .

The genes for the remaining rRNAs, Sp.c-j (excluding Sp.h), followed (in a 5' to 3' direction) the coding region for the SSU rRNA. Their arrangement is as follows: SSU rRNA-ITS1-Sp.i-ITS2-Sp.c-ITS3-Sp.e-ITS4-Sp.d-ITS5-Sp.f-ITS6-Sp.j-ITS7-Sp.g-IGS. The LSU rRNA is encoded by seven coding segments that are interrupted by six internal transcribed spacers (ITS). Secondary-structural modelling suggested that this LSU rRNA complex is held together by long-range intermolecular base-pairing interactions which are

intramolecular in the covalently contiguous LSU rRNA of E. coli. The discontinuities between the rRNAs map to highly variable regions in covalently continuous LSU rRNAs.

The 5' to 3' arrangement of the LSU rRNAs corresponds to their position in more conventional LSU rRNAs. This fact coupled with what is known in other eukaryotic nuclear systems implies that the individual rRNAs are most likely produced by complex post-transcriptional processing of a pre-rRNA whose synthesis is initiated from a single promoter situated in the IGS. This complexity is not only exemplified by a requirement for removal of many spacer sequences but also by the presence or absence of 5' phosphate groups on the individual LSU rRNAs and by the post-transcriptional addition of A residues to the 3'-end of Sp. g, the RNA encoded at the distal end of the rDNA repeat (see ref. 166, 178).

The multiply fragmented LSU rRNA of Crithidia illustrates that ribosomal RNAs need not be covalently continuous to be functional, although the sites of the discontinuities are vitally important, as pointed out by the example of ribosome inactivation produced by disruption of the LSU rRNA by fungal cytotoxins at the $\alpha$ -sarcin site.

At the time this work was completed, Crithidia fasciculata provided the most dramatic example of a fragmented rRNA. Although the most extreme, Crithidia was not the only example of a fragmented rRNA. Indeed the 5.8S rRNA has been shown to be the homologue of the covalently continuous 5'-end of the LSU rRNA in E. coli, and is therefore in essence a fragment of the LSU rRNA whose coding region is separated from the rest of the LSU rRNA by an internal transcribed spacer. Other examples exist of further splitting of LSU rRNA, including the 4.5S rRNA in plant chloroplasts and bipartite LSU rRNAs in Acanthameoba and Drosophlia, which result from the post-transcriptional

excision of ITS from the pre-rRNA. The number of rRNAs is therefore dependent on the number of ITS present in the rDNA.

As previously noted, Crithidia appears to belong to an evolutionary lineage that separated from the main line of eukaryotic descent at a very early age of eukaryotic nuclear evolution. This raises the question, "Does the presence of ITS represent an ancient condition, subsequently lost in most other eukaryotes but retained in Crithidia, or does it represent a condition acquired by the lineage giving rise to Crithidia?"

Crithidia fasciculata is a member of the Kinetoplastida, a group of organisms which includes species responsible for serious diseases in humans and domestic animals. The most important of these pathogenic species are Leishmania donovanii (responsible for visceral leishmaniasis), Trypanosoma cruzi (responsible for Chagas' disease), and Trypanosoma brucei (responsible for nagana disease in cattle and horses). Like Crithidia, all these organisms possess an unusual complement of nuclear-encoded rRNAs. In each case, the cytoplasmic large subunit rRNA is present as a multiply split RNA in the mature ribosome. These RNA components have been mapped to the rDNA in various trypanosomatid species. In L. donovanii (167), which has 166 copies of a 13.5 kbp rDNA repeat arranged in a conventional head-to-tail fashion and integrated into the nuclear chromatin, coding regions for the two high molecular weight pieces are separated by a 587 bp gap. To date the small rRNA components have not been mapped in this species.

White et al. (168) have studied the seven rRNA components of the T. brucei cytoplasmic ribosome. In this organism, the LSU rRNA is made up of six pieces: the LSU $\alpha$ and LSU $\beta$ rRNAs, three small rRNAs (180, 70 and 140 nt in length) whose genes are located 3' to the LSU $\beta$ coding region, and a 140 nt

rRNA, the gene for which is located in the gap between the LSU α and LSU β cistrons.

The T. cruzi rDNA map has been established by Hernandez and Castaneda (169,170). These workers reported the existence of five rather than six small rRNAs (the 70 nt equivalent of Sp. j was not detected in this study). The S3 (presumed 5.8S) rRNA mapped between the SSU (18S) and the LSU (28S) rRNA genes, while the S1 rRNA (the equivalent of the 220 nt rRNA in T. brucei) was placed in the same region as the presumed 5.8S rRNA (based on the data presented, however, it could just as well have been placed in the gap between the coding regions for the two halves of the 26S rRNA, as is the case in T. brucei). Further work will undoubtedly clarify the precise organization of the rDNA in this organism.

The presence of a highly fragmented LSU rRNA thus seems to be a characteristic of the trypanosomatid flagellates. Interestingly, Euglena gracilis, an organism which possesses a chloroplast and flagellum, contains within its cytoplasmic ribosome 16 separate components, of which 15 are located in the LSU. Fourteen of these 15 rRNAs (excluding the 5S rRNA) map to a 7221 bp region of rDNA which has been completely sequenced. The rRNAs are encoded in a linear array as they are found in a conventional LSU rRNA and are separated by 14 ITS, three of which are in the same positions as in Crithidia (173,174). Primary sequence comparison of the SSU rRNA has demonstrated that Euglena diverged from the main line of eukaryotes at about the same time as the trypanosomatid protoctists (173). The fact that Euglena diverged from the main line of eukaryotic descent at or before Crithidia and that they share some of the same ITS lends weight to the proposition that multiply fragmented rRNAs are a primitive condition in eukaryotes.

## 4. 5S rDNA

In Crithidia the 5S rRNA (Sp. h) hybridizes to very high molecular weight nDNA in all digests tested. The smallest-sized fragments are found in a PstI digest and are ca. 11-12 kp in length. Such digests also contain two less intense bands much larger in size (20 and 25 kp in length). In a SalI digest, the 5S rRNA hybridizes to a single tight band of ca. 30 kbp but a BamHI digest reveals a prominent tight band of ca. 25 kbp and a less intense band of ca. 30 kbp. Taken together the evidence seems to indicate that the 5S rDNA genes are physically unlinked to the other rRNAs and may very well be on a different chromosome. The data also suggest that at a minimum there are three or four 5S rRNA genes, although for all eukaryotes in which it has been studied, this gene is found hundreds of time, often in tandem arrays. In the related trypanosome, T. brucei, the 5S rRNA gene is physically unlinked to the remaining rRNA genes and found on a 750-bp AluI tandem repeat. Each repeat unit possesses only one gene (174). Xenopus contains 750-bp tandem repeats but each one possesses two or three 5S rRNA genes (175-177). Repeat-sequence heterogeneity was noted between one of the sequenced T. brucei repeats and three other sequenced 5S repeats.

## 5. pCfH1 rDNA INSTABILITY

When the pCfH1 plasmid was propagated in the $rec^+$ strain JM83 for large-scale plasmid extraction, a large deletion of the insert DNA occurred, resulting in a plasmid of altered size. pCfH1 contains the large 6.9 kbp HindIII rDNA fragment and has within it the Cf3-4 boundary, Cf5, and the Cf5-1

boundary. As will be discussed below, CfH1 contains the rDNA IGS which encompasses the region responsible for length heterogeneity (due to variable numbers of tandemly repeated sequences) in Crithidia rDNA. In two instances large-scale preparations of pCfH1 (propagated in E. coli strain JM83) contained truncated inserts of the same size. Restriction analysis of one of these mutated pCfH1 inserts revealed that only the left-hand PstI site of the homologous pCf4 DNA sequence was retained, and that the unique restriction endonuclease sites 3' to it (see Figure 22) were missing. Interestingly, intact CfH1 could be isolated in minipreps from JM83. To obtain usable quantities of intact pCfH1 plasmid the DNA had to be transferred to a rec$^-$ strain of E. coli , DH1.

Other investigators have examined rDNA spacer stability in E. coli. Morgan and McMahon (181) cloned the Xenopus rDNA spacer into pBR322 and transformed a strain of E. coli (JC8679 recBC SbcA) which exhibits an unusually high level of plasmid recombination. The changes in the plasmid studied were confined to the promoter-related regions and consisted for the most part of precise deletions of the tandemly repeated BamH1 super repeats, which are composed of a spacer promoter and multiple 60/81-bp repeats. Alteration due to recombination in this engineered plasmid was an infrequent event in recA strains of E. coli. In contrast the Crithidia rDNA spacer DNA did not require a hyper-recombinational strain of E. coli to be altered. It seems that sequences in CfH1 participate in homology-dependent, intra-plasmidic recombination involving a region of the Crithidia IGS which possesses tandemly repeated DNA. This rearrangement is abolished or greatly reduced by a recA genotypic background. The fact that intact plasmid could be obtained in small-scale preparations of DNA suggests that the event(s) involved in the alterations is (are) dependent, at least in part, on growth or culture conditions. The large size of

CfH1 may also be a contributing factor. JM83 had been used to prepare the plasmid DNA of pCf1, pCf2, pCf3 and pCf4 with no observed insert alterations.

## 6. rDNA LENGTH HETEROGENEITY

In genomic-blot analysis of EcoRI-digested nDNA, it was observed that the rDNA was cut in at least two sites, such that the SSU rRNA gene was liberated from the remaining portion of the rDNA to which all the other rRNAs (except Sp. h) hybridized. It could also be seen for the first time that there existed some size heterogeneity in the rDNA. Instead of hybridizing to a single rDNA fragment, Sp. c-j (except Sp. h) hybridized to a broad band of rDNA fragments differing in size by 1.0 kbp at the extremes of the size range. This implied that the rDNA was heterogeneous in size. This size heterogeneity was more dramatically demonstrated when PstI-, HindIII- or BglII-digested nDNA was probed with $^{32}$P-labelled rRNA (iRNA in the case of HindIII and BglII and Sp. g and j in the case of PstI). In each case, a ladder of fragments was seen. The size difference between the smallest and the largest fragments was 1.0 kbp. The banding pattern was identical in all digestions, as was the pattern of band intensities. The individual rDNA fragments which make up the "struts" of the ladder did not differ from one another by a simple size multiple (i.e. 100, 200 or 300 bp intervals). The actual size differences between one strut and the next (either above or below) was the same for the different enzymes used to generate the ladder.

It was possible at this point to make some assumptions about the location of that portion of the rDNA that was responsible for the length heterogeneity. Experiments in which EcoRI-digested nDNA was probed with the individual rRNAs demonstrated that all of the rRNAs (except perhaps Sp. b)

hybridized to the 8.5-9.5 kbp size fragments, indicating that their coding regions were physically linked to the heterogeneous rDNA. Sp. b weakly hybridized to the same group of fragments, suggesting that a small portion of its coding region was linked with the other rRNA genes and hence the heterogeneous region. (Alternatively, this weak hybridization could be due to some cross-contamination of Sp. b with Sp. c and/or d). The combined sizes of the Sp. b-hybridizing rDNA fragments approximate (11-12 kbp) the size of the rRNA hybridizing rDNA fragments in the SalI or BamHI digests, suggesting that the two fragments constitute the entire rDNA. If the 11-12 kbp BamHI/SalI-hybridizing fragments are indeed the entire rDNA, then it should be possible to discern the length heterogeneity. On re-inspection of the blots the bands do appear to be rather broad but because of the low degree of resolution at this size range in 1% agarose gels, it would not be possible to see individual bands differing by only two or three hundred bp.

Whereas all the rRNAs, (in the case of SalI and BamHI,) and Sp. c-j, excluding Sp. h (in the case of EcoRI digests), hybridized to nDNA fragments showing length heterogeneity, only Sp. g and j did so in the PstI-digested nDNA. Therefore it was possible to localize the heterogeneous region to within 2000 bp of the genes for Sp. g and j.

As indicated previously, repeat size periodicity was not constant (the different repeats differing from one another by 200 or 300 bp) and repeat number (as expressed in the intensity of hybridization to the labelled rRNA probe) also varied substantially. These size differences are not easily explained by arguing that they are due to simple variations in the number of the individual small tandemly repeated sequences in Crithidia spacer DNA (to be discussed later), which are only ca. 20 bp long. It was not possible to demonstrate any rDNA repeats which differed by 20, 40, 60, 80 or 100 bp. It seems, then, that if

the length difference is due to these repeats, as indirect evidence suggests, it must be as a result of blocks of 10 or 15 of the small direct repeats. This could be readily answered by sequence analysis of the region in pCfH1 that houses the DNA responsible for the difference in size between its Cf4 homologue and Cf4 itself. This region resides between the unique SalI and NotI sites of CfH1 (as depicted in Figure 27) and is approximately 1800 bp in length.

As was mentioned in the Introduction, length heterogeneity of rDNA repeats is a common finding. It is almost always the result of differences in the size of the 'NTS' or 'IGS'. In Drosophila, some of the rDNA repeats differ in size due to intervening sequences which interrupt the LSU rRNA gene. It is thought that these rDNAs do not give rise to functional RNAs. In Crithidia the length heterogeneity is confined to the same spacer region. Although the length heterogeneity has not been precisely localized, the size differences do map to a position of the spacer DNA containing the directly repeated DNA and bounding unique-sequence DNA within pCf4/pCfH1.

The question of whether the length heterogeneity seen in Crithidia is unique to any individual cell or is a population phenomenon was studied. As shown here, the pattern of length heterogeneity is a stable, reproducible one, found within each individual cell. In the case of multi-cellular organisms (in which almost all studies of length heterogeneity have been done) the heterogeneity is an individual phenomenon. Indeed it is possible to use markers of length heterogeneity to study the relatedness of species, sub-species or sympatric populations of individuals. It is generally thought that this length heterogeneity is the product of an unequal crossing-over event during mitosis or meiosis, or gene conversion. Because Crithidia is most likely asexual, it seems that this heterogeneity is the result of mitotic unequal crossing-over or gene conversion. The frequency of conversion or unequal crossing-over must be quite

low. DNA extracted from cells which were separated by seven years and presumably thousands of mitotic generations was identical in the pattern of hybridization of the spacer DNA. Interestingly, the coding region in one of fifteen clonally selected rDNA did exhibit some difference from the standard pattern. This was seen when Sp. f was hybridized to the clonally derived DNAs after digestion with PstI. The aberrant clone contained, in addition to the same PstI fragment seen in the original rDNA (Cf3), a second much larger fragment which hybridized with lower intensity. This altered pattern was not due to incomplete digestion because the Cf3 homolog was present in the same quantity (as measured by the degree of hybridization of the probe to the rDNA fragment) as in the other 14 clones. When the same probe was hybridized to BglII digests of six different clonally derived nDNAs, no variation in hybridization pattern could be seen, although the large size of the BglII fragments and consequently poorer resolution may have masked such a change.

The most complex arrangement of IGS/NTS DNA so far described is in Xenopus, whose rDNA contains four regions of repetitive DNA designated 0, 1, 2 and 3. Region 0 contains a variable number of 34 bp perfect repeats while region 1 contains a variable number of near perfect repeats ca. 100 bp in length. Regions 2 and 3 contain tandemly repeated mixtures of two related sequences designated "60" and "81", one differing from the other by an insertion/deletion of 21 bp. The 60-bp repeats and 81-bp repeats are very well conserved within each class, differing by very few bp. The 60-bp and 81-bp sequences (excluding the 21-bp insertion/deletion) are also very similar to one another. The 60-bp and 81-bp sequences possess a 42-bp sequence that is homologous to 42-bp within the pre-rDNA promoter region found at the boundary of the NTS/IGS and ETS.

By contrast, individual Crithidia tandem direct repeats do not exhibit any sequence similarity with the region around the experimentally determined site of pre-rRNA transcription initiation. Nor do the Crithidia direct repeats or site of pre-rDNA transcription initiation share any obvious similarity with the Xenopus spacer promoters, which is not surprising given the observed rate of evolution in this DNA. The Xenopus 60/81-bp sequences have been shown to possess enhancer-like activity. Demonstration of such activity by the direct repeats in Crithidia awaits the development of a quantitative in vitro rDNA transcriptional assay.

Drosophila also possesses a complex NTS/IGS. Travelling in a 5' or 3' direction there are three regions of repetitive DNA. Rep I is made up of three sequences (A, B, C) repeated tandemly five to nine times. Rep II is immediately after Rep I and is also known as the 330-bp region. It possesses sequences 2, 3 and 4, in variable number, arranged tandemly. Sequence 2 possesses sequence similarity with the pre-rDNA promoter. Rep II also contains sequences c and a. Rep III is composed of sequences 1, 2, 3, 4 and 5 and is obviously related to Rep II sequences. These sequences also possess sequences homologous to the pre-rRNA promoter. The number of Rep III repeats is variable from rDNA repeat to rDNA repeat. The 240- and 330-bp repeats (Rep II and III) have been shown to enhance rDNA transcription in co-transfection experiments. In addition the 240-bp repeats are known to promote transcription, resulting in spacer transcripts differing in length by 240 nt up to a maximum length of 1680 nt. The direct repeats in Crithidia spacer DNA do not share any sequence similarity with any of the sequences in the Drosophila rDNA spacer.

Repeated sequences of a very similar size to those in Crithidia IGS/NTS have been identified in the mouse NTS/IGS and are known to be important in transcription termination and promotion. These so-called Sall boxes

are 18 bp in length and are highly conserved. They possess a SalI recognition site (hence the name) and are found near the two sites of pre-rRNA transcript termination. They are repeated eight times but not in a tandem fashion, and have been shown to bind protein factors. When compared to the direct repeats (Dr i-x) in Crithidia there is no sequence conservation. Whereas the Crithidia direct repeats are tandemly arranged, the SalI boxes are separated from one another by a variable number (18 to 123 bp) of unique sequence DNA. Whether the Crithidia direct repeats are involved in transcription termination remains to be demonstrated.

The NTS/IGS of the three related hypotrichous ciliates, T. thermophila, T. pyriformis and Glaucoma chattoni (201), have been sequenced and analyzed. Interestingly, they possess short repeat sequences both 5' and 3' to the transcribed portion of the rDNA. The sequences in the 5' region are of three unrelated types and are interspersed with unique sequence DNA. The Type I repeat is 33 bp long and extremely rich in A. Type II repeats are variable in length but most are 22 bp long. In T. thermophila and G. chattoni these repeats are tandemly arranged and conserved in position but not sequence. They are repeated eleven and eight times, respectively. These Type II repeats resemble Crithidia's tandem direct repeats, Dr i-x, in arrangement and position, but not sequence. Type III is about 20 bp long and conserved in sequence but not location. It is not tandemly repeated. The Type I repeat may serve a special r' 'e in Tetrahymena rDNA replication and transcription. It is found four times in the spacer DNA: twice 5' to the site of transcription initiation, once grouped with three copies of type III repeats in Domain I (a DNase I hypersensitivity site and region known to be important in rDNA replication) and once again with three copies of type III repeats in Domain II. Larson et al. (180) determined the primary sequence of an rDNA molecule isolated from a mutated Tetrahymena

strain which exhibits an alteration in the ability to form or maintain amplified rDNA, and found that the new genotype was due to a single base change in the Domain II, type I repeat.

## 7. SITE OF TRANSCRIPTION INITIATION

A detailed comparison of the nucleotide sequence surrounding the transcription initiation site from several organisms highlights sequences that are both highly and poorly conserved in evolutionary terms (see Fig. 44). Overall, RNAP I promoter sequences are poorly conserved in position and sequence. This is in stark contrast to promoters for RNAP II and III, which not only are highly conserved in sequence, but are rather precisely fixed in position with respect to the site of initiation. Nucleotides -16 and +1 are the only ones that show any degree of evolutionary conservation. Such strict conservation implies an important functional role. As noted earlier in this thesis, nucleotide -16 has been shown to be absolutely essential for transcription initiation in human, mouse and Acanthamoeba. A similar role for nucleotide +1 has not been established although it does reside in a stretch of DNA that binds RNAP I in mouse and Xenopus. Nucleotide -7 shows a lesser degree of conservation, but still fairly high over a broad spectrum of organisms. This residue is an absolute requirement in mouse as judged by complete lc ; of transcription when G at -7 is replaced by A. There are positions that exhibit a purine and/or pyrimidine preference. At positions -8, +5, +7 and +9, G or A is preferred, while C or T is preferred at -1. The functional significance, if any, remains unclear for these positions. Although there is no conservation of extended sequences over great evolutionary distances, sequence similarity is evident between closely related organisms. Mouse and rat share the same sequences between -38 and -32, -30 and -16, and -1 and +10. Humans have

the -38 to -32 region (shifted 15 nt upstream relative to mouse and rat) and -1 to +10 sequences in common with mouse and rat. The -30 to -16 region is not shared, an interesting observation when viewed from the perspective that rat, but not human, TIF will support transcription of wild type mouse rDNA with mouse RNAP I in vitro, and that mouse TIF-IB protects a portion of this motif in in vitro experiments.

The Xenopus species share a high degree of similarity over an extended portion surrounding their rDNA transcription initiation sites. This includes a region (-10 to +15) protected by a protein factor in vitro. A similar degree of sequence similarity is seen between two ciliate protozoans, Tetrahymena thermophila and Glaucoma chattoni.

The six fungal species share a region from -9 to -3, but Schizosaccharomyces pombe and Neurospora crassa do not share a region extending from -9 to +10 with the four Saccharomyceloidae. S. pombe shares with K. lactis an additional 11/12 nt between -26 and -15 which are not shared with the remaining fungi.

Grondal et al. (195) determined the putative site of RNAP I transcription initiation in Crithidia by $S_1$ nuclease mapping and nuclear run-on experiments. Crithidia does not share any sequence similarity with any of the other determined RNAP I promoter sequences, even with respect to Trypanosoma brucei, a fellow trypanosomatid. In all other organisms listed in Figure 44, the important +1 nucleotide is A or G, whereas Crithidia possesses a C. Interestingly, if the Crithidia sequence is shifted one nucleotide to the left, the +1 position becomes a G and the -7 position becomes an A, more in keeping with the convention (i.e., a purine at position +1 and -7). RNAP I promoter-identification experiments using a quantitative in vitro transcriptional assay and deletion mutants plus single-site mutants generated by the sodium bisulfite technique

carried out to determine what sequences are important in Crithidia rDNA transcription initiation. This initial sequence comparison seems to indicate that these sequences may be unlike regulatory sequences in the rDNA of other eukaryotes.

Niles et al. (89) mapped the presumed site(s) of transcription termination in Tetrahymena sp. using a 35S pre-rRNA molecule whose 3'-end was positioned 15 bp downstream of the 3'-end of the mature LSU rRNA. As has been noted previously in this thesis, interpretation of the site of transcription termination using these transcripts must be made with caution because pre-rRNA processing may be so rapid that the true transcription termination site cannot easily be identified. These authors comment on sequences near the $S_1$-mapped 3'-end which can theoretically assume hairpin-loop structures followed by stretches of unique nucleotides in the DNA. This kind of arrangement has been shown to be responsible for rho-independent RNA transcription termination in E. coli.

Downstream (286 nt) from the 3'-end of the coding region for the small rRNA Sp. j is a 12-nt palindromic sequence. More impressive is a 17-nt perfectly matched hairpin with a 7-nt loop (Bloc A). This sequence ends 8 bp from the start of the tandemly repeated 19-bp sequences. A sequence bearing 80% identity with the A repeat and also capable of forming a hairpin and loop structure is found immediately downstream of the tandemly repeated region, such that this region is flanked by repeats A and B. A third related sequence (85% identical to A), is located about 120 bp upstream from A. Repeats A and C are separated by DNA sequence containing two identical 12-bp sequences bearing no resemblance to the direct repeats or sequences A, B and C.

The palindromic repeats A, B, and C resemble rho-independent sites in their potential secondary structure, position (downstream of the coding regions) and possession of T residues immediately downstream of the hairpin-

loop structures. What role, if any, they play in transcription awaits direct experimental evidence. It is also possible that these sequences have some role in rDNA replication and/or the process of rDNA homogenization.

The only remaining repeated DNA is found 5' to the start of the coding sequence for the SSU rDNA at positions 3476 and 3560. These repeats differ by a 2-bp insertion/deletion 4 bp from the 5'-end. These repeats are unique in sequence and bear no resemblance to any of the other repeats in the Crithidia rDNA spacer.

In summary, Crithidia fasciculata possesses a novel complement of rRNAs. The SSU and 5.8S rRNAs are unusually long but by far the most interesting aspect of this system is the highly fragmented LSU rRNA which is made up of seven separate species. With the exception of the 5S rRNA, all the rRNAs are encoded together in an rDNA repeat unit ranging in size from 11-13 kbp. The coding regions for the rRNAs are separated from one another by eight internal transcribed spacers. In spite of the number of rRNAs the rDNA conforms to the conventional nuclear rDNA arrangement of 5'-ETS-18S-5.8S-"28S"-IGS-3'.

The IGS differs in base composition from that of the transcribed portion of the rDNA and possesses a more complex substructure. The most conspicuous part of the IGS is the tandemly repeated 19 bp repeats. The region implicated in rDNA length heterogeneity maps to the tandem repeats. These repeats are not identical but display variable degrees of relatedness; sequences homologous to them are not found outside this region. What role, if any, they have in rRNA transcription initiation and/or termination, rRNA processing, rDNA replication, or rDNA evolution remains to be determined. The spacer DNA also possesses other repeated elements of unknown significance, as well as motifs that could potentially form hairpin-loop structures (in an RNA

transcript). Their involvement, if any, in transcription awaits experimental evidence. The length heterogeneity discovered in this rDNA was determined to be an individual and not a population phenomenon. It was also shown to be a rather stable arrangement, not having changed in over seven years of serial culture of this organism. The degree of methylation in the rDNA, as determined by digestion with methylation sensitive/insensitive restriction endonucleases, appears minimal, keeping in mind the limitation that such experiments do not sample all possible methylatable sites.

## 8. EVOLUTIONARY SIGNIFICANCE OF THE STRUCTURAL ORGANIZATION OF THE rDNA REPEAT

As a starting point for discussion about the evolutionary significance of the Crithidia rDNA organization, I would like to review the common features of eukaryotic rDNA. As has been noted elsewhere in this thesis, most eukaryotes possess a characteristic rDNA structure. Commonly the rDNA is tandemly repeated and chromosomally integrated, with a sequence arrangement as follows: IGS-5'-ETS-18S-ITS1-5.8S-ITS2-28S-3'-ETS-IGS. In all eukaryotes studied to date, transcription of the pre-rRNA starts at the boundary between the IGS and 5'-ETS and within an RNAP I core promoter sequence. Termination of transcription often occurs at specific sequences either 3' to the coding region of the 28S rRNA gene or further downstream in the IGS. The rDNA may exhibit length heterogeneity which, when present, is usually due to a variable number of repeated sequences in the IGS. In some cases it is attributable to the presence of IVS in a sub-population of rDNA cistrons.

When one compares the primary sequence of rDNA cistrons from divergent organisms, homology is confined to the coding region for the

rRNAs. The degree of similarity is a measure of organismal evolutionary relatedness. In closely related species (i.e., X. laevis and X. borealis) some sequence conservation is seen in portions of the 5'- or 3'-ETS and in portions of the IGS. No significant sequence identity can be detected when the IGSs of distantly related eukaryotes are compared but some homology is evident when closely related species are analyzed (i.e., Xenopus spp. and Drosophila spp.) (80,135). A region which might be anticipated to be under evolutionary pressure to remain conserved in sequence due to its importance in rDNA transcription, the RNAP I promoter, is now known to be remarkably divergent in sequence and position, exhibiting only imperfect homology in a cohesive group like the mammals (106).

Sequences which can form secondary structures outside of the rRNA coding regions have been poorly conserved. There are sequences in certain IGS, (i.e., Tetrahymena spp. and yeast) (41,84) which have the potential to form secondary structures but as has been stated elsewhere, they are not conserved in sequence or configuration. Generally speaking all eukaryotes possess the same overall rDNA arrangement, with some notable exceptions to be discussed below.

In most eukaryotes the LSU rRNA consists of a complex between the 5.8S and 28S rRNAs. These molecules together constitute the homologue of the 23S rRNA in eubacteria and archaebacteria. The LSU rRNA in some eukaryotes (i.e., Crithidia, Euglena, Trypanosoma, and Leishmania) (167,169,170,171,172) can consist of multiple rRNAs which arise as a result of the post-transcriptional removal of multiple ITS sequences from the pre-rRNA. At the other extreme, there is now an example of a eukaryote (a microsporidian) (208) which does not possess a 5.8S rRNA, due to the absence of an ITS2. In essence, the LSU rRNA is of the eubacterial/archaebacterial configuration. This is particularly interesting because this organism appears to have branched off from

the main eukaryotic lineage well before the kinetoplastids and euglenoids (208). This raises the question, "Is the presence of an intact LSU rRNA gene a primitive condition, as its phylogenetic placement might suggest, or is it a peculiarity of the microsporidians?" If a deeper branching organism could be identified, it might be possible to answer this question. In this regard the recent determination of the Giardia rDNA sequence and analysis of the SSU rRNA gene (209) suggests that this organism branched even earlier than the microsporidians did from the main eukaryotic lineage. The Giardia ribosome possesses a 5.8S rRNA, albeit shorter than normal, whose gene is separated from the 28S rRNA gene by an ITS2. It should be noted that the 28S rRNA is a covalently continuous molecule inthis organism.

The fact that Crithidia and Giardia have an IGS-SSU-LSU-IGS rDNA arrangement suggests this pattern was present very early in the evolution of eukaryotes. This organization contrasts with that found in eubacteria. Whereas the rDNA in eukaryotes is usually chromosomally integrated and tandemly repeated, or exists as extrachromosomal circles, eubacteria contain a small number of dispersed rDNA cistrons (seven in E. coli) on their single chromosome (13). The gene arrangement in these cistrons is as follows: 16S-tRNA-23S-5S-(tRNA). Superficially, but perhaps significantly, this resembles the eukaryotic arrangement. As mentioned earlier there is no separate 5.8S rRNA sequence in bacteria and in contrast to eukaryotic rDNA, a gene(s) for a tRNA(s) is (are) found in the spacer between the SSU and LSU rRNA genes in many but not all eubacteria. The presence of a 5S rRNA gene mapping with the other rRNA genes has been reported in a number of eukaryotes (47,86), but as in the cases where it is separated from the rDNA repeat, this 5S rRNA gene is transcribed by RNAP III and not RNAP I.

It has become increasingly apparent that the IGS is extremely important in rDNA transcription, enhancement of promotion, termination and replication. It also has become apparent that these functions are dependent upon the presence of blocks of sequences and/or specific nucleotides which are found in the IGS outside of the ETS regions. Two of the best-studied systems in this regard are Xenopus and Drosophila. In both cases the IGS possesses a complex array of tandem repeats interspersed with imperfectly duplicated RNAP I promoters. In both systems these repeated sequences are involved in the enhancement of rDNA transcription. In other organisms (yeast, human, and rat) (83,86,107,119,120,122,123,134) the IGS does not possess a highly repeated structure but does contain single or double duplications of the RNAP I promoters which function in enhancement.

The Crithidia tandem direct repeats are similiar to the IGS repeats in Drosophila and Xenopus in location and form, but not in sequence. It is tempting to speculate that they are enhancers but there is no sequence identity with the recently determined RNAP I site in Crithidia to support such a claim. If this could be established, it would imply that the rDNA transcriptional apparatus present in eukaryotes became fixed, in evolutionary terms, early in the evolution of this group. On the other hand the hairpin-loop structures (Blocs A, B and C) which flank the tandem array in the Crithidia IGS (and which have not been described in other eukaryotic rDNA) resemble eubacterial rho-independent termination sites, suggesting that the Crithidia transcription initiation/termination apparatus may possess properties like that found in eubacterial systems.

When one compares and contrasts rDNA transcription in eukaryotes with that in eubacteria or archaebacteria, similarities and differences come to light. In eubacteria the rRNAs usually arise from the processing of a 30S pre-rRNA transcript which is initiated from one of two possible promoters found

upstream of the mature end of the 16S rRNA gene. Termination occurs downstream of the 5S rRNA gene (10). The situation in archaebacteria is more interesting in many ways. The rRNA operons of the halophiles and methanogens resemble typical eubacterial rRNA operons, having closely linked 16S-, 23S- and 5S-rRNA encoding genes with a tRNA gene located in the 16S-23S spacer and (in some operons) a tRNA gene located immediately downstream of the 5S rRNA gene. These operons can be found one to four times per genome (for a review of this topic see 210). The situation in the thermoacidophiles (the third main group of archaebacteria) is very reminiscent of eukaryotic rDNA. In this case, there is variable linkage of the 5S rRNA gene with the 16S-23S rRNA genes and no tRNA genes are present in the 16S-23S spacer. An arrangement also found in eukaryotes but not in eubacteria (nor, apparently in the other groups of archaebacteria) is the presence in a thermoacidophile of an operon that contains multiple repeated promoters upstream of the true promoter (210). Surprising, also, is the location of an RNAP termination site in conjunction with the first upstream promoter, suggesting that the multiple promoters may act to enhance transcription, as in Xenopus and mouse (75,130).

The Crithidia IGS direct tandem repeats seem to have arisen by unequal crossing-over of less highly repeated DNA generated, perhaps, by a DNA slippage mechanism at the time of DNA replication. Unequal crossing-over is thought to be responsible for the production and maintenance of many eukaryotic tandem repeats (14,80). Presumabley such a process is also reponsible for the production and maintenance of Crithidia rDNA, given its tandemly repeated organization. This process can also work to produce and maintain much smaller repeated DNAs, such as the Crithidia IGS tandem direct repeats. Given that these tandem repeats are not identical in sequence, it seems that mutations have been introduced at various sites within the repeats since the production of the array by

the successive rounds of unequal crossing-over (or by the unequal crossing-over process itself). It would be of some interest to determine the sequence details of similiar arrays in other copies of the size-heterogeneous rDNA repeats found in Crithidia, and to make comparisons about qualitatve and quantitative differences among them. With this type of information it may be possible to better work out the details of how the array was generated and is maintained.

## 9. FUTURE RESEARCH DIRECTIONS

The work completed to date and the insights it provides suggest several future areas of experimentation.

(1) The exact cause of the length heterogeneity could be better understood by determination of the DNA sequence of pCfH1, which contains an rDNA spacer of different length than that sequenced in pCf4.

(2) Spacer DNA sequence comparison between pCf4+pCf1 and pCfH1 would allow determination of conserved and non-conserved nucleotides in the spacer rDNA of different rDNA cistrons, and hence allow some comment on the importance of various identified sequence motifs in rRNA transcription or evolution.

(3) Perhaps the most interesting experiments which could evolve from the work outlined in this thesis would be the development of a cell-free system for in vitro rDNA transcription and processing. Early experiments of this type have been done with the related flagellate, T. brucei. The sites of transcription initiation and termination could be mapped precisely and then the substrates for DNA mutagenesis would be available for detailed studies of the rDNA sequences necessary for rDNA initiation and termination. Using what is known from studies of in vitro rDNA transcription in Xenopus, Drosophila,

mouse, and Acanthameoba, it should be possible to develop such a system in Crithidia.

## BIBLIOGRAPHY


1. Crick, F.H.C. "On Protein Synthesis", *In* The12th Symposium of the Society for Experimental Biology, Symp. Soc. Exp. Biol. 12, 138-163 (1957).

2. Crick, F.H.C. Central Dogma of Molecular Biology. Nature, 227, 461-463 (1970).

3. Mayr, E. *In* "The Growth of Biological Thought" Harvard Univ. Press, Cambridge, MA (1982).

4. Darwin, C. The Origin of Species. The New American Library, Inc., NewYork (1958).

5. Ochoa, S. and de Haro, C. Regulation of Protein Synthesis in Eukaryotes. Ann. Rev. Biochem., 48, 549-580 (1979).

6. Wool, I. The Structure and Function of Eukaryotic Ribosomes. Ann. Rev. Biochem., 48, 710-754 (1979).

7. Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L., and Nomura, M. (eds), Ribosomes - Structure, Function and Genetics. University Park Press, Baltimore, MD (1980).

8. Wittmann, H.G. Architecture of Prokaryotic Ribosomes. Ann. Rev. Biochem., 52, 35-66 (1983).

9. Noller, H.F. Structure of Ribosomal rRNA. Ann. Rev. Biochem., 53, 119-162 (1984).

10. McClure, W.R. Mechanisms and Control of Transcription Initiation in Prokaryotes. Ann. Rev. Biochem., 54, 171-204 (1985).

11. Moldave, K. Eukaryotic Protein Synthesis. Ann. Rev. Biochem., 54, 1109-1149 (1985).

12. Wittmann, H.G. Components of The Bacterial Ribosome. Ann. Rev. Biochem., 51, 155-184 (1982).

13. Morgan, E.A., Ikemura, T., and Nomura, M. Identification of Spacer tRNA Genes in Individual Ribosomal RNA Transcription Units of Escherichia coli. Proc. Natl. Acad. Sci. USA, 74, 2710-2714 (1977).

14. Long, E.O. and Dawid, I.B. Repeated Genes in Eukaryotes. Ann. Rev. Biochem., 49, 727-764 (1980).

15. Pavlakis, G.N., Jordan, B.R., Wurst, R.M., and Vournakis, J.N. Sequence and Secondary Structure of Drosophila melanogaster 5.8S and 2S rRNAs and the Processing Site Between Them. Nucl. Acids Res., 7, 2213-2238 (1979).

16. Ware, V.C., Renkawitz, R., and Gerbi, S.A. rRNA Processing: Removal of Only Nineteen Bases at the Gap Between 28Sα and 28Sβ rRNAs in Sciara coprophila. Nucl. Acids Res., 13, 3581-3597 (1985).

17. MacKay, R.M. The Origin of Plant Chloroplast 4.5S Ribosomal RNA. FEBS Letters, 123, 17-18 (1981).

18. Rochaix, J.B. and Darlix, J.L. Composite Structure of the Chloroplast 23S Ribosomal RNA Genes of Chlamydomonas reinhardii. J. Mol. Biol., 159, 383-395 (1982).

19. Leaver, C.J. and Gray, M.W. Mitochondrial Gene Organization and Expression in Higher Plants. Ann. Rev. Plant Physiol, 33, 373-402 (1982).

20. Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., and Woese, C.R. The Phylogeny of Prokaryotes. Science, 209, 457-463 (1980).

21. Doolittle, W.F. Genome Structure in Archaebacteria, *In* "The Bacteria",Vol. VIII, pp. 545-560, Academic Press, Orlando, FL, USA (1985).

22. Tu, J. and W. Zillig. Organization of rRNA Structural Genes in The Archaebacterium, Thermoplasma acidophilum, Nucl. Acids Res., 10, 7231-7238 (1982).

23. Gray, M.W., Sankoff, D., and Cedergren, R.J. On The Evolutionary Descent of Organisms and Organelles: A Global Phylogeny Based on a Highly Conserved Structural Core in Small Subunit Ribosomal RNA. Nucl. Acids Res., 12, 5837-5852 (1984).

24. Kimura, M. and T. Ohta. Eukaryotes-Prokaryotes Divergence Estimated by 5S Ribosomal RNA Sequences. Nature New Biol., 243, 199-200 (1973).

25. Hori, H. and Osawa, S. Evolutionary Change in 5S Ribosomal RNA Secondary Structure and a Phylogenic Tree of 54 5S RNA Species. Proc. Natl. Acad. Sci. USA, 76, 381-385 (1979).

26. Whittaker, R.H. New Concepts of Kingdoms of Organisms. Science, 163, 150-160 (1969).

27. Copeland, H.F. The Classification of Lower Organisms. Pacific Books, Palo Alto, CA (1956).

28. Dillan, L.S. The Science of Life. MacMillan, NY (1964 edn).

29. Gray, M.W. and Doolittle, W.F. Has The Endosymbiont Hypothesis Been Proven? Microbiol. Rev., 46, 1-42 (1982).

30. Pace, N.R., Olsen, G.J., and Woese, C.R. Ribosomal RNA Phylogeny and The Primary Lines of Evolutionary Descent. Cell, 45, 325-326 (1986)

31. Margulis, L. and Schwartz, K. Five Kingdoms. (2nd edition) W.H. Freeman and Company, NY(1988).

32. D'Alessio, J.M., Harris, G.H., Perna, P.J., and Paule, M.R. Ribosomal Ribonucleic Acid Repeat Unit of Acanthamoeba castellanii: Cloning and Restriction Endonuclease Map. Biochemistry, 20, 3822-3827 (1981).

33. Swanton, M.T., Greslin, A.F., and Prescott, D.M. Arrangement of Coding and Noncoding Sequences in the DNA Molecules Coding for rRNAs in Oxytricha Sp. Chromosoma (Berl.), 77, 203-215 (1980).

34. Lawn, R.M., Herrick, G., Heumann, J., and Prescott, D.M. Structural Organization of the Gene-Sized Pieces of DNA in the Ciliate Oxytricha. Cold Spring Harbor Symp. Quant. Biol., 42, 483-492 (1977).

35. Lipps, H.J. and Steinbrück, G. Free Genes For rRNAs in the Macronuclear Genome of the Ciliate Stylonychia mytilus. Chromosoma (Berl.), 69, 21-26 (1978).

36. Rae, P.M.M. and Spear, B.B. Macronuclear DNA of the Hypotrichous Ciliate Oxytricha fallax. Proc. Natl. Acad. Sci. USA, 75, 4992-4996 (1978).

37. Swanton, M.T., McCarroll, R.M., and Spear, B.B. The Organization of Macronuclear rDNA Molecules of Four Hypotrichous Ciliate Protozoans. Chromosoma (Berl.), 85, 1-9 (1982).

38. Blackburn, E.H. Characterization and Species Differences of rDNA: Protozoans. *In* "The Cell Nucleus Part A" (H. Busch, and L. Rothblum, eds) Vol. X, pp.145-170. Academic Press, Toronto (1982).

39. Katzen, A.L., Cann, G.M., and Blackburn, E.H. Sequence-Specific Fragmentation of Macronuclear DNA in a Holotrichous Ciliate. Cell, 24, 313-320 (1981).

40. Findly, R.C. and Gall, J.G. Organization of Ribosomal Genes in Paramecium tetraurelia. J. Cell Biol., 84, 547-559 (1980).

41. Engberg, J. The Ribosomal RNA Genes of Tetrahymena: Structure and Function. Eur. J. Cell Biol., 36, 133-151 (1985).

42. Din, N. and Engberg, J. Extra-Chromosomal Ribosomal RNA Genes in Tetrahymena: Structure and Evolution. J. Mol. Biol., 134, 555-574 (1979).

43. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. Cell, 31, 147-157 (1982).

44. Oka, K., Shiota, S., Nakai, S., Nishida, Y., and Okubo, S. Inverted Terminal Repeat Sequence in the Macronuclear DNA of Stylonychia pustulata. Gene, 10, 301-306 (1980).

45. Sakaki, Y., Nakamura, M., Nomiyama, H., and Takagi, Y. Cloning and Characterization of Ribosomal RNA Gene of Physarum polycephalum. J. Biochem., 89, 1693-1698 (1981).

46. Ferris, P.J. Nucleotide Sequence of the Central Non-transcribed Spacer Region of Physarum polycephalum rDNA. Gene, 39, 203-211 (1985).

47. Frankel, G., Cockburn, A.F., Kindle, K.L., and Firtel, R.A. Organization of the Ribosomal RNA Genes of Dictyostelium discoideum. Mapping of the Transcribed Region. J. Mol. Biol., 109, 539-558 (1977).

48. Ozaki, T., Hoshikawa, Y., Iida, Y., and Iwabuchi, M. Sequence Analysis of the Transcribed and 5' Non-transcribed Regions of the Ribosomal RNA Gene in Dictyostelium discoideum. Nucl. Acids Res., 12, 4171-4184 (1984).

49. Miller, O.L. and Bakken, A.M. Morphological Studies of Transcription. Acta. Endocr. Suppl., 168, 155-177 (1972).

50. Berger, S. and Schweiger, H.G. Ribosomal DNA in Different Members of a Family of Green Algae (Chlorophyta: Dasyclodaceae): an Electron Microscopal Study. Planta (Berl.), 127, 49-62 (1975).

51. Berger, S. and Schweiger, H.G. 80S Ribosomes in Acetabularia major: Redundancy of rRNA Cistrons. Protoplasma. 83, 41-50 (1975).

52. Berger, S. and Schweiger, H.G. An Apparent Lack of Non-Transcribed Spacers in the rDNA of a Green Alga. Mol. Gen. Genet., 139, 269-295 (1975).

53. Marco, Y. and Rochaix, T.D. Comparison of the Nuclear Ribosomal Limits of Five Chlamydomonas Species. Chromosoma (Berl). 81, 629-640 (1981).

54. Unnasch, T.R. and Wirth, D.F. The Avian Malaria Plasmodium lophurae Has a Small Number of Heterogeneous Ribosomal RNA Genes. Nucl. Acids Res., 23, 8443-8459 (1983).

55. Unnasch, T.R. and Wirth, D.F. The Cloned rRNA Genes of P. lophurae: A Novel rDNA Structure. Nucl. Acids Res., 23, 8461-8472 (1983).

56. Unnasch, T.R., McLafferty, M., and Wirth, D.F. The Primary Structure of the rRNA Insertions of Plasmodium lophurae. Mol. Biochem. Parasitol., 16, 149-161 (1985).

57. Dame, J.B. and MacCutchan, T.F. The Four Ribosomal DNA Units of the Malaria Parasite Plasmodium berghei. J. Biol. Chem., 258, 6984-6990 (1983).

58. Dame, J.B., Sullivan, M., and McCutchan, T.F. Two Major Sequence Classes of Ribosomal RNA Genes in Plasmodium berghei. Nucl. Acids Res., 12, 5943-5952 (1984).

59. Langsley, G., Hyde, J.E., Goman, M., and Scaife, J.G. Cloning and Characterization of the rRNA Genes From the Human Malaria Parasite Plasmodium falciparum. Nucl. Acids Res., 11, 8703-8717 (1983).

60. Boseley, P., Moss, T., Machler, M., Portmann, R , and Birnstiel, M. Sequence Organization of the Spacer DNA in a Ribosomal Gene Unit of Xenopus laevis. Cell, 17, 19-31 (1979).

61. Moss, T., Boseley, P.G., and Birnstiel, M.L. More Ribosomal Spacer Sequences From Xenopus laevis, Nucl. Acids Res., 8, 467-485 (1980).

62. Maden, B.E.H., Moss, M., and Salim, M. Nucleotide Sequence of an External Transcribed Spacer in Xenopus laevis rDNA: Sequences Flanking the 5' and 3'-Ends of 18S rRNA Are Non-complementary. Nucl. Acids Res., 10, 2382-2398 (1982).

63. Moss, T. Transcription of Cloned Xenopus laevis Ribosomal DNA Microinjected into Xenopus Oocytes and the Identification of an RNA Polymerase I Promoter. Cell, 30, 835-842 (1982).

64. Moss, T. A Transcriptional Function for the Repetitive Ribosomal Spacer in Xenopus laevis, Nature, 302, 223-228 (1983).

65. DeWinter, R.F.J. and Moss, T. Spacer Promoters are Essential for Efficient Enhancement of X. laevis Ribosomal Transcription. Cell, 44, 313-318 (1986).

66. DeWinter, R.F.J. and Moss, T. The Ribosomal Spacer in Xenopus laevis is Transcribed as Part of the Primary Ribosomal RNA. Nucl. Acids Res., 14, 6041-6051 (1986).

67. Firek, S., Read, C., Smith, D.R., and Moss, T. The Xenopus laevis Ribosomal Gene Terminator Contains Sequences That Both Enhance and Repress Ribosomal Transcription. Mol. Cell. Biol., 9, 3777-3784 (1989).

68. Firek, S., Read, C., Smith, D.R., and Moss, T. Point Mutation Analysis of the Xenopus laevis RNA Polymerase I Core Promoter. Nucl. Acids Res., 18, 105-109 (1989).

69. Bakken, A., Morgan, G., Sollner-Webb, B., Roan, J., Busby, S., and Reeder, R.H. Mapping of Transcription Initiation and Termination Signals on Xenopus laevis Ribosomal DNA. Proc. Natl. Acad. Sci. USA, 79, 56-60 (1982).

70. Labhart, P. and Reeder, R.H. Enhancer-Like Properties of the 60/81 bp Elements in the Ribosomal Gene Spacer of Xenopus laevis. Cell, 37, 285-289 (1984).

71. Reeder, R. H. Enhancers and Ribosomal Gene Spacers. Cell, 38, 349-351 (1984).

72. Dunaway, M. and Reeder, R.H. DNase I Footprinting Shows Three Protected Regions in the Promoter of the rRNA Genes of Xenopus laevis, Mol. Cell. Biol., 5, 313-319 (1985).

73. Labhart, P. and Reeder, R.H. Xenopus Ribosomal Gene Enhancers Function When Inserted Inside The Gene They Enhance. Nucl. Acids Res.,13, 8999-9009 (1985).

74. Labhart, P. and Reeder, R.H. Characterization of Three Sites of RNA 3'-End Formation in the Xenopus Ribosomal Gene Spacer. Cell, 45, 431-443 (1986).

75. McStay, B. and Reeder, R.H. A Termination Site of Xenopus RNA Polymerase I Also Acts as an Element of an Adjacent Promoter. Cell, 47, 913-920 (1986).

76. Labhart, P. and Reeder, R.H. Ribosomal Precursor 3'-End Formation Requires a Conserved Element Upstream of the Promoter. Cell, 50, 51-57 (1987).

77. Simeone, A., deFalco, A., Macino, G., and Boncinelli, E. Sequence Organization of the Ribosomal Spacer of D. melanogaster, Nucl. Acids Res., 10, 8263-8284 (1982).

78. Simeone, A., La Volpe, A., and Boncinelli, E. Nucleotide Sequence of a Complete Ribosomal Spacer of D. melanogaster. Nucl. Acids Res., 13, 1089-1101 (1985).

79. Coen, E.S. and Dover, G.A. Multiple Pol I Initiation Sequences in rDNA Spacers of Drosophila melanogaster. Nucl. Acids Res., 10, 7017-7026 (1982).

80. Tautz, D., Tautz, C., Webb, D., and Dover, G.A. Evolutionary Divergence of Promoters and Spacers in the rDNA Family of Four Drosophila Species. J. Mol. Biol., 195, 525-542 (1987).

81. Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., and Dover, G.A. Complete Sequences of the rRNA Genes of Drosophila melanogaster. Mol. Biol. Evol., 5, 366-376 (1988).

82. Mian, A. and Dover, G.A. Promoter Variation in the Ribosomal RNA Genes in Drosophila melanogaster Strains. Nucl. Acids Res., 18, 3795-3801 (1990).

83. Swanson, M.E., Yip, M., and Holland, M.J. Characterization of an RNA Polymerase I-Dependent Promoter Within the Spacer Region of Yeast Ribosomal Cistrons. J. Biol. Chem., 260, 9905-9915 (1985).

84. van der Sande, C.A.F.M., Kulkens, T., Kramer, A.B., de Wijs, I.J., van Heerikhuizen, H., Klootwijk, J., and Planta, R.J. Termination of Transcription by Yeast RNA Polymerase I. Nucl. Acids Res., 17, 9127-9147 (1989).

85. Musters, W., Knol, J., Maas, P., Dekker, A.F., van Heerikhuizen, H., and Planta, R.J. Linker Scanning of the Yeast RNA Polymerase I Promoter. Nucl. Acids Res., 17, 9661-9678 (1989).

86. Mestel, R., Yip, M., Holland, J.P., Wang, E., Kang, J., and Holland, M.J. Sequences Within the Spacer Region of Yeast rRNA Cistrons That Stimulate 35S rRNA Synthesis In Vivo Mediate RNA Polymerase I-Dependent Promoter and Terminator Activities. Mol. Cell. Biol., 9, 1243-1254 (1989).

87. Moss, T., *In* "Oxford Surveys on Eukaryotic Genes Vol. 2", (M. MacLean, ed) Oxford, Oxford University Press.

88. Niles, E.G., Sutiphong, J., and Haque, S. Structure of the Tetrahymena pyriformis rRNA Gene. Nucleotide Sequence of the Transcription Initiation Region. J. Biol. Chem., 256, 12849-12856 (1981).

89. Niles, E.G., Cunningham, K., and Jain, R. Structure of the Tetrahymena pyriformis rRNA Gene. Nucleotide Sequence of the Transcription Termination Region. J. Biol. Chem., 256, 12857-12860 (1981).

90. Grummt, I. Mapping of a Mouse Ribosomal DNA Promoter by In Vitro Transcription. Nucl. Acids Res., 9, 6093-6102 (1981).

91. Grummt, I. Nucleotide Sequence Requirements for Specific Initiation of Transcription by RNA Polymerase I. Proc. Natl. Acad. Sci. USA, 79, 6908-6911 (1982).

92. Wandelt, C. and Grummt, I. Formation of Stable Preinitiation Complexes is a Prerequisite for Ribosomal DNA Transcription in vitro. Nucl. Acids Res., 11, 3795-3809 (1983).

93. Skinner, J.A., Ohrlein, A., and Grummt, I. In Vitro Mutagenesis and Transcriptional Analysis of a Mouse Ribosomal Promoter Element. Proc. Natl. Acad. Sci. USA, 81, 2137-2141 (1984).

94. Grummt, I. and Skinner, J.A. Efficient Transcription of a Protein-Coding Gene From the RNA Polymerase I Promoter in Transfected Cells. Proc. Natl. Acad. Sci. USA, 82, 722-726 (1985).

95. Clos, J., Buttgereit, D., and Grummt, I. A Purified Transcription Factor (TIB-1B) Binds to Essential Sequences of the Mouse rDNA Promoter. Proc. Natl. Acad. Sci. USA, 83, 604-608 (1986).

96. Mishima, Y., Financsek, I., Kominami, R., and Muramatsu, M. Fractionation and Reconstitution of Factors Required for Accurate Transcription of Mammalian Ribosomal RNA Genes: Identification of a Species-Dependent Initiation Factor. Nucl. Acids Res., 10, 6659-6670 (1982).

97. Yamamoto, O., Takakusa, N., Mishima, Y., Kominami, R., and Muramatsu, M. Determination of the Promoter Region of Mouse Ribosomal RNA Gene by an In Vitro Transcription System. Proc. Natl. Acad. Sci. USA, 81, 299-303 (1984).

98. Kishimoto, T., Nagamine, M., Sasaki, T., Takakusa, N., Miwa, T., Kominami, R., and Muramatsu, M. Presence of a Limited Number of Essential Nucleotides in the Promoter Region of Mouse Ribosomal RNA Gene. Nucl. Acids Res., 13, 3515-3532 (1985).

99. Safrony, G., Tanaka, N., Kishimoto, T., Ishikawa, Y., Kato, H., Kominami, R., and Muramatsu, M. Structural Determinant of the Species-Specific Transcription of the Mouse rRNA Gene Promoter. Mol. Cell. Biol., 9, 349-353 (1989).

100. Grummt, I., Roth, E., and Paule, M.R. Ribosomal RNA Transcription In Vitro is Species Specific. Nature, 296, 173-174 (1982).

101. Miesfeld, R. and Arnheim, N. Species-Specific rDNA Transcription is Due to Promoter-Specific Binding Factors. Mol. Cell. Biol., 4, 221-227 (1984).

102. Kownin, P. Iida, C.T., Brown-Shimer, S., and Paule, M.R. The Ribosomal RNA Promoter of Acanthamoeba castellanii Determined by Transcription in a Cell-Free System. Nucl. Acids Res.,13, 6237-6248 (1985).

103. Bateman, E., Iida, C.T., Kownin, P., and Paule, M.R. Footprinting of Ribosomal RNA Genes by Transcription Initiation Factor and RNA Polymerase I. Proc. Natl. Acad. Sci. USA, 82, 8004-8008 (1985).

104. Sollner-Webb, B., Wilkinson, J.A.K., Roan, J., and Reeder, R.H Nested Control Regions Promote Xenopus Ribosomal RNA Synthesis by RNA Polymerase I. Cell, 35, 199-206 (1983).

105. Windle, J.J. and Sollner-Webb, B. Two Distant and Precisely Positioned Domains Promote Transcription of Xenopus laevis rRNA Genes: Analysis With Linker-Scanning Mutants. Mol. Cell. Biol., 6, 4585-4593 (1986).

106. Financsek, I., Mizumoto, K., Mishima, Y., and Muramatsu, M. Human Ribosomal RNA Gene: Nucleotide Sequence of the Transcription Initiation Region and Comparison of Three Mammalian Genes. Proc. Natl. Acad. Sci. USA, 79, 3092-3096 (1982).

107. Learned, R.M., Smale, S.T., Haltiner, M.M., and Tjian, R. Regulation of Human Ribosomal RNA Transcription. Proc. Natl. Acad. Sci. USA, 80, 3558-3562 (1983).

108. Learned, R.M., Cordes, S., and Tjian, R. Purification and Characterization of a Transcription Factor That Confers Promoter Specificity to Human RNA Polymerase I. Mol.Cell. Biol., 5, 1358-1369 (1985).

109. Learned, R.M., Learned, T.K., Haltiner, M.M., and Tjian, R.T. Human rRNA Transcription is Modulated by the Coordinate Binding of Two Factors to an Upstream Control Element. Cell, 45, 847-857 (1986).

110. Bell, S.P., Learned, R.M., Jantzen, H.-M., and Tjian, R. Functional Cooperativity Between Transcription Factors UBF1 and SL1 Mediates Human Ribosomal RNA Synthesis. Science, 241, 1192-1197 (1988).

111. Kohorn, B.D. and Rae, P.M.M. Nontranscibed Spacer Sequences Promote In Vitro Transcription of Drosophilia Ribosomal DNA. Nucl. Acids Res., 10, 6879-6886 (1982).

112. Kohorn, B.D. and Rae, P.M.M. A Component of Drosophila RNA Polymerase I Promoter Lies Within the rRNA Transcription Unit. Nature, 304, 179-181 (1983).

113. Miller, J.R. Transcription of the 'Non-Transcribed' Spacer of Drosophila melanogaster rDNA. Nucl. Acids Res., 11, 11-19 (1983).

114. Grimaldi, G. and DiNocera, P.P. Transient Expression of Drosophila melanogaster rDNA Promoter In Cultured Drosophilia Cells. Nucl. Acids Res., 14, 6417-6432 (1986).

115. Grimaldi, G. and DiNocera, P.P. Multiple Repeated Units in Drosophila melanogaster Ribosomal DNA Spacer Stimulate rRNA Precursor Transcription. Proc. Natl. Acad. Sci. USA, 85, 5502-5506 (1988).

116. Mandel, R.K. and Dawid, I.B. The Nucleotide Sequence of the Transcription Termination Site of Ribosomal RNA in Drosophila melanogaster. Nucl. Acids Res., 9, 1801-1811 (1981).

117. Yang-Yen, H.F., Subrahmanyam, C.S., Cassidy, B., and Rothblum, L.I. Characterization of Rat Ribosomal DNA II. J. Mol. Biol., 184, 389-398 (1985).

118. Haglund, R.E. and Rothblum, L.I. Isolation, Fractionation and Reconstitution of a Nuclear Extract Capable of Transcribing Ribosomal DNA. Mol. Cell. Biochem., 73, 11-20 (1987).

119. Cassidy, B.G., Haglund, R., and Rothblum, L.I. Regions Upstream From the Core Promoter of the Rat Ribosomal Gene Are Required for the Formation of a Stable Transcription Initiation Complex by RNA Polymerase I In Vitro. Biochim. Biophys. Acta, 909, 133-144 (1987).

120. Cassidy, B.G., Yang-Yen, H.-F., and Rothblum, L.I. Additional RNA Polymerase I Initiation Site Within the Non-Transcribed Spacer Region of the Rat rRNA Gene. Mol. Cell. Biol., 7, 2388-2396 (1987).

121. Yang-Yen, H-F. and Rothblum, L.I. Purification and Characterization of a High-Mobility-Group-Like DNA Binding Protein That Stimulates rRNA Synthesis In Vitro. Mol. Cell Biol., 8, 3406-3414 (1988).

122. Smith, S.D., Oriahi, E., Yang-Yen, H.-F., Xie, W., Chen, C., and Rothblum, L.I. Interaction of RNA Polymerase I Transcription Factors With a Promoter in the Non-Transcribed Spacer of Rat Ribosomal DNA. Nucl. Acids Res., 7, 1677-1685 (1990).

123. Dixit, A., Garg, L.C., Chao, W., and Jacob, S.T. An Enhancer Element in the Far Upstream Spacer Region of Rat Ribosomal RNA Gene. J. Biol. Chem., 262, 11616-11622 (1987).

124. Kermekchiev, M.B. and Grummt, I. Natural Point Mutations Within Rat rDNA Transcription Terminator Elements Reveal The Functional Importance of Single Bases for Factor Binding and Termination. Nucl. Acids Res., 10, 4131-4143 (1987).

125. Labhart, P. and Reeder, R.H. Heat Shock Stabilizes Highly Unstable Transcripts of the Xenopus Ribosomal Gene Spacer. Proc. Natl. Acad. Sci. USA, 84,56-60 (1987).

126. Labhart, P. and Reeder, R.H. A 12-Base-Pair Sequence is an Essential Element of the Ribosomal Gene Terminator in Xenopus laevis. Mol. Cell. Biol., 7, 1900-1905 (1987).

127. Kominami, R., Mishima, Y., Urano, Y., Sakai, M., and Muramatsu, M. Cloning and Determination of the Transcription Termination Site of Ribosomal RNA Gene of the Mouse. Nucl. Acids Res., 10, 1963-1979 (1982).

128. Grummt, I., Rosenbauer, H., Niedermeyer, I., Maier, U., and Ohrlein, A. A Repeated 18-bp Sequence Motif in the Mouse rDNA Spacer Mediates Binding of a Nuclear Factor and Transcription Termination. Cell, 45, 837-846 (1986).

129. Grummt, I., Kuhn, A., Bartsch, I., and Rosenbauer, H. A Transcription Terminator Located Upstream of the Mouse rDNA Initiation Site Affects rRNA Synthesis. Cell, 47, 901-911 (1986).

130. Henderson, S. and Sollner-Webb, B. A Transcriptional Terminator is a Novel Element of the Promoter of the Mouse Ribosomal Gene. Cell, 47, 891-900 (1986).

131. Miwa, T., Kominami, R., Yoshikura, H., Sudo, K., and Muramatsu, M. Transcription Termination and RNA Processing in the 3'-End Spacer of Mouse Ribosomal RNA Genes. Nucl. Acids Res., 15, 2043-2058 (1987).

132. Sollner-Webb, B., Windle, J., Henderson, S., Tower, J., Culotta, V., Kass, S., and Craig, N. The Action of the Sequences and Factors That Direct Ribosomal RNA Transcription and Processing of the Primary rRNA Transcript, *In* "Transcriptional Control Mechanisms", Alan R. Liss, Inc., pp.113-122 (1987).

133. Tautz, D. and Dover, G.A. Transcription of the Tandem Array of Ribosomal DNA in Drosophila melanogaster Does Not Terminate at any Fixed Point.The EMBO Journal, 5, 1267-1273 (1986).

134. Mestel, R., Yip, M., Holland, J.P., Wang, E., Kang. J., and Holland, M.J. Sequences Within the Spacer Region of Yeast rRNA Cistrons That Stimulate 35S rRNA Synthesis In Vivo Mediate RNA Polymerase I-Dependent Promoter and Terminator Activities. Mol. Cell. Biol., 9, 1243-1254 (1989).

135. Brown, D.D., Wensink, P.C. and Jorden, E. A Comparison of the Ribosomal DNA's of Xenopus laevis and Xenopus mulleri: the Evolution of Tandem Genes. J. Mol. Biol., 63, 57-73 (1972).

136. Dover, G. A Molecular Drive Through Evolution. Bioscience, 32, 526-533 (1982).

137. Doerfler, W. DNA Methylation and Gene Activity. Ann. Rev. Biochem., 52, 93-124 (1983).

138. Adams, R.L.P. and Burdon, R.H. DNA Methylation in Eukaryotes. CRC Crit. Rev. Biochem. 13, 349-384 (1982).

139. Razin, A. and Cedar, H. DNA Methylation in Eukaryotic Cells. Int. Rev. Cytol., 92, 159-185 (1984).

140. Razin, A. and Szyf, M. DNA Methylation Patterns: Formation and Function. Biochim. Biophys. Acta, 782, 331-342 (1984).

141. Yisraeli, J. and Szyf, M. DNA Methylation, Biochemistry and Biological Significance (A. Razin, H. Cedar, and A.D. Riggs, eds), Springer Verlag Inc. (1984).

142. Bird, A., Taggart, M., and Macleod, D. Loss of rDNA Methylation Accompanies the Onset of Ribosomal Gene Activity in Early Development of X. laevis. Cell, 26, 381-390 (1981).

143. Weintraub, H. and Groudine, M. Chromosomal Subunits in Active Genes Have an Altered Conformation. Science, 193, 848-856 (1976).

144. Gray, M.W. Unusual Pattern of Ribonucleic Acid Components in the Ribosome of Crithidia fasciculata, a Trypansomatid Protozoan. Mol. Cell Biol., 1, 347-357 (1981).

145. Cordingley, J.S. and Turner, M.J. 6.5S RNA: Preliminary Characterization of Unusual Small RNAs in Trypanosoma brucei. Mol. Biochem. Parasitol., 1, 91-96 (1980).

146. Spencer, R. and Cross, G.A. Lability of RNA From the Large Cytoplasmic Ribosomal Subunit of the Protozoan Crithidia oncopelti. J. Gen. Microbiol., 93, 82-88 (1978).

147. Castro, C., Hernandez, R., and Castaneda, M. Trypanosoma cruzi Ribosomal RNA: Internal Break in the Large-Molecular-Mass Species and Number of Genes. Mol. Biochem. Parasitol., 2, 219-233 (1981).

148. Maclean, F.I. and Amiro, E.R. The Effect of Solute Concentration on the Shape of the Trypanosomatid Flagellate Crithidia fasciculata. Can. J. Microbiol., 19, 878-880 (1973).

149. Gray, M.W. The Ribosomal RNA of the Trypanosomatid Protozoan Crithidia fasciculata: Physical Characteristics and Methylated Sequences. Can. J. Biochem., 57, 914-926 (1979).

150. Maxam, A.M. and Gilbert, W. A New Method For Sequencing DNA. Proc. Natl. Acad. Sci. USA, 74, 560-564 (1977).

151. Donis-Keller, H., Maxam, A.M., and Gilbert, W. Mapping Adenines, Guanines, and Pyrimidines in RNA. Nucl. Acids Res., 4, 2527-2538 (1977).

152. Walseth, T.F. and Johnson, R.A. The Enzymatic Preparation of [ $\alpha$-$^{32}$P] Nucleoside Triphospates, Cyclic [$^{32}$P] AMP, and Cyclic [$^{32}$P] GMP. Biochim. Biophys. Acta., 526, 11-31 (1979).

153. Southern, E.M. Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis. J. Mol. Biol., 98, 503-517 (1975).

154. Bonen, L. and Gray, M.W. Organization and Expression of the Mitochondrial Genome of Plants I. The Genes for Wheat Mitochondrial Ribosomal and Transfer RNA: Evidence for an Unusual Arrangement. Nucl. Acids Res., 8, 319-335 (1981).

155. Williamson, S.E. and Doolittle, W.F. Genes for tRNA$^{Ile}$ and tRNA$^{Ala}$ in the Spacer Between the 16S and 23S rRNA Genes of a Blue-Green Alga: Strong Homology to Chloroplast tRNA Genes and tRNA Genes of the E. coli rrnD Gene Cluster. Nucl. Acids Res., 11, 225-235 (1983).

156. Messing, J., Crea, R., and Seeburg, P.H. A System For Shotgun DNA Sequencing. Nucl. Acids Res., 9, 309-321 (1981).

157. Hanahan, D. Studies on Transformation of Escherichia coli With Plasmids. J. Mol. Biol., 166, 557-580 (1983).

158. Grunstein, M. and Hogness, D.S. Colony Hybridization: A Method for the Isolation of Cloned DNAs That Contain a Specific Gene. Proc. Natl. Acad. Sci. USA, 72, 3961-3965 (1975).

159. Denhardt, D.T. A Membrane-Filter Technique for the Detection of Complementary DNA. Biochem. Biophys. Res. Commun., 23, 641-646 (1966).

160. Holmes, D.S. and Quigley, M. A Rapid Boiling Method for the Preparation of Bacterial Plasmids. Anal. Biochem., 114, 193-197 (1981).

161. Sehleit, R.F. and Wensink, P.C.. *In* "Practical Methods in Molecular Biology", pp. 106-107, Springer-Verlag, NY (1981).

162. Spencer, D.F., Schnare, M.N., and Gray, M.W. Pronounced Structural Similarities Between the Small Subunit Ribosomal RNA Genes of Wheat Mitochondria and Escherichia coli. Proc. Natl. Acad. Sci. USA, 81, 493-497 (1984).

163. Schnare, M.N. and Gray, M.W. Nucleotide Sequence of an Exceptionally Long 5.8S Ribosomal RNA from Crithidia fasciculata. Nucl. Acids Res., 10, 2085-2092 (1982).

164. Hughes, D.E., Schneider, C.A., and Simpson, L. Isolation and Characterization of Drug Resistant Mutants of Crithidia fasciculata. J. Parasitol., 68, 642-649 (1982).

165. Schnare, M.N. and Gray, M.W. 3'-Terminal Nucleotide Sequence of Crithidia fasciculata Small Ribosomal Subunit RNA. FEBS Lett., 128, 298-304 (1981).

166. Schnare, M.N., Spencer, D.F., and Gray, M.W. Primary Structures of Four Novel Small Ribosomal RNAs from Crithidia fasciculata. Can. J. Biochem. Cell. Biol., 61, 38-45 (1983).

167. Lean, W., Fouts, D.L., and Manning, J. Sequence Arrangement of the 16S and 26S rRNA Genes in the Pathogenic Haemoflagellate, Leishmania donovanii. Nucl. Acids Res., 5, 491-504 (1978).

168. White, T.C., Rudenko, G., and Borst, P. Three Small RNAs Within the 10 kb Trypanosome rRNA Transcription Unit are Analogous to Domain VII of Other Eukaryotic 28S rRNAs. Nucl. Acids Res., 14, 9471-9489 (1986).

169. Hernandez, R., Nava, G., and Castaneda, M. Small-Size Ribosomal RNA Species in Trypanosoma cruzi. Mol. Biochem. Parasitol., 8, 297-304 (1983).

170. Hernandez, R. and Castaneda, M. An Endonuclease Restriction Analysis of the Ribosomal RNA Genes of Trypanosoma cruzi. Mol. Biochem. Parasitol., 8, 305-315 (1983).

171. Schnare, M.N. and Gray, M.W. Sixteen Discrete RNA Components in the Cytoplasmic Ribosome of Euglena gracilis. J. Mol. Biol., 215, 73-83 (1990).

172. Schnare, M.N., Cook, J.R., and Gray, M.W. Fourteen Internal Transcribed Spacers in the Circular Ribosomal DNA of Euglena gracilis. J. Mol. Biol., 215, 85-91 (1990).

173. Sogin, M.L., Elwood, H.J., and Gunderson, J.H. Evolutionary Diversity of Eukaryotic Small Subunit rRNA Genes. Proc. Natl. Acad. Sci. USA, 83, 1383-1487 (1986).

174. Cordingley, J.S. Nucleotide Sequence of the 5S Ribosomal RNA Gene Repeat of Trypanosoma brucei. Mol. Biochem. Parasitol., 17, 321-330 (1985).

175. Fedoroff, N.V. and Brown, D.D. The Nucleotide Sequence of Oocyte 5S DNA in Xenopus laevis I: The AT-Rich Spacers. Cell, 13, 701-716 (1978).

176. Miller, J.R., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V., and Brown, D.D. The Nucleotide Sequence of the Oocyte 5S DNA in Xenopus laevis II: The GC-Rich Region. Cell, 13, 717-725 (1978).

177. Korn, L.J. and Brown, D.D. Nucleotide Sequence of Xenopus borealis Oocyte 5S DNA: Comparison of Sequences That Flank Several Related Eukaryotic Genes. Cell, 15, 1145-1156 (1978).

178. Spencer, D.F., Collings, J.C., Schnare, M.N., and Gray, M.W. Multiple Spacer Sequences in the Nuclear Large Subunit Ribosomal RNA Gene of Crithidia fasciculata. EMBO J., 6, 1063-1071 (1987).

179. Morgan, G.T. and McMahon, H. Alterations in Cloned Xenopus Ribosomal Spacers Generated by High-Frequency Plasmid Recombination. Gene, 49, 389-394 (1986).

180. Larson, D.D., Blackburn, E.H., Yaeger, P.C., and Orias, E. Control of rDNA Replication in Tetrahymena Involves a Cis-Acting Upstream Repeat of a Promoter Element. Cell, 47, 229-240 (1986).

181. Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D., and Warner, J.R. (eds) The Ribosome: Structure, Function, and Evolution. American Society of Microbiology, Washington, DC (1990).

182. Gray, M.W. The Bacterial Ancestry of Plastids and Mitochondria. Bioscience 33.693-699 (1983).

183. Gray, M. W. Organelle Origins and Ribosomal RNA. Biochem. Cell Biol., 66, 325-348 (1988).

184. Turmel, M., Boulanger, J., Schnare, M.N., Gray, M.W., and Lemieux, C. Six Group I Introns and Three Internal Transcibed Spacers in the Chloroplast Large Subunit Ribosomal RNA Gene of the Green Alga Chlamydomonas eugametos. J. Mol. Biol., 218, 293-311 (1991).

185. Woese, C. R., Kandler, O., and Wheelis, M.L. Towards a Natural System of Organisms: Proposal For the Domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA, 87, 4576-4579 (1990).

186. Iwabe, N., Kuma, K-I., Hasegawa, M., Osawa, S., and Miyata, T. Evolutionary Relationship of Archaebacteria, Eubacteria, and Eukaryotes Inferred From Phylogenetic Trees of Duplicated Genes. Proc. Natl. Acad. Sci. USA, 86, 9355-9359 (1989).

187. Culotta, V. C., Wilkinson, J.K., and Sollner-Webb, B. Mouse and Frog Violate the Paradigm of Species-specific Transcription of Ribosomal RNA Genes. Proc. Nalt. Acad.Sci.USA 84,7498-7502 (1987).

188. Hillis, D. M., Moritz, C., Porter, C.A., and Baker, R.J. Evidence For Biased Gene Conversion in Concerted Evolution of Ribosomal DNA. Science, 251 308-310 (1991).

189. MacKay, R.M., Gray, M.W., and Doolittle, W.F. Nucleotide Sequence of Crithidia fasciculata Cytosol 5S Ribosomal Ribonucleic Acid. Nucl. Acids Res., 8, 4911-4917 (1980).

190. Van der Ploeg, L.H.T., Cornelissen, A.W.C.A., Barry, J.D., and Borst, P. Chromosomes of Kinetoplastida. EMBO J., 3, 3109-3115 (1984).

191. Schnare, M.N., Collings, J.C., and Gray, M.W. Structure and Evolution of the Small Subunit Ribosomal RNA Gene of Crithidia fasciculata. Curr. Genet. 10, 405-410 (1986).

192. Clark, C.G. and Cross, G.A.M. rRNA Genes of Naegleria gruberi Are Carried Exclusively on a 14-Kilobase-Pair Plasmid. Mol. Cell. Biol., 7, 3027-3031 (1987).

193. Huber, M., Koller, B., Gitler, C., Mirelman, D., Revel, M., Rozenblatt, S., and Garfinkel, L. Entamoeba histolytica Ribosomal RNA Genes Are Carried On Palindromic Circular Molecules. Mol. Biochem. Parasitol., 32, 285-296 (1989).

194. Lee, J.J., Hutner, S.H., and Bovee, E.C. (eds) *In* "An Illustrated Guide To The Protozoa". Society of Protozoologists, Lawrence, Kansas, pp.144-155 (1985).

195. Grondal, E.J.M., Evers, R., and Cornelissen, A.W.C.A. Identification and Sequence Analysis of the Ribosomal DNA Promoter Region of Crithidia fasciculata. Nucl. Acids Res., 18, 1333-1338 (1990).

196. Boyes, J. and Bird, A. DNA Methylation Inhibits Transcription Indirectly via a Methyl-CpG Binding Protein. Cell, 64, 1123-1134 (1991).

197. Chan,Y.-L., Endo, Y., and Wool, I.G. The Sequence of the Nucleotides at the α-Sarcin Cleavage Site in Rat 28S Ribosomal Ribonucleic Acid. J. Biol. Chem., 258, 12768-12770 (1983).

198. Dubin, D.T., HsuChen, C.C., Timko, K.D., Azzolina, T.M., Prince, D.L., and Ranzini, J.L. *In* "Mitochondrial Genes." (Slonimski, P., Borst, P., and Attardi, G., eds), pp. 89-98, Cold Spring Harbor Laboratory, NY, (1982).

199. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R., and Woese, C.R. Secondary Structure Model for the 23S Ribosomal RNA. Nucl. Acids Res., 9, 6167-6189 (1981).

200. Michot, B., Hassouna, N. and Bachellerie, J.-P. Secondary Structure of Mouse 28S rRNA and General Model for the Folding of the Large rRNA in Eukaryotes. Nucl. Acids Res., 12, 4259-4279 (1984).

201. Challoner, P.B., Amin, A.A., Pearlman, R.E., and Blackburn, E.H. Conserved Arrangement of Repeated DNA Sequences in Nontranscribed Spacers of Ciliate Ribosomal RNA Genes: Evidence for Molecular Coevolution. Nucl. Acids Res., 13, 2661-2680 (1985).

202. Higashinakagawa, T., Saiga, H., Shintani, N., Narushima-Iio, M., and Mita, T. Localization of Putative Transcription Initiation Site on the Cloned rDNA Fragment of Tetrahymena pyriformis. Nucl. Acids Res., 9, 5905-5916 (1981).

203. Saiga, H., Mizumoto, K., Matsui, T., and Higashinakagawa, T. Determination of the Transcription Initiation Site of Tetrahymena pyriformis rDNA using in vitro Capping of 35S pre-rRNA. Nucl. Acids Res., 10, 4223-4236 (1982).

204. Engberg, J., Din, N. Saiga, H., and Higashinakagawa, T. Nucleotide Sequence of the 5'-Terminal Coding Region for pre-rRNA and Mature 17S rRNA in Tetrahymena thermophila rDNA. Nucl. Acids Res., 12, 959-972 (1984).

205. Matsuura, T., Matsui, T., Saiga, H., Mita, T., and Higashinakagawa, T. Faithful Initiation of the in vitro Transcription of a Cloned rDNA from Tetrahymena pyriformis. Gene, 49, 225-233 (1986).

206. Engberg, J. and Nielsen, H. Complete Sequence of the Extra-Chromosomal rDNA Molecule From the Ciliate Tetrahymena thermophila strain B1868VII. Nucl. Acids Res., 18, 6915-6919 (1990).

207. Judson, H.F. The Eighth Day of Creation: Makers of the Revolution in Biology. Simon and Schuster, NY, (1980).

208. Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A. and Woese, C.R. Ribosomal RNA Sequence Suggests Microsporidia are Extremely Ancient Eukaryotes. Nature, 326, 411-414 (1987).

209. Sogin, M.L., Gunderson, J.H, Elwood, H.J., Alonso, R.A. and Peattie, D.A. Phylogenetic Meaning of the Kingdom Concept: An Unusual Ribosomal RNA from Giardia lamblia. Science, 243, 75-77 (1989).

210. Brown, J.W., Daniels, C.J. and Reeve, J. N. Gene Structure, Organization, and Expression in Archaebacteria. CRC Crit. Rev. Microbiol., 16, 287-338 (1989).