To the Graduate Council:

I am submitting herewith a dissertation written by Arthur E. Roberson entitled "The 5S Ribosomal RNA Gene of Euplotes." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Donald E. Olins, Major Professor

We have read this dissertation
and recommend its acceptance:

Salil K. Niyogi

Kai-Lin Lee

Ada W. Olins

Gerald P. Bunick

K Bruce Jacobson

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

THE 5S RIBOSOMAL RNA GENE OF EUPLOTES

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Arthur E. Roberson

March 1988

DEDICATION

This work is dedicated to God.

ACKNOWLEDGMENTS

The author wishes to thank the following people and organizations for their assistance in the completion of this project:

Lisa Balogh
Jean Barwick
The Biomedical School Staff
Gerard Bunick
Lucia Cacheiro
Carmen Cadilla
James Finch
Richard Fowler
Loren Hauser
Elke Helftenbein
Adria Herrman
K. Bruce Jacobson
Frank Larimer
Kai Lin Lee
Richard Mural
The National Institutes of Health
The Oak Ridge National Laboratory,
Biology Division
The University of Tennessee--
Oak Ridge Graduate School
of Biomedical Sciences
Kave Nikbakht
Salil Niyogi
Ada Olins
Donald Olins
David Pederson
David Prescott
Twilah Roberson
Charles Stiles
Greta Stiles
Bruce Suttle
Marshal Swanton
Ray Thompson
Thelma Thompson
Michael Tindal

ABSTRACT

A library of Euplotes eurystomus macronuclear DNA was constructed in the pUC plasmid vector and cloned in E. coli. Screening of the library with the 5S ribosomal RNA gene of Tetrahymena thermophila resulted in 19 apparently identical isolates of the Euplotes 5S gene. The gene was subcloned into bacteriophage M13 and its sequence was determined by the Sanger method. The Euplotes macronuclear 5S rRNA gene is 930 bp in length, is flanked by telomeres, and contains a single 5S rRNA coding region. This gene serves as the template for a single transcript of 120 nucleotides in vivo and in vitro; this transcript is generated by RNA polymerase III. The 5S gene is amplified in the Euplotes macronucleus approximately 120-fold compared to a gene of typical abundance, amounting to about 960,000 copies per macronuclear genome. The Euplotes macronuclear 5S rRNA gene can be enriched as a native, intact chromatin structure by sucrose gradient centrifugation; in the peak fraction about 1 in 30 chromatin molecules harbors a 5S gene.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Opening Remarks

The primary interest of the Olins lab is the relationship of chromatin structure to the various physiologic states of genes. We are interested in the chromatin structure of genes which are actively being transcribed, the chromatin structure of genes which are repressed, and the structure of chromatin as it is being replicated. We are interested in how gene activity can be affected by modulation of chromatin architecture. Our goal is to attain both a descriptive and a mechanistic explanation of the role of chromatin structure in the storage and utilization of genomic information. Why study chromatin structure? The molecular mechanism of regulation of gene expression is clearly the central problem of biology today. In the following discussion it will be demonstrated that the expression of genomic information is regulated by the structure of chromatin as well as by the DNA sequence itself. The diversity and complexity of chromatin forms causes us to believe that the best approach to this problem is to investigate the chromatin structure of specific genes. We have chosen the 5S ribosomal RNA gene of Euplotes eurystomus as our experimental system.

The Molecular Biology of the 5S rRNA Gene

We must be cognizant of the molecular mechanisms of the differential regulation of gene expression if fundamental biological phenomena

such as development and differentiation are to be understood. Elucidation of the interaction of proteins with DNA to modulate the utilization of genomic information is paramount to this end. The 5S ribosomal RNA gene of the African frog Xenopus has proven to be an excellent model system with which to address this problem since the Xenopus genome contains two types of 5S rRNA genes which are differentially regulated during development (1-4). The diploid Xenopus nucleus contains about 40,000 oocyte-type 5S rRNA genes and about 800 somatic-type 5S rRNA genes (1,5). These genes are arranged in many clusters of tandemly repeating units and are found near the telomeres of most Xenopus chromosomes (5,6). A cluster can contain from hundreds to thousands of repeat units, in each of which typically resides a single 120 bp 5S rRNA gene plus a nontranscribed spacer that is from two to six times the length of the coding region (7). In E. coli the genes for 5S, 18S, and 28S ribosomal RNA occur as a single transcription unit which is processed to yield the mature RNAs. In yeast all three rRNA genes are found within the same repeating unit but the 5S rRNA gene is transcribed separately. In eukaryotes the 18S and 28S rRNA genes are transcribed by RNA polymerase I while the 5S rRNA gene is transcribed by RNA polymerase III. In multicellular eukaryotes the 5S rRNA genes are clustered in their own separate multigene families apart from the 18S and 28S genes (7).

Oocytes have the ability to synthesize ribosomes at 1,000 times the rate of somatic cells. Oocytes need to accumulate large amounts of ribosomes (or ribosomal precursors) to allow for very rapid growth

following fertilization. To accomplish this, in developing oocytes of amphibians and fish the genes for 18S and 28S rRNA (these genes are called "rDNA") are amplified 1000-fold on extrachromosomal fragments. These genes are transcribed during oogenesis and are discarded at meiosis. The 5S rRNA genes (called "5S DNA") are not amplified. Instead, to enable oocytes to synthesize 5S rRNA at a rate much faster than that needed by somatic cells, the genome contains a large family of 5S rRNA genes whose transcription is specific to the oocyte. These are called oocyte-type 5S rRNA genes. While oocytes synthesize both oocyte-type and somatic-type 5S rRNA, somatic cells synthesize almost exclusively somatic-type 5S rRNA (1,8-10). To summarize, the large number of 5S rRNA gene copies and the amplification of rDNA allow oocytes to synthesize ribosomal components very quickly, and the two types of 5S rRNA genes (oocyte and somatic) provide a mechanism whereby most of these genes can be repressed following oogenesis. It is interesting to note that different means are used by the oocyte for accumulating vast amounts of 5S versus 18S and 28S rRNAs (1,7).

The developmental pattern of expression of 5S rRNA genes is precisely regulated (1-4). During the first seven hours following fertilization, the Xenopus egg undergoes twelve cleavages and thus develops from one cell to a 4000-cell blastula (or stage 9 embryo). This period is characterized by rapid DNA synthesis and cell cleavage; little or no RNA or protein synthesis occurs at this time. Materials stored in the oocyte are thus diluted by partitioning into the dividing cells. (As a

matter of interest, asymmetric partitioning of cytoplasmic components during this period of rapid cleavage is known to be a central causal mechanism in determination of various cell types as a single egg cell develops the body pattern of a complex multicellular organism.) An event called the "mid-blastula transition" (MBT) occurs at about the 4000 cell stage and is characterized by the onset of RNA synthesis. This RNA synthesis is relatively resistant to α-amanitin, indicating that it is mediated by RNA polymerase I or III (2,4,11). Nascent 5S rRNA can be detected at this time. Thus during oogenesis oocytes synthesize and accumulate in their cytoplasm vast amounts of 5S rRNA (12). Oocytes synthesize both oocyte-type and somatic-type 5S rRNA but accumulate mainly oocyte-type because the oocyte-type genes are in vast excess over somatic-type 5S rRNA genes. In the mature egg, and in early embryogenesis, RNA and protein syntheses are generally repressed. When 5S rRNA synthesis is resumed following the MBT, nearly an equal mixture of oocyte-type and somatic-type 5S rRNAs are transcribed from both the maternal and paternal genes (2). This is the only time somatic cells are seen to synthesize appreciable amounts of oocyte-type 5S rRNA. Since there is a 50:1 ratio of oocyte-type to somatic-type 5S DNA, and since the two types of 5S rRNA are synthesized in equal abundance at MBT, most of the oocyte-type 5S genes must be inactive at this time. This inactivation of the oocyte-type genes was demonstrated to be at the level of transcriptional regulation by in vitro transcription of stage 9 embryo chromatin (2). Three cell divisions later the animal has completed gastrulation and nascent 5S

rRNA is approximately 90% somatic-type (2,4). In somatic cells of the adult, more than 95% of the 5S rRNA is somatic-type, representing a 1000-fold preference of somatic-type over oocyte-type genes on a per-gene basis. What is the molecular mechanism by which these genes are differentially regulated during development?

Posing the question above would be completely impossible without the powerful, new techniques of molecular biology. One such technique which has proven invaluable in the analysis of the regulation of gene expression is the development of cell-free systems capable of accurate and specific transcription in vitro. In 1978 Birkenmeier et al. (13) reported that a soluble, cell-free supernatant of Xenopus oocyte nuclei was able to accurately transcribe cloned 5S genes. The proportion of nascent RNA which is 5S RNA depends on the concentration of DNA in the reaction, the ionic strength, and the $MgCl_2$ concentration. Under optimal conditions, about 40% of the newly made RNA is 5S RNA. Inhibition of transcription by intermediate concentrations of α-amanitin indicates that more than 90% of the RNA is synthesized by RNA polymerase III. In 1979 Weil et al. (14) discovered extracts of human, murine, and amphibian kidney cells which could mediate accurate transcription in vitro dependent on addition of purified RNA polymerase III and exogenous DNA templates. In 1982 Shastry and coworkers (15) reported that cell-free extracts from Xenopus oocytes (ovarian tissue), unfertilized eggs, embryos, and cultured kidney cells can direct accurate transcription of cloned 5S rRNA genes. It is worthy of notice that purified RNA polymerase III cannot accurately transcribe purified DNA

templates, but it can accurately transcribe chromatin templates with which it is presented (16). Therefore additional protein components besides the polymerase, which are present in chromatin and in the cell-free extracts, must be required. Fractionation of these extracts by classical ion-exchange chromatography revealed that multiple factors are required for accurate transcription by RNA polymerase III (15,17). Two chromatographically distinct fractions plus RNA polymerase III (RNA pol III) are required for transcription of tRNA genes. Transcription of 5S rRNA genes requires RNA pol III, the same two fractions mentioned above, plus a third fraction. The fractions from different tissue extracts are interchangeable, indicating that these factors are general and not tissue-specific (15). In 1984 it was observed that the transcription factors from different organisms (such as frog and human) can work together, indicating that the transcriptional mechanism has been well conserved (18). Using these in vitro transcription assays the DNA sequences required for control of accurate transcription were delineated. A series of deletion mutants approaching the 5S coding region from both the 5' and 3' ends were constructed and tested for their ability to serve as templates for transcription of unit length (120 nucleotides) 5S rRNA. It was thus demonstrated that a sequence within the coding region from position +50 to position +83 (with respect to the transcription start site) is required for accurate initiation of transcription (19,20). This sequence is referred to as the internal promoter or the internal control region (ICR). A 5S gene ICR, when subcloned by itself in a plasmid without flanking 5S DNA, can

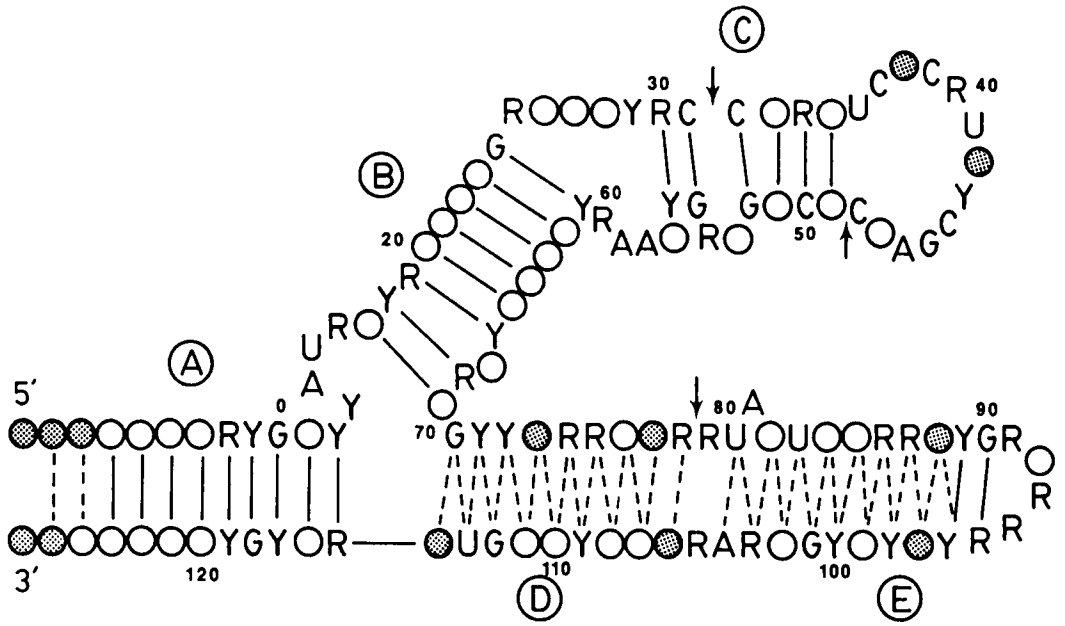
accurately direct initiation of transcription at a position 50 bases 5' to the 5' border of the ICR (i.e., within the plasmid DNA). In 1985 Bogenhagen constructed linker substitution mutants within the ICR and found that the sequence in the center of the ICR could be replaced with foreign DNA resulting in a mutant promoter which was still functional (21). The DNA sequences at the borders of the ICR and the spacing between them could not be altered. Although such a two-domain (A-box and B-box) model is well established for tRNA gene ICRs, it is less well accepted for 5S genes (46). Accurate termination of transcription of all class III genes (i.e., those genes transcribed by RNA pol III) is mediated by a consensus sequence at the termination site. This signal has the structure of a block of four or five T's nested within two pairs of GC dinucleotides, such as: GCTTTTGC (1). This sequence occurs on the DNA strand which is homologous to the RNA and transcription terminates after the first two T's, thus most 5S rRNAs end with ...UU-3'. In summary, initiation of transcription of 5S rRNA genes (and class III genes in general) is mediated by a control region within the coding sequence and termination is controlled by a terminator sequence recognized by RNA polymerase III. The primary transcript of 5S genes is identical in length to the mature RNA, that is, 5S rRNA is not processed (Fig. 1).

A 38,000 D protein, TFIIIA, which specifically binds to the internal control region of 5S genes thereby positively regulating their transcription by RNA polymerase III, was purified from Xenopus ovaries by ion exchange chromatography (23,24). Binding of this

Figure 1: Consensus sequence and structure of 5S rRNA and its gene.

Top: Consensus sequence and secondary structure of 5S rRNA. Nucleotides conserved at ≧80% frequency are indicated (R = purine, Y = pyrimidine). Group-specific deletions are marked by shaded circles, group-specific insertions by arrows. Potential base pairs A/U, G/C, and G/U conserved at ≧95% frequency are shown in straight lines, alternative base pair interactions in broken lines [from Kuntzel, H., Piechulla, B., and Hahn, U. (1983) Consensus structure and evolution of 5S RNA. Nuc. Acids Res. 11, 893-900].

Bottom: Consensus structure of the 5S rRNA gene schematically indicating several functional sequence elements. In the helical regions of the RNA structure, A' pairs with A, B' with B, etc.



CONSENSUS SEQUENCE AND STRUCTURE OF 5S rRNA AND ITS GENE

Figure 1

factor to the ICR is necessary but not sufficient for transcription of both oocyte and somatic type 5S rRNA genes. It is not required for transcription of tRNA genes. In addition to TFIIIA, 5S gene transcription is dependent upon the two other chromatographically distinct fractions mentioned earlier (now called TFIIIB and TFIIIC) as well as RNA pol III. That TFIIIA binds to the ICR of 5S genes was demonstrated by DNA footprint analysis (12). TFIIIA was the first eukaryotic transcription factor to be purified. It was noticed that 5S rRNA can inhibit transcription of 5S rRNA genes in vitro. This is explained by the observation that TFIIIA can bind to 5S rRNA itself as well as to its gene (12). Thus the steady state level of 5S rRNA synthesis in somatic cells may be controlled by feedback inhibition. As 5S rRNA accumulates it sequesters TFIIIA, thereby inactivating transcription of the gene. Accumulation of 5S rRNA in a cell may or may not be coupled to ribosome synthesis. When it is, then amounts of TFIIIA are limiting and feedback inhibition occurs. As the 5S rRNA is incorporated into ribosomes, TFIIIA is released, freeing it to bind to the gene and stimulate transcription again. Transcription of 5S rRNA can be uncoupled from ribosome assembly by synthesizing massive amounts of TFIIIA, thereby allowing 5S rRNA to accumulate in the cytoplasm, as occurs in immature oocytes. The ribonucleoprotein complex of TFIIIA with 5S rRNA is called the 7S particle, and is the storage form of 5S rRNA in the cytoplasm (12).

The regulation of 5S rRNA gene expression is at the level of transcription (2,25). The developmental pattern of 5S gene expression

correlates with the cellular concentration of TFIIIA (4). TFIIIA comprises about 15% of the total soluble protein in immature oocytes (12), amounting to 10^{12} molecules per cell. This is in vast excess over the number of 5S genes, thus allowing oocytes to accumulate 5S rRNA in their cytoplasm to a level of 10^{11} to 10^{12} molecules per cell. When RNA synthesis is first seen in the embryo at around the 4000-cell stage there are about ten copies of TFIIIA for each 5S gene (3). (Remember that during early embryogenesis cell cleavage occurs in the absence of protein synthesis, thus TFIIIA is diluted by partitioning into the dividing cells.) At this time both oocyte and somatic type 5S rRNAs are made in about equal amounts. Three cell divisions later when the animal has completed gastrulation the concentration of TFIIIA has been reduced to about one copy per 5S gene. At this time only somatic-type 5S rRNA is made. Although there are still enough TFIIIA molecules in each cell to activate all of the 5S genes, many of them are sequestered in the cytoplasm in the 7S RNP particle. The cellular concentration of TFIIIA is further reduced during cell division to a steady state level of about 10^{4} molecules per cell, or about one TFIIIA per four 5S genes (3). In this condition the TFIIIA available for DNA binding is preferentially associated with the somatic-type 5S genes resulting in expression of these genes and repression of the oocyte-type 5S genes. This preferential expression of the somatic-type genes in the presence of limiting amounts of TFIIIA is explained by a two base difference in the ICR between the oocyte and somatic genes which results in a higher binding affinity of TFIIIA to the ICR of somatic-type 5S genes (4,26).

In summary, modulation of the general level of 5S rRNA production during development of the oocyte and the embryo is mediated by changing the concentration of the positive transcription factor of the 5S gene (TFIIIA) and the differential expression of the two forms of the 5S gene during development is caused by differential binding affinities of that factor for the two genes (4,26,27). The concentration of TFIIIA protein has been observed to correlate with the concentration of its cognate RNA, thus expression of TFIIIA itself may also be regulated at the level of transcription (28). Interestingly, a TFIIIA cDNA clone was found to contain sequences homologous to the internal promoters of some class III genes. Perhaps TFIIIA gene and 5S rRNA gene transcription are coordinately regulated by binding a common factor. Overall gene control in cells seems to have a hierarchial structure wherein products of a small number of master genes control expression of batteries of regulatory genes whose products control expression of multitudes of structural genes. Products of these structural genes then directly mediate the phenotype and metabolic physiology of the cell.

The protein structure of TFIIIA and the nature of its interaction with 5S DNA is currently under investigation (22,28-34). It was found that the first 281 amino acids of TFIIIA contain nine repeats of a conserved 30 amino acid (aa) stretch (31,33). These 30 aa stretches are basic, histone-like sequences and are believed to fold into finger-like projections upon complex formation with Zn^{+2}. Detailed footprint analysis showed that these "fingers" interact with the major groove of the ICR on one face of the double helix (22). That protection by the

fingers of TFIIIA is observed along the entire length of the ICR causes doubt about whether the split promoter model for tRNA genes can be equally well applied to 5S genes. Since TFIIIA contains nine tandem repeats of a 30 aa sequence block, it has been proposed that perhaps TFIIIA evolved by duplication of a primordial 30 aa residue unit (33). Analysis of the genomic TFIIIA clone lends support to this notion. The gene was found to be 11 kbp in length and is comprised of nine coding segments separated by eight introns. Interestingly, the intron/exon boundries correspond closely to the repeating unit, implying that each exon originated from duplication of a primordial 30 aa unit (32).

As discussed earlier, in addition to TFIIIA two other chromatographically distinct fractions (TFIIIB and TFIIIC) plus RNA pol III are required for accurate transcription of 5S rRNA genes. Formation of a stable transcription complex by binding of the three positive transcription factors to the ICR of 5S genes is believed to be a prerequisite to initiation of transcription (15,17,35 for reviews see 4,36). Such complexes are stable to dilution and to addition of competitor DNA and are capable of supporting many rounds of transcription without dissociation (25). Binding of the components of a stable transcription complex is cooperative in that the binding of the complex is tighter than the binding of the individual components. This cooperativity accounts for the stability of the complex (36). Stable transcription complexes have been reported for all three classes of eukaryotic genes (for references see 4). RNA polymerase III recognizes and binds to transcription complexes but is not itself part of the stable complex (3). Once the

transcription complex is assembled, the only other components required for transcription are RNA polymerase and NTPs (3,25). If histones are added to a preformed transcription complex they are unable to inhibit in vitro transcription of 5S rRNA, but if 5S DNA is complexed with histones prior to the addition of transcription factors then no transcription is observed (25,37,38). That is, complexes between 5S DNA and the transcription factors or 5S DNA and histones are stable to competitor proteins as well as to competitor DNA; histones and transcription factors seem to be unable to displace one another from 5S DNA. (This last point is not universally accepted however as in the literature there is a report of a triple complex between 5S DNA, TFIIIA, and a histone octamer (39).) Nucleosome formation at the control region of a gene is believed to repress expression of that gene by inhibiting binding of transcription factors to the control region (4). Donald Brown has proposed a model of development based on this concept of stably activated and stably repressed genes (4,25). In somatic cell chromatin the somatic-type 5S genes are assembled in stable, active transcription complexes while the oocyte-type 5S genes are believed to be configured in nucleosomes. Somatic cell chromatin must be washed in salt or with Biorex 70 cation exchange resin before the oocyte-type 5S genes become usable templates for transcription in vitro. Repression of the oocyte-type 5S genes in somatic cell chromatin was found to be due to a structure dependent on histone H1 (3,25,38).

As we have seen earlier, two features involved in the differential regulation of two genes which share the same transcription factor are

the concentration of the factor (is it present in saturating or limiting amounts) and the binding affinity of the factor for each of the two genes (which becomes important when the factor is present in limiting amounts). A third feature which now becomes evident is whether a particular gene has been exposed to activator or repressor molecules first. If activator molecules (TFIIIA) and repressor molecules (histones) are not in equilibrium (37), then gene activity may be programmed by formation of stably activated or stably repressed gene complexes (4,25). The relative concentration of activators and repressors at the time of DNA replication and chromatin assembly and the relative binding affinities of these molecules for the control region of a gene could constitute a mechanism by which genes are programmed to be on or off (4). It is important to realize that a gene so programmed to be "on" by its involvement in a stable, active transcription complex is not necessarily constantly being transcribed. Rather it is in a chromatin conformation rendering it potentially active. Its transcription can still be modulated by diffusible effector molecules such as hormone-receptor complexes, etc.

It has been observed that purified RNA pol III cannot accurately initiate transcription of 5S genes present as naked DNA (40). A similar result was observed for RNA pol II (41). Thus it is believed that eukaryotic RNA polymerases recognize, bind to, and are directed by transcription complexes, not naked DNA. Brown states (4) that prokaryotic RNA polymerase can recognize DNA sequences directly while eukaryotic RNA polymerases recognize transcription complexes. In

E. coli the in vitro binding affinity of RNA polymerase to promoter DNA sequences reflects the strength of the promoter in vivo (42). In eukaryotes transcription complexes mark relevant genes for transcription and reduce the number of binding sites in chromatin for RNA polymerase to a smaller number of high affinity sites (43). Most DNA in eukaryotic cells is rendered invisible to RNA polymerase by complex formation with nucleosomes. This helps RNA polymerase find the relevant control regions in the eukaryotic genome, which contains much more DNA (up to 1000 times more) than the genome of prokaryotes (4).

Thus a fundamental, general concept in the regulation of gene expression is reached: the regulatory information is built into the chromatin structure and is not contained in the DNA sequence alone. Chromatin structural domains corresponding to active or repressed states are established early in development (by mechanisms such as stable complex formation) and may be propagated or changed during chromatin replication and cell division (by mechanisms dependent on factors such as the relative concentrations and binding affinities of activators and repressors during replication). One can see that the reductionist philosophy has its limits since every cell in every tissue of an organism has the same DNA content (ignoring small, discrete exceptions such as polyteny in Drosophila salivary glands) yet expression of this genomic information is different in different tissues. Therefore the DNA sequence alone is insufficient to explain various biological phenomena such as tissue-specific gene expression, or differentiation and development of a complex multicellular organism from a single cell.

Complex hierarchies of control within the system become evident which could not be predicted from the DNA sequence alone. As the system under consideration grows larger (from DNA to chromatin to cells to tissues to organisms) properties emerge with each step which could not be expected based on understanding the previous step. Complex systems contain properties not present in the individual components of those systems.

The temporal order of molecular events in the assembly of stable transcription complexes has been determined using a combination of techniques. Formation of a competent complex is assayed by its ability to serve as a template for transcription in vitro. In one approach, cloned 5S genes were immobilized on cellulose and this insoluble resin was used as substrate for assembly of transcription complexes (44). The chromatographically resolved transcription fractions (TFIIIA, TFIIIB, and TFIIIC) were added separately and in combinations in various orders. It was observed that TFIIIA and TFIIIC could bind 5S DNA in the absence of other factors while TFIIIB stably interacts with 5S genes only after they have been previously incubated with TFIIIA and TFIIIC (44,45). TFIIIA (29,30), TFIIIB, and TFIIIC (44) have been shown to act stoichiometrically. In another approach, a combination of preincubation experiments, order of addition experiments, and competition experiments were used to investigate transcription complex assembly (46). The temporal order of events in transcription complex assembly is as follows. As a prerequisite to complex formation, histone H1 must be absent. Then TFIIIA, TFIIIC, TFIIIB, and RNA pol III

bind in that order. One molecule of each transcription factor binds to a single 5S gene to form a stable complex. RNA pol III is the least tightly associated component. In Brown's terminology (44), a "stable transcription complex" consists of one molecule of each of the transcription factors bound to a 5S rRNA gene and this complex requires only RNA pol III for transcriptional activity. The polymerase is not considered part of the complex. Upon addition of RNA pol III to such a complex a lag period of about 15 minutes is observed before a steady state rate of transcription is attained. In Roeder's terminology (46), a "rate limiting complex" consists of one molecule of each of the transcription factors plus a molecule of RNA pol III all bound to a 5S gene. Such a complex requires only nucleotides for transcriptional activity. Upon addition of NTPs to a rate limiting complex no lag period is observed. It was found that Mg^{+2} and ATP are required for formation of the rate limiting complex (46). Wolffe has described a phenomenon called "transcriptional rate enhancement" wherein as total DNA concentration is kept constant, transcriptional efficiency (defined as the number of transcripts generated per gene per hour) is seen to increase as the amount of 5S DNA is decreased (47). This is explained by sequestration of repressor molecules (such as histones) by the carrier DNA. Under rate enhanced conditions, in vitro transcriptional efficiency approaches that in vivo (about 300 transcripts per gene per hour for the 5S gene). Through some elegant experiments Wolffe has demonstrated that SP6 RNA polymerase can transcribe both strands of a 5S rRNA gene while it is involved in a stable transcription complex

without disrupting the complex. The transcription complex remains stably associated with the gene after multiple transits of the RNA polymerase (47). Wolffe also found that a stable transcription complex on the 5S gene does not impede passage of a DNA replication fork in either direction. However, the transcription complex is disrupted by fork passage with no apparent memory of the preexisting transcription complex being transmitted to the daughter duplexes (48). This does not prove that transcriptional programming cannot be preserved through replication, since housekeeping genes such as the 5S rRNA gene may not require memory. Memory of transcriptional programming would be more important for regulatory genes. At any rate, programming of transcriptional activity could be reestablished by mechanisms such as those discussed earlier.

Most of the work discussed to this point has come from two labs, that of Donald Brown at the Carnegie Institute of Washington in Baltimore, Maryland, and that of Robert Roeder at The Rockefeller University in New York, New York. Another principal investigator who has made a significant contribution to the literature of the 5S rRNA gene is Abraham Worcel at the University of Rochester in New York. In a series of several papers Worcel develops the model that in the presence of Xenopus oocyte supernatants TFIIIA induces concerted gyration (supercoiling) of 5S gene-containing plasmids and that this torsional stress is required for transcription (49-54). This DNA gyration and transcriptional activation are inhibited by novobiocin, implying the involvement of DNA topoisomerase II (55,56). Although Worcel's gels showing isomerization of 5S gene plasmids seem clear

enough, his model has at least three serious problems. First, the TFIIIA-induced gyration of DNA does not show an absolute sequence requirement. Although it is specific for 5S DNA-containing plasmids at low TFIIIA:DNA ratios, it is observed in all plasmids tested at high TFIIIA:DNA ratios, including negative-control plasmids which do not contain a 5S gene (51). In fact, this gyration is also observed in the absence of TFIIIA, but the reaction is slower. Second, the role of topoisomerase-mediated gyration in the transcriptional activation of 5S DNA was eliminated in a carefully controlled experiment by Gottesfeld (57). Gottesfeld demonstrated that another specific inhibitor of DNA topoisomerase II, VM26, and anti-DNA topo II antibodies fail to inhibit 5S gene transcription in vitro under conditions where topo II activity is inhibited. He obtained similar results for DNA topoisomerase I. Furthermore, Gottesfeld showed that novobiocin will inhibit transcription from linear as well as circular DNA templates, so the inhibition by novobiocin of transcription by RNA polymerase III must be by a mechanism distinct from topoisomerase II. Third, the dependence of transcription on torsional stress seems highly unlikely since at least two independent investigators have demonstrated specific and accurate transcription of 5S rRNA from linear DNA templates (57,58). Thus it is difficult to explain the correlation of DNA gyration and transcriptional activation observed by Worcel.

The discovery of the nucleosome in 1973 (59-61) advanced the study of chromosome structure from the level of observational phenomenology to the level of molecular mechanics. Since then the role of

proteins in genome organization and regulation has been studied extensively. (For reviews see 62-73.) The chromatin structure of the 5S rRNA genes of Xenopus (74-76) and Drosophila (77,78) has been investigated by digesting nuclei or isolated chromatin with reagents which preferentially cleave in the linker DNA between nucleosome particles. The most commonly used cleavage reagent is the enzyme micrococcal nuclease, although recently a new chemical reagent, methidiumpropyl EDTA•Fe(II), has become popular since it is reported to exhibit less sequence specificity on naked DNA. In addition to these structural probes, sites in chromatin which are hypersensitive to attack by DNase I are interpreted to be in an extremely accessible conformation. In general, the 5S genes of Drosophila and the oocyte-type 5S genes of Xenopus are found to have the same chromatin repeat length as bulk chromatin and the nucleosomes are found to be phased on the DNA. Regular or specific nucleosome arrangement simply means that nucleosomes are positioned nonrandomly on a DNA sequence while nucleosome phasing refers to a situation where the periodicity of the repeating nucleosome structure coincides with the periodicity of a repeating DNA sequence. Thus while any DNA sequence can exhibit regular nucleosome positioning, nucleosome phasing can only occur in DNAs with tandemly reiterated sequence blocks. In Xenopus it was found that the phase relationship can be adjusted to accommodate varying lengths of nontranscribed spacers between coding regions among different 5S DNA repeats (75). This implies the involvement of a DNA sequence-specific element in positioning nucleosomes (75,79). It was also found that

TFIIIA-binding to the ICR of Xenopus oocyte-type 5S genes induces a DNase I hypersensitive site at the 5' border of the ICR and that supercoiling induces an S1 nuclease site at the same position (76). Although a regular, periodic chromatin structure is also observed in Drosophila, it is different than that generated by a typical nucleosome array. It is suggested that involvement of the 5S gene in a stable transcription complex may alter the local nucleosome structure (78; also see 39).

Although much is known about the overall mechanism of protein synthesis, less is known about the structure and function of individual ribosomal components. Ribosomes contain one molecule each of 5S, 18S, and 28S rRNA with which are complexed some 70 ribosomal proteins. The complicated structural arrangement of these components to form a mature ribosome is mediated by a vast array of RNA-RNA (probably both intramolecular and intermolecular), RNA-protein, and protein-protein interactions. A central theme in biochemistry and cell biology is the interdependency of structure and function; the validity of this relationship is demonstrated clearly in the ribosome since individual ribosomal proteins exhibit no enzymatic activities (80). Although the role of 5S rRNA in ribosome structure and function is unknown, it has been postulated that 5S rRNA interacts with tRNA or with the 16S-18S rRNA component during protein synthesis (for references see 80). It is known that 5S rRNA is specifically bound by three ribosomal proteins in E. coli and by one in yeast. The yeast 5S rRNA binding protein has a molecular weight of 38,000 (81). The primary RNA-binding site is located at the C-terminal end of the protein but binding specificity

requires elements in the N-terminal portion of the molecule as well. This protein binds mainly to the 3' end of the 5S rRNA together with a small portion of the 5' end (82). (The 5' and 3' termini of 5S rRNA are base paired.) It is believed that the single eukaryotic 5S rRNA binding protein has evolved through a fusion of genes for the multiple 5S rRNA binding proteins of prokaryotes (80-82). Several models for the secondary structure of 5S rRNA have been presented (83-88). Generally these differ in only minor details.

The 5S rRNA gene occupies a prominent position in the field of molecular evolution and phylogenetics (83-89; for a general review on biochemical evolution see 90). Several reasons for this are that all organisms have a 5S rRNA gene and the gene product has the same function in all organisms, that its sequence is highly conserved throughout evolution, that 5S RNA and 5S DNA are easy to isolate, and that the 5S sequence is short and thus easily determined. It is believed that the secondary structure of 5S rRNA has been conserved during evolution as has the primary structure. (Predictive models of 5S rRNA secondary structure take into account this notion.) It has been observed that the nontranscribed spacers between 5S genes diverge much more rapidly than do the coding regions (6). In general, the 5S gene is evolving more slowly than are protein-coding genes (80). Numerical algorithms are available for the alignment of sets of sequences and the construction of phylogenic trees from sequence data (91-93, 215). To date, the nucleotide sequences of at least 353 species of 5S rRNA have been determined (86,87, this manuscript).

In summary, the 5S rRNA gene has had considerable impact on molecular biology. Its developmental regulation, differential expression, and transcriptional machinery are described in molecular detail. The 5S gene has proven useful in the investigation of chromatin structure. Ribosomal 5S RNA is a model system for RNA-protein interactions and for the study of ribosome structure. The 5S rRNA sequence library has yielded valuable insights in molecular evolution. The 5S rRNA gene is the best understood of any eukaryotic gene.

Genome Structure of Hypotrichous Ciliated Protozoa

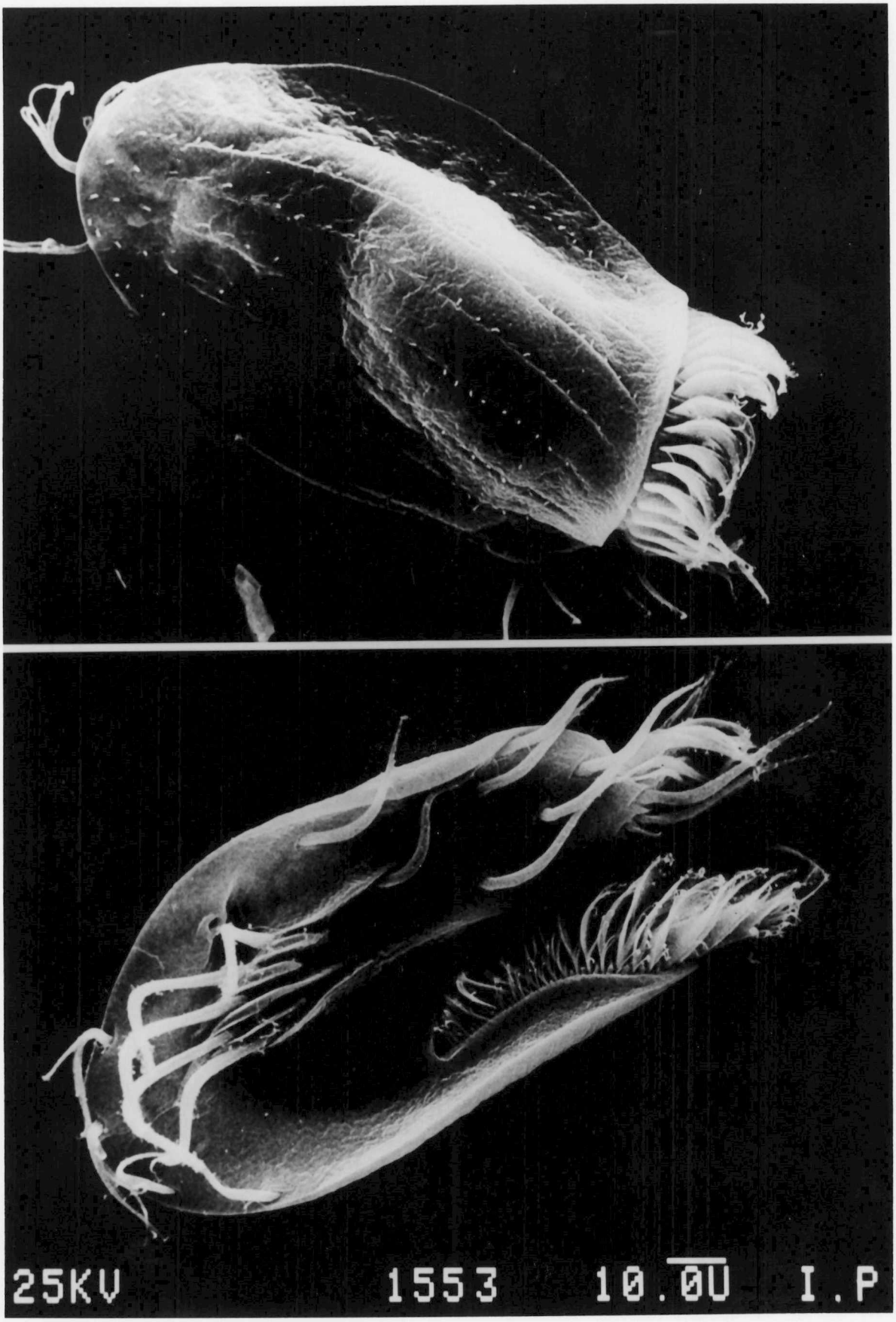
Hypotrichous ciliated protozoa, such as Euplotes, Oxytricha, and Stylonychia, are unicellular eukaryotic organisms dwelling in ponds the world over (Fig. 2). They feed on bacteria, algae, and smaller protozoa. These cells contain two morphologically and functionally distinct types of nuclei within the same cyctoplasm (reviewed in 94-97; see Table 1). The micronucleus is the germ line nucleus and functions in the sexual exchange of genetic information during conjugation. The micronuclear genome represents the complete genetic complexity of the organism. Micronuclear DNA is very high molecular weight and is arranged in a fashion typical of eukaryotic chromosomes. The diploid micronucleus divides mitotically during vegetative growth and is capable of meiotic division during conjugation. Little or no RNA synthesis occurs in the micronucleus.

The macronucleus functions as the somatic nucleus and produces all of the nuclear-derived RNA required for vegetative growth. The

Figure 2: Euplotes eurystomus.

2A: Scanning electron micrograph of Euplotes eurystomus; top: dorsal view, bottom: lateral view.

2B: Illustration of Euplotes eurystomus depicting the macronucleus and replication bands.


25KV
1553
10.0U
I.P

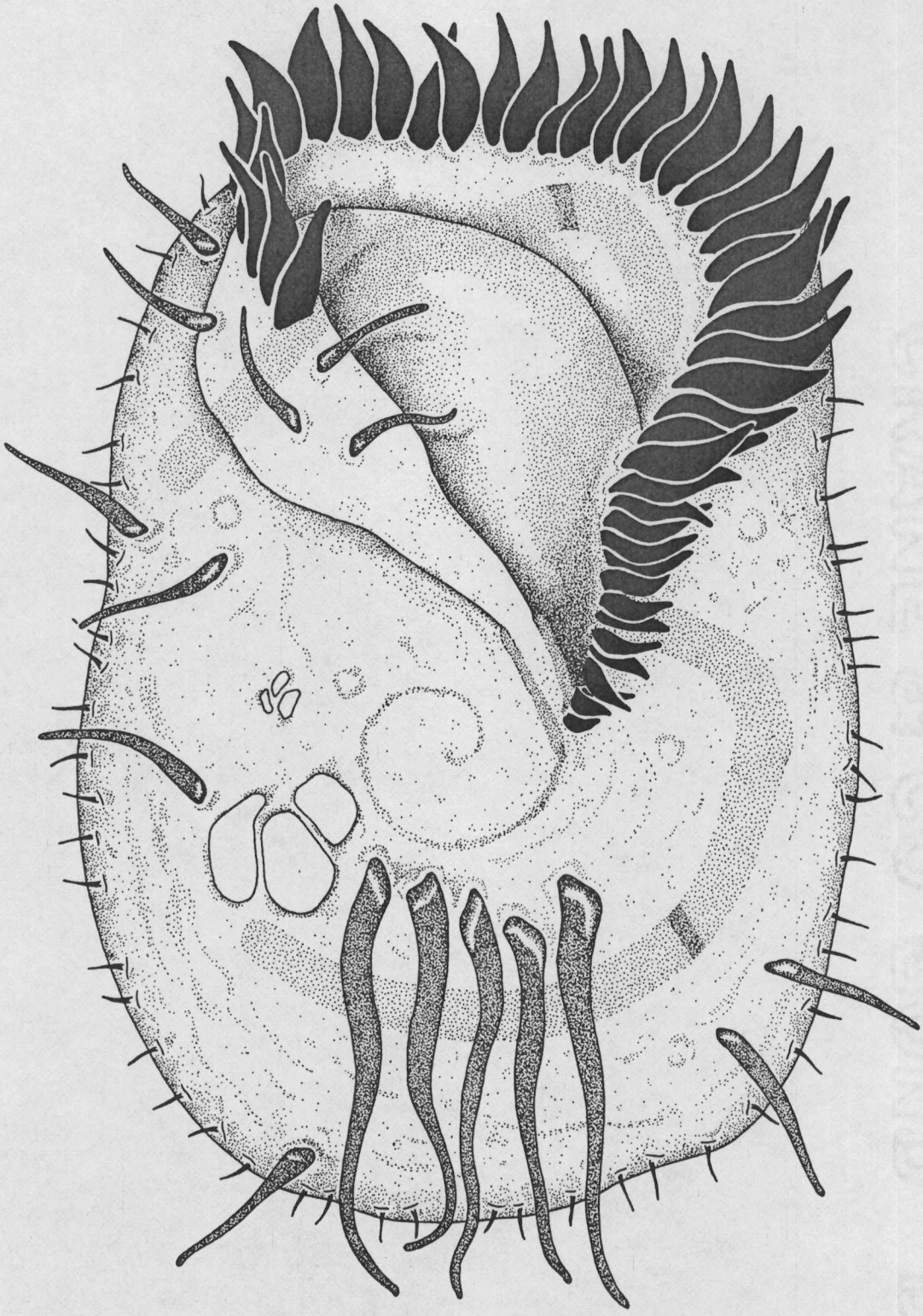


Figure 2B

Table 1. Comparison of Euplotes nuclei.

Macronucleus	Micronucleus
Linear, gene-sized DNA	Chromosomal DNA
Chromatin freely soluble	Chromatin is insoluble
Transcïptionally active	Transcriptionally silent
Genome complexity = 43 Mbp	Complexity = 50X macronucleus
Contains 380 pg DNA	Contains 1.96 pg DNA (diploid)
Replicates by replication band	No replication bands
Amitotic division	Mitotic division
Contains only unique sequence DNA	Contains repetitive DNA
Genes do not contain introns	Genes do contain introns
Genes have telomeres	Genes not flanked by telomeres

macronucleus divides amitotically (without segregation of chromosomes) following replication. Macronuclear chromatin replication is mediated by a cytologically distinct structure called a replication band (98,99). Replication bands originate at one or both ends or at defined positions in the middle of a macronucleus (depending on the species) and migrate through the macronucleus, leaving nascent chromatin in their wake. Replication bands are comprised of two discernible domains (104). The forward zone is believed to be involved in relaxation of chromatin structure to prepare it for replication. The rear zone is the site of DNA synthesis, as monitored by incorporation of tritiated thymidine. Thus the replication band contains all of the enzymatic activities required to decondense and replicate chromatin. The replication band can be enriched as an insoluble particle (99). This represents perhaps the only example of eukaryotic chromatin replication being localized in a cytologically distinct structure, permitting visualization and isolation of the replication apparatus.

Native macronuclear genomic DNA occurs as short, linear fragments (100-106). Within each of these fragments is believed to reside a single gene, since most are large enough to contain only one coding function. These linear DNAs range in size from ∿400 to ∿20,000 bp and when electrophoresed through agarose migrate as a continuum of discrete bands. Some bands are more prominent than the bulk of those in the continuum, representing hyperamplification of DNA molecules in that particular size class. The precise size distribution and banding pattern of macronuclear DNA is an inherent genetic property of the species and strain (107). Although the general properties of macronuclei of the

various hypotrichous ciliates are quite similar, the exact parameters of DNA structure are characteristic of each species (Table 2). As an example, the macronucleus of Euplotes aediculatus is observed to contain 380 pg of DNA (by microspectrophotometry (108)) with a number average molecular size of 1836 bp (by EM (109)) and a genetic complexity of 2.8×10^{10} da (by reassociation kinetics (110)). (Also see 95-97). This genetic complexity corresponds to (2.8×10^{10} da)(1 bp/649 da) = 43.1×10^{6} bp. Thus there are (43.1×10^{6} bp/MAC)(1 molecule/1836 bp) = 23,500 different DNA molecules per macronucleus. A DNA molecule of 1836 bp has a mass of 1.98×10^{-18} g. Therefore a single Euplotes macronucleus contains (380×10^{-12} g DNA/MAC)/(1.98×10^{-18} g DNA/molecule) = 192×10^{6} DNA molecules. A typical DNA molecule is thus present at (192×10^{6} total molecules)/(23,500 different molecules) = 8170 copies per macronuclear genome. Some genes are present at levels above this typical abundance, such as the 19S + 25S rRNA gene (rDNA), which is amplified approximately 100-fold above typical abundance. Although the macronucleus contains much more DNA than the micronucleus (12-200 times more depending on the species), it represents only a small fraction of the micronuclear genome complexity (generally around 2%). As a macronucleus develops from a micronucleus during the sexual phase of the life cycle, all of the repetitive DNA and 95% of the unique sequence DNA is eliminated. Only those sequences required for vegetative growth are retained in the macronucleus.

During conjugation, which can be induced by starvation, a macronucleus develops from a haploid micronucleus by a complex series of

Table 2. Parameters of hypotrichous ciliate genome structure.

Organism	Oxytricha	Stylonychia	Euplotes	References
MAC complexity (da)	3.6×10^{10}	3.1×10^{10}	2.8×10^{10}	95,97,110,112
MAC complexity (bp)	55×10^{6}	48×10^{6}	43×10^{6}	95,96,calculated
MIC complexity (da)	8.5×10^{11}	1.45×10^{12}	--	95,97
MIC complexity (bp)	1.3×10^{9}	2.2×10^{9}	--	calculated
Amt. DNA MAC (pg)	58	788	380	95,97,108,110,112
Amt. DNA hap. MIC (pg)	0.66	6.5	0.98	95,97,108,112
No. Ave. Size MAC DNA (bp)	2,200	2,514	1,836	109
No. Diff. DNAs/MAC	24,000	19,100	23,500	calculated
Total No. DNAs/MAC	25×10^{6}	290×10^{6}	192×10^{6}	calculated
Ave. abundance/MAC	1,000	15,000	8,200	calculated
% GC MAC DNA[1]	26	32	37	110
Size range MAC DNA (KBP)	0.4-20	0.4-20	0.4-20	109
Length cell cycle (h)	--	12	21	108
No. chromosomes hap. MIC	∿120	∿100	∿100	95,96

[1]Reported values vary; also see p. 194.

(For additional information and references, see 216.)

events (94-97 and 111-118; see Fig. 3). Two cells join and their micronuclei undergo meiosis. Each cell donates a haploid micronucleus (called a pronucleus) to the other through a cytoplasmic bridge. The migratory micronucleus fuses with a resident haploid micronucleus to produce a new diploid micronucleus (called a synkaryon) in each cell. All of the unused haploid micronuclei and the old macronuclei completely degenerate. The new diploid micronucleus divides mitotically and one of the daughters becomes the new micronucleus while the other develops into a new macronucleus. The micronucleus which will give rise to a macronucleus is first identified by an increase in its DNA content and the development of polytene chromosomes. Some of the satellite DNA of the micronucleus is not replicated during polytenization and this represents the first step in the diminution of sequence complexity. At 25°C the polytene chromosomes reach their peak of development at about 48 hours after fusion of haploid micronuclei. At this point the polytene chromosomes are rapidly destroyed by the formation of septa through all interbands. The septa subsequently extend to envelop separately each of the bands of all the chromosomes, forming vesicle-like structures. The nucleus is thus converted from a bag of polytene chromosomes to a bag of vesicles, each vesicle containing the material formerly present in a band plus material from the two adjacent interbands. If the nucleus is broken open at this point the vesicles float away from each other, showing that the chromosomes have been, in fact, transected band by band. Immediately after vesicle formation, most of the DNA in the nucleus is destroyed. Next, the vesicle organization disappears,

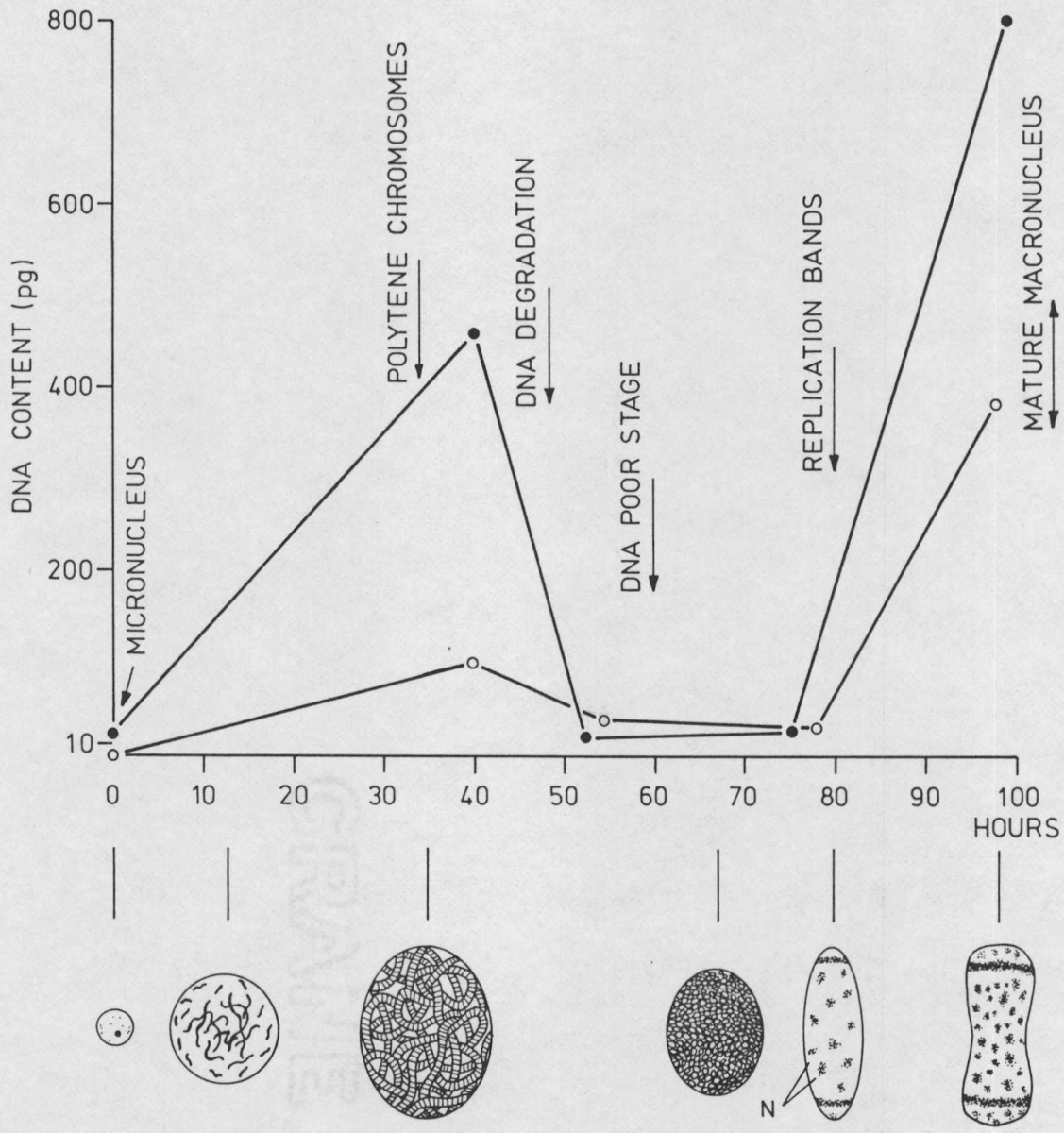


Figure 3: Time course of macronuclear development in hypotrichous ciliated protozoa.

Solid circles: Stylonychia lemnae; open circles: Euplotes aediculatus [from Kraut, H., Lipps, H. J., and Prescott, D. M. (1986) The genome of hypotrichous ciliates. International Review of Cytology 99, 1-28 as modified from Ammermann et al. (1974) Chromosoma 45, 401-429].

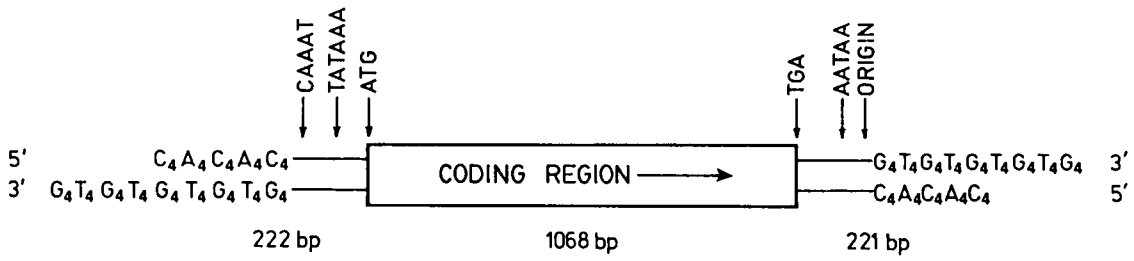
nucleoli are formed, and the DNA, which now exists as linear, gene-sized fragments, is replicated five rounds by the replication band mechanism, resulting in a mature macronucleus. The entire process of macronuclear development takes about 100 hours and results in a profound reduction of DNA length and sequence complexity. During the stage of DNA elimination, those genes destined to be found in the mature macronucleus are modified and processed in some very interesting ways. First, intervening sequences (IVS, also referred to as internally eliminated sequences (IES), or introns) are removed by some nucleic acid splicing process (117). It has been observed that the coding region of some genes is interrupted by IESs as the gene resides in the micronucleus but these IESs are absent in the macronuclear gene. Although reverse transciptase activity has been detected in the macronucleus of Stylonychia (119), the involvement of an RNA intermediate in macronuclear DNA maturation has yet to be proved. (It could be that macronuclear genes are cDNAs.) Second, macronuclear DNA fragments are bounded by telomeres (120-125). These telomeric sequences are not flanking the genes as they reside in the micronucleus (116-118). A telomere-specific terminal transferase activity has been reported for Tetrahymena (126), so the telomeres may be synthesized de novo. Alternatively, it has been postulated that the telomeres may be added by a transposition event from banks of telomeric sequence blocks which are found in the micronuclear genome (127-130). Although the origin of the telomeres flanking macronuclear DNA fragments is obscure, their functionality has been clearly demonstrated. Restriction

fragments containing hypotrich telomeres were ligated onto linear plasmids and these constructs were introduced into yeast by transformation. It was observed that the telomeres conferred stability and replicative capacity to the linear plasmids. The hypotrich telomeres were recognized by yeast and extended by addition of yeast telomeric sequences (131-134). Recently, two proteins from Oxytricha, 26 and 55 kDa, have been identified which specifically recognize and bind to Oxytricha telomeres (135). Thus each macronuclear DNA fragment represents a functional genetic unit containing, in addition to a single coding region, those sequences required for its transcription, translation, regulation, replication, and stability (Fig. 4). This system represents a unique opportunity to study eukaryotic genes which have been removed from the complexity of a typical eukaryotic genome because the gene has not been excised by experimental manipulation but rather by the organism itself as a normal part of its life cycle.

Macronuclear DNA is organized in a typical nucleosomal structure containing the four core histones, a host of nonhistone proteins, and a H1-like protein (except in the case of Oxytricha nova which seems to lack H1) (136-138). Without the use of nucleases, macronuclear chromatin is freely soluble in low ionic strength buffers and specific genes can be enriched as native, intact chromatin molecules (138-140). It is our intention to exploit this property. Although the general characteristics of bulk macronuclear chromatin have been described in detail, very little is known about the chromatin structure of specific genes. It has been suggested that nucleosome positioning is mediated inwardly

Figure 4: Consensus structure of the canonical macronuclear gene.

Schema of the macronuclear actin gene of Oxytricha fallax indicating several functional sequence elements.



STRUCTURE OF A REPRESENTATIVE MACRONUCLEAR DNA FRAGMENT:
THE ACTIN GENE OF OXYTRICHA FALLAX (1511 bp)

Figure 4

along the macronuclear chromatin fragments by the protein-telomeric complexes at the ends (141). The nucleosome repeat length of Euplotes bulk macronuclear chromatin was found to be 186 bp. The biophysical parameters of macronuclear chromatin resemble those of chicken erythrocyte whole chromatin. Electron microscopic analysis demonstrated a linear relationship between molecule length and number of nucleosomes per molecule. Chromatin molecules electrophoretically migrate in proportion to their DNA length (138).

The processing of DNA during macronuclear development, resulting in the linear, gene-sized fragments found in the mature macronucleus, is specific, precise, and reproducible. When total macronuclear genomic DNA is fractionated by agarose gel electrophoresis, transferred to a solid phase membrane, and analyzed by hybridization with specific gene probes (Southern analysis), only one or a few DNA bands in the continuum are observed to react with the probe. That is, a particular gene in the macronucleus is always found to reside on a DNA fragment of discrete size. This means that genome fragmentation is carefully controlled. If genome fragmentation was random, without sequence specificity, then a given gene probe would hybridize to DNA throughout the continuum. To date, only a handful of macronuclear genes have been examined in any detail (Table 3). Not including the data in this manuscript, among all the hypotrichous ciliated protozoa the complete nucleotide sequences of only three macronuclear fragments have been thus far determined. (Two other macronuclear fragments have been partially sequenced.) The dogma of macronuclear development and

Table 3. Specific genes from hypotrichous ciliate macronuclei.

Gene	Organism	Length MAC fragment	Length coding region	Introns	Telomeres	No. of transcripts	No. of coding regions	Transcribed	Regulatory signals*	Sequenced	References
rRNA	Oxytricha nova	7.5 KB	5.7 KB				3	yes		no	94,143,145
rRNA	Oxytricha fallax	7.5 KB	5.7 KB				3	yes		no	101,144,145
rRNA	Stylonychia mytilus	7.5 KB	5.7 KB				3	yes		no	142
rRNA	Stylonychia postulata	7.5 KB	5.7 KB				3	yes		no	109,145
rRNA	Euplotes aediculatus	7.5 KB	5.7 KB				3	yes		partial	145,146
rRNA	Euplotes eurystomus	7.5 KB								no	here,140
5S rRNA	Stylonychia mytilus	200-300 bp						yes		no	142
5S rRNA	Oxytricha fallax	690 bp						yes		no	101
5S rRNA	Euplotes eurystomus	930 bp	120 bp	no	yes	1	1	yes	ICR	yes	here,140,147
Actin	Oxytricha fallax	1.6 KB	1.1 KB	no	yes	1	1	yes	1,2,3,4	yes	148,149
Actin	Oxytricha fallax	1.4 KB								no	148,149
α tubulin	Stylonychia lemnae	1.85 KB	1.3 KB	no	yes	1	1	yes	1,2,3,4	yes	150,151
α tubulin	Stylonychia lemnae	1.73 KB		no		1	1	yes		partial	150,151
α tubulin	Euplotes eurystomus	1.6 KB								no	here,140
α tubulin	Oxytricha fallax	2.05 KB								no	96,97
α tubulin	Oxytricha fallax	2 KB								no	96,97

Table 3 (continued).

Gene	Organism	Length MAC fragment	Length coding region	Introns	Telomeres	No. of transcripts	No. of coding regions	Transcribed	Regulatory signals*	Sequenced	References
β tubulin	Oxytricha fallax	2.03 KB								no	96,97
β tubulin	Oxytricha fallax	1.88 KB								no	96,97
C2	Oxytricha nova	810 bp	246 bp	no	yes	1	1	yes	1,2,3,?	yes	117
tRNAs		400-3200 bp								no	96,97
hsp 70	Euplotes eurystomus	2.1 KB								no	L. Hauser
ubiquitin	Euplotes eurystomus	950 bp								no	L. Hauser
H1	Stylonychia mytilus	7.58 KB								no	152
H1	Stylonychia mytilus	730 bp								no	152
H2A	Stylonychia mytilus	6.36 KB								no	152
H2A	Stylonychia mytilus	4.12 KB								no	152
H2A	Stylonychia mytilus	3.64 KB								no	152
H2A	Stylonychia mytilus	1.36 KB								no	152
H2B	Stylonychia mytilus	7.58 KB								no	152
H2B	Stylonychia mytilus	4.32 KB								no	152
H2B	Stylonychia mytilus	3.06 KB								no	152
H2B	Stylonychia mytilus	2.55 KB								no	152

Table 3 (continued).

Gene	Organism	Length MAC fragment	Length coding region	Introns	Telomeres	No. of transcripts	No. of coding regions	Transcribed	Regulatory signals*	Sequenced	References
H3	Stylonychia mytilus	3.94 KB								no	152
H3	Stylonychia mytilus	2.55 KB								no	152
H3	Stylonychia mytilus	2.30 KB								no	152
H3	Stylonychia mytilus	1.59 KB								no	152
H3	Stylonychia mytilus	1.21 KB								no	152
H3	Stylonychia mytilus	990 bp								no	152
H4	Stylonychia mytilus	2.88 KB								no	152
H4	Stylonychia mytilus	1.79 KB								no	152

*1 = TATA; 2 = CAAT; 3 = AATAA; 4 = replication origin.

(For additional information and references, see 216.)

genome structure is supported by a copious paucity of information about specific genes. Hybridization analysis has resulted in the identification of 19S + 25S rRNA genes, 5S rRNA genes, actin genes, α and β tubulin genes, tRNA genes, and histone genes in macronuclei of various hypotrichs. The nucleotide sequences of an actin gene, an α tubulin gene, and a gene of unknown function called C2 have been reported. (See Table 3 for more information and references.) In those genes which have been sequenced typical eukaryotic regulatory signals have been observed, such as CAAT and TATA (transcription initiation signals), AATAA (polyadenylation signal), and sequences similar to replication origins. The work presented here contributes significantly to this body of information. I report the identification of three genes in the macronucleus of Euplotes eurystomus and the complete nucleotide sequence of the 5S rRNA gene.

Experimental Proposal

It is our eventual goal to isolate specific genes as native, intact chromatin molecules and to characterize the proteins which are associated with DNA as it is being transcribed and as it is silent. It is our eventual goal to mechanistically relate chromatin structure to gene function. To this end we have undertaken study of the 5S ribosomal RNA gene of Euplotes eurystomus. My contribution to this project is to isolate the 5S gene and to characterize it at the DNA level. In addition, various methodologies useful in the analysis of chromatin structure such as chromatin fractionation and in vitro transcription will be applied to this system.

Following is a summary of the reasons for choosing the 5S rRNA gene for analysis: (1) The 5S rRNA gene is well characterized in other systems. Its transcriptional machinery is the best understood of any eukaryotic gene. (2) The 5S gene is located on a small macronuclear fragment in Euplotes thus facilitating its resolution from the bulk of macronuclear chromatin. (3) Since the 5S gene resides on a small piece of DNA, the binding of regulatory proteins will have a more pronounced effect on its biophysical properties than would the interaction of a protein with a large DNA molecule. Thus it will be easier to detect and resolve various chromatin forms of the 5S gene. (4) Since the gene is small its chromatin structure will be simpler and easier to characterize. (5) In vitro transcription systems are available which will transcribe 5S genes input as DNA or chromatin. (6) The 5S gene seems to be amplified relative to a gene of typical abundance in the Euplotes macronucleus, thereby enhancing its enrichment as chromatin. (7) The positive transcription factor specific for 5S genes, TFIIIA, has been purified from other systems and is well characterized. (8) Anti-TFIIIA antibodies are available. (9) We know that the 5S gene must be transcribed during vegetative growth. It is possible that its expression could be regulated by starvation and refeeding.

Following is a summary of the reasons for choosing the Euplotes cell system: (1) The genome is precisely fragmented by the organism into linear, gene-sized pieces so that no manipulation is necessary to solubilize the chromatin. Thus genes can be enriched as native, intact chromatin molecules. (2) In the macronucleus, genome complexity has

been reduced about 50-fold. (3) As a natural biological property of the cell, the macronucleus is enriched in actively transcribing sequences relative to a typical eukaryotic nucleus. (4) Within the same cell a given gene exists in an active chromatin conformation in the macronucleus and in an inactive chromatin conformation in the micronucleus. (5) Except for molecular weight, the general properties (such as protein content and biophysical parameters) of macronuclear chromatin are typical of eukaryotic chromatin. (6) Each macronuclear gene is a discrete, functional genetic unit containing, in addition to a single coding region, all those sequences required for transcription, translation, regulation, replication, and stability. Thus macronuclear DNA fragments are highly enriched for regulatory and functional DNA sequences. This represents an excellent system for the examination of the minimum requirements of a eukaryotic gene. (7) This cell presents a unique opportunity to observe and isolate a eukaryotic replication apparatus. (8) The macronucleus is extremely enriched in telomeric sequences and proteins and is the system of choice for investigation of telomere structure and function. (There are $\sim 384 \times 10^6$ telomeres per Euplotes macronucleus as compared to 92 in a human nucleus.) (9) There are several advantages of using Euplotes in particular as opposed to another hypotrichous ciliate: it can be synchronized by starvation, it does not senesce upon prolonged cultivation in the laboratory, and it is commercially available. In addition, Euplotes macronuclear chromatin is more typical of eukaryotic chromatin than is Oxytricha's, since Euplotes contains H1.

CHAPTER II

EXPERIMENTAL PROCEDURES

In all cases only the highest purity, best quality reagents, enzymes, and related biologicals were used. Whenever possible, solutions were Millipore-filtered and sterilized by autoclaving or microfiltration. All glassware and other equipment was rinsed extensively with glass-distilled water before use and autoclaved when indicated. Perishable or labile materials were made fresh as needed. Biological materials were always kept cold, except during chemical reactions. Of course, standard laboratory safety procedures and the NIH guidelines for recombinant DNA experimentation were followed.

Euplotes Cell Culture

Stock cultures of Euplotes eurystomus were obtained from Carolina Biological Supply Company and were maintained in 150 x 15 mm culture dishes containing 6-7 previously boiled wheat seeds in Carolina Spring-water. Large-scale cultures of Euplotes were grown in trays containing 3 liters of Pringsheim salt solution composed of 0.02 g/L Na_2HPO_4 (0.140 mM), 0.02 g/L $MgSO_4 \cdot 7H_2O$ (0.166 mM), 0.2 g/L $Ca(NO_3)_2 \cdot 4H_2O$ (1.22 mM), and 0.02 g/L KCl (0.268 mM) (pH 7.2). The cells were fed three times weekly with live Chlorogonium elongatum (algae) which were grown in a medium consisting of 2.67 g/L yeast extract and 1.33 g/L anhydrous sodium acetate (16.20 mM). Algae were resuspended in Pringsheim solution before being presented to Euplotes. The

Chlorogonium strain was kindly provided by David Prescott (University of Colorado, Boulder).

Preparation of Euplotes Macronuclear Soluble Chromatin

Euplotes macronuclear soluble chromatin was isolated by the method of Cadilla et al. (138). Euplotes were starved for 2 days to reduce food organisms. Euplotes cells were concentrated by filtration over a 1 μm nylon mesh and lysed by addition of a concentrated stock of cell lysis buffer. The final composition of the lysis buffer was 10 mM Pipes, 5 mM $MgCl_2$, 1 mM spermidine phosphate, 1 mM PMSF (phenylmethylsulfonyl fluoride; Sigma), 1 mM TLCK (N-α-p-tosyl-L-lysine chloromethyl ketone; Sigma), and 0.5% (v/v) NP-40 (pH 6.75). Cell lysis occurred on ice for 30 minutes with occasional pipetting. Macronuclei were pelleted from the lysate by a brief, gentle spin at 4°C. The macronuclear pellet was resuspended in a small volume of cold cell lysis buffer and layered over a 10 ml linear 10-60% (w/v) metrizamide gradient cast in cell lysis buffer. Buoyant density centrifugation was carried out in an SW41 rotor at 25,000 rpm for 1 hour at 4°C. The white macronuclear band was collected from the gradient and washed extensively in cell lysis buffer to remove metrizamide. Macronuclei were lysed by resuspending the final pellet in nuclear lysis buffer A consisting of 10 mM Na_2EDTA, 5 mM triethanolamine hydrochloride (TEA·HCl), 1 mM PMSF, and 1 mM TLCK (pH 7.0). Nuclear lysis was carried out on ice for 1 hour with occasional pipetting. Insoluble debris was removed by centrifugation of the

nuclear lysate at 3000 x g for 30 minutes at 4°C. The supernatant, containing soluble chromatin, was dialyzed exhaustively against 0.25 mM Na_2EDTA, 5 mM TEA·HCl, 0.1 mM PMSF, and 0.1 mM TLCK (pH 7.0), and was stored at 4°C. When required, the chromatin was concentrated using an Amicon Centricon apparatus or by magnesium precipitation.

Preparation of Euplotes Macronuclear DNA

Euplotes macronuclear DNA was isolated by a modification of Swanton's procedure (143). Euplotes were starved for 2 days to reduce food organisms. Cells were concentrated to a volume of ∿50 ml by filtration over a nylon screen. Cells were then pelleted in a 50 ml Falcon tube by spinning in an IEC clinical centrifuge at 1000 rpm (250 x g) for 5 minutes at 4°C. The cell pellet was resuspended in 10 ml of ice-cold cell lysis buffer consisting of 10 mM Tris, 0.01% (w/v) spermidine phosphate, and 0.5% (v/v) Triton X-100 (pH 6.8). Lysis occurred on ice for 20-30 minutes with occasional, gentle agitation (rocking) of the tube. The 10 ml cell lysate was layered over a discontinuous sucrose gradient composed of 15 ml of 70% (w/v) sucrose and 15 ml of 40% (w/v) sucrose in cell lysis buffer. Centrifugation was carried out in the clinical centrifuge at 900 rpm for 5 minutes at 4°C. Macronuclei, being the heaviest constituent of the cell lysate, penetrated the 40% sucrose layer and came to rest on the 70% sucrose cushion. The top 10 ml of the gradient, containing most of the material from the cell lysate except the macronuclei, was aspirated off and discarded. The 40% and 70% sucrose layers were retained and brought to

a volume of 50 ml by addition of fresh, cold cell lysis buffer. The contents of the tube were mixed by inversion and the nuclei were pelleted by centrifugation at 1000 rpm for 15 minutes at 4°C in the clinical centrifuge. The macronuclear pellet was resuspended in 10 ml of 10% (w/v) sucrose in cell lysis buffer and layered onto a fresh gradient identical to that described above. Another cycle of centrifugation was carried out as previously described. Again, the top 10 ml were discarded and the nuclei were recovered. Alternative to two tandem cycles of discontinuous sucrose gradient centrifugation, macronuclei can be isolated by buoyant density centrifugation in metrizamide as described for preparation of soluble chromatin (except the protease inhibitors PMSF and TLCK are omitted). In any case, the final macronuclear pellet was resuspended in 5 ml of ice-cold nuclear lysis buffer B consisting of 0.1 M Na_2EDTA, 0.25 M NaCl, 10 mM Tris·HCl, and 0.5% (w/v) sodium sarcosinate (pH 8.0). Nuclear lysis occurred on ice for 30 minutes with occasional, gentle agitation. The nuclear lysate was transferred to a 15 ml Corex tube and insoluble debris was removed by centrifugation. To the supernatant was added solid proteinase K (Boeringer Mannheim) to a final concentration of 1 mg/ml. The mixture was incubated at 50°C for 12 hours. The reaction was then diluted to a volume of ∿20 ml by addition of water, 1 g of CsCl was added per ml of solution, and ethidium bromide was added to a final concentration of 400 µg/ml. The refractive index of the resulting solution was adjusted to 1.3865 by addition of water or solid CsCl. Cesium chloride equilibrium centrifugation was performed in a Ti60 rotor at 35,000 rpm for

48 hours at 20°C. The tube was illuminated with UV and the fluorescent DNA band, near the middle of the tube, was collected with a syringe. Ethidium bromide was removed from the DNA by repeated extractions with butanol. Cesium chloride was removed by exhaustive dialysis against 10 mM Tris·HCl, 1 mM Na_2EDTA (pH 8.0) (TE). Sodium acetate (pH 5.3) was added to a concentration of 0.3 M followed by addition of 2.5 volumes of absolute ethanol. DNA was precipitated by incubation at -20°C for 1 hour or more and was recovered by centrifugation. The pellet was rinsed with 80% (v/v) ethanol at -20°C, dried, and resuspended in sterile TE. (If desired the sample may be treated with RNase A.) DNA was stored at 4°C or -20°C.

Preparation of Euplotes Total Cellular RNA

Euplotes total cellular RNA was isolated essentially by the method described by Chirgwin et al. (153) and by Maniatis et al. (154). When working with RNA one must be aware of the ubiquitous presence of ribonucleases. All materials and reagents to be used must be sterilized. Glassware, weighing spatulas, tubes, etc. should be baked at 240°C overnight. Gloves must be worn not only during the isolation procedure but also during the preparation of materials and reagents to avoid contamination with ribonuclease from the skin. Endogenous ribonucleases must be inactivated as rapidly as possible following cell lysis. The following procedure was selected because cells are lysed by addition of a potent chaotropic agent (guanidinium thiocyanate) which

rapidly denatures endogenous ribonuclease. The isolation should be executed as quickly as possible.

Euplotes were starved for 2 days to reduce food organisms. Residual algae, when present, was reduced by filtration through several layers of gauze. Euplotes were removed from media by filtration over a sterile 5 μm nylon screen. When the media had drained, the nylon filter was immediately placed over a sterile beaker on ice and Euplotes were washed from it with GSCN solution. GSCN is 5 M guanidinium thiocyanate, 0.1 M β-mercaptoethanol, 25 mM sodium citrate, 0.5% (w/v) sodium sarcosinate, and 0.1%(v/v) antifoam A (Sigma) (pH 7.0). Guanidinium thiocyanate is an ionic compound, both ions of which are protein denaturants. This solution simultaneously lyses the cells and inactivates endogenous nucleases. Approximately 50 ml of GSCN solution were used. One gram of solid CsCl was added per 2.5 ml of lysate. The CsCl was dissolved by swirling and the solution was allowed to warm to room temperature. In sterile SW41 tubes, 8 ml aliquots of lysate were overlayed on 2.5 ml cushions of 5.7 M CsCl in 0.1 M Na_2EDTA (pH 7.4). Centrifugation was performed in an SW41 rotor at 35,000 rpm for 12 hours at 20°C. RNA is the only cellular constituent more dense than 5.7 M CsCl, thus the pellet so formed represents pure RNA. The supernatant was carefully aspirated and discarded. The RNA pellets were resuspended in a small volume (1-2 ml total) of 1 mM Na_2EDTA (pH 7.4); samples were heated at 65°C to facilitate solution. Care was taken not to contaminate the preparation with material residual on the inner tube wall. The RNA

solutions were pooled in a clean, sterile Corex tube; sodium acetate (pH 5.3) was added to 0.2 M; and 2.5 volumes of absolute ethanol were added. The sample was stored at -20°C overnight. Precipitated RNA was recovered by centrifugation in an SS34 rotor at 10,000 rpm for 1 hour at 4°C. The pellet was washed with 80% (v/v) ethanol at -20°C, dried by vacuum, resuspended in sterile water, transferred to a sterile 1.5 ml Eppendorf tube, and precipitated with ethanol a second time as above. Two cycles of precipitation from sodium acetate/ethanol were performed to ensure salt exchange (that is, the final product was desired to be the sodium salt of RNA, not the guanidinium or cesium salt of the acid). RNA was stored in 70% (v/v) ethanol at -70°C until use, at which time it was recovered as described above.

Bacterial Cell Strains and Culture

For cloning and propagation of the genomic DNA plasmid library, Escherichia coli strain DH5 (a $recA^-$, $endAI^-$ derivative of strain MM294 crossed with strain KL16-99) was used as the bacterial host cell. For large-scale plasmid preparations either DH5 or RR1ΔM15 (a $recA^+$, $lacZ^-$ derivative of HB101, an E. coli K-12 x E. coli B hybrid) were used. When bacteriophage M13 was used as the DNA vector, E. coli strain JM107 (an F^+, $lacZ^-$ derivative of K-12) served as host. On rare occasions strain HB101 ($recA^-$) was used to maintain and prepare plasmid DNAs.

For strains DH5, RR1ΔM15, and HB101 Luria-Bertani broth (LB) served as culture medium. LB consists of 10 g/L (1%) bactotryptone,

5 g/L (0.5%) yeast extract, and 10 g/L (170 mM) NaCl, the pH being adjusted to 7.5 by addition of NaOH. Agar medium (for plates or stabs) is prepared by addition of 15 g/L (1.5%) agar to this solution, and top agar by addition of 7.5 g/L (0.75%) agar. For liquid cultures of strain JM107 either LB or SOB media were used. SOB medium is 20 g/L (2%) bactotryptone, 5 g/L (0.5%) yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 2.5 mM KCl (pH ~6.9). Solid phase cultures (plates) of JM107 were grown on B agar which is prepared by addition of 10 g/L (1%) bactotryptone, 8 g/L (140 mM) NaCl, and 15 g/L (1.5%) agar. B top agar is the same except only 6 g/L (0.6%) agar is added. Selection for cells harboring plasmids which contain antibiotic resistance markers was achieved by addition of tetracycline (15 μg/ml; stock in 50% (v/v) ethanol) or ampicillin (50 μg/ml in liquid cultures or 100 μg/ml in plates) to media as indicated. Aliquots of sterile antibiotic stock solutions were added to autoclaved media which had cooled to 50°C or less. Amplification of plasmids was achieved by supplementation of media with chloramphenicol (170 μg/ml; stock in 100% ethanol) or with spectinomycin (Trobicin, UpJohn) (280 μg/ml). Agar for use in colorimetric assay of pUC9 or M13 clones (in $lacZ^-$ hosts) was supplemented with 40 μM isopropylthiogalactoside (IPTG) and 0.004% (w/v) 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal; stock in dimethylformamide). All media components except Xgal, tetracycline, and chloramphenicol were sterilized by autoclaving or microfiltration. Media supplemented with antibiotics or colorimetric indicators were made fresh as needed. Solutions and media containing tetracycline were

stored in the dark, as this antibiotic is light-sensitive. Liquid cultures were incubated at 37°C with gyrotory oscillation (shaking) at 250 rpm while plates were incubated inverted at 37°C.

Preparation of Plasmid DNA

Plasmid DNA was isolated from bacteria by a modification of the method of Clewell (155). A single colony of cells containing the desired plasmid was used as inoculum for 50 ml of LB containing the appropriate antibiotic. This culture was incubated overnight at 37°C, 250 rpm. Of this overnight culture, 10 ml were used to inoculate a fresh 1 L portion of antibiotic-supplemented LB (a 1/100 dilution of cells). This subculture was grown at 37°C, 250 rpm until an optical density (proportional to cell density) of A_{600} = 0.8 was attained (typically 3½-4 hours). This corresponds to late log phase. At this time cell division was halted and plasmids were amplified by addition of spectinomycin to 280 µg/ml or chloramphenicol to 170 µg/ml. Incubation was continued at 37°C, 250 rpm for 12-16 additional hours. The culture was then briefly chilled on ice. Cells were pelleted in sterile GSA bottles by spinning at 5000 rpm, 3 minutes, 4°C in a cold Sorvall GSA rotor. The medium was decanted and the pellets drained by inversion of the tubes. The cell pellets were placed on ice and resuspended in a total volume of 15 ml of sterile, ice-cold 15% (w/v) sucrose, 50 mM Tris·HCl, and 50 mM Na_2EDTA (pH 8.0). The suspension was pipetted until homogeneous; that is, until no clumps of cells remained. Aliquots, 5 ml, were distributed among sterile type 30 tubes (Oak Ridge tubes).

Two ml of a freshly prepared lysozyme solution (10 mg/ml stock in the above sucrose buffer) were added to each tube. The reaction ingredients were mixed by gently inverting the tubes 10 times and digestion was on ice for 5 minutes. Five ml of sterile, ice-cold 50 mM Tris·HCl, 50 mM Na_2EDTA, and 1% (v/v) triton X-100 (pH 8.0) were then added to each tube. Cell lysis was achieved by constant, gentle inversion of the tubes for 5 minutes. Chromosomal DNA and cell debris were pelleted by centrifugation at 30,000 rpm, 1 hour, 4°C in a type 30 rotor. The clear supernatant, containing the plasmid DNA, was carefully decanted into a sterile 50 ml Falcon tube. Solid cesium chloride was added to 1 g per ml of supernatant and dissolved by gentle inversion. Following complete solution of the crystals, the mixture was allowed to warm to room temperature. Ethidium bromide was added from a stock solution (10 mg/ml in H_2O) to a final concentration of 400 µg/ml (400-600 µg/ml is acceptable). The refractive index of the resulting solution was carefully adjusted to 1.3865 by addition of H_2O or CsCl. This material was equally distributed among an even number of Ti60 tubes. Tubes were filled completely by overlaying the samples with mineral oil and were balanced precisely, keeping in mind that one cesium chloride gradient can only be balanced by a second, identical gradient. Equilibrium buoyant density ultracentrifugation was performed in a Ti60 rotor at 35,000 rpm for 48 hours at 20°C. The tubes were illuminated with ultraviolet and the fluorescent DNA bands visualized. The lower of the two bands, representing supercoiled DNA, was extracted from the gradient by

puncturing the tube (just below the band) with a syringe needle. If desired, the plasmid DNA was further purified by a second round of CsCl gradient ultracentrifugation as above. Ethidium bromide was removed from the DNA by repeated extraction with butanol (n-butanol equilibrated with CsCl-saturated TE). Cesium chloride was removed by exhaustive dialysis of the sample against TE (pH 8.0) at 4°C. The dialysate was transferred to a sterile tube and the DNA was precipitated by addition of 1/9 volume of 3 M NaAc (pH 5.3; final concentration 0.3 M) and 2.5 volumes of absolute ethanol. The precipitate was incubated at -20°C for at least 1 hour following which time the DNA was recovered by centrifugation. The pellet was washed with 80% (v/v) ethanol at -20°C, dried by vacuum, and resuspended in sterile H_2O or TE. Working DNA solutions were stored at 4°C while archival stocks were kept at -20°C.

Precipitation of Nucleic Acids from Ethanol

Nucleic acids may be concentrated and separated from certain contaminants by precipitation from solutions of ethanol. This is also a rapid and convenient means of exchanging salts (the counterion bound to the acid) and/or the buffer in which the nucleic acid is dissolved. In most cases recovery is quantitative, but the efficiency of precipitation is reduced in the case of very small molecules (<200 bp) or in very dilute solutions (<0.1 µg/ml). To precipitate nucleic acids from aqueous solutions using ethanol, monovalent cations must be present to neutralize the charge of the acid (DNA and RNA are polyanions at

neutral pH); it is the salt of the acid that is insoluble in ethanol. The salt which was added to the DNA or RNA solution prior to precipitation (to provide the cation) was selected based on factors such as the volume of the solution and the presence of various contaminants (especially proteins or mononucleotides) that were desired to be removed. From sterile, concentrated stock solutions, sodium acetate (NaAc) (pH 5.3) was added to 0.3 M, sodium chloride to 0.2 M, or ammonium acetate (NH_4Ac) to 2.5 M. (If the DNA was originally present in a solution containing a higher salt concentration than above, it was diluted with water or TE prior to addition of ethanol.) In the work presented here, sodium acetate was most commonly used, as it is quite soluble in ethanol. Sodium chloride proved a prudent choice when the DNA or RNA was extremely dilute or when space was limiting and it was necessary to supply the salt in a very small volume. When contaminants such as proteins or mononucleotides were present (such as after restriction enzyme digests or nick translation reactions) ammonium acetate was used, as this results in less coprecipitation of these materials. The ammonium acetate salts of proteins and the ammonium salts of mononucleotides are more soluble in ethanol than are other salts of those compounds. Also, ammonium acetate is extremely soluble in ethanol and thus can be used at high concentration, thereby disrupting ionic protein-nucleic acid interactions and resulting in a cleaner precipitate. Addition of SDS to 0.5% (w/v) prior to precipitation also increases solubility of proteins in ethanol. Following addition of salt, 2.5 times the resulting volume of ice-cold 95% (v/v) or 100% ethanol were added

and mixing was achieved by several inversions of the tube. Precipitate formation was enhanced by incubation at -20°C for at least 1 hour or at -70°C for at least 15 minutes. In situations that involved very small amounts of DNA or RNA, or when the molecules were very small, -70°C was chosen. (Also in such conditions 1-2 μg of tRNA may be added as carrier to increase recovery.) After chilling, the nucleic acid was recovered by centrifugation, typically for 10 minutes in a microfuge (12,000 x g). The supernatant was decanted or aspirated and discarded. The pellet was then washed with 80% (v/v) ethanol at -20°C to remove excess salt and other solutes trapped within the pellet. After a quick spin, this supernatant was removed and the pellet briefly drained by inversion. The pellet was then dried by vacuum, typically for 5 minutes. The nucleic acid was then brought to the desired concentration by resuspension in the proper amount of the appropriate buffer.

Extraction with Phenol

Removal of proteins from DNA in complex mixtures such as cell lysates usually requires digestion with proteinase K. However in many cases proteins may be efficiently extracted with phenol. While DNA is only slightly soluble in phenol, proteins accumulate at the organic/aqueous interface, since they contain both hydrophobic and hydrophilic moieties. Besides proteins, other contaminants such as agarose and polyethylene glycol may also be removed from DNA solutions by extraction with phenol.

Most frequently used was an organic reagent, PCA, prepared in the following manner. Ultrapure, redistilled phenol (BRL; stored under N_2 at -20°C) was neutralized and saturated with water by extraction and equilibration with 10 mM Tris•HCl (pH 8.0). One part of neutralized phenol was added to an equal volume of chloroform-isoamyl alcohol (24:1 v/v) resulting in an organic phase with the final composition 25 parts phenol:24 parts chloroform:1 part isoamyl alcohol (v/v/v). Aqueous solutions of DNA were extracted using an equal volume of PCA by brief mixing achieved by tapping, inversion, or vortex. The phases were separated by centrifugation and the upper, aqueous phase aspirated and transferred to a fresh tube. To enhance recovery, the organic phase was back-extracted with fresh TE and the two aqueous phases combined. When more rigorous protein removal was required, the original aqueous solution was extracted first with neutralized phenol, then with PCA, and finally with chloroform-isoamyl alcohol. In any case, DNA was precipitated from the final aqueous phase by addition of sodium acetate (pH 5.3) to 0.3 M and ethanol to 70% (v/v) as previously described.

Proteinase K Digestion

Proteinase K (ultrapure, Boehringer Mannheim), a potent serine protease active in a broad spectrum of reaction conditions, was used in the purification of DNA from complex mixtures such as cell or nuclear lysates. Digestion was typically at 50°C for 12 hours in 0.1 M Na_2EDTA, 0.25 M NaCl, 10 mM Tris•HCl, and 0.5% (w/v) sarcosyl

(pH 8.0) by addition of 1 mg solid enzyme per ml of solution. Following digestion, the DNA was further purified by extraction with phenol or by CsCl gradient ultracentrifugation (or both) and by precipitation from ethanol.

RNase A Digestion

In addition to banding in CsCl, treatment with RNase A was also used to remove RNA from DNA preparations. RNase A (ultrapure, Boehringer Mannheim) was prepared as a 10 mg/ml stock solution in TE (pH 8.0). The solution was boiled for 10 minutes to inactivate any contaminating DNase, allowed to slowly cool to room temperature, and stored at 4°C. For digestion, 1/500 volume was added to a DNA solution (usually in TE) resulting in a final enzyme concentration of 20 µg/ml. The reaction was incubated at 65°C for 10 minutes. (The high temperature helps destabilize RNA duplex structures which are relatively resistant to RNase A.) Afterwards, when pure DNA was required, the enzyme was removed by extraction with phenol and the mono- and oligoribonucleotides by precipitation of the DNA from ammonium acetate/ethanol as previously described.

Electrophoresis of DNA in Agarose Gels

Apparati for agarose gel electrophoresis were supplied by Bethesda Research Laboratories (BRL) and International Biotechnologies Inc. Molecular weight standards (λDNA digested to completion with Hind III and ØX174 RF DNA digested to completion with Hae III) were obtained

from BRL. Horizontal, submarine gels from 7 to 20 cm in length were used, depending on the resolution required. Gels were run at constant voltage between 50 and 150 V. During prolonged electrophoresis experiments, cooling water was used and the running buffer was recirculated with a peristaltic pump. Gels were 0.8-1.25% (w/v) agarose (ultrapure, BRL or Biorad) cast in 1X TBE buffer (89 mM Tris, 89 mM boric acid, and 2 mM Na_2EDTA; pH 8.3). Most commonly used was the BRL unit with gel tray dimensions of 11 cm x 14 cm, containing a 60 ml gel of 1% (w/v) agarose in 1X TBE, electrophoresed at 100 V for 3-3½ hours in 1 L of 1X TBE running buffer. Typically, DNA samples of 0.5-3.0 µg in a final volume of 10 µl were loaded into 1 mm wells. Aqueous DNA samples were prepared by addition of 1/4 volume of 5X DNA gel sample loading buffer (12.5% (w/v) Ficoll-400 DL (Sigma), 2.5 x TBE, 1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue; 0.22 µm microfilter sterilized) followed by brief heating at 65°C to dissociate aggregates. Alternatively, dry DNA pellets were resuspended in 10 µl of 1X sample loading buffer and heated.

Following electrophoresis, DNA bands within the gel were visualized by staining with the fluorochrome ethidium bromide. Usually, 0.5 µg/ml ethidium bromide was included in the gel and running buffer during electrophoresis; if not, following the run the gel was immersed in a solution of 1 µg/ml ethidium bromide for 30 minutes at room temperature on a rocking platform. To enhance visualization of very faint DNA bands, nonspecific backround fluorescence may be reduced by washing the ethidium bromide-stained gel in 1 mM $MgSO_4$ for 1 hour at

4°C on a rocking platform. Illumination was with UV irradiation, 300 nm emission maxima, from a Fotodyne transilluminator. Photography was with Polaroid film, type 55 P/N; exposure was typically at f5.6 for 1-2 minutes using a red filter.

Electrophoresis of DNA in Polyacrylamide Gels

Polyacrylamide gel electrophoresis (PAGE) was used to monitor and preparatively purify small DNA molecules (<1000 bp). PAGE was performed in an apparatus manufactured by Biorad. Before use, the glass plates were scrubbed with detergent, rinsed thoroughly with distilled water, and then washed with methanol. Usually, gels measuring 20 cm x 16 cm x 1 mm (or 2 mm) were prepared. Gels in 1X TBE buffer consisted of 5% (w/v) total acrylamide (29 parts acrylamide:1 part bisacrylamide), 0.13% (w/v) ammonium persulfate, and 0.05% (v/v) tetramethylenediamine (TEMED). A 100 ml portion of polyacrylamide gel solution (enough for 2 gels) was prepared by combining 16.67 ml 30% (w/v) acrylamide solution (29:1), 20 ml 5X TBE, 1.33 ml of freshly prepared 10% (w/v) ammonium persulfate, and 62 ml H_2O. This mixture was degassed by vacuum for 3-5 minutes (to remove oxygen, an inhibitor of acrylamide polymerization) and briefly chilled on ice. Per 100 ml gel solution, 50 µl TEMED was added, the reaction mixed by swirling, and the gel quickly poured. Polymerization occurred at room temperature for at least 1 hour. Before loading, gels were prelectrophoresed for 15 minutes at 200 V. Typically, 0.5-1.0 µg DNA in 10 µl of the above-mentioned sample loading buffer was loaded per lane.

Electrophoresis was at 200 V for 2 hours in 1X TBE running buffer with cooling water circulation. Following the run, the gels were stained for 10-15 minutes in 1 μg/ml ethidium bromide and photographed as previously described.

Electrophoresis of DNA in Agarose Gels Containing Sarcosyl

The DNA component of chromatin samples was monitored by electrophoresis in agarose gels containing sarcosyl. Inclusion of detergent in the gel alleviated the need to purify the DNA whenever quick analysis of chromatin samples was required. Gels of 1.25% (w/v) agarose (ultrapure, BRL or Biorad) were cast in a running buffer composed of 40 mM Tris•HCl, 5 mM sodium acetate, 1 mM Na_2EDTA, and 1% (w/v) sarcosyl (pH 7.9) and were electrophoresed at 100 V for 4 hours using cooling water and buffer exchange. Staining and photography was as previously described.

Electrophoresis of RNA in Agarose Gels

RNA was analyzed by electrophoresis through 1.2% (w/v) agarose gels containing 2.2 M formaldehyde in 10 mM NaH_2PO_4 buffer (pH 6.5). A 60 ml gel was prepared by combining 0.75 g agarose (ultrapure, BRL or Biorad), 0.6 ml sterile 1 M NaH_2PO_4 (pH 6.5), and ~45 ml sterile H_2O. After heating in a microwave oven to dissolve agarose, 10.8 ml 37% (w/w) formaldehyde (neutralized to pH 7.0) was added and the volume adjusted to 60 ml in a sterile graduate cylinder by addition of sterile water. The mixture was allowed to cool to 50-55°C before

casting the gel; the gel was allowed to harden for at least 1 hour before use. Samples were prepared by dissolving dry RNA pellets in sterile water and then adding 3 volumes of 1.5X sample buffer (prepared by combining 300 µl deionized formamide, 108 µl 37% (w/w) neutralized formaldehyde, and 6 µl 1 M sterile NaH_2PO_4 (pH 6.5); made fresh each time) and heating at 50°C for 15 minutes. Afterwards, 1/9 volume of 10X sample loading buffer was added (25% (w/v) Ficoll 400 DL (Sigma), 0.25% (w/v) xylene cyanol, and 0.25% (w/v) bromophenol blue; 0.22 µm microfilter sterilized). Typically, 5 µg of Euplotes RNA in a final volume of 11.1 µl was loaded per lane (5 µg RNA in 2.5 µl H_2O + 7.5 µl sample buffer + 1.1 µl loading buffer). RNA molecular weight standards (BRL) were loaded 3 µg per lane. Electrophoresis was at 80 V for 3 hours in sterile 10 mM NaH_2PO_4 (pH 6.5) running buffer, with buffer exchange.

Following the run, RNA was visualized by staining with the fluorochrome acridine orange (156). The gel was first washed with ∿500 ml fresh, sterile 10 mM NaH_2PO_4 (pH 6.5) for 30 minutes at room temperature on a rocking platform. The gel was then stained in a plastic photographic tray by immersion in a solution prepared by combining 485 ml of fresh, sterile 10 mM NaH_2PO_4 (pH 6.5) with 15 ml acridine orange stock solution (1 mg/ml in H_2O; 0.22 µm microfilter sterilized) (final concentration is 30 µg/ml) for 30 minutes at room temperature in the dark (cover tray with foil) with rocking. The gel was then destained in an enameled pan using 500 ml of fresh phosphate buffer at 4°C, in the dark, overnight, with gentle rocking. (The enamel adsorbs

the dye as the gel destains.) Illumination with UV irradiation and photography were as previously described. In the presence of acridine orange, single-stranded nucleic acid species fluoresce red-orange while double-stranded species appear green. Although ethidium bromide may also be used to stain RNA, acridine orange is more sensitive.

Electrophoresis of Proteins by SDS-PAGE

Euplotes macronuclear chromatin-associated proteins were analyzed by polyacrylamide gel electrophoresis in the presence of SDS. This yielded not only information about the histone and nonhistone chromosomal proteins of the Euplotes macronucleus, but also served as a sensitive indicator of the structural integrity of chromatin samples. Electrophoresis, with water cooling, was performed in 15% (w/v) SDS-polyacrylamide slab gels measuring 14.0 cm x 17.8 cm x 1.0 mm according to the method of Laemmli (157) using the modified electrode buffer of Thomas and Kornberg (158). Protein samples were prepared by precipitation of chromatin by addition of 0.2 M sodium phosphate buffer (pH 7.6) to a final concentration of 5 mM, then by $MgCl_2$ to 6 mM, followed by 0.7 volumes of 95% (v/v) ethanol and incubation at -20°C for at least 30 minutes (159). The precipitates were recovered by microcentrifugation for 10 minutes at 4°C, the supernatants discarded, and the pellets dried by vacuum. Samples were resuspended in 1X SDS sample buffer (157) and boiled for 1 minute prior to loading. Following the run, proteins were visualized by staining with silver

according to the method of Wray et al. (160) or by staining with Coomasie brilliant blue R250 (Biorad).

Restriction Endonuclease Digestion

Restriction enzymes were obtained from BRL, New England Biolabs, and Promega, and were used according to the manufacturers' directions. Great care was taken in assembling reactions to provide each enzyme with its optimal chemical milieu so as to ensure accurate and specific performance. The most important reaction parameters are the purity of the DNA substrate, the assay buffer, the assay volume, the incubation temperature, and the incubation time (161). Best results were obtained using ultrapure DNA prepared by proteinase K treatment, CsCl density gradient ultracentrifugation, dialysis, RNase A treatment, extraction with phenol, molecular exclusion chromatography, and precipitation from ammonium acetate/ethanol as previously described. For optimum results, the DNA must be free of proteins, detergents, organic solvents, heavy metals, EDTA, and certain other unidentified contaminants (probably sulfonated polysacharrides) commonly contained in commercial preparations of agarose and acrylamide. All digestions were carried out in sterile microfuge tubes using sterile pipette tips and sterile buffer components. The paradigmatic reaction was digestion of 1 μg DNA with 10 units of enzyme in 100 μl of the appropriate buffer for 2 hours at the appropriate temperature. All reaction components except the enzyme were combined and the reaction was initiated by addition of enzyme followed by gentle tapping of the tube to mix the

ingredients. This represented a 10-fold excess digest to ensure that the cleavage reaction went to completion. Whenever possible, the DNA was provided in a small volume (1-5 µl) so as to dilute any inhibitors carried along with it in the final reaction. In no case was the volume of enzyme stock solution added greater than 10% (v/v) of the final reaction volume (usually 10 units of enzyme were provided in a volume of 1 µl). (This is because restriction enzymes are stored at -20°C in a buffer containing 50% (v/v) glycerol, and glycerol concentrations greater than 5% (v/v) in the reaction can promote secondary activity (inaccurate site recognition and cleavage) of the enzyme.) Following reaction, the cleavage products were purified by one of several methods depending on the subsequent application of the DNA. For analysis by gel electrophoresis, the digestion was stopped by addition of EDTA to 10 mM (add 5.3 µl sterile 0.2 M Na_2EDTA (pH 8.0) stock per 100 µl reaction). This chelates Mg^{+2} (a requirement of all restriction endonucleases) thereby inactivating the enzyme and assisting in its dissociation from the DNA substrate. Afterwards, 15 µl of sterile 10% (w/v) SDS, 80 µl of sterile H_2O, and 100 µl of sterile 7.5 M NH_4Ac were added. The resulting solution had a volume of 300 µl and was 3.5 mM in EDTA, 0.5% (w/v) in SDS, and 2 M in NH_4Ac. DNA was precipitated from this mixture by addition of 2.5 volumes (750 µl) of ice-cold absolute ethanol followed by chilling at -70°C for at least 15 minutes or at -20°C for at least one hour. (Although SDS is insoluble in 2.5 M NH_4Ac, it dissolves immediately upon addition of ethanol.) After freezing, the sample was allowed to warm to room

temperature, ∿10 minutes. The DNA precipitate was recovered by microcentrifugation for 10-15 minutes at room temperature. The supernatant was decanted and discarded, the pellet was washed first with 1 ml 80% (v/v) ethanol at -20°C, then with 1 ml absolute ethanol at -20°C, allowed to drain inverted for ∿1 minute, and finally dried by vacuum (∿5 minutes). The dry DNA pellet was then resuspended in 1X DNA gel sample loading buffer, water, TE, or whatever buffer was appropriate for subsequent use. (The purification strategy described above was designed to minimize coprecipitation of the enzyme along with the DNA.) Alternatively, when more rigorous protein-removal was required following digestion (as when the DNA was to be further manipulated by other DNA modifying enzymes), the DNA was purified by extraction with phenol (PCA) and precipitation from ethanol as previously described.

Southern Hybridization Analysis

Hybridization analysis allows identification, localization, and quantification of specific nucleic acid sequences immobilized on a two-dimensional, solid-phase support matrix by staining (hybridization) with labelled, sequence-specific nucleic acid probes. Both the immobilized species under investigation and the hybridization probe may be either DNA or RNA. In "Southern analysis" the immobilized species is DNA while in "northern analysis" the immobile phase is RNA. Support materials are commonly thin membranes of nitrocellulose or nylon. Application of nucleic acids to the blotting membrane is generally by

transfer from an agarose gel via capillary action, electrophoresis, or vacuum. Alternatively, nucleic acid samples may be directly spotted onto the membrane in a vacuum manifold (dot blots). Hybridization probes are commonly labelled by incorporation of radioactivity (and detected by autoradiography) or biotin (and detected by reaction with labelled avidin, streptavidin, or anti-biotin antibodies). The overall process consists of the following chronological order of steps: construction of the blot, prehybridization (blocking) of the blot, hybridization with the probe, washing (to remove nonspecifically bound probe), and detection. Much, if not most, of what is known today about gene structure and function has been derived from such experiments.

DNA was transferred from agarose gels to nitrocellulose (Schleicher and Schuell Inc., BA 85, 0.45 μm pore size; nitrocellulose is always handled using gloves and forceps) essentially as described by Southern (162). Following electrophoresis, staining with ethidium bromide and photography, the gel was illuminated with UV irradiation (300 nm maxima) for an additional 5 minutes. UV, especially in the presence of ethidium bromide, will nick the DNA and this fragmentation of the DNA into smaller pieces assists in its transfer out of the gel. When especially large DNA fragments were to be transferred, the gel was then immersed twice, 15 minutes each, in 500 ml of 0.25 M HCl in a glass baking dish at room temperature on a rocking platform. (Between all immersions the gel was briefly rinsed with distilled water.) This treatment, which was usually omitted, causes partial depurination of

the DNA and enhances transfer efficiency of large molecules. Subsequently, DNA within the gel was denatured by immersion twice, 15 minutes each, in 500 ml of 0.5 M NaOH, 1.5 M NaCl at room temperature with rocking. Next the gel was neutralized (hydroxide attacks nitrocellulose) by immersion twice, 30 minutes each, in 500 ml of 1 M sodium acetate (pH 4.2) at room temperature with rocking. Finally the gel was immersed for 10-15 minutes in 500 ml of 10X SSC at room temperature with rocking. (Standard saline citrate, SSC, is prepared as a 20X concentrated stock solution which is 3 M NaCl, 0.3 M trisodium citrate.) The nitrocellulose was hydrated by floating on the surface of distilled water (this technique expels gases from within the membrane) followed by immersion in distilled water (it is necessary to remove the glycerol with which the nitrocellulose is impregnated during manufacture). The nitrocellulose and the wick (Whatman 3MM paper) were then equilibrated with 2X SSC (DNA adheres to the nitrocellulose more strongly in the presence of salt). Transfer of the DNA from the gel to the nitrocellulose was by capillary action for 24 hours at room temperature with 10X SSC buffer using the geometry described by Maniatis et al. (154). Following elution, the nitrocellulose was demarcated to allow future recollection of its orientation (with respect to the bound DNA). The membrane was air-dried (DNA-side-up) for several minutes and then was baked at 80°C for 2 hours in vacuo to fix the DNA. The dry blot may be stored at room temperature or 4°C if not used immediately. Customarily the gel was restained with ethidium bromide after transfer to monitor efficiency.

Northern Hybridization Analysis

As RNA was denatured prior to electrophoresis, the transfer protocol is simplified relative to that of DNA (163-165). The part of the gel to be transferred was not stained and photographed; lanes of the gel were removed with a razor blade for that purpose. The remainder of the gel was immersed in 15X SSC for 15 minutes at room temperature with rocking. Transfer was as described previously except using 15X SSC as elution buffer. Following transfer, efficiency was monitored by staining the gel with acridine orange or ethidium bromide.

Dot Blot Hybridization Analysis

Dot blots were performed using a Schleicher and Schuell Minifold Micro-sample Filtration Manifold and Schleicher and Schuell BA85 nitrocellulose membranes (0.45 μm pore size). The apparatus was assembled and used according to the manufacturer's directions. The nitrocellulose was hydrated by floating on distilled water followed by immersion in distilled water as usual. The nitocellulose and filter paper (used for supporting the nitrocellulose) were then equilibrated with 10X SSC. After assembly of the apparatus, low vacuum was applied and each 500 μl filtration well to be used was flushed with 10X SSC. Afterwards, samples containing known amounts (typically 0.1-1.0 μg) of denatured DNA in a volume of 500 μl were applied per well. Following application of samples, the apparatus was disassembled and the nitrocellulose was allowed to air dry for ∿5 minutes. Finally, the DNA

was fixed on the membrane by baking at 80°C for 2 hours in vacuo. Samples were prepared from 100 µl aliquots containing the desired amount of DNA by addition of 1/3 volume (33 µl) of 2N NaOH (final concentration = 0.5 N). To ensure that the DNA was completely denatured (and that any contaminating RNA was destroyed), the basic solution was incubated at 65°C for 15 minutes. Afterwards the sample was diluted by addition of 334 µl of sterile 20X SSC and neutralized by addition of 33 µl of 2 N HCl. The neutralized, denatured DNA sample, now in a volume of 500 µl of 13X SSC, was briefly chilled on ice before application to the nitrocellulose with a sterile P1000 pipette tip. Since nucleic acids adhere more strongly to nitrocellulose in the presence of high salt, in the above strategy the DNA sample was denatured in a small volume so as to allow subsequent dilution using a concentrated salt solution; the solution was neutralized following dilution in order to reduce the rate of reassociation. When chromatin samples were to be analyzed, the DNA was deproteinized using proteinase K as previously described and then treated as above (166). Following hybridization, washing, and autoradiography (to be described in detail later), hybridization intensities were quantitated by scanning the autoradiograms with an LKB 2202 Ultrascan Laser Densitometer interfaced with an Apple IIe computer.

Preparation of Gene Probes

Several methods exist for labelling cloned DNA fragments. DNA may be labelled at the 5' end by polynucleotide kinase or at the 3' end

by terminal deoxynucleotidyl transferase. DNA with recessed 3' ends may be labelled with the Klenow fragment of DNA polymerase I. Alternative to labelling only at the ends, a DNA (or RNA) molecule may be labelled along its entire length by de novo synthesis from labelled DNA (or RNA) precursors. The advantage of this approach is that probes of higher specific activity may be generated, since each nucleic acid molecule can contain many labelled residues. The most common such method is "nick translation," so named because a single-strand nick, immediately preceding the nascent strand, is linearly translated along a DNA molecule. The reaction mixture includes a double-stranded DNA substrate, DNase I, E. coli DNA polymerase I, and all 4 deoxynucleoside triphosphates. The nucleotides used for this purpose may be labelled with ^{32}P, ^{35}S, or biotin. DNase I attacks the DNA substrate, creating single-strand nicks, thus allowing binding and synthesis by DNA pol I. As the polymerase locomotes along the template strand its 5'→3' exonuclease activity removes the nucleotide immediately before it (on the sister strand) while the polymerase activity replaces the nucleotide immediately following the enzyme. Thus a nick, created by DNase I, is linearly translated along the DNA molecule by DNA pol I. When labelled deoxynucleoside triphosphates are present, the nascent strand so synthesized will incorporate the label. RNA probes may be prepared by incorporation of labelled ribonucleotides into nascent transcripts by bacteriophage SP6 RNA polymerase.

Nick translation reagent kits were obtained from BRL, Worthington, and Amersham. Radioactive nucleotides (usually [$\alpha^{32}P$]-dCTP;

specific activity = 3200 Ci/mmole) were purchased from ICN Radiochemicals. Nick translation reactions were assembled in accordance with the instructions accompanying each reagent kit. Reactions of 25-100 µl containing 0.2-1.0 µg of linear or circular DNA were used. The paradigmatic reaction was labelling of 0.2 µg DNA in a volume of 50 µl by incorporation of [α^{32}P]-dCTP. Following incubation at 15°C for 1 hour, the reaction was terminated by addition of EDTA to 20 mM and the DNA was separated from unincorporated nucleotides by molecular exclusion chromatography (Worthington minispin columns). To the eluate was added ∿20 µg of tRNA (as carrier), H_2O to 200 µl, 100 µl of 7.5 M NH_4Ac (final concentration = 2.5 M), and 750 µl (2.5 volumes) of absolute ethanol. After chilling, the pellet was collected, rinsed, dried, resuspended, and precipitated from ethanol a second time. The final DNA pellet was rinsed, dried, and resuspended in 100 µl of sterile H_2O. By Cerenkov counting of a 1 µl aliquot the specific activity of the probe was determined, typically between 0.5-2.0 x 10^8 cpm/µg DNA. The ethanolic supernatants and rinses were also frequently monitored to follow diminution of free counts. The radioactive DNA probe was stored in a lead vial at -20°C for up to two weeks.

Hybridization and Washing of Nucleic Acid Transfers

Southern, northern, and dot blots are analyzed by hybridization with sequence-specific gene probes. This allows precise determination of nucleic acid species immobilized on the membrane which are cognate to the probe. Parameters affecting hybridization stringency (that is,

the tolerance of unpaired bases in a duplex of imperfect complementarity) and kinetics include the concentration and nature of reactants, the reaction temperature, the reaction time, and the nature of the solvent (for example, the ionic strength of the solvent and the inclusion of denaturants such as formamide). Usually the hybridization conditions are kept constant and the specificity of the reaction is adjusted by systematic alteration of the washing conditions until the optimum for each probe is empirically determined. Sequences of only slight homology to the probe may be detected or discrimination may be for exact matches.

After baking, dry nitrocellulose membranes were hydrated by floating on 6X SSC until wet and then were submerged in 6X SSC for several minutes. Prior to exposure to the probe, the blots were prehybridized so as to block nonspecific interaction of the probe with the nitrocellulose. Blocking was with the same buffer as used for hybridization except for omission of the probe. Hybridization buffer was prepared as a 2X stock containing 10X SSC; 100 mM NaH_2PO_4; 0.5 mg/ml sheared, denatured salmon sperm DNA; 20 µg/ml polyadenylate; 0.04% (w/v) bovine serum albumin; 0.04% (w/v) ficoll 400 DL; and 0.04% (w/v) polyvinylpyrrolidone (pH 6.8) and was stored at 4°C. All components of the 2X hybridization buffer were 0.22 µm filter sterilized, except for the DNA and the poly(A) which were 0.45 µm filter sterilized. Hybridization buffer was prepared for use by boiling the appropriate amount of the 2X stock for 5 minutes (in a boiling water bath) followed by quick chilling via transfer to an ice/water bath and

immediate addition of an equal volume of ice-cold, deionized formamide. The resulting solution was thus 1X hybridization buffer, 50% (v/v) formamide. Formamide (BRL, redistilled and stored under N_2 at -20°C) was deionized using mixed bed ion exchange resin (Biorad). Prehybridization using 1X hybridization buffer, 50% (v/v) formamide was for at least 12 hours (usually ∿24 hours) at 42°C in a hermetically sealed plastic bag with occasional, gentle agitation. The volume used for prehybridization varied from 2-50 ml depending on the size and number of blots being probed (up to 42 separate nitrocellulose membranes were simultaneously screened in a single bag) and was generally 2-4 times the volume of the subsequent hybridization reaction. Following blocking, the prehybridization solution was expelled from the bag and discarded. A small aliquot of the radioactive probe, in H_2O, was added to the appropriate amount of 2X hybridization buffer and was boiled for 5 minutes; the mixture was then chilled on ice by addition of an equal volume of ice-cold, deionized formamide as before. Hybridization was in 1-25 ml of 1X hybridization buffer, 50% (v/v) formamide in a hermetically sealed plastic bag at 42°C for at least 12 hours with occasional, gentle agitation. Afterwards, unbound and nonspecifically bound probe was removed by washing the nitrocellulose in SSC + SDS. Typically about 6 washes of 15 minutes each were performed in a glass baking dish at room temperature on a rocking platform followed by 10 or more washes each of an hour or longer in a fresh, sealed plastic bag in a water bath with occasional, gentle agitation; about 100 ml of SSC + SDS were used per wash. An aliquot

of each wash (∿20 ml) was radioquantitated by Cerenkov counting to follow diminution of backround. After washing, the blot was dried under a heat lamp for ∿5 minutes and radiation from the probe was detected by autoradiography.

The concentration (volume specific activity) of probe used during the hybridization reaction and the washing stringency varied with each particular probe. In general, as homology of the probe to Euplotes eurystomus macronuclear DNA increased, the concentration of probe decreased and the washing stringency increased. The 5S rRNA gene probe from the macronucleus of Tetrahymena thermophila (provided by David Pederson (167)) was used at a concentration of 10×10^6 cpm/ml and washing was with 6X SSC, 0.1% (w/v) SDS at 42°C. The macronuclear α-tubulin gene from Stylonichia lemnae (donated by Elke Helftenbein (150, 151)) and the rRNA gene from the macronucleus of Euplotes aediculatus (supplied by Marshal Swanton (145)) were used at 2×10^6 cpm/ml and were washed with 2X SSC, 0.1% (w/v) SDS at 50°C. The macronuclear 5S rRNA gene from Euplotes eurystomus (the homologous system) (58, 147, and this manuscript) was used at 1×10^6 cpm/ml and washing was with 0.1 x SSC, 0.1% (w/v) SDS at 65°C.

Autoradiography and Fluorography

Radioactive DNA probes localized on nitrocellulose membranes were imaged by exposure to Kodak XAR-5 x-ray film. (XAR-5 contains photograhic emulsion on both sides of the film and thus is quite sensitive.) For Southern, northern, and dot blot hybridizations, signal

enhancement was commonly achieved by exposure between two Cronex Ligthning-Plus intensifier screens (Dupont) at -70°C. [Intensifier screens are thin sheets of plastic impregnated with the phosphor calcium tungstate. Energy absorbed by the screen is dissipated by phosphorescent decay, a quantum energy transfer modality characterized by a long half-life. Energy emitted by the sample (in the form of β particles) is absorbed by the screen and is number-amplified via dispersion into smaller quanta (photons) which are reemitted back toward the film resulting in signal intensification up to 400-fold. Since both the sample and the screen emit energy in all directions, concomitant with increased sensitivity is decreased resolution. For this reason intensifier screens are not used in the radiodetection of DNA sequencing gels except in desperate circumstances. Autoradiographic exposures using intensifier screens are always performed at ultracold temperatures to reduce the kinetic energy, and thus the mean flight path, of radionuclide decay particles so as to minimize the loss of resolution. Also, the efficiency of photographic light detection is improved since individual silver atoms within a silver halide crystal have a longer half-life at ultracold temperatures. That is, the back reaction (thermal reversion) of the light-induced conversion of a silver ion to a silver atom is decreased (168, 169). Thus the latent image has an increasd probability of accumulation. It seems unclear if the efficiency of phosphorescence itself, the quantum yield, is enhanced in the cold (170).] In the absence of signal intensification, exposure was typically at -20°C. Exposure times varied from 20 minutes to several days. Following exposure, film was developed for 2-3 minutes at room

temperature with occasional agitation, briefly rinsed in water, fixed for 5 minutes at room temperature with occasional agitation, rinsed with distilled water for 15-60 minutes, and air dried. Signals were quantitated by scanning laser densitometry as previously described.

Reuse of Hybridization Membranes

Gene probes may be eluted from nucleic acid transfers to allow subsequent rehybridization with a different probe (or with the same probe under different conditions). For this purpose the method of Gatti et al. (171) was applied. Blots intended to be reused were not dried following washing (as was usually done) but instead were simply enclosed in Saran wrap during autoradiography. Blots were washed 4 times, 15 minutes each, with 500 ml of 0.1X SSC, 0.1% (w/v) SDS at 95°C in a pyrex dish on a rocking platform. This elution buffer was heated to boiling in a microwave oven and immediately poured onto the nitrocellulose. Of each wash an aliquot was quantitated by Cerenkov counting to monitor diminution of solubilized radioactivity. It was observed that 98% of the probe was eluted in the first wash and the remainder in the second. Upon completion of washing, the membrane was dried and exposed to film overnight with screens at -70°C to confirm dehybridization.

Construction of Chimeric Plasmids and Generation of the Genomic DNA Library

A DNA library is a collection of DNA molecules containing either a complete or partial representation of the genetic information of an

organism. A genomic DNA library is constructed with DNA isolated directly from an organism while a cDNA (complementary DNA) library is made using reverse transcription products of RNA isolated from the organism. Thus a genomic DNA library represents the entire genetic complexity of the organism while a cDNA library contains only those sequences which are transcribed. The DNA molecules comprising the library are carried in extrachromosomal genetic vehicles (vectors) providing their maintenance, replication, and selection in bacterial host cells. Introduction of the library into bacteria allows the isolation and production of specific sequences in quantities sufficient for biochemical analysis. Three general sorts of vectors have been developed to permit cloning of foreign DNA in bacteria: plasmids, viruses, and cosmids. Selection of the appropriate vector depends on the application at hand. Plasmids are useful for cloning and ultrastructural analysis of small DNA fragments, viruses (specifically bacteriophage λ) are useful for routine generation of eukaryotic genomic DNA libraries, and cosmids are used for cloning very large DNA fragments. Additional vectors designed for specific purposes include bacteriophages M13 (for nucleotide sequence determination and generation of single-stranded DNA probes) and λgt11 (allowing expression of foreign DNA in bacteria). Two strategies are available for insertion of foreign DNA fragments into host vectors in the construction of recombinant DNA molecules. First the vector is opened using a restriction endonuclease and then DNA is inserted either by ligation or by annealing comlementary homopolymer nucleotide tracts.

In this work a library of Euplotes eurystomus macronuclear genomic DNA was constructed essentially by the technique of Swanton et al. (143). Briefly, the 3' termini of purified macronuclear DNA were extended with poly(dC) using terminal deoxynucleotidyl transferase and this DNA was annealed to the plasmid vector pUC9 which had been linearized by Pst I and similarly tailed with poly(dG). The resulting recombinant plasmids were introduced into E. coli DH5 by the Hanahan procedure. Screening of the library with the 5S rRNA gene of Tetrahymena resulted in the isolation of the Euplotes 5S rRNA gene.

Nucleotide Polymerization by Terminal Deoxynucleotidyl Transferase

Euplotes macronuclear DNA was isolated as previously described. The 3'-OH termini of the DNA molecules were extended by reaction with dCTP catalyzed by terminal transferase (143, 172-174). Terminal transferase and bovine serum albumin (BSA; utrapure, nuclease-free) were obtained from BRL, dCTP from Sigma, and [α^{32}P]-dCTP (specific activity = 3200 Ci/mmole) from ICN Radiochemicals. Reaction was at 22°C in a final volume of 300 µl containing 10 µg DNA, 0.5 mg/ml BSA, 50 µM dCTP (final specific activity = 3.33 Ci/mmole), and 150 units of enzyme in a buffer of 100 mM potassium cacodylate, 2 mM $CoCl_2$, and 0.2 mM dithiothreitol (pH 7.2). Besides the buffer composition and temperature, important reaction parameters include the concentration of DNA 3' termini, the concentration of enzyme, the concentration of free nucleotide, the ratio of free nucleotide to 3' ends, and the ratio of

enzyme to 3' ends. Euplotes macronuclear DNA has a number-average size of 1836 bp, thus 10 µg represents 16.5 pmoles of 3' ends; in a 300 µl reaction this is equivalent to a final DNA concentration of 33.3 µg/ml or 55.0 pmoles 3' ends/ml. Enzyme was used at a final concentration of 500 units/ml and dCTP at 50 µM. The general strategy for using terminal transferase involves assembling a reaction in which enzyme and free nucleotides are in excess over 3' DNA termini so as to optimze the proportion of DNAs that are substrates (that is, the proportion of DNA molecules which receive tails) (see Ray Wu's work: 173, 174). Terminal transferase operates processively (as opposed to distributively) and the rate limiting step is initiation of the polymerization reaction; thus the reaction is constructed such that all 3' DNA termini are tailed simultaneously. If the enzyme was present in limiting amounts, then one of two undesirable results would occur: either only small proportion of DNAs would be tailed, or, if the reaction was allowed to proceed, a high proportion of tailed DNA molecules would be obtained but those tails would be much too long. Therefore reaction conditions are chosen so that all of the DNAs are tailed simultaneously (requiring an excess of enzyme and nucleotides over 3' termini) and the tail length is adjusted by stopping the reaction at the appropriate time. For this purpose a ratio of ∿10 enzyme units/pmole 3' ends was empirically observed to be adequate. The ratio of moles dNTP/moles 3' ends should exceed 20. In this work, ∿9.1 enzyme units/pmole 3' ends were used with a ∿910-fold excess of dNTP/3' ends.

A pilot assay was assembled as described above and reaction aliquots were removed at various time points and terminated by quenching in an excess of ice-cold 0.1 M Na_2EDTA (pH 8.0). A trace of [α^{32}P]-dCTP was included in the reaction to monitor incorporation as a function of time. At various intervals, radioactivity incorporated into the DNA was quantitated by spotting a portion of the quenched reaction aliquot onto each of four Whatman DE-81 paper disks (154). After air drying for several minutes, two of the disks were washed six times, 5 minutes each, in 500 ml of 0.5 M Na_2HPO_4 at room temperature. These disks were then rinsed in distilled water for 1 minute followed by 95% (v/v) ethanol for 1 minute. All disks were dried using an infrared heat lamp and were radioquantitated by aqueous scintillation counting. The unwashed disks measured the total radioactivity in the sample and the washed disks measured the radioactivity incorporated in DNA. Duplicate measurements were made for each time point and the results were averaged. Since the proportion of radioactivity incorporated is equivalent to the proportion of dCTP incorporated, the number of residues added per 3' end was readily apparent. These calculations were corrected by subtracting the background from washing (clean, blank disks were included in the washes and the radioactivity adsorbed to them was determined) and the backround at time zero (an aliquot of reaction was removed prior to addition of enzyme and was spotted, washed, and counted as above). (Implicit in these calculations is the assumption that 100% of the DNA molecules are substrates for the enzyme and do in fact receive tails. However, even under optimum

conditions only 70-80% of the molecules are observed to receive 3' extensions, thus the true tail length is always greater than that predicted by theory.) The corrected number of residues added per 3' end was plotted as a function of time.

From this pilot experiment it was evident that a reaction time of 1 minute would produce the desired homopolymer tail length of 10-15 residues. Therefore a preparative tailing reaction was performed identically to that above except that the entire reaction was terminated after 1 minute by addition of an equal volume of ice-cold 0.2 M Na_2EDTA (pH 8.0). An aliquot of the quenched reaction was removed and incorporation was measured as above to verify that the proper tail length was generated. The tailed DNA was purified from the other reaction products by extraction with phenol, with phenol/chloroform, and with chloroform followed by three cycles of precipitation from ethanol as previously described. The clean DNA pellet was resuspended in 100 µl of sterile TE (pH 8.0) and stored at -20°C.

Annealing Complementary DNAs

The plasmid vector pUC9 was chosen as the cloning vehicle because of its small size and the utility of the multiple cloning site (the MCS, also known as the polylinker) (175). This plasmid contains a β-lactamase coding function (conferring resistance to ampicillin and thus providing a selectable marker), an origin of replication, and a β-galactosidase (lac Z) gene α-complementation unit. Within the 5' coding region of the lac Z gene is the polylinker, a set of nine unique

restriction enzyme sites. Introduction of DNA into any of these enzyme sites disrupts the β-galactosidase coding activity thereby providing a colorimetric assay for insertion when cells are plated on media supplemented with IPTG (an inducer of the gene) and Xgal (a chromogenic substrate which turns blue when acted on by the enzyme). The polylinker is very useful also because it provides a simple and rapid means of determining the restriction enzyme map of inserted DNA, of generating fragments for subcloning or sequencing, and of performing deletion mutagenesis.

pUC9 DNA which had been linearized with Pst I (in the polylinker) and tailed with poly(dG) by terminal transferase was obtained from Pharmacia (average tail length = 12 residues). Recombinant plasmids were constructed by annealing linear, poly(dC)-tailed Euplotes macronuclear DNA with linear, poly(dG)-tailed pUC9 DNA (143, 154). Although the immediate product of the annealing reaction contains single-strand gaps because (in most cases) the complementary homopolymer tracts are of unequal length, these are filled in and ligated by the bacteria. This procedure results in the restoration of Pst I sites immediately flanking the G•C tracts surrounding the insert, thus the insert (plus flanking G•C tracts) can be retrieved from the vector by digestion of the recombinant plasmid with Pst I. A reaction was assembled containing 115 ng Euplotes DNA and 85 ng pUC9 DNA in 200 µl of 100 mM NaCl, 10 mM Tris•HCl, and 1 mM Na_2EDTA (pH 8.0) by combining concentrated stock solutions of DNA and buffer components. (All reaction ingredients except DNA were sterilized by

microfiltration.) The resulting solution contained a molar ratio of insert:vector = 2:1 and a total DNA concentration of 1 μg/ml. The annealing reaction was warmed to 75°C for 5 minutes, slow-cooled and incubated at 42°C for at least 2 hours, slow-cooled to room temperature for at least 2 hours, and slow-cooled to 4°C for at least 1 hour. Working solutions of chimeric plasmids so generated were stored at 4°C while library archives were kept at -20°C.

Transformation and Transfection of Bacteria

Central to genetic engineering is the introduction of recombinant DNA molecules into cells. Transformation is the introduction of plasmid DNA into cells and transfection is the introduction of viral DNA. Competent cells are those rendered able to accept foreign DNA. Many techniques have been developed for the introduction of exogenous DNA into cells including for bacteria calcium chloride-mediated transformation, rubidium chloride-mediated transformation, infection with bacteriophage λ, and infection with bacteriophage M13, and for eukaryotic cells precipitation with calcium phosphate, precipitation with polyamines, precipitation with hydroxyl appetite, protoplast fusion, liposome encapsulation, osmotic shock, electroporation, retroviral infection, and microinjection. With the exception of viral infection, all of these techniques are extremely inefficient. In general, only a small fraction of cells survive the process (∿10%) and of those only a very small fraction (∿0.1%) are competent. This problem is overcome by using enormous numbers of cells and DNA molecules. To reduce the large background

of nontransformed cells, transformants are maintained under selective pressure so that only cells which harbor the vector are able to survive.

For the present study, the rubidium chloride-mediated transformation procedure of Hanahan (176) was adopted because this method is about two orders of magnitude more efficient than the calcium chloride technique. (Although transformation efficiency is not as crucial for subcloning, it is very important during library generation; a large number of clones are usually required in order to isolate the particular sequence of interest.) During library screening, frozen competent E. coli DH5 cells (library efficiency), commercially obtained from BRL, were used. (These cells are rendered competent by the Hanahan procedure and frozen at -70°C.) For other purposes, competent cells were prepared in the following manner. A 5 ml culture in SOB medium (described previously) was inoculated with a single bacterial colony from a plate and incubated overnight at 37°C, 250 rpm. The next day a subculture in SOB was inoculated with 1/100 volume of the fresh overnight culture (typically 0.15 ml of the overnight culture was added to 15 ml of fresh SOB). (When the competent cells were to be used for transfection by bacteriophage M13 DNA, a 1/20 dilution subculture in SOB was begun simultaneously for use as an indicator lawn.) The 1/100 subculture was grown at 37°C, 250 rpm until an A_{600} of 0.15-0.20 was attained (1½-2 hours). A portion of the subculture (typically 9 ml--this is enough for 4 transformation reactions) was then transferred to a sterile Falcon 2057 tube (Falcon 2059 tubes may also be used) and chilled on ice for 10-15 minutes (cold cells pellet better).

The chilled cells were pelleted by spinning in an SS34 rotor, 5000 rpm, 10 minutes, 4°C. The medium was decanted and the cell pellet was resuspended by vortexing in 1/3 volume (3 ml here) of ice-cold transformation buffer (TFB). TFB is 100 mM RbCl, 10 mM K-MES, 45 mM $MnCl_2$, 10 mM $CaCl_2$, and 3 mM $HACoCl_3$ (hexamine cobalt trichloride) (pH 6.3). After incubation on ice for 15 minutes, cells were centrifuged again as above. The cells were resuspended by vortexing in 1/12.5 times their original volume (720 µl here) of ice-cold TFB. Dimethyl sulfoxide, DMSO, was added to 3.5% (v/v) (25.2 µl per 720 µl TFB or 35 µl per ml TFB); the solution was mixed by gentle swirling or tapping and incubated on ice for 5 minutes. (Cells may not be vortexed after addition of DMSO as they are very fragile in the presence of this organic solvent. Spectral grade DMSO was obtained from Schwartz Mann and was frozen at -20°C in small aliquots upon arrival. An aliquot was removed from the freezer, used once, and the remainder was discarded. A fresh portion of ultrapure DMSO was used for each experiment.) Dithiothreitol (DTT) was added from a 1 M stock to a final concentration of 68 mM (54 µl per 720 µl TFB or 75 µl per ml TFB). The solution was mixed by gentle tapping or swirling and incubated on ice for 10 minutes. A second equal portion of DMSO (25.2 µl here) was then added, gently mixed, and incubation on ice was continued for 5 additional minutes. To prechilled Falcon 2057 tubes on ice were added 200 µl aliquots of the competent cells and 1-5 µl of a DNA solution. After gentle mixing, the cells were incubated in the presence of the DNA for 30 minutes on ice. Afterwards the cells were heat

shocked at 42°C for 90 seconds without agitation. Cells were subsequently chilled on ice for 1-2 minutes. When competent cells were transformed by plasmid DNA, to each tube then was added 0.8 ml of SOC (SOC medium is SOB + 20 mM glucose) and cells were allowed to recover for 1 hour at 37°C, 250 rpm before plating on antibiotic-supplemented media. (The recovery period allows cells to begin expression of their antibiotic resistance genes.) When competent cells were transfected with bacteriophage M13 DNA, following the heat pulse and chilling, portions of the reaction (typically 1/100, 1/10, and the rest) were combined with 200 µl of JM107 indicator cells (from the 1/20 dilution subculture), 10 µl of 100 mM IPTG, 50 µl of 2% (w/v) Xgal, and 3 ml of B top agar at 50°C (all described previously). These components were quickly mixed by swirling, poured onto B agar plates, and allowed to solidify. (The indicator cells form a confluent lawn over the plate on which to visualize bacteriophage plaques. The IPTG and Xgal are included for the colorimetric assay of insertion.) All plates (for both plasmid and phage vectors) were kept inverted in a moist incubator at 37°C overnight. Plates may be stored inverted in plastic bags at 4°C for a few weeks.

Library Screening by In Situ Colony Hybridization

Several methods exist for the isolation of a particular sequence from a library of DNA molecules. The gene of interest may be identified by in situ hybridization of bacterial colonies or bacteriophage plaques with a labelled, sequence-specific DNA or RNA probe (or by

plus-minus screening). A cDNA clone may be identified by hybridization selection, in which detection is by recognition of the in vitro translation product of mRNA cognate to the clones. A bacteriophage λ library may be screened by recombination in E. coli with plasmid πVX. Finally, when the gene is carried in an expression vector, its clone may be located by immunodection of the gene product. The first of these is the most straight-forward and general technique. The only prerequisite of this approach is possession of a probe with sequence complementary to the gene of interest. The procedure involves replicating the recombinant colonies or plaques onto a solid support matrix, denaturing and fixing the DNA on the membrane, and hybridizing with the labelled sequence probe.

After transformation of E. coli with the recombinant DNA construction, cells were plated on agar containing 100 μg/ml ampillicin and colonies were allowed to form overnight at 37°C. The next day, nitrocellulose replicas were made of the transformants essentially as described by Maniatis et al. (154). The master plates containing the original colonies were numbered and chilled inverted at 4°C for about an hour. Nitrocellulose disks (Schleicher and Schuell BA85, 0.45 μm) were numbered and washed by floating on distilled water followed by submersion in distilled water. Washed disks were air-dried by placement on a sheet of Whatman 3MM paper for several minutes. Nitrocellulose disks were layed on the surface of the master plates with care not to smear the colonies and orientation marks were placed on the disks and the plates to allow future realignment. Using blunt-end Millipore

forceps the disks were carefully peeled off of the master plates and transferred, cell side up, to fresh AMP plates. In this process some of the cells of each colony adhere to the nitrocellulose while part of each colony remains on the original agar plate. Two such nitrocellulose replicas were made of each master plate. Colonies were regenerated on the master plates by their brief incubation at 37°C. Master plates were sealed with parafilm and stored inverted at 4°C until the screening results were obtained. On the replicas at this time were placed, using sterile toothpicks, colonies of DH5 cells containing the plasmids pUC9 and pDP5 (harboring the 5S rRNA gene of Tetrahymena). These colonies respectively served as negative and positive controls to aid in distinguishing bona fide hybridization signals from nonspecific background. Colonies were allowed to form directly on the nitrocellulose disks by incubation on fresh AMP plates inverted at 37°C for ∿5 hours. After visible colonies appeared, but before they grew to maturity, the disks were transferred, colony side up, to fresh agar plates containing 100 μg/ml ampicillin and 170 μg/ml chloramphenicol. Incubation of inverted plates continued at 37°C for 12-16 additional hours to allow amplification of plasmids, thereby greatly increasing the signal to noise ratio.

Afterwards, the colonies were chilled at 4°C inverted for about an hour. During this time two sheets of Whatman 3MM paper were placed in each of three large photographic trays and were saturated with the following solutions. The paper in the first tray was saturated with 10% (w/v) SDS; the paper in the second tray was soaked with a solution of

0.5 M NaOH, 1.5 M NaCl; and the third with 1 M NaAc (pH 4.2). Excess liquid was poured off. Using Millipore forceps the chilled nitrocellulose replicas were removed from the plates and placed, colony side up, on the SDS-impregnated paper for 5 minutes at room temperature. This treatment lyses the cells and seems to limit diffusion of the plasmid DNA during subsequent steps. The disks were then transferred, colony side up, to the sheet of paper containing NaOH and NaCl and allowed to remain for 5 minutes at room temperature. This causes denaturation of the DNA. When transferring disks from one tray to another, the edge of the first tray was used as a scraper to remove as much fluid as possible from the underside of the filters. Next the disks were placed, colony side up, on the NaAc-saturated paper for 5 minutes at room temperature. This neutralizes the base used in the previous step. Finally the disks were air-dried, colony side up, on a dry piece of 3MM paper for 5-10 mintues. The replica nitrocellulose filters were sandwiched between paper towels and the denatured DNA was fixed by baking at 80°C for 2 hours in vacuo.

The baked filters were hydrated by floating on 6X SSC followed by submersion in 6X SSC in the usual manner. To reduce nonspecific backround adsorption of the probe, cellular debris was removed from the disks by washing in a solution of 1 M NaCl, 50 mM Tris·HCl, 1 mM Na_2EDTA, and 0.1% (w/v) SDS (pH 8.0) at 50°C. Several washes were performed in a glass baking dish on a rocking platform followed by several washes with gentle agitation in a hermetically sealed plastic bag in a 50°C water bath. One such wash was allowed to proceed

overnight. After prewashing, disks were stacked and prehybridized in a small plastic bag in a volume of 50 ml. Hybridization was with the gel-purified insert of pDP5, the 5S rRNA gene of Tetrahymena, in a volume of 25 ml and a probe concentration of 1×10^6 cpm/ml. Pre-hybridization, hybridization, washing, and autoradiography were as previously described.

When positive hybridization signals were obtained, the original master plates were aligned with the autoradiograms and the colonies overlying the signals were transferred to fresh AMP plates using sterile toothpicks. Colonies were formed on this second generation of master plates by incubation at 37°C for 8-10 hours. Nitrocellulose replicas were made and the clones were screened a second round in a manner identical to the first. Colonies which generated positive signals in the second round of screening were streaked out on fresh AMP plates to ensure that colonies originating from single cells (clones) were obtained. From these plates single, well-isolated colonies were picked and used to make the third generation of master plates. These plates were replicated and screened as above. Colonies generating a positive signal here were streaked on fresh AMP plates again to be certain that individual clones (originating from a single bacterial cell) were obtained. From these colonies, which had generated positive signals in three consecutive rounds of screening by in situ hybridization and which had been resolved (monodispersed) by two consecutive rounds of streaking, plasmids were isolated in 1 L cultures as previously described. In the ultimate round of screening, Euplotes macronuclear

DNA inserts were retrieved by digestion of the purified recombinant plasmids with Pst I. These purified DNAs were electrophoresed through agarose gels, transferred to nitrocellulose by the technique of Southern, and probed with the gel-purified 5S rRNA gene of Tetrahymena. Clones testing positive here were considered putitive isolates of the Euplotes 5S rRNA gene and were subjected to further analysis (see Results and Discussion).

Recovery of DNA from Agarose Gels

In addition to its analytical application, gel electrophoresis may also be used to preparatively purify specific DNA fragments. This is particularly helpful in the isolation of DNA sequences to be used as probes or for subcloning. A variety of methods exist for the recovery of DNA from gels; these are mostly based on diffusion or electrophoresis of the DNA out of the gel. DNA may be recovered from agarose gels by electroelution onto DE-81 paper, electroelution onto a dialysis membrane, electroelution within a dialysis bag, electroelution into a trough, electroelution into a salt bridge, by the freeze and squeeze technique, by diffusion and extraction with phenol, or by liquefaction of low-melting-temperature agarose followed by extraction with phenol.

In this work DNA was recovered from agarose gels by electroelution into a salt bridge using an electroelution chamber supplied by IBI. The device is constructed so that as DNA exits the gel by electromotive force it encounters a conduit of 7.5 M ammonium acetate (a salt bridge). The DNA stops migrating as it enters the conduit, as there

current is carried by the salt. Gel electrophoresis was performed as previously described. When it was important to preserve the structural integrity of the DNA (as for subcloning), two adjacent lanes were run for each DNA sample: a preparative lane and an analytical lane. Following electrophoresis the analytical lane was removed using a razor blade and stained with ethidium bromide. The gel was then reassembled on a UV transilluminator and the stained DNA was visualizd by fluorescence. The region of the gel in the preparative lane containing the fragment of interest was identified by correspondence to the analytical lane and was excised using a razor blade. It is advantageous to retrieve the DNA band in the smallest possible volume of gel, therefore thin gels were used and care was taken to excise only the region of the gel occupied by the desired DNA. The strategy outlined above was designed to minimize UV-induced nicking, to which ethidium bromide-stained DNA is especially susceptible. When maintainance of structural integrity was less important (as when the DNA fragment was to be used as a nick translated probe), only a preparative lane was run and the entire gel was stained. A gel slice containing the DNA band of interest was placed in the electroelution cell and electrophoresis was at 100 V for 1 hour in 0.5X TBE buffer. Following elution, the 7.5 M NH_4Ac, now containing the DNA, was aspirated from the conduit using a Pasteur pipette and diluted 3-4 fold with water used to rinse the channel. DNA was precipitated by addition of 2.5 volumes of absolute ethanol and chilling. The pellet was collected by centrifugation, rinsed, dried, and resuspended in 100 µl of sterile TE (pH 8.0) in the

customary fashion. The DNA was further purified by molecular exclusion column chromatography (Worthington minispin columns) and another cycle of precipitation from ethanol as above. Gel-purified DNA fragments were resuspended in sterile TE (pH 8.0) and stored at 4°C or -20°C.

Recovery of DNA from Polyacrylamide Gels

The geometry of polyacrylamide gels (the fact that they must be contained between two glass plates to exclude oxygen, an inhibitor of acrylamide polymerization) precludes use of some of the techniques of DNA recovery mentioned above for agarose gels. DNA may be recovered from polyacrylamide gels by electroelution of a gel slice within a dialysis bag, electroelution into a salt bridge, or by diffusion. In the present study a diffusion protocol known as "crush and soak" was applied (Mike Tindal, personal communication). Polyacrylamide gels were electrophoresed in the usual manner and gel slices containing the desired DNA fragments were excised as described for agarose gels. One gel slice was placed in a sterile Eppendorf tube and ground to a paste with a tight-fitting conical teflon pestle (Kontes). The paste was suspended in 400 µl of sterile 0.5 M NH_4Ac by vortexing and incubated at 37°C (or 42°C) overnight. The polyacrylamide particles were pelleted in a microfuge (12,000 x g) for 5 minutes at room temperature. The aqueous phase was aspirated using a Pasteur pipette with care to withdraw absolutely no polyacrylamide and was transferred to a clean, sterile tube on ice. The gel paste was reextracted with 300 µl of

fresh, sterile 0.5 M NH_4Ac for 1-2 hours at 37°C. After centrifugation as above, this aqueous layer was combined with the first. The gel was extracted a third time using 200 µl of 0.5 M NH_4Ac at 37°C for 1 hour followed by centrifugation as before. Insoluble debris was removed from the pooled aqueous phases by microcentrifugation for 5 minutes. Soluble material was carefully aspirated using a Pasteur pipette, transferred to a clean tube, recentrifuged, and reaspirated. In another fresh tube, DNA was precipitated from the clear solution by addition of 2.5 volumes of absolute ethanol and chilling. The pellet was collected by centrifugation, rinsed four times with 1 ml of 80% (v/v) ethanol at -20°C, dried, and resuspended in 100 µl of sterile water. The DNA was further purified by molecular exclusion column chromatography (Worthington minispin columns) and another cycle of precipitation from ethanol. Gel-purified DNA fragments were resuspended in sterile TE (pH 8.0) and stored at 4°C or -20°C.

M13 Subcloning

To prepare templates for nucleotide sequence determination, the Euplotes 5S rRNA gene was subcloned into bacteriophage M13. M13 is a single-stranded, filamentous bacteriophage of male *E. coli* which has been modified by Messing and coworkers (177-184) to render it an ideal cloning vector for the generation of DNA templates for sequence analysis by the dideoxy technique. The M13 phage vectors, as do the pUC plasmid vectors, contain the polylinker inserted within the 5' region of

a portion of the E. coli lac operon. This lac sequence will produce a peptide capable of α-complementation with a mutant form of the lac Z gene, lac Z ΔM15, carried on the F episome. Thus in a lac Z^- host cell which contains the F episome, M13 infection will generate a functional β-galactosidase enzyme visually detectable in the colorimetric assay using indicator plates supplemented with IPTG (inducer) and Xgal (chromogenic substrate) as previously mentioned for the pUC system. Insertion mutagenesis, by introduction of foreign DNA into the polylinker, will disrupt the coding function of the lac region of M13, α-complementation will not occur, and bacteriophage plaques will remain colorless. Thus the polylinker incorporates a convenient assay for insertion as well as providing a simple means of forced-orientation cloning and generation of specific deletion mutants of inserted DNA. In addition, a nucleotide sequence adjacent to the polylinker constitutes a primer binding site for the initiation of DNA synthesis. In different M13 constructions the polylinker has been installed in reverse orientations, allowing determination of the complete nucleotide sequence of both strands of the DNA fragment of interest. In vitro genetic manipulations are performed on the replicative form (RF) of the viral genome, a duplex DNA structure. M13 RF is isolated as a plasmid from infected E. coli, the polylinker is opened by restriction endonuclease digestion, a linear DNA fragment is inserted by ligation, and the recombinant double-stranded DNA construct is reintroduced in bacteria by transfection. Phage particles are then isolated from supernatants

of liquid cultures of infected bacteria, and purified phage DNA (single-stranded) serves as template for in vitro DNA synthesis using chain terminating dideoxynucleotides.

Insertion Mutagenesis and DNA Ligation

DNA ligase from phage T4 was used to seal nicks between two abutting DNA termini (that is, to catalyze the formation of a phosphodiester bond between the 3'-hydroxyl of one DNA terminus and the 5'-phosphate of another DNA terminus) and thereby generate covalently linked molecules (154, 184). Two general sorts of reactions were performed, according to the nature of the DNA termini being joined. In cohesive end ligation the termini contain complementary single-strand overhangs, as are generated by most restriction enzymes. In this case the DNA ends are capable of transiently annealing thus stabilizing the interaction of DNA termini and promoting the reaction. Such ligation reactions are carried out at low temperature to enhance the hybridization of the cohesive ends. In blunt-end ligation the DNA termini undergoing joining are blunt, that is both strands within a duplex DNA terminus are flush. In this situation the DNA ends are not able to stably interact and the reaction requires the simultaneous collision of two DNA termini and an enzyme molecule (plus other cofactors and substrates) with sufficient energy and in a suitable orientation. Thus blunt-end ligation reactions are substantially less efficient than cohesive-end ligations. The Km of the enzyme for

blunt-end substrates is 100 times greater than for cohesive-end substrates (154). A blunt-end ligation reaction is therefore usually assembled containing higher concentrations of DNA termini and enzyme. Since the DNA ends cannot anneal, there is no cause to incubate the reaction in the cold. Although the enzyme's optimum temperature is 37°C, one of the substrates, ATP, is unstable in aqueous solution at that temperature, so a compromise is struck and the reaction is carried out at 25°C.

The 5S rRNA gene of Euplotes was retrieved from the plasmid vector pUC9 by digestion of the recombinant plasmid p320 with the restriction endonuclease Pst I (as previously described) and purified from the cloning vehicle by gel electrophoresis (as previously described). Replicative form DNAs of M13mp18 and M13mp19 (differing only in the orientation of the polylinker) were linearized with Pst I and purified by extraction with phenol and precipitation from ethanol. The gel-purified Pst I-insert of p320 (that is, that 5S rRNA gene of Euplotes) was inserted into the Pst I site of M13mp18 and M13mp19 by cohesive end ligation. Two sets of reaction conditions were used, differing only in the molar ratio of insert:vector (both were successful). In one set of reactions (referred to as L1 and L2 when containing M13mp18 and M13mp19, respectively), 0.406 µg of vector DNA were combined with 0.054 µg of insert DNA and 1.4 Weiss units of T4 DNA ligase (BRL) in a volume of 20 µl of fresh 50 mM Tris·HCl, 10 mM $MgCl_2$, 1 mM ATP, 1 mM dithiothreitol, and 5% (w/v) polyethylene

glycol-8000 (pH 7.6) followed by incubation at 11°C overnight (12-24 hours). (Fresh buffer is necessary not only because the enzyme requires ATP but also because the enzyme is inhibited by phosphate, a hydrolysis product of ATP.) This represents a final DNA concentration of 23 μg/ml, a molar ratio of insert:vector = 1:1, and an enzyme concentration (volume specific activity) of 70 units/ml. Under these reaction conditions, where i (the concentration of DNA termini) is equal to j (the effective concentration of one end of a DNA molecule in the volume of the other end of the same molecule), intermolecular interactions and intramolecular interactions have an equal probability of occurrence (see 154). Since the sought-after ligation product must be formed in a two-step process involving the intermolecular joining of an insert molecule to a vector molecule and then the intramolecular circularization of this concatemer, reaction conditions were selected in which both intermolecular and intramolecular processes could proceed. In another set of reactions (referred to as L3 and L4 when containing M13mp18 and M13mp19, respectively), 0.363 μg vector DNA were combined with 0.097 μg insert DNA and 1.4 Weiss units of ligase in 20 μl of the above-mentioned buffer followed by incubation as before. This represents a final DNA concentration of 23 μg/ml, a molar ratio of insert: vector = 2:1, and an enzyme concentration of 70 units/ml. Following ligation, the reaction products were monitored by agarose gel electrophoresis and the chimeric DNA constructs were introduced into E. coli JM107 by RbCl-mediated transfection (previously described).

White (colorless) plaques were picked from indicator plates using sterile Pasteur pipettes and RF was isolated by the miniprep protocol (to be described later). The construction was verified by digestion of recombinant M13 RFs with Pst I and analysis by agarose gel electrophoresis. The orientation of the inserts within the vectors was determined by digestion of the recombinant RFs with other restriction enzymes which asymmetrically cleaved the insert and also cut the polylinker (see Results and Discussion section for details). Four parental M13 subclones were thus isolated: the intact Euplotes 5S gene was introduced in both orientations into both M13mp18 and M13mp19. These parental subclones were then plaque-purified to ensure that pure clones were obtained. This was achieved by transfection of JM107 by the various constructions (RFs) followed by isolation of single, well-resolved plaques. RFs were then prepared from the plaque-purified parental M13 subclones and were stored in TE (pH 8.0) at -20°C.

Deletion Mutagenesis and DNA Ligation

To generate templates allowing initiation of the sequencing reactions within the interior of the Euplotes 5S gene, a series of deletion mutants were constructed from the four parental M13 subclones (see Results and Discussion section for more details). This was achieved by digestion of the recombinant M13 RFs with a restriction enzyme (or two enzymes) which cleaved the insert (the intact Euplotes

5S gene) once and the M13 vector once (in the polylinker). This double cut converted the circular RF into two linear molecules: a long piece containing the vector plus part of the insert, and a short piece composed of insert sequences spanning the polylinker to the restriction enzyme site within the insert. These linear DNAs were purified by extraction with phenol and precipitation from ethanol. A cohesive end DNA ligation reaction was assembled containing approximately 0.5 μg of the linear RF fragments described above and 2.8 Weiss units of T4 DNA ligase in 100 μl of the above-mentioned buffer and was incubated over night at 11°C. This represents a final DNA concentration of 5 μg/ml and an enzyme concentration of 28 units/ml. Under these reaction conditions, where i is less than j, intramolecular interactions are statistically preferred over intermolecular interactions, therefore unimolecular circularization (and thus deletion) is favored. Simply stated, in concentrated solutions of linear DNA molecules the termini of two different DNAs are likely to collide and be ligated whereas in dilute solutions one end of a DNA molecule is more likely to collide with the opposite end of the same molecule. Therefore when $i > j$ intramolecular circles are favored, when $i < j$ intermolecular concatamers are favored, and when $i = j$ the two processes proceed at equal rates (see 154). Since in this application the sought-after product was a deletion mutant formed by intramolecular recircularization of the large fragment, reaction conditions were chosen such that $i > j$. The many reaction products generated included the recircularized large restriction

fragment of the parental RF, containing the vector and part of the original insert. Thus the small fragment of the parental RF, containing the remainder of the insert, was deleted. A family of deletion mutants constructed in this manner were reintroduced into JM107 by RbCl-mediated transfection. Colorless plaques were isolated and new RF prepared as before. The accuracy of the constructions (deletions) were verified by digestion of the deleted recombinant RFs with appropriate restriction enzymes and analysis by agarose gel electrophoresis. Mutant RFs, as the parental subclones, were stored in TE (pH 8.0) at -20°C.

For the construction of certain deletion mutants (to be described in detail later), it was necessary to join flush-ended DNA molecules. DNA was cleaved with the appropriate restriction enzymes and purified by extraction with phenol and precipitation from ethanol. The single-stranded overhangs left by the restriction enzymes were filled in using the Klenow fragment of E. coli DNA polymerase I (see next section) resulting in blunt-ended DNA molecules. (This was necessary because two different restriction enzymes, which generated non-compatible ends, were used to excise the fragment being deleted.) The DNA was again purified by extraction with phenol and precipitation from ethanol in the usually manner. For blunt-end DNA ligation, approximately 0.5 μg of linear DNA was combined with 2.8 Weiss units of T4 DNA ligase in 100 μl of the above-mentioned buffer plus 0.025% (v/v) Noniodet P-40 (NP-40) and incubated overnight at 25°C. The

nonionic detergent NP-40 was included to increase the hydrophobicity of the solvent and thereby promote interaction of DNA termini. The above reaction contained a final DNA concentration of 5 μg/ml and an enzyme concentration of 28 units/ml, favoring circularization over concatamerization as discussed above. The ligation reaction products were used to transfect JM107, clear plaques were used to infect liquid cultures of JM107 from which RF was isolated, and the accuracy of the construction was verified as before. Purified RF DNA was stored in TE (pH 8.0) at -20°C.

Generation of Blunt DNA Termini

For joining of noncohesive DNA termini generated by different restriction enzymes producing incompatible (noncomplementary) ends, blunt-ended DNA termini were formed and these were ligated as described above. Single-strand DNA overhangs may be removed by S1 nuclease digestion or by exonucleolysis or, conversely, a 3' gap may be filled in using the Klenow fragment of E. coli DNA polymerase I (154, 184). Since it was desired in this case not to compromise informational content, the latter method was utilized. Circular recombinant RF DNA was double-digested with the appropriate restriction enzymes and purified in the usual manner. A reaction was assembled containing ∿0.5 μg DNA, 50 μM each dNTP, and one unit of Klenow enzyme (BRL) in 20 μl of 50 mM Tris•HCl, 10 mM $MgCl_2$, and 50 mM NaCl (pH 8.0) and was incubated at 37°C for 30 minutes. When synthesis was

complete, the reaction was diluted to a volume of 100 µl by addition of sterile TE (pH 8.0), extracted with phenol, and the DNA was precipitated from ethanol. Deletion mutants were constructed by ligation as described above for blunt end DNA molecules. The mutants were screened and stored as usual.

In summary, M13 subclones were constructed by insertion mutagenesis via cohesive end DNA ligation under reaction conditions equally favoring concatamer and circle formation and by subsequent site-directed deletion mutagenesis using both cohesive end and blunt end ligation reactions under conditions favoring intramolecular circularization. Blunt-end DNA termini were generated using the Klenow fragment.

Small Scale Preparations of Viral RF and Plasmid DNAs

Replicative form DNA of bacteriophage M13 (a covalently closed circular duplex DNA structure) was isolated as a plasmid from infected bacteria by a modification of the alkaline-SDS method described by Maniatis (154), a modification of the original method of Birnboim and Doly (185). Using sterile Pasteur pipettes, colorless plaques were plugged from indicator plates generated in transfection experiments and were transferred to 5 ml aliquots of LB medium in Falcon 2059 tubes. To each tube was added 50 µl (1/100 volume) of a fresh overnight JM107 culture (inoculated from a single colony). Cultures were incubated at 37°C for 3 hours with vigorous shaking (∿250 rpm). The

infected cells were pelleted in a desk top clinical centrifuge at full speed for 10 minutes at room temperature. Media supernatants, containing phage particles, were decanted into fresh Falcon tubes and stored at 4°C (without chloroform) for future use. Pellets were drained by inversion for several minutes and transferred to ice. Bacterial cell pellets were resuspended in 100 µl of cold GTE (50 mM glucose, 25 mM Tris·HCl, and 10 mM Na_2EDTA (pH 8.0)) containing 4 mg/ml freshly added lysozyme, were transferred to sterile 1.5 ml microfuge tubes, and were incubated at room temperature for 5 minutes (or on ice for 10 minutes depending on the number of samples). (This brief lysozyme treatment does not completely lyse the cells but rather permeabilizes the cell membrane thereby allowing small soluble molecules, such as plasmids, to diffuse out.) Two hundred µl of ice-cold 1% (w/v) SDS, 0.2 M NaOH were then added; the sample was mixed by gentle tapping; and incubation was on ice for 5 minutes. One hundred fifty µl of an ice-cold potassium acetate solution were then added, the sample was mixed by gentle tapping, and incubation was continued on ice for an additional 5 minutes. (The potassium acetate solution was prepared by combining 60 ml of 5 M potassium acetate, 11.5 ml of glacial acetic acid, and 28.5 ml of H_2O resulting in a solution 3 M in potassium, 5 M in acetate, and with pH ∿4.8.) [The strategy is to liberate the small plasmid DNA molecules by permeabilizing the cell with lysozyme, to denature the linear chromosomal DNA with hydroxide, to complex the proteins with SDS, to rapidly neutralize the solution with acetate, and

to precipitate the detergent-protein complex with potassium thereby forming an insoluble matrix which entraps cell debris and large chromosomal DNA molecules.] The solution was then clarified by Eppendorf microcentrifugation at 12,000 x g for 5-10 minutes at room temperature. The clear supernatant, containing soluble plasmid (or viral RF) DNA, was carefully decanted into a clean Eppendorf tube so as not to disturb the pellet. The solution was then extracted with an equal volume (~450 µl) of PCA by vortexing until emulsification, phases were separated by centrifugation for 5 minutes, and the aqueous fraction was aspirated and transferred to a fresh tube. DNA was precipitated by addition of 2 volumes of absolute ethanol followed by chilling. (Additional salt is unnecessary here as the solution is already rich in potassium acetate.) The DNA pellet was collected by centrifugation, washed, and dried in the usual manner. The dry DNA pellet was then resuspended in 50 µl of sterile TE (pH 8.0) containing 20 µg/ml RNase A (added from a previously boiled 10 mg/ml stock solution of Boehringer Mannheim ultrapure pancreatic RNase) and was incubated at 65°C for 10 minutes. (The elevated temperature is to relax double-helical RNA structures which are relatively resistant to degradation.) The DNA was commonly then used directly (5 µl of the 50 µl sample for analysis by restriction enzyme digestion and agarose gel electrophoresis) or alternatively the DNA was purified from the enzyme and small oligoribonucleotides by extraction with phenol followed by two cycles of precipitation from ammonium acetate/ethanol. This

procedure yields approximately 2.5-5.0 μg of viral RF per 5 ml culture. DNAs prepared in this manner were stored in sterile TE (pH 8.0) at -20°C.

Preparation of Bacteriophage M13 Genomic DNA

For use as template in nucleotide sequence determination by the dideoxy technique, genomic DNA from bacteriophage M13 (a single-stranded DNA structure) was isolated from the virion essentially as described in the BRL manual (184). After the accuracy of a given recombinant M13 construction had been verified by digestion of its RF DNA with restriction enzymes, viral DNA of the clone was prepared using as inoculum its cognate phage supernatant put aside during the RF isolation. One hundred μl of phage supernatant and 100 μl of a fresh overnight JM107 culture (inoculated with a single colony) were combined in a sterile Falcon 2057 tube. To allow absorption of the phage onto the bacterial cell surface (specifically, onto the F pilus), incubation was at 37°C for 10 minutes without agitation. After infection, 5 ml of LB medium was added (a 50-fold dilution of the original overnight culture) followed by incubation at 37°C, 250 rpm for 6 hours. (As the culture develops during this period, phage are produced within the cell and extruded into the media. Bacteriophage M13 does not cause cell lysis, but merely reduces the growth rate of the host.) Afterwards, 1.5 ml of the culture was transferred to a sterile microfuge tube. (From the remainder of the 5 ml culture, cells

were cleared by centrifugation and the phage supernatant was stored in a fresh tube at 4°C (without chloroform) to retain a virus stock of the clone in the event that additional template DNA be required in the future.) Infected cells were precipitated from the 1.5 ml aliquot by microcentrifugation (12,000 x g) for 5 minutes at room temperature. The top 1.2 ml of the culture supernatant was aspirated using a mechanical pipet with care to avoid the cell pellet and was transferred to a fresh microfuge tube. Then was added to the supernatant 1/4 volume (300 µl) of sterile 20% (w/v) PEG 4000-6000 (polyethylene glycol of molecular weight range 4000-6000), 2.5 M NaCl. After thorough vortexing, the phage were precipitated at room temperature (or on ice) for 15 minutes. (Ethylene glycol polymers interact with viral coats so as to bridge physically disparate phage particles and thereby form an insoluble PEG-phage matrix.) The viral pellet was collected by microcentrifugation for 10 minutes at room temperature. The supernatant was decanted and the pellet drained by inversion for several minutes; small droplets of solvent residual on the inner tube surface were removed by blotting with a Kimwipe. The small viral pellet (white when visible) was resuspended in 100 µl of sterile STE (100 mM NaCl, 10 mM Tris•HCl, and 1 mM Na_2EDTA (pH 8.0)) and was extracted with 50 µl of PCA by vortexing 15-20 seconds. This treatment lyses the virion and removes proteins from the viral DNA. The phases were separated by microcentrifugation for 5 minutes at room temperature. Using a mechanical pipet, the upper 80 µl (and no more)

of the clear aqueous phase were aspirated with care to avoid the inter face and were transferred to a clean tube. DNA was precipitated by addition of 200 µl (2.5 volumes) of absolute ethanol and chilling at -20°C overnight. The pellet was collected by centrifugation, rinsed with 1 ml of 80% (v/v) ethanol at -20°C, and dried as usual. The dry DNA pellet was resuspended in 20 µl of sterile TE (pH 8.0) and was stored at -20°C. Single-stranded M13 phage genomic DNA prepared in this manner served as template in the sequencing reactions.

DNA Sequencing Reactions

The first level of storage of genomic information, and the easiest to investigate, is the primary nucleotide sequence of DNA. The first practical method of sequencing DNA was developed by Maxam and Gilbert (188, 189). In this tripartite technique, end-labelled DNA is modified at positions randomly distributed along the molecule by chemical reactions specific for one or two bases. Four sets of limited modification reactions are performed (G, G+A, T+C, C) in each of which about 1% of the positions are substituted. The base specificity and the statistical nature of the distribution of modified bases reside in this first reaction. The subsequent cleavage reactions are quantitative. In the second reaction, modified bases are displaced from their sugars by piperidine. In the third reaction, piperidine goes on to catalyze β-elimination of the phosphates from the empty sugars resulting in strand scisson. In summary, the three consecutive steps are (from

189): "modification of a base, removal of the modified base from its sugar, and DNA strand scisson at that sugar . . . . DNA will break only at a sugar without a base, and only an appropriately modified base can be removed from its sugar." A second convenient approach to DNA sequencing was invented by Sanger (184, 190-193). In this technique, single-stranded DNA is used as template for the in vitro synthesis of the labelled complementary strand by the Klenow subfragment of E. coli DNA polymerase I. Four reactions are performed (A, G, C, T), each containing all four dNTPs plus one 2',3'-dideoxynucleoside triphosphate analog (ddNTP). When incorporated into a nascently growing deoxy nucleotide polymer, these dideoxynucleotides cause chain termination (base-specific cessation of elongation) because they provide no 3' hydroxyl group on which to form the next phosphodiester bond (the 3' terminius is no longer a substrate for elongation). The positions of chain termination are randomly distributed with the probability of termination depending on the ddNTP/dNTP ratio. The advantage of the chemical technique is that it is immune to sequence-specific topological constraints which can hinder the enzyme (secondary structures of the template or the nascent strand can impede locomotion of the enzyme or cause it to make errors). The advantages of the dideoxy technique include that it is much faster and simpler to perform, it requires much less radioactivity, and it involves the use of no hazardous chemicals (such as dimethyl sulfate, piperidine, and hydrazine). Both techniques utilize high resolution denaturing polyacrylamide gel electrophoresis for

analysis of the reaction products. After autoradiography, the nucleotide sequence of the DNA substrate is simply read from the series of bands in the base-specific lanes of the gel.

Single-stranded genomic DNA from phage M13 served as template for sequencing by the Sanger (dideoxy) technique. Of the 20 µl sample of template DNA in TE (pH 8.0) (see Preparation of Bacteriophage M13 Genomic DNA), 5 µl (0.5-1.0 µg) were combined in a 1.5 ml Eppendorf tube with 2 µl (4 ng) of primer DNA (BRL; a 15 base universal M13 primer which hybridizes to a sequence on M13 adjacent to the polylinker), 1 µl of 10X polymerase buffer, and 2 µl of H_2O resulting in a volume of 10 µl. (The 10X polymerase buffer is 70 mM Tris•HCl, 70 mM $MgCl_2$, and 500 mM NaCl (pH 7.5).) These components were mixed by brief centrifugation (∿1 sec) and the tube was placed in a boiling water bath for 4 minutes. The bath was then removed from heat and in it the sample was slowly cooled to room temperature (∿30 minutes) to allow annealing of the primer to the template. Afterwards, 3 µl of [α^{32}P]-dTTP (NEN; 32 µM in 10 mM tricine, specific activity = 3000 Ci/mmol) was added. In each of four fresh tubes (corresponding to the four base-specific synthesis reactions) were then deposited 3 µl of the above template--primer mix (containing label), 2 µl of a nucleotide solution specific for each particular base (see Tables 4 and 5), and 1 µl (∿1 unit) of Klenow enzyme resulting in a reaction volume of 6 µl (see Table 5 for final concentration of all reaction components). (The Klenow enzyme BRL was diluted 1:20 in 50% (v/v)

Table 4. Summary of DNA sequencing reactions.

Annealing Reaction

1. combine: 5 µl template in TE (0.5-1.0 µg)
   2 µl prmer (4 ng)
   1 µl 10X polymerase buffer
   2 µl H_2O
   10 µl total volume
2. mix, spin
3. boil 4 minutes, spin
4. slow cool to room temperature ∿30 minutes, spin
5. add 3 µl [α^{32}P]-dTTP (NEN, 3.32 µM, 3000 Ci/mmol)
   13 µl total volume
6. mix, spin
7. store on ice

   (note: mix by tapping, spin 1 second)

Synthesis Reaction

1. for each template label four tubes: A, G, C, T
2. on the inner surface of each tube deposit:

   3 µl annealed template:primer mix
   2 µl appropriate nucleotide mix (see below)
   1 µl diluted enzyme (1 unit) (see below)
   6 µl reaction volume

3. spin to combine ingredients and initiate reaction
4. incubate 37°C 15 minutes, spin
5. add 1 µl cold chase, spin (volume = 7 µl)
6. incubate 37°C 15 minutes, spin
7. add 3 µl stop dye mix, spin (volume = 10 µl)
8. boil 4-5 minutes, spin

Table 4 (continued).

9. quick chill in ice-water bath

10. load sample on gel (1-2 µl per lane)

11. store remaining sample at -20°C (2-3 days maximum)

12. reboil and chill sample immediately before each use

(note: mix by tapping, spin 1 second)

Buffers for Sequencing Reactions

1. 10X Polymerase Buffer:

   70 mM Tris·HCl (pH 7.5)
   70 mM $MgCl_2$
   500 mM NaCl

2. Klenow Enzyme Dilution Buffer:
   (dilute BRL Klenow enzyme 1/20 and use 1 µl)

   100 mM KPO_4 (pH 7.5)
   50% (v/v) glycerol

3. Cold chase: 0.5 mM each dNTP in H_2O (store at -70°C)

4. Stop Dye Mix A:
   (make fresh monthly; store at 4°C)

   95% (v/v) deionized formamide
   10 mM Na_2EDTA (pH 8.0)
   0.1% (w/v) bromophenol blue
   0.1% (w/v) xylene cyanol

5. Stop Dye Mix B:

   7 M urea (ultra pure)
   0.1% (w/v) bromophenol blue
   0.1% (w/v) xylene cyanol

6. dNTP, ddNTP stocks are 10 mM in H_2O (store at -70°C) dilute 1/20 in H_2O as needed (see below)

7. Nucleotide mixes (see below)

(note: 0.22 µm filter sterilize whenever possible)

Table 5. Nucleotide mixes for DNA sequencing reactions (using T label).

	A Reaction			C Reaction			G Reaction			T Reaction		
Reagent	[Stock]	For 100 µl Stock	[Reaction][3,4]	[Stock]	For 100 µl Stock	[Reaction][3,4]	[Stock]	For 100 µl Stock	[Reaction][3,4]	[Stock]	For 100 µl Stock	[Reaction][3,4]
ddNTP[1,2]	1 mM	10 µl 10 mM	333 µM	0.5 mM	5 µl 10 mM	167 µM	0.5 mM	5 µl 10 mM	167 µM	0.5 mM	5 µl 10 mM	167 µM
dATP[2]	20 µM	4 µl (1/20)	6.67 µM	100 µM	1 µl 10 mM	33.3 µM	100 µM	1 µl 10 mM	33.3 µM	75 µM	15 µl (1/20)	25 µM
dCTP[2]	100 µM	1 µl 10 mM	33.3 µM	20 µM	4 µl (1/20)	6.67 µM	100 µM	1 µl 10 mM	33.3 µM	75 µM	15 µl (1/20)	25 µM
dGTP[2]	100 µM	1 µl 10 mM	33.3 µM	100 µM	1 µl 10 mM	33.3 µM	30 µM	5 µl (1/20)	10 µM	75 µM	15 µl (1/20)	25 µM
dTTP[2]	2.5 µM	0.5 µl (1/20)	1.22 µM	2.5 µM	0.5 µl (1/20)	1.22 µM	2.5 µM	0.5 µl (1/20)	1.22 µM	2.5 µM	0.5 µl (1/20)	1.22 µM
Polymerase buffer	1.5X	15 µl 10X	0.885X[5]	1.5X	15 µl 10X	0.885X[5]	1.5X	15 µl 10X	0.885X[5]	1.5X	15 µl 10X	0.885X[5]
DTT	10 mM	10 µl 0.1 M	3.33 mM	10 mM	10 µl 0.1 M	3.33 mM	10 mM	10 µl 0.1 M	3.33 mM	10 mM	10 µl 0.1 M	3.33 mM
H_2O	--	59 µl	--	--	65 µl	--	--	63 µl	--	--	24 µl	--

1. For A reaction use ddATP, C reacton ddCTP, G reaction ddGTP, T reaction ddTTP.
2. dNTP, ddNTP stocks are 10 mM in H_2O; dilute 1/20 in H_2O as needed.
3. Includes contribution from anneal mix. Values are for 6 µl reaction before chase.
4. Enzyme contributes 16.7 mM KPO_4 (pH 7.5) to reaction.
5. 0.885X pol buffer is 6.19 mM Tris (pH 7.5), 6.19 mM $MgCl_2$, 44.2 mM NaCl.

glycerol, 100 mM KPO_4 (pH 7.5) prior to use.) Pipetting accuracy and precision are critical here. These reaction components were deposited on the inner tube surface as three separate droplets using a Pipetman P20 or a Hamilton syringe; the ingredients were combined and the synthesis reaction initiated by centrifugation for 1 second. Incubation was at 37°C for 15 minutes. At this time 1 µl of cold chase solution was added to each tube. (Cold chase is 0.5 mM each dNTP in H_2O. This replenishes the depleted nucleotide supply in the reaction providing for extended synthesis.) Incubation at 37°C was continued for an additional 15 minutes. The reaction was terminated by addition of 3 µl of stop dye mix (95% (v/v) deionized formamide, 10 mM Na_2EDTA (pH 8.0), 0.1% (w/v) bromophenol blue, and 0.1% (w/v) xylene cyanol) and was boiled for 4-5 minutes to denature the DNA and dissociate the enzyme from the nucleic acid. After the sample was quickly chilled in an ice/water bath, 1-2 µl were loaded per lane on a DNA sequencing gel. Samples prepared in this manner may be stored at -20°C for 2-3 days before analysis and are reboiled and chilled immediately before each loading.

DNA Sequencing Gels

Products of the DNA sequencing reactions were analyzed by high resolution denaturing polyacrylamide gel electrophoresis using a Poker Face DNA Sequencer electrophoresis apparatus manufactured by Hoefer, with gel dimensions of 40 cm x 35 cm x 0.35 mm, and an LKB power

supply. Both square-well and shark's tooth-type combs were used. The glass plates were meticulously cleaned by rigorous scrubbing with detergent followed by extensive rinsing with distilled water. The plates were then dried with paper towels, washed with methanol, and polished to dryness with Kimwipes. The small glass plate was siliconized with liberal amounts of 5% (v/v) dichlorodimethylsilane in chloroform, polished to dryness, washed with 95% (v/v) ethanol, and polished to dryness again using Kimwipes. (When the small glass plate is siliconized, then the gel preferentially adheres to the large glass plate, resulting in restoration of its original orientation when it is applied to Whatman 3MM paper following the run; see below.) The glass plate assembly was sealed using 1% (w/v) agarose. Gels of 7% (w/v) poly acrylamide (60 parts acrylamide:1 part bisacrylamide; ultrapure preparations obtained from BRL or Biorad) were cast in 1X TBE buffer (pH 8.3) containing 8 M urea (ultrapure, BRL), 0.05% (w/v) fresh ammonium persulfate, and 0.025% (v/v) tetramethylenediamine (TEMED). (A protocol for preparing such a gel follows.) Gels were poured vertically and allowed to polymerize horizontally from 3 hours to overnight. Gels were preelectrophoresed in prewarmed 1X TBE running buffer (defined previously) at 1500 V for 15-30 minutes prior to loading. Immediately before loading, samples were boiled and chilled as previously described and the wells were flushed with running buffer to rinse out urea (to permit samples to layer on the bottom of the wells). In each well 1-2 μl of sequencing reaction products was loaded, and

electrophoresis was in 1X TBE running buffer at constant 1500 V at ∿50°C. (To prevent renaturation of the samples, DNA sequencing gels are kept warm by electrical resistance heating at high voltage.) Optimum run times were empirically determined. Typically each gel was loaded three times a t = 0, 2½ hours, and 5 hours, and the experiment was terminated at t = 7 hours corresponding to run times of 7 hours, 4½ hours, and 2 hours, respectively. This resulted in resolution of approximately the first 350 bases 3' to the primer with about 20-50 bases overlap between loads. Following electrophoresis, the gel was applied to a sheet of Whatman 3MM paper, covered with saran wrap, and dried at 80°C for 45-60 minutes under vacuum in a Biorad gel drier. The dry gel was exposed to Kodak XAR-5 x-ray film at -20°C (without intensifier screens) as previously described; autoradiographic exposure times varied from 4 hours to 1 week. A description of the preparation of 160 ml of polyacrylamide gel solution, enough for two sequencing gels measuring 40 cm x 35 cm x 0.35 mm, appears in Table 6.

Enrichment of Specific Genes as Chromatin

In collaboration with Carmen Cadilla, soluble Euplotes macronuclear chromatin was prepared as previously described and was concentrated to a final $A_{260} \cong 150$ using an Amicon Centricon microcentrator (molecular weight retention ≧ 10 KDa) in an SS34 rotor at 6000 rpm for 1 hour at 4°C. Between 5-10 A_{260} units of chromatin in a volume of

Table 6. Preparation of DNA sequencing gels.

1.	add 37.3 ml 30% (w/v) acrylamide stock (60:1) → 7% (w/v)
2.	add 76.88 g urea (ultrapure, BRL) → 8 M
3.	add 16 ml 10X TBE (pH 8.3) → 1X
4.	add H_2O to ∿150 ml
5.	heat on magnetic stirrer to dissolve urea
6.	adjust volume to 160 ml with H_2O, cool
7.	add 800 µl fresh 10% (w/v) ammonium persulfate → 0.05% (w/v)
8.	degas under vacuum ∿3 minutes (removes O_2)
9.	filter through dispo Nalgene 0.45 µm filter unit
10.	divide into two 80 ml portions (for two gels)
11.	chill briefly on ice (to ∿15-20°C)
12.	add 20 µl TEMED per 80 ml portion, swirl → 0.025% (v/v)
13.	pour slowly from beaker into vertical assembly
14.	lay gel horizontally, insert comb, allow to polymerize ≧ 3 hours

55-100 µl were layered onto a ∿10 ml 5-27% (w/v) isokinetic sucrose density gradient in 50 mM NaCl, 1 mM Na_2EDTA, 1 mM TEA•HCl, 0.1 mM PMSF, and 0.1 mM TLCK (pH 7.0) with Ct = 5%, Cr = 27%, Vm = 9.4 ml, and a particle density of 1.51 g/ml. Ultracentrifugation was in a Beckman SW41 rotor at 28,000 rpm for 8 hours at 4°C under vacuum (138, 140, 197-199). Using an ISCO gradient fractionator 0.4 ml fractions were collected and their A_{260} was monitored with a Zeiss spectrophotometer. Individual chromatin fractions were analyzed by DNA and protein gel electrophoresis, southern and dot blot hybridization (all as previously described), and electron microscopy.

Electron Microscopy

Chromatin samples were analyzed by electron microscopy in collaboration with Ada Olins using a modification of the Miller spreading technique as described by Carmen Cadilla (197):

> Soluble MAC chromatin fractions were diluted with 1 mM Na_2EDTA, 1 mM TEA•HCl, 50 mM NaCl, pH 7.0 to an A_{260} of about 1.0, then were placed on a carbon film which had been freshly glowed in the presence of amylamine vapor for 30-60 seconds, rinsed in dilute photoflo, air-dried, and stained with 0.1% uranyl acetate in glass distilled water (200). Chromatin was visualized in the darkfield mode on a Siemens Elmiskop 102 electron microscope operated at 80 kV (201). Nucleosomes were counted on 3 times magnified prints of negatives taken at an instrument magnification of 60,000 x.

In Vitro Transcription

The functionality of the Euplotes macronuclear 5S rRNA gene as template for the accurate in vitro synthesis of 5S rRNA was

investigated in collaboration with Alan Wolffe in Don Brown's lab at the Carnegie Institute of Washington in Baltimore. A soluble, cell-free extract of Xenopus laevis oocyte nuclei was prepared as described by Birkenmeier, Brown, and Jordon (13):

> Ovaries were removed from adult Xenopus laevis, rinsed briefly with ice-cold distilled water to remove blood, blotted on a paper towel to remove excess liquid, and stored in a beaker in ice until dissected. Fragments of ovary were "primed" for isolation by incubation at room temperature for 30-60 min in 5 mM Tris (pH 7.8) and 10 mM $MgCl_2$. This step changes the consistency of the oocyte cytoplasm (mainly the yolk) and appears to precipitate nuclear contents. For germinal vesicle isolation, fragments of ovary were then transferred to ice-cold J buffer [70 mM NH_4Cl, 7 mM $MgCl_2$, 0.1 mM EDTA, 2.5 mM dithiothreitol, 10% (v/v) glycerol and 10 mM HEPES (pH 7.4)]. In experiments performed to measure divalent metal ion effects on transcription, $MgCl_2$ was omitted from the J buffer. Each GV was transferred in 1 µl of J buffer to a test tube and stored in ice until the desired number of GVs was obtained.
>
> GVs were either stored intact at -70°C indefinitely or used immediately to prepare the extract. GVs were disrupted by pipetting 5-10 times in a plastic mircopipette. The resulting lysate was centrifuged at 500 x g for 10 min at 4°C to pellet nuclear debris. The supernatant fluid (GV supernatant) was carefully removed with a micropipette and could be stored at 4°C for up to 24 hr without loss of activity. Protein concentration in the GV supernatant was determined by the method described by Lowry et al. (1951).

(For more information on in vitro transcription systems see 13-17 and 202-206. The Lowry protein determination method is contained in 207.)

The in vitro RNA synthesis reactions and subsequent analysis of nascent transcription products were performed as described by Wolffe (personal communication):

> The standard transcription reaction mixture was prepared by mixing 10 µl of GV supernatant with 2 µl J buffer containing 30 units RNasin (Promega Biotec), 2 µl of a mix of nucleoside triphosphates (final concentration = 0.5 mM ATP, CTP, and

UTP; 25 μM GTP and 10 μCi radioactive α^{32}P-GTP, specific activity = 2000 Ci/mmol (New England Nuclear)) and DNA as indicated below in 1 μl J buffer. Reactions occurred in capped conical tubes at 22°C for 1 hour. Reactions were stopped by the addition of 300 μl SETS (100 mM NaCl, 10 mM Tris, 1 mM EDTA, 1% (w/v) SDS (pH 8.0)), phenol extracted, chloroform extracted, and nucleic acids precipitated by removing the aqueous phase and adding to it 1/10 volume of 3 M Na acetate and 3 volumes of ethanol. The labelled RNA was redissolved in 80% (v/v) formamide, 25 mM Tris•borate, 0.5 mM EDTA (pH 8.3). Samples were analyzed by autoradiography following electrophoresis on a 10% (w/v) polyacrylamide gel containing 50 mM Tris•borate, 1 mM EDTA, and 8 M urea. 5S RNA synthesis was quantified by excising a gel fragment containing the 5S RNA band and measuring Cerenkov radiation. Backrounds were measured by excising and counting gel fragments on either side of the 5S RNA band.

The various Euplotes DNA templates tested included native linear Euplotes total macronuclear genomic DNA, covalently closed circular supercoiled plasmid p320 (this is the recombinant construct containing the Euplotes macronuclear 5S rRNA gene in the Pst I site of the plasmid vector pUC9), and the gel-purified Pst I-insert of p320 (that is, the cloned, linear Euplotes 5S gene). These were prepared as previously described. A Xenopus laevis somatic-type 5S rRNA gene-containing plasmid (pXls 11) served as positive control. A series of transcription reactions containing increasing concentrations of α-amanitin was performed to determine which RNA polymerase transcribes the Euplotes 5S gene (11).

Western Blot Analysis

Immuno-crossreactivity of Euplotes macronuclear chromosomal (chromatin-associated) proteins with polyclonal antibodies (anti-serum)

directed against Xenopus laevis TFIIIA was demonstrated in collaboration with Matt Andrews in Don Brown's lab at the Carnegie Institute of Washington in Baltimore. Soluble Euplotes macronuclear chromatin was isolated as previously described and brought to 1X in SDS gel sample loading buffer by addition of a concentrated stock. After boiling, proteins were resolved by polyacrylamide gel electrophoresis in the presence of SDS as previously described. Proteins were transferred from the gel to nitrocellulose by electroblotting. The western blot was probed with ^{125}I-labelled anti-Xenopus TFIIIA antibodies and Euplotes cross-reacting proteins were visualized by autoradiography.

Quantification of Nucleic Acids

Nucleic acids in aqueous solution were quantified by measurement of optical density (absorbance of light) at 260 nm (A_{260}) assuming mass extinction coefficients (ε mg/ml; dimensions $(mg/ml)^{-1}cm^{-1}$ or cm^2mg^{-1}) of 20 for DNA (duplex) and 25 for RNA (single-stranded). That is, in a cell of 1 cm path length, a 1 mg/ml solution of double-stranded DNA was assumed to have an A_{260} = 20 (or 1 A_{260} = 50 µg/ml) and a 1 mg/ml solution of single-stranded RNA was assumed to have an A_{260} = 25 (or 1 A_{260} = 40 µg/ml). The molar extinction coefficients (ε_M; dimensions $M^{-1}cm^{-1}$ or cm^2mole^{-1}) adopted for ribo- and deoxyribomono nucleotide solutions are shown in Table 7. Cuvettes were always cleaned immediately before and after each use with a solution prepared

Table 7. Extinction coefficients of nucleotides.[1]

Base	λ_{max} (nm)	ϵ_M ($M^{-1}cm^{-1}$)
A	259	1.54×10^4
G	253	1.37×10^4
C	271	9.1×10^3
U	262	1.0×10^4
T	260	7.4×10^3

[1]From Molecular Cloning--A Laboratory Manual (p. 449) by T. Maniatis, E. F. Fritsch, and J. Sambrook, 1982, Cold Spring Harbor Laboratory.

by combining 600 ml 95% (v/v) ethanol and 320 ml concentrated hydrochloric acid, then washed extensively with distilled water and dried by a brief methanol rinse. In rare instances when it was necessary to quantify a very small amount (<1 μg) of a homogeneous DNA species (a DNA species of specific size which will migrate as a discrete band on a gel), this was done by electrophoresis of a portion of the sample in an agarose gel and comparison of its ethidium bromide fluorescence intensity to that of known amounts of DNA standards in an adjacent lane.

CHAPTER III

RESULTS AND DISCUSSION

Identification of Specific Genes in the Euplotes Macronucleus

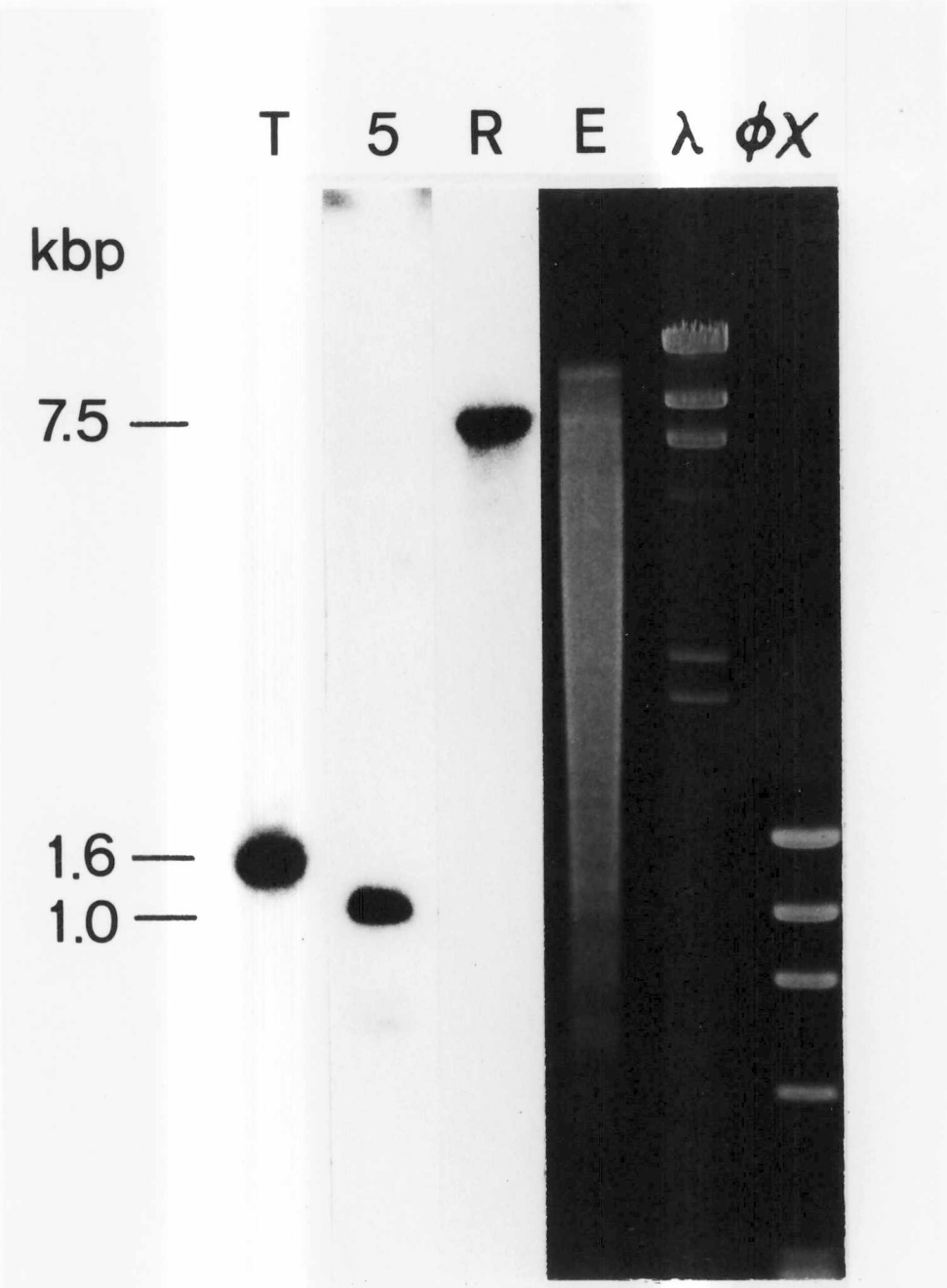
Banding Pattern of Total Macronuclear Genomic DNA

Euplotes eurystomus total macronuclear genomic DNA was purified from isolated macronuclei, fractionated by electrophoresis in agarose gels, and visualized by fluorescence of the intercalating chromophore ethidium bromide under ultraviolet illumination (Fig. 5, lane E). Native macronuclear genomic DNA so resolved is observed to migrate as a continuum of multitudinous discrete linear bands ranging in size from ∿400 bp to >20,000 bp. Superimposed on this continuum are several bands of augmented fluorescent intensity representing DNA species hyperamplified in number relative to those of typical abundance (average ploidy is ∿8000 copies per macronuclear genome in Euplotes). These presumably are DNA sequences whose transcription products are required by the cell in great numbers, such as structural RNAs or RNAs coding for structural proteins. One such example characterized in some detail is the hyperamplified band at 7.5 kbp which harbors genes coding for the large ribosomal RNAs, 25S and 19S, plus 5.8S rRNA (142-146). These large rRNA genes are number amplified ∿100-fold compared to genes of typical abundance (97, 109). Numerous examples exist from a diversity of organisms in which specific DNA sequences have been amplified to allow a cell to generate a greater

Figure 5: Hybridization analysis of Euplotes macronuclear DNA using specific gene probes.

Left: Autoradiogram of a Southern blot of Euplotes macronuclear DNA probed with: lane T, the α tubulin gene of Stylonychia; lane 5, the 5S rRNA gene of Tetrahymena; lane R, the large rRNA gene of Euplotes.

Right: Ethidium bromide staining of Euplotes macronuclear DNA and molecular weight markers λ and ØX.


T
5
R
E
λ
ϕX
kbp
7.5
1.6
1.0

number of transcripts per unit time than would otherwise be possible (amplification of template copy number is one mechanism of gene regulation (7)). It has been speculated (Pierluigi Donini, University of Rome, personal communication) that perhaps differential relative abundance (copy number) of macronuclear genes constitutes the sole (or major) means of expression regulation. Whether the transcription rates of macronuclear genes are discriminately, individually regulated or are purely the consequence of template copy number (en masse regulation in which the probability of transcription from a particular template is statistical) remains to be determined. (The molecular mechanisms which regulate gene expression in hypotrichous ciliated protozoa are entirely unknown.) In addition to the presence of hyperamplified bands (genes) in macronuclear DNA, it is also apparent from the staining pattern that the low molecular weight DNA is size-quantized. Superimposed on the continuum from ∿400 bp to ∿1000 bp are regions (not sharp bands) of enhanced intensity which vary in size by integral multiples of ∿200 bp, the approximate repeat length of bulk macronuclear chromatin (187 bp for Euplotes eurystomus as determined by Cadilla; 138, 197). Possible explanations of this observation include that this happens by chance, that the fragmentation of the macronuclear genome is controlled so as to produce DNAs in quantum multiples of the repeat length (DNAs which will fit neatly onto polynucleosomes), that (more likely) during nucleolytic processing of the genome in the developing macronucleus the linker DNA is preferentially cleaved due to its greater accessibility (again resulting in DNAs of quantum polynucleosomal length), or (the

trivial explanation) that this is an artifact resulting from degradation of the DNA during isolation. (The distinction between the second and third possible explanations above is a subtle one and is in essence the difference between active and passive cleavage site direction.)

Hybridization Analysis of Euplotes Macronuclear DNA Using Specific Gene Probes

Euplotes macronuclear DNA so resolved by agarose gel electrophoresis was transferred to a nitrocellulose membrane by the technique of Southern (162) and analyzed by hybridization with radiolabelled, specific gene probes. Homologous sequences were detected by autoradiography and thus several genes in the macronucleus of Euplotes eurystomus were identified (Fig. 5). This represents the first report of the identification of specific genes in this species (140, 147). The gene coding for large ribosomal RNAs (25S and 19S) in E. eurystomus was found to reside on a macronuclear fragment 7.5 kbp in length as is also the case for Euplotes aediculatus (145, 146), Stylonychia pustulata (109, 145), Stylonychia mytilus (142), Oxytricha fallax (101, 144, 145), and Oxytricha nova (94, 143, 145) (see Table 3, p. 39). Obviously the length, as well as the overall structure, of this gene has been highly conserved during the evolution of these protozoa. Restriction fragment length polymorphism analysis revealed that of the above the Euplotes gene sequence is the least similar, with the position of only one of the seventeen restriction sites tested conserved (145). Nonetheless the gross gene structure (positions and sizes of coding regions and spacers) was nearly identical in all organisms. These genes each

contain single copies of regions coding for 25S, 19S, and 5.8S rRNAs; it was found that the nontranscribed spacers exhibit greater sequence divergence than do the coding regions (143, 145). The macronuclear α tubulin gene in E. eurystomus is 1.6 kbp in length in contrast to genes of 1.85 kbp and 1.73 kbp in Stylonychia lemnae (150, 151), and 2.00 kbp and 2.05 kbp in Oxytricha fallax (96, 97). The 1.85 kbp gene of Stylonychia was found to harbor a single coding region 1.3 kbp in length. It is interesting to note that in Euplotes this gene is monomorphic whereas the other two organisms both have two size classes of the gene. It is of course possible that Euplotes also contains multiple versions of the α tubulin gene but that the heterologous hybridization probe used was only able to detect one of these. Euplotes eurystomus was found to contain a macronuclear 5S rRNA gene 930 bp in length as compared to genes of 690 bp in Oxytricha fallax (101) and 200-300 bp in Stylonychia mytilus (142). This report describes the first detailed characterization of a 5S rRNA gene in the hypotrichous ciliated protozoa. Loren Hauser (personal communication) has recently identified in the macronucleus of Euplotes eurystomus a gene of 2.1 kbp coding for the hsp 70 protein and a gene of 950 bp coding for ubiquitin. This represents the first report of either of these genes in the hypotrichs.

The novel results described above for Euplotes eurystomus lend support to the dogma of hypotrichous ciliated protozoan macronuclear genome development and structure. Without mechanical shearing or treatment with nucleases, native macronuclear genomic DNA is freely

soluble and migrates through agarose gels as short, linear fragments. Strikingly, sequence-specific gene probes hybridize to sharp, discrete bands in native total macronuclear genomic DNA. This means that during macronuclear development the genome (a polytenized micronuclear genome; see Introduction for a description of the organism's life cycle) is processed by specific, accurate, precise, and reproducible fragmentation such that a particular gene will always come to reside on a mature macronuclear DNA molecule of characteristic length. This is remarkable. (This may lend favor to Donini's hypothesis that macronuclear gene regulation is by differential template abundance; perhaps the extreme control exerted by the organism during macronuclear development is stringent enough to program the future regulation of the mature genome. This seems unlikely however since it would appear not to provide the organism with sufficient flexibility to cope with rapidly changing circumstances. The counterargument that the role of conjugation is to provide a mechanism for the animal to adapt or adjust itself to changing environments would not be able to explain provision for tolerance of sudden stress such as heat shock.) In addition, these results support the dogma that the canonical macronuclear DNA molecule represents only a single genetic function, since the fragments identified here are long enough to accommodate only one coding region (with the exception of the 5S rRNA gene). The 5S rRNA gene was selected for further study for reasons discussed earlier (refer to the description of research objectives in the Introduction).

Construction of the Macronuclear Genomic DNA Library

Recombinant DNA technology provides a means of purifying specific DNA sequences in amounts sufficient to allow detailed biochemical analysis of gene structure and function. This is achieved by attaching the sequence of interest to a retrievable genetic vehicle which provides for its selection and propagation in bacteria. From a single culture of the recombinant bacterial clone enough of the gene may be isolated for many experiments. In practice, the entire genomic complement of an organism is usually cloned en masse and the particular sequence of interest is then purified from this collection of molecules. Different types and applications of recombinant DNA libraries, cloning vectors, construction methods, and screening techniques, as well as general uses of genetic engineering, were discussed previously.

A library of Euplotes eurystomus total macronuclear genomic DNA was constructed essentially as described by Swanton et al. (143). The plasmid vector pUC9 (conferring ampicillin resistance) was linearized with Pst I and its 3' ends were extended with poly(dG) by terminal transferase; annealed to this was Euplotes macronuclear DNA which had been similarly tailed with poly(dC) (Fig. 6). This cloning strategy results in the restoration of Pst I sites flanking the insert so that it may be recovered from the vector by digestion of the recombinant plasmid with Pst I. pUC9 was selected as the cloning vehicle because of its small size and the utility of the polylinker in the subsequent analysis and manipulation of the insert. By the Hanahan procedure, E. coli DH5 was transformed with the chimeric construct and selected on ampicillin.

Figure 6: pUC 9 cloning strategy.

In the construction of the genomic DNA library, terminal transferase was used to extend the 3' ends of Euplotes macronuclear DNA with poly(dC) and those of Pst I-linearized pUC9 with poly(dG). These tailed, linear DNAs were annealed to generate chimeric, circular plasmids with which E. coli was transformed. The clone bank was screened by in situ colony hybridization with the 5S rRNA gene of Tetrahymena.

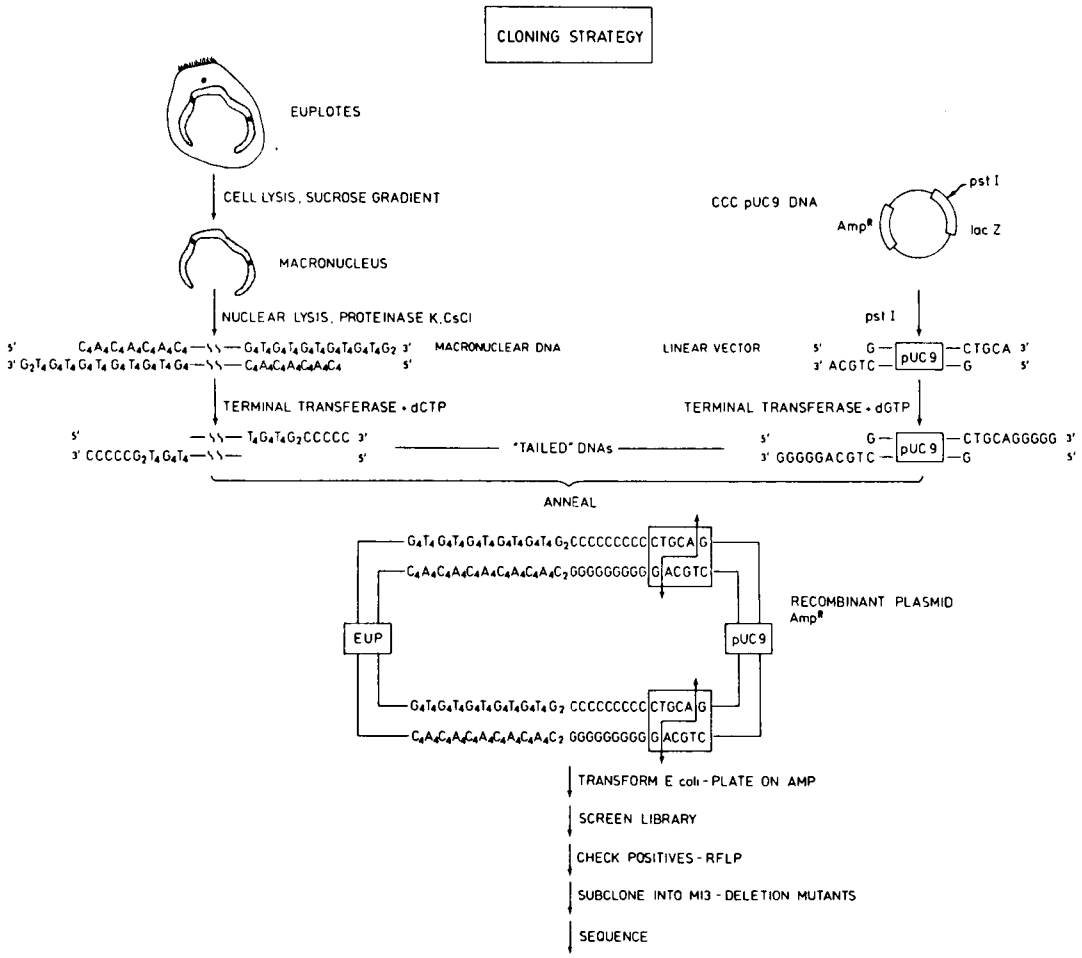

CLONING STRATEGY
EUPLOTES
CELL LYSIS, SUCROSE GRADIENT
MACRONUCLEUS
NUCLEAR LYSIS, PROTEINASE K,CsCl
MACRONUCLEAR DNA
TERMINAL TRANSFERASE + dCTP
CCC pUC9 DNA
pst I
Amp
lac Z
pst I
LINEAR VECTOR
pUC9
TERMINAL TRANSFERASE + dGTP
"TAILED" DNAs
ANNEAL
EUP
pUC9
RECOMBINANT PLASMID
Amp
TRANSFORM E coli - PLATE ON AMP
SCREEN LIBRARY
CHECK POSITIVES - RFLP
SUBCLONE INTO MI3 - DELETION MUTANTS
SEQUENCE

Figure 6

During this procedure it was found that Euplotes macronuclear DNA which was less than absolutely pure was refractory to end modification by terminal transferase. This is explained by Gottschling's report of specific proteins which bind to the termini of the linear macronuclear DNA molecules (135). (The function of these telomeric sequences and proteins will be discussed in more detail later.)

Screening the Recombinant DNA Library

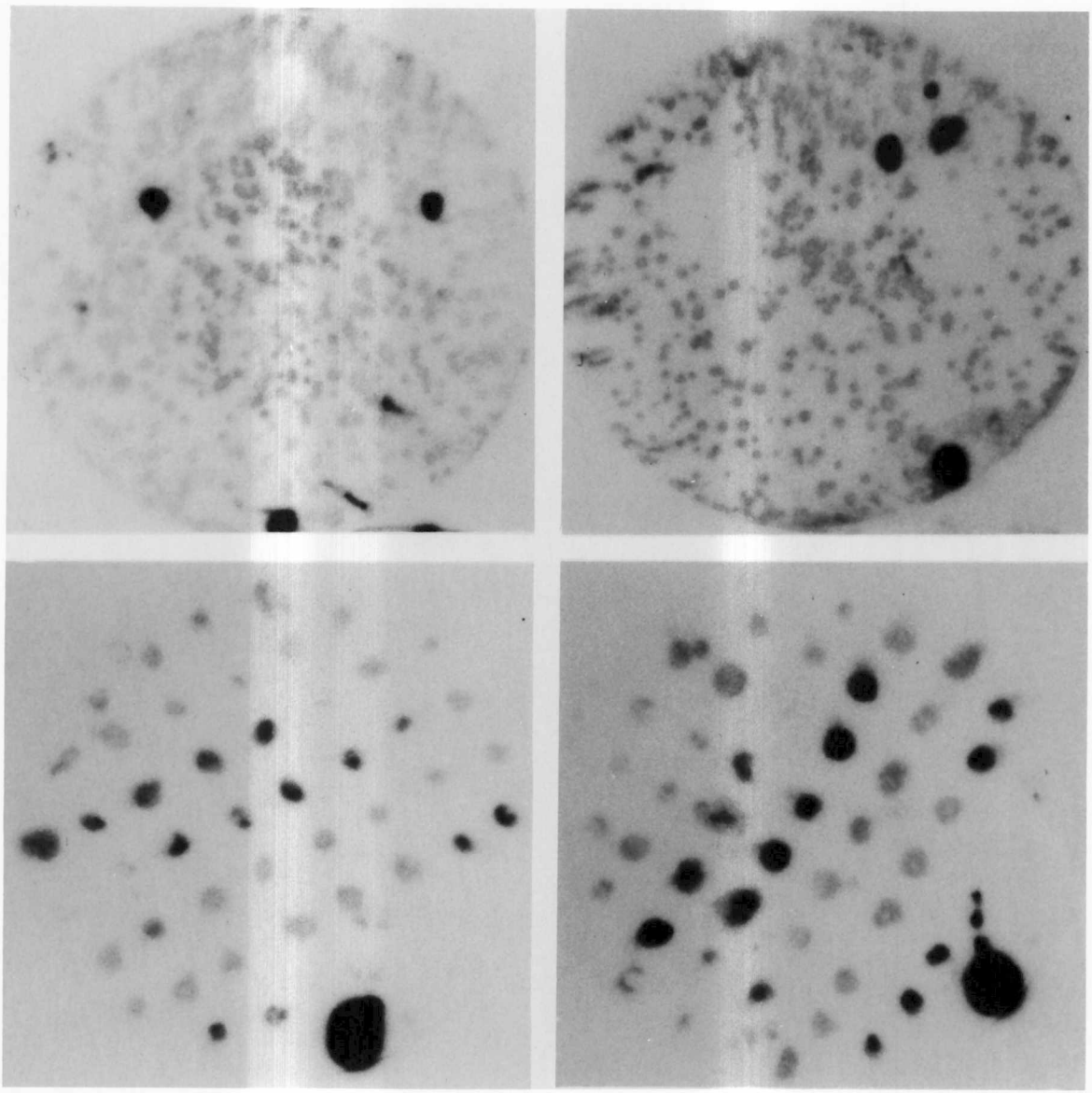
Chimeric plasmids harboring Euplotes macronuclear DNA inserts were introduced into E. coli DH5 in a grand transformation reaction. The mixture was divided among 42 LB agar plates containing 100 µg/ml ampicillin. The library was screened by in situ colony hybridization using the gel-purified 5S rRNA gene of Tetrahymena (167) as probe (Fig. 7).

Calculation of Efficiency

The efficiency of a transformation reaction is defined as the number of transformants generated per unit mass DNA and its quantitation is based on the amount of vector DNA put into the reaction, since it is the vector which transforms the cells to antibiotic resistance (the selection criterion). The ideal substrate for transformation is a small, supercoiled plasmid (small DNAs can penetrate the permeabilized cell membrane more easily than can large DNAs) and under optimum conditions efficiencies exceeding 1×10^8 transformants/µg-vector-DNA can be attained. Bulky, relaxed circular plasmids formed by annealing DNAs with complementary single-stranded homopolymer tracts (still

Figure 7: Screening the recombinant DNA library by in situ colony hybridization.

Bacterial colonies harboring recombinant plasmids containing Euplotes macronuclear DNA inserts were screened by hybridization with the Tetrahymena 5S rRNA gene probe. Top: initial round of screening; bottom: final round of screening.



containing single-strand gaps due to unequal tail lengths) are among the poorest transformation substrates. In this work, using such constructs, efficiencies between 1.6×10^6 and 2.1×10^6 transformants/μg were reproducibly achieved. This is in comparison to results of 1.5×10^4 to 2.5×10^4 transformants/μg reported by Swanton for the same construction technique (143). That efficiency was improved by two orders of magnitude is because here cell competency was mediated by RbCl whereas Swanton used $CaCl_2$ and because the vector pUC9 (2686 bp) used here is smaller than pBR322 (4362 bp) used by Swanton.

Calculation of Background

The background of a transformation reaction is defined as the relative abundance of transformants without inserts (transformants harboring the nonrecombinant, wild-type plasmid vector) and is quantified here based on transformation frequencies of the linear, tailed vector alone. Background is caused by contamination of the linear tailed vector preparation with a trace amount of native uncut supercoiled vector and by (exceedingly rarely) recircularization of the vector without an insert. Although the amount of supercoiled vector contaminating the linearized preparation is quite small, it is a much more efficient transformation substrate than is the annealed recombinant construct. E. coli DH5 was chosen as the bacterial cell host because it is transformable at high efficiency, because it is $recA^-$ (deficient in homologous recombination which can lead to cloning artifacts) and because competent DH5 cells are commercially available. Since,

however, DH5 is $lacZ^+$, the colorimetric assay for insertion mutagenesis incorporated in the pUC system could not be applied in this instance. For this reason, control experiments were very carefully designed and executed to determine background. Three parallel reactions were assembled in which the only variant was the nature of the DNA. In one reaction transformation was with the native supercoiled vector simply to monitor proper cell performance (positive transformation control). Background was determined from the other two transformation reactions, one of which was with the annealed recombinant construct (experimental) and the other of which was with the linear tailed vector alone (negative transformation control). Transformation frequencies were quantitated based on the amount of vector DNA put into the reactions. Background was thus reproducibly determined to be 12-14% (this means that 12-14% of the transformed colonies contain plasmids without Euplotes macronuclear DNA inserts).

Additional Controls

In addition to the positive and negative control transformation reactions described above, an aliquot of each transformation mixture was plated on LB agar without ampicillin as a positive control to monitor cell survival (to allow precise interpretation of negative results). The transformation efficiency and background results described above were considered adequate (excellent in fact) to justify continuation of the experiment. (For example, it would not be worthwhile to screen a library in which only 10% of the transformants were bonafide

insert-bearing recombinants.) As the transformants were transferred to nitrocellulose for screening by in situ colony hybridization, separate colonies of DH5 containing plasmids pUC9 and pDP5 (which harbors the 5S rRNA gene of Tetrahymena used as probe; 167) were also applied to the membrane to serve respectively as negative and positive hybridization controls. This provided an internal standard criterion of hybridization intensity to assist in evaluating Euplotes clones as negative or positive. Recombinant plasmids within the bacterial clones were amplified by treatment with chloramphenicol to increase the hybridization signal to noise ratio. The pDP5 insert (the 5S rRNA gene of Tetrahymena) was gel-purified for use as nick translated probe in order to avoid artifactual signals generated by hybridization of the pDP5 vector (pBR322) to the colonies. Thus during construction and screening of the DNA library positive and negative control experiments were performed and results were quantitatively analyzed at every step.

Calculation of N

The number of clones required to achieve a given probability of representing a particular sequence in a DNA library is given by:

$$N = \frac{\ln(1 - P)}{\ln(1 - f)}$$

where

N = number of clones required,

P = desired probability, and

f = fractional proportion of the genome complexity represented by a single recombinant clone (154, pp. 225, 271).

Since the Euplotes macronucleus contains ∿23,500 different DNA molecules, f here is 1/23,500 (this is identical to the number-average size of a macronuclear DNA fragment (insert size) divided by the total macronuclear genome complexity: 1836 bp/43 x 10^6 bp). If the 5S rRNA gene is present in the macronucleus at typical abundance, then to achieve a 99% probability of isolating a single 5S gene clone (substituting and solving the above equation) ∿108,000 clones would have to be screened. Since as a normal part of its life cycle the organism eliminates from the macronucleus ∿98% of the micronuclear genome complexity (the total genetic complexity of the organism), substantially fewer clones are required to attain a given statistical probability of representing a particular sequence than would be otherwise. This saving of labor is reduced however because the number-average insert size (1836 bp here) is much less than is the case for typical genomic DNA libraries (say ∿20 kbp for a λ library), so although in Euplotes much less DNA has to be screened it must be dealt with in much smaller bytes (less DNA per clone). Thus the required number of clones is about 10 times greater than would be expected based solely on consideration of the reduced complexity of the macronuclear genome.

Screening Results

In the initial screening 42 plates were analyzed, each containing an average of about 800 colonies, amounting to ∿33,000 total recombinant clones. The number of clones required to achieve a 50% probability

level of isolating a particular gene is 16,300, therefore 33,000/16,300 = 2 positive results were expected by random chance. In fact, 19 copies of the putative 5S rRNA gene were isolated from these 33,000 recombinant clones providing the first indication that this gene is present at greater than typical abundance in the macronucleus. These 19 putative 5S rRNA gene clones tested positive in three consecutive rounds of screening (by colony hybridization) and were purified by two tandem cycles of streaking (to resolve single bacterial cells). Master stocks of the recombinant bacterial clones were stored at 4°C on sealed, inverted agar plates containing 100 µg/ml ampicillin. Plasmids were prepared from all of the clones for analysis and archival storage.

Preliminary Characterization of the Euplotes 5S rRNA Gene

Hybridization of the Recombinant Plasmids with the 5S rRNA Gene of Tetrahymena

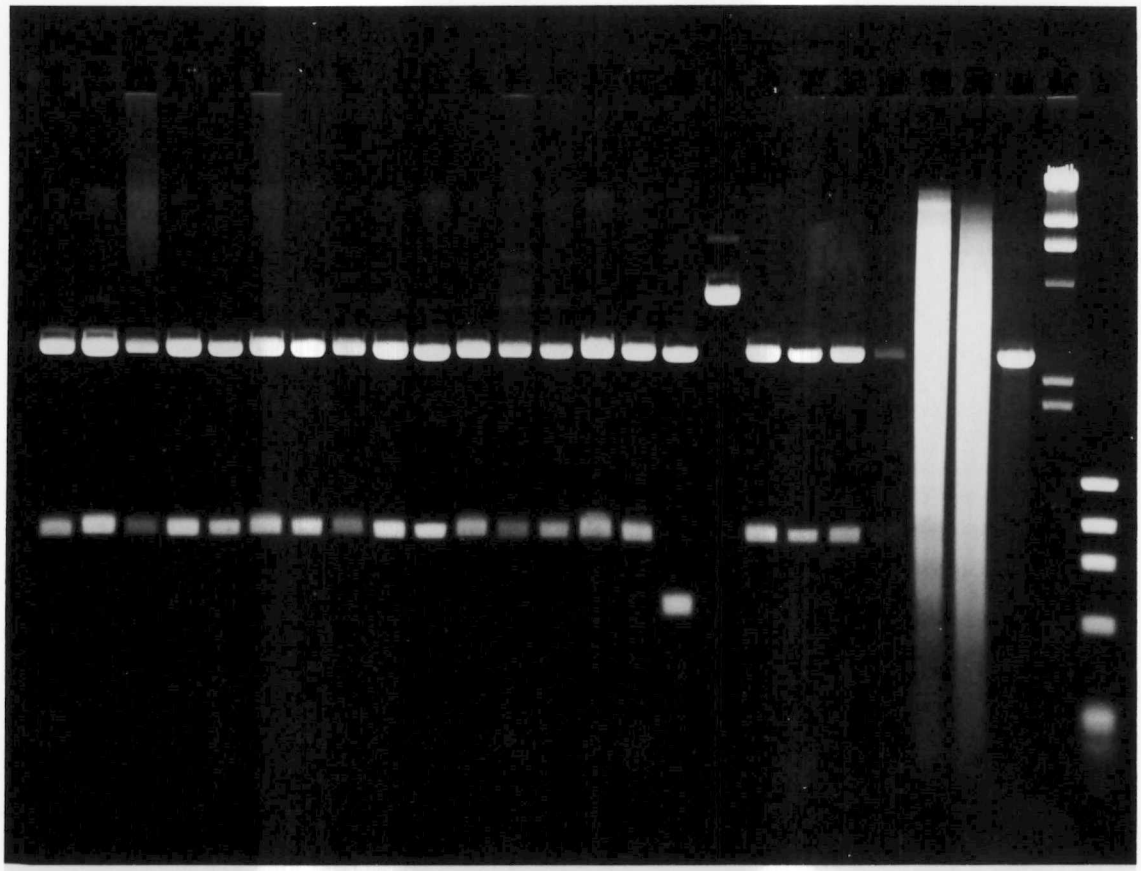
For initial analysis of the putative Euplotes macronuclear 5S rRNA gene clones, plasmids were prepared by a modification of Clewell's procedure (155) from 1 L cultures of each of the 19 positive bacterial colonies (plus from a few negative colonies). After one round of cesium chloride gradient ultracentrifugation plasmids were further purified by extraction, dialysis, and precipitation (see Materials and Methods). The Euplotes inserts were examined by digestion of the chimeric constructs with the restriction endonuclease Pst I and resolution of the cleavage products by agarose gel electrophoresis. Following visualization by ethidium bromide staining, the DNA was transferred to

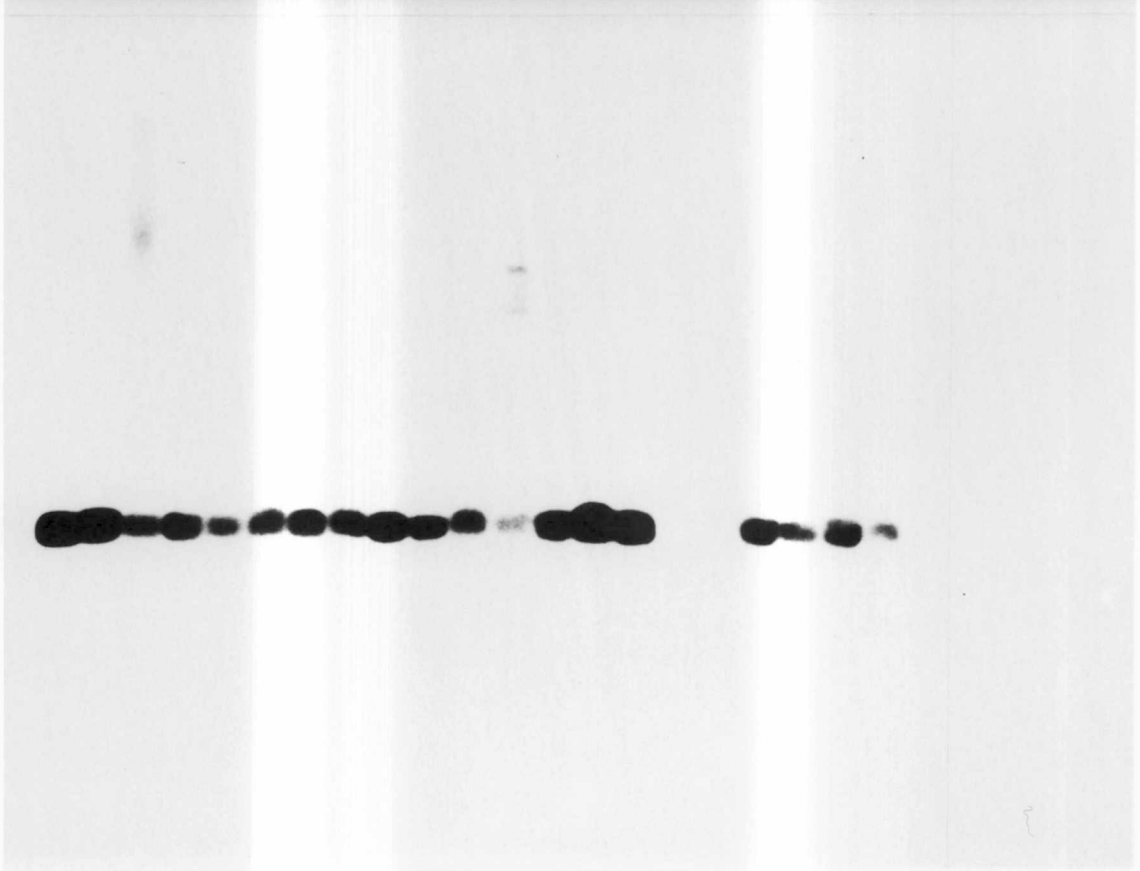
nitrocellulose by the technique of Southern (162) and probed with the gel-purified 5S rRNA gene of Tetrahymena (Fig. 8). From left to right, this gel contains the following samples: 21 recombinant plasmids harboring Euplotes macronuclear DNA inserts cleaved from the Pst I site of pUC9 (note the aberrant migration of samples in lanes 16 and 17; the sample in lane 17 contains only one successfully reconstructed Pst I site so it migrates as a single linear band), native Euplotes total macronuclear genomic DNA, Euplotes total macronuclear genomic DNA digested to completion with Pst I, the native parental plasmid vector pUC9 linearized with Pst I (for size comparison with lanes containing recombinant plasmids), phage λ DNA digested with Hind III (molecular weight standard), and phage ØX174 RF DNA digested with Hae III (molecular weight standard). In all of the plasmid x Pst I lanes the vector pUC9 is seen to migrate as unit size (except in lane 17 where the insert is still attached); 19 of the Euplotes macronuclear DNA inserts are also observed to comigrate. Several specific conclusions can be drawn from this experiment: (1) 19 of the Euplotes macronuclear DNA inserts are identical in size, ∿1 kbp; (2) all of the 1 kbp inserts hybridize to the 5S rRNA gene of Tetrahymena (and thus are considered putative positives, see autoradiogram); (3) the two samples displaying aberrant migration do not hybridize to the 5S rRNA gene of Tetrahymena (and thus are negatives); (4) all of the 1 kbp inserts lack internal Pst I sites; (5) the 5S rRNA gene of Tetrahymena hybridizes specifically to a band in native Euplotes total macronuclear genomic DNA of size identical to the 19 comigrating 1 kbp plasmid inserts, therefore these inserts are

Figure 8: Hybridization of the recombinant plasmids with the 5S rRNA gene of Tetrahymena.

Top: Euplotes macronuclear DNA inserts were excised by digestion of the chimeric constructs with Pst I. Ethidium bromide staining of an agarose gel loaded left to right as follows: 21 recombinant plasmids x Pst I, native Euplotes macronuclear DNA, Euplotes macronuclear DNA x Pst I, pUC9 x Pst I, molecular weight markers λ and ØX.

Bottom: Autoradiogram of a Southern blot of the above gel probed with the 5S rRNA gene of Tetrahymena resulting in the identification of 19 isolates of the Euplotes 5S rRNA gene.





the same size as the native Euplotes macronuclear 5S rRNA gene (apparent from a longer autoradiographic exposure); (6) the 5S rRNA gene of Tetrahymena also hybridizes to exactly the same size band in Euplotes total macronuclear genomic DNA which has been digested with Pst I, so the native (uncloned) Euplotes macronuclear 5S rRNA gene also lacks an internal Pst I site (also apparent in a longer exposure). For these reasons the plasmid clones harboring 1 kbp inserts were tentatively considered to be 19 independent isolates of the Euplotes macronuclear 5S rRNA gene. A technical observation is that of 24 total recombinant plasmids tested (21 are presented in Figure 8), 23 had inserts excisable with Pst I and the one which did not migrated as a single linear molecule, therefore 47/48 or 98% of the Pst I sites were successfully reconstructed during the cloning procedure.

Hybridization of the Recombinant Plasmids With the 5S rRNA Gene of Euplotes

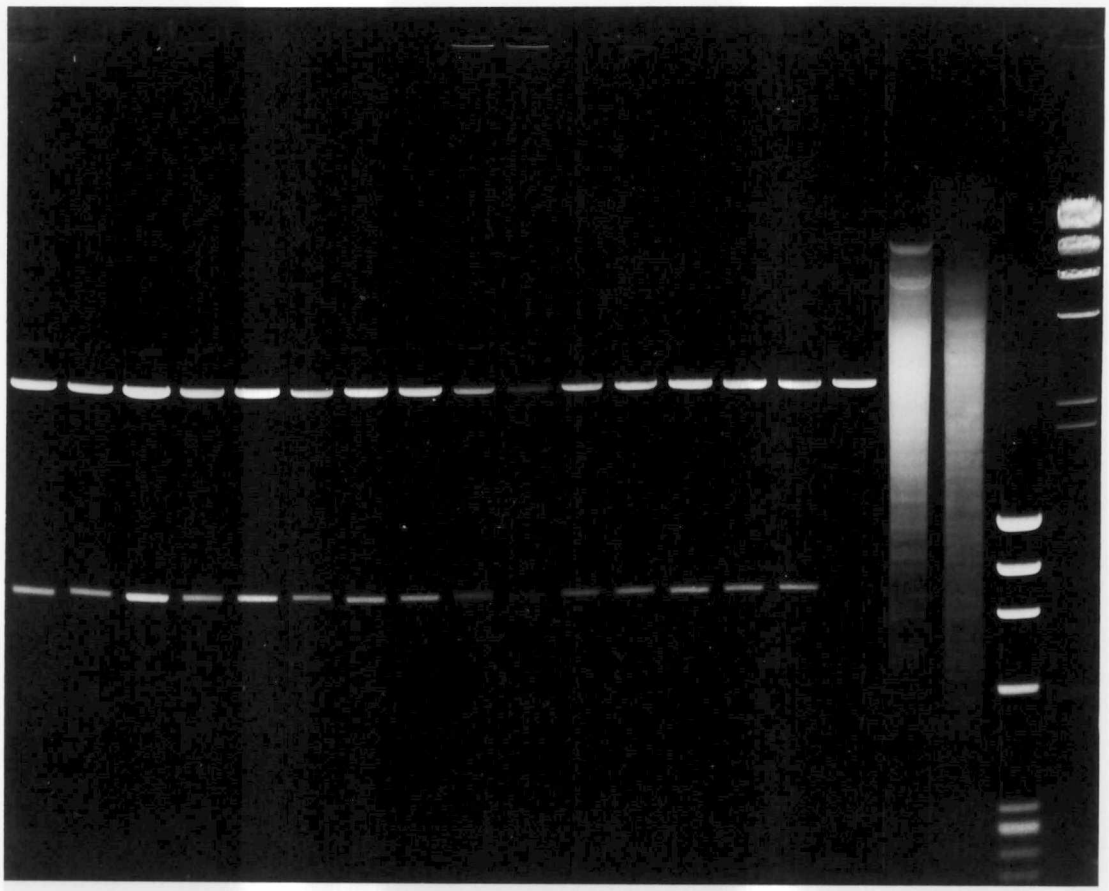
The 19 putative Euplotes macronuclear 5S rRNA gene clones were further purified by a second round of cesium chloride gradient ultracentrifugation to remove traces of *E. coli* genomic DNA and tRNA prior to more detailed analysis. (In a sense, the Southern blot in Figure 8 constituted the fourth round of screening--screening the clones as pure DNA instead of as colonies--as well as representing the initial characterization.) Due to leaky dialysis clips, only 15 of the samples were recovered in yields sufficient to allow complete characterization. These 15 plasmids were digested with Pst I, electrophoresed through an agarose gel, and transferred to nitrocellulose as before. This time the

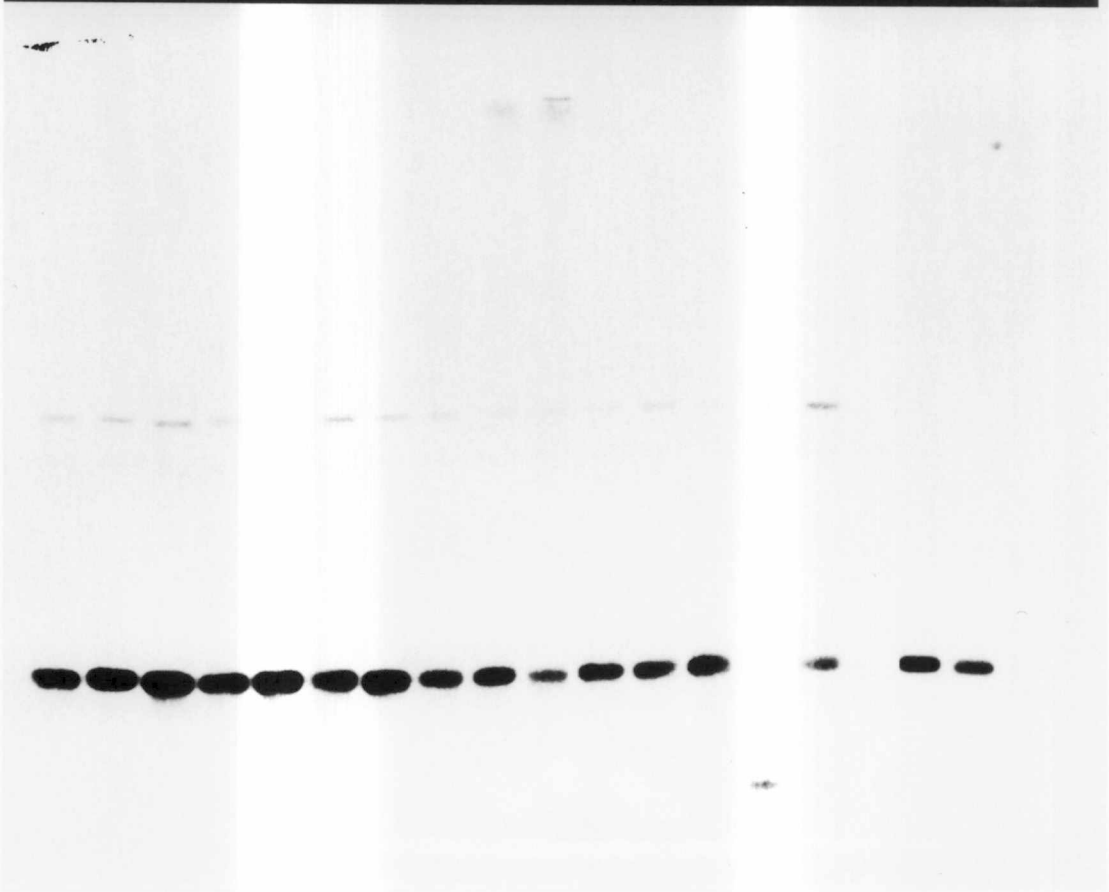
blot was probed with the gel-purified Pst I-insert of one of the 19 putative Euplotes 5S rRNA gene clones: p320 (this was later proven to be a true 5S rRNA gene by sequence analysis and other studies; see below) (Fig. 9). [Definition: "p320" is recombinant plasmid isolate number 320, a chimeric construct containing the intact Euplotes macronuclear 5S rRNA gene in the Pst I site of the plasmid vector pUC9 (see below for proof). The gel-purified Pst I-insert of p320 will be hereafter called "the p320 insert" or "the cloned Euplotes macronuclear 5S rRNA gene."] This gel contains the following samples from left to right: 15 recombinant plasmids harboring Euplotes macronuclear DNA inserts which have been excised with Pst I (p320 is the plasmid in lane 2 of this gel and of the gel in Figure 8), the native parental plasmid vector pUC9 linearized with Pst I, native Euplotes total macronuclear genomic DNA, Euplotes total macronuclear genomic DNA digested with Pst I, and molecular weight markers as before. Again, all of the vectors of the recombinant clones migrate as unit-size native vector and all of the Euplotes macronuclear DNA inserts comigrate as a single band of 1 kbp. This experiment was designed to demonstrate only a few specific conclusions in addition to those of the previous investigation: (1) The Euplotes DNA insert of one of the clones (p320) hybridizes to the Euplotes DNA inserts of all of the clones under very stringent conditions, therefore all of these cloned insert sequences are highly homologous (see below for elaboration); (2) all of the inserts hybridize to the cloned Euplotes macronuclear 5S rRNA gene (the p320

Figure 9: Hybridization of the recombinant plasmids with the 5S rRNA gene of Euplotes.

Top: Euplotes macronuclear DNA inserts were excised by digestion of the chimeric constructs with Pst I. Ethidium bromide staining of an agarose gel loaded left to right as follows: 15 recombinant plasmids x Pst I, pUC9 x Pst I, native Euplotes macronuclear DNA, Euplotes macronuclear DNA x Pst I, molecular weight markers ØX and λ.

Bottom: Autoradiogram of a Southern blot of the above gel probed with the gel-purified p320 insert.





insert) under very stringent conditions (as well as to the Tetrahymena 5S rRNA gene under less stringent conditions) providing further evidence that all of the clones represent independent isolates of the Euplotes macronuclear 5S rRNA gene (see below for proof of the identity of the p320 insert); (3) the p320 insert hybridizes specifically to a single band in native Euplotes total macronuclear genomic DNA which is identical in size to all of the cloned plasmid inserts and to the band detected by the Tetrahymena 5S rRNA gene probe; (4) the previous finding is unaltered by treatment of Euplotes DNA with Pst I, providing additional evidence that the native Euplotes macronuclear 5S rRNA gene lacks an internal Pst I site. The Southern blot presented in Figure 9 was probed with the nick translated p320 insert under usual hybridization conditions and washed exhaustively with 0.1X SSC, 0.1% (w/v) SDS at 65°C. This elution regimen is quite stringent so that only sequences of exact (or very nearly exact) complementarity may hybridize. Thus it is concluded that all of the inserts are highly homologous (their sequences are identical or nearly identical) to each other and to a fragment of the same size in native Euplotes total macronuclear genomic DNA.

All of the findings from the previous two experiments (Figs. 8 and 9) taken together support the more general conclusions that there is only one size class of 5S rRNA gene in the macronucleus of Euplotes eurystomus and the cloned isolates of this gene are accurate, faithful representations of the native, wild-type gene.

Restriction Enzyme Mapping of p320

Of the 19 putative Euplotes macronuclear 5S rRNA gene clones, construct ps320 was arbitrarily selected for intense scrutiny. A low resolution restriction endonuclease cleavage site map was determined in a single experiment. The pUC system of cloning vectors carries the polylinker (also called the multiple cloning site, the MCS), a set of several adjacent unique restriction enzyme sites (see 161, 175, 183, or 184 for a restriction map of the pUC plasmid). A restriction enzyme map of foreign DNA inserted into any of the enzyme sites in the polylinker may be rapidly determined by digestion of the recombinant construct with the other enzymes in the polylinker and analysis of the products by gel electrophoresis. For a particular enzyme, the sizes of the cleavage fragments generated are defined by the distance from the enzyme site in the polylinker to the site (or sites) for that enzyme within the insert. Since the enzyme sites in the polylinker (including the insertion site) are known, sites within the insert are mapped simply by visual inspection of the gel: the size of a fragment generated by a particular enzyme is equal to the distance spanning sites for that enzyme in the polylinker and in the insert.

In the case of recombinant plasmid p320, Euplotes macronuclear DNA was introduced into the Pst I site of pUC9. p320 was digested with several enzymes cutting in the polylinker, and the sites for those enzymes within the Euplotes macronuclear DNA insert were mapped by inspection of the cleavage products via agarose gel electrophoresis

(Fig. 10). From left to right this gel contains the p320 digestion fragments of Pst I (to monitor the intact, unit-size insert), Bam HI, Hind III, Eco RI, and Ava I, a double digestion of p320 with Pst I and Asu II, the native parental plasmid vector pUC9 linearized with Pst I (for size comparison), and the usual molecular weight markers ØX174 RF x Hae III and λ x Hind III. By this experiment internal sites of the p320 insert for the enzymes Bam HI, Hind III, Eco RI, and Ava I were unambiguously localized; these enzymes were each observed to cut the insert once (Fig. 11). Since pUC9 contains no Asu II sites, a double digestion of p320 with Pst I and Asu II was used to demonstrate that the insert contains a single Asu II site ∿630 bp from one of its ends (although from which end cannot be discerned from this experiment). The p320 insert was again observed to contain no Pst I sites; in other experiments the insert was demonstrated to also be refractory to the restriction enzymes Acc I, Hae II, Sal I, Sma I, Sst I, and Xho I (data not shown). [Definition: The order of restriction enzyme sites of the p320 insert as diagrammed in Figure 11 define orientation "A" of this duplex DNA structure. The top strand in this diagram is called the (+) strand and is written in the orientation 5'→3', as for all duplex DNA structures reported in this manuscript. The contrapositive (the inverse of the converse) of orientation A is orientation "B" wherein the order of restriction enzyme sites (from left to right) is now reversed and the top strand is now the (-) strand, also written 5'→3'. All unihelical polynucleotide structures are oriented 5'→3'. Sequence information and sequence elements will be presented as 5'→3' and the

Figure 10: Restriction enzyme digestion of chimeric construct p320.

Ethidium bromide staining of an agarose gel loaded left to right as follows: p320 digested with Pst I, Bam HI, Hind III, Eco RI, Ava I, and Pst I x Asu II; pUC9 x Pst I; molecular weight markers ØX and λ.

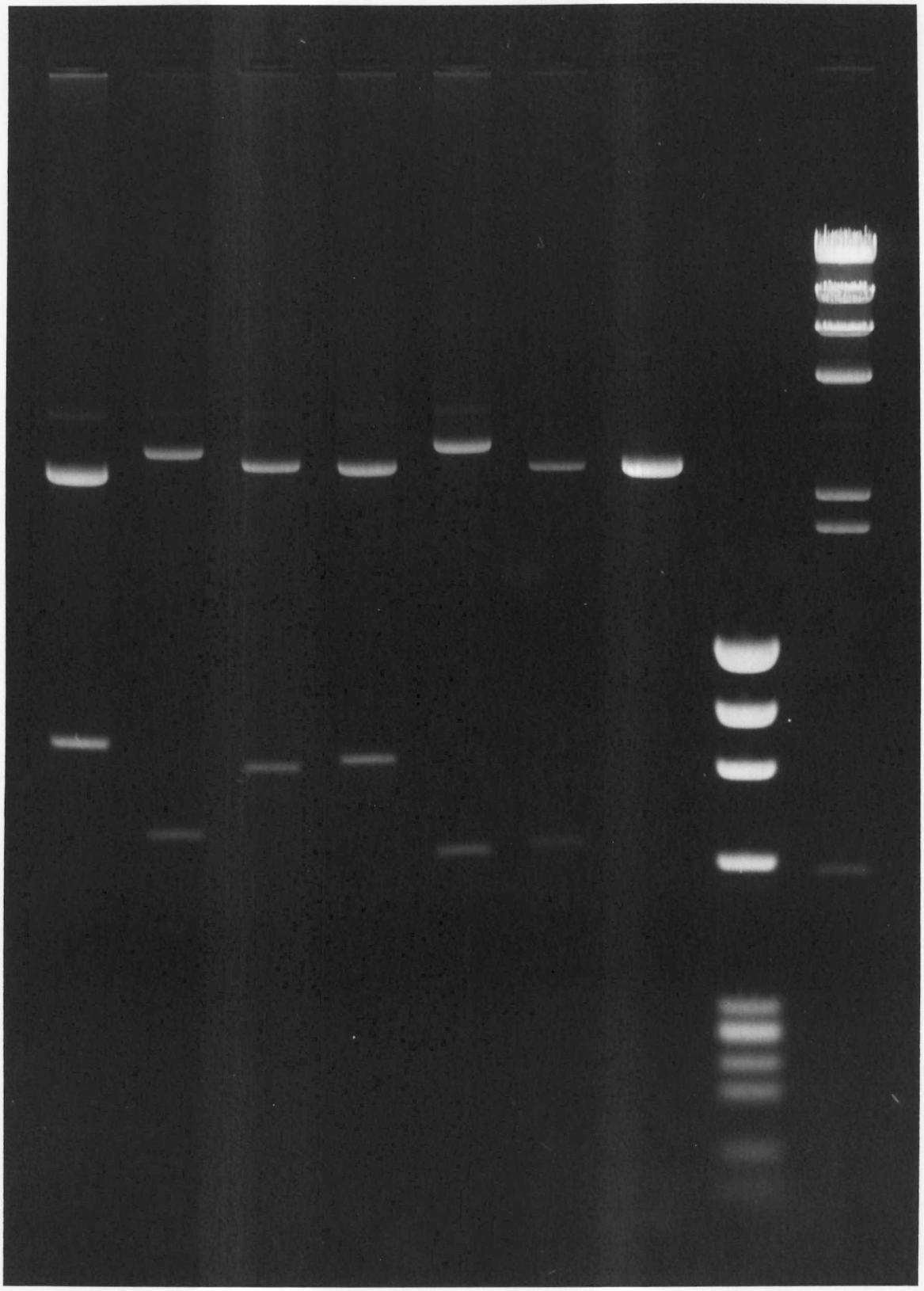


Figure 11: Restriction enzyme map and sequencing strategy for the p320 insert, the macronuclear 5S rRNA gene of Euplotes.

The p320 insert is shown in orientation A with the (+) strand on top written as 5'→3' (see text). Above, the distances between adjacent restriction enzyme sites are given in bp with the overall fragment length being 930 bp. Below, arrows indicate initiation positions and directions of the DNA sequencing reactions such that both strands are sequenced entirely within overlap between adjacent start sites.

RESTRICTION MAP and SEQUENCING STRATEGY

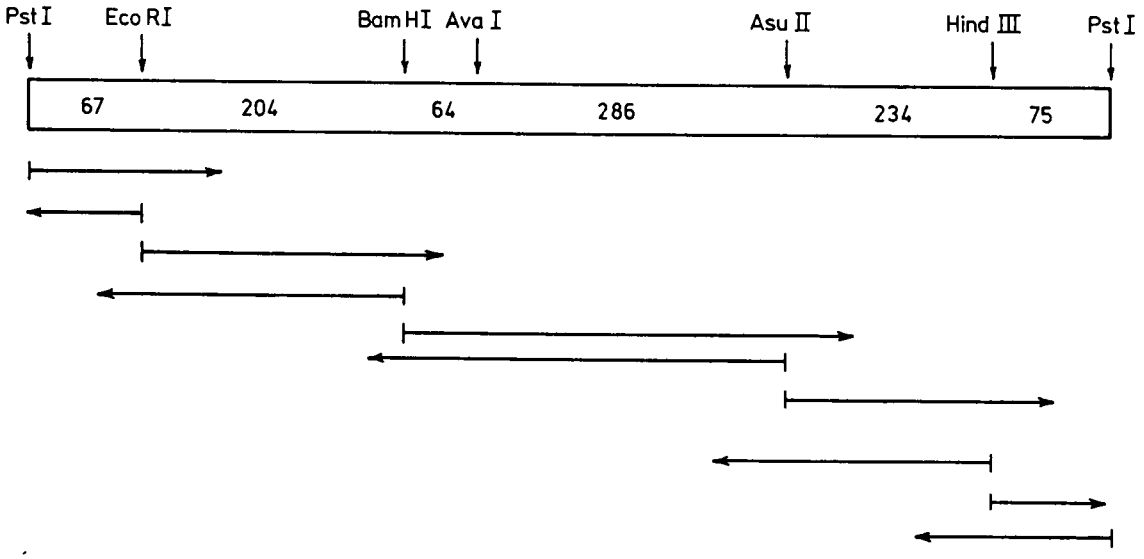


Figure 11

positions of sequence elements will be reported as the first nucleotide of those elements. Sequence positions within the p320 insert will be reported in reference to orientation A; position 1 is the first nucleotide of the (+) strand.]

Restriction Fragment Length Polymorphism Analysis

Restriction fragment length polymorphism (RFLP) is a simple and commonly used assay for demonstrating that two (or more) DNA samples are different. If distinct sets of cleavage products are generated by digestion of two DNA samples with the same restriction enzyme then the DNAs must be different. (A rare exception includes the case when one of two identical DNA sequences is methylated at the site of the restriction enzyme chosen, rendering it inert to the enzyme.) Axiomatically, this technique cannot be used to prove that two DNAs are the same (except to a low level of resolution); two DNA samples must both be sequenced to demonstrate exact correspondence. Therefore the only sorts of definitive conclusions which can be supported by such an analysis are that two DNAs are different or that two DNAs share the same gross structure. However, the greater the density of matching restriction enzyme sites (in the absence of negative data) the higher the probability that the samples are in fact the same.

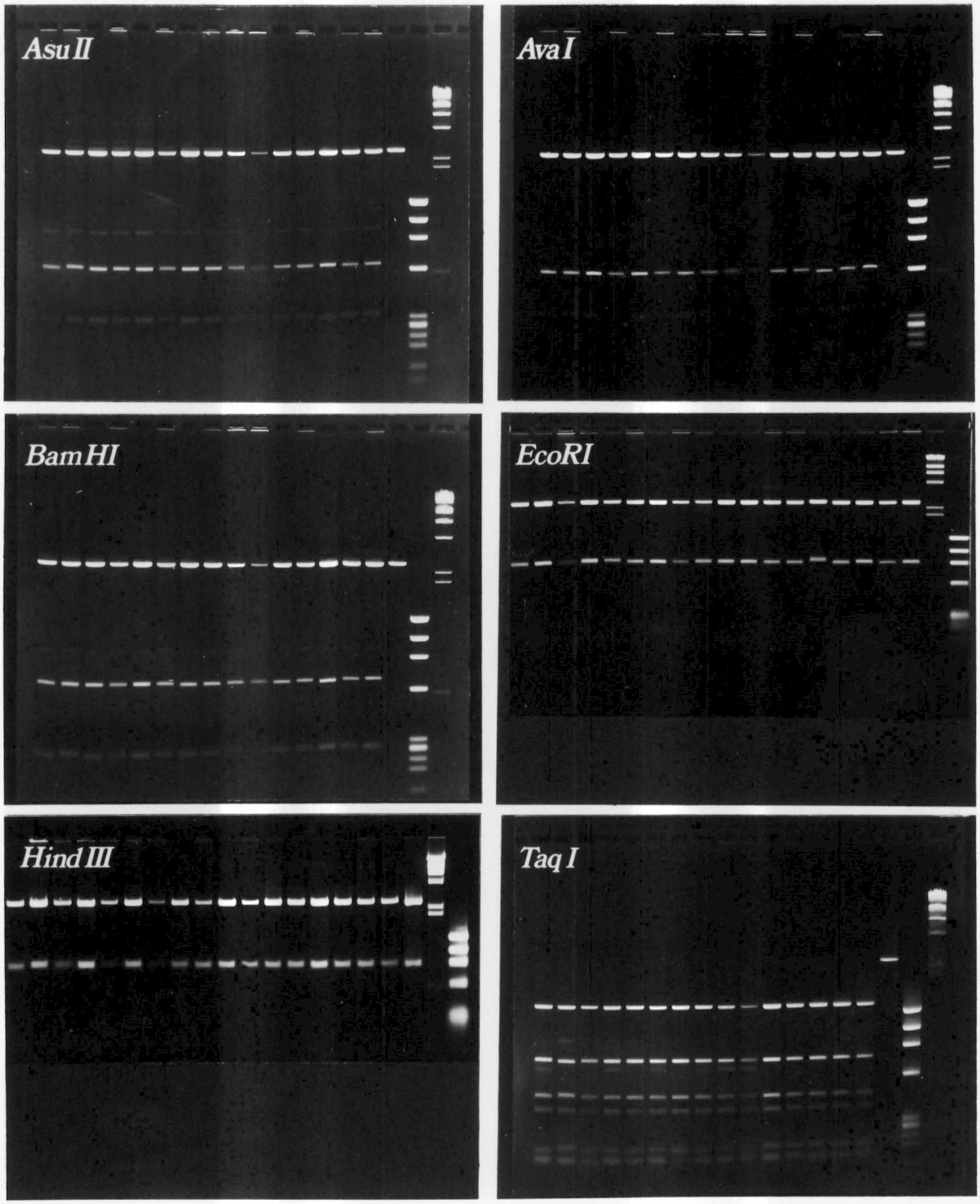
To compare the various putative isolates of the Euplotes macronuclear 5S rRNA gene, six experiments were performed in each of which the plasmid clones were double-digested with Pst I and another restriction enzyme. (Since during construction of the macronuclear

genomic DNA library inserts were cloned randomly into the vector in either of the two possible orientations (A or B), it was necessary to excise them with Pst I to allow direct comparison of fragmentation patterns generated by other restriction enzymes.) The Pst I-treated samples were individually digested with Asu II, Ava I, Bam HI, Eco RI, Hind III (the same enzymes used in the p320 mapping experiment), and Taq I. The DNA cleavage products were purified from the reactions and analyzed by electrophoresis in agarose gels, one gel per enzyme (Fig. 12). Although the execution of these experiments was mildly tedious, the interpretation of the results is straightforward: for each restriction enzyme tested, no fragment length heterogeneity was observed among the various isolates. (Recall the clones are similar also in that they all lack internal Pst I sites, although in this kind of analysis a negative result is not very informative.) Therefore the clones do not differ by global insertions, deletions, or rearrangements, although discrete base changes cannot be excluded; all of the isolates appear to share a common gross structural organization. This result, taken together with the strict degree of homology among the inserts as demonstrated by hybridization, supports the notion that all of the clones represent multiple, identical isolates of the same macronuclear DNA molecule. Upon demonstration that the p320 insert is truely a Euplotes macronuclear 5S rRNA gene, it can thus be concluded that the Euplotes macronucleus probably harbors only one sort of 5S rRNA gene.

From a rational viewpoint, this result was not unexpected. Recall that Xenopus contains two types of 5S rRNA genes which are differentially regulated during development. Modulation of the general level of

Figure 12: Restriction fragment length polymorphism analysis.

The putative Euplotes macronuclear 5S rRNA gene clones were compared by restriction endonuclease digestion. Ethidium bromide staining of six agarose gels containing the cleavage fragments of the various recombinant plasmids double digested with Pst I and Asu II, Ava I, Bam HI, Eco RI, Hind III, or Taq I as indicated. The gels are loaded left to right as follows: the restriction fragments of the various recombinants, pUC9 x Pst I (included on the Asu II, Ava I, Bam HI, and Taq I gels only), molecular weight markers ØX and λ. (The absence of unit size vector in the recombinant plasmid lanes of the Taq I gel is because pUC9 contains 3 Taq I sites whereas the other enzymes cleave pUC9 only in the polylinker.)


Asu II
Ava I
Bam HI
EcoRI
Hind III
Taq I

5S rRNA production during development is mediated by changing the concentration of TFIIIA and the differential expression of the two forms of the 5S gene during development is caused by differential binding affinities of this factor for the two genes (4, 26, 27). The reason that the Xenopus genome contains so many 5S rRNA genes (∿40,000 oocyte-type and ∿800 somatic-type per diploid nucleus) is to allow oocytes to synthesize ribosomal components at a rate faster than is required by somatic cells and the reason that there are two types of 5S rRNA genes is to provide a mechanism whereby most of these genes (the oocyte-type genes) may be permanently (stably) repressed following oogenesis. Thus Xenopus 5S rRNA gene expression is regulated during development at (at least) two levels: modulation of TFIIIA concentration and differential binding affinity of this positive effector for the two classes of 5S genes. [Here is advanced the hypothesis that Xenopus uses these two interrelated levels of control for (at least) three interrelated reasons. First, the greater the number of levels of control the more finely can control be adjusted (in general) and the more sensitive is control to a greater number of variables. Second, a low steady state level of a particular transcription product can be more accurately and precisely regulated from a smaller number of genes (it is easier to regulate 800 genes than it is to regulate 40,800 genes). Third, less energy needs to be expended by a cell to control a smaller number of genes. (Remember that entropy and information are thermodynamic opposites (see 186, 187). Chaos is the manifestation of entropy, order is the manifestation of information. Chaos happens spontaneously,

order requires the expenditure of energy. Gene regulation is the regulation of genomic information (in fact, gene regulation per se is a form of information in itself) and this is work which requires energy. Less energy is required to regulate 800 genes than is required to regulate 40,800 genes. A general principle obvious from biochemistry is that cells are usually very economical in their use of energy. After oogenesis, a low steady state level of 5S rRNA synthesis can be maintained more economically from a small number of templates than if all the genes were kept potentially active. Thus following oogenesis most of the genes are permanently shut off (stably repressed) so that no more energy is required for their regulation.)] Since Euplotes is a unicellular organism and generates no oocytes and undergoes no embryonic development as does Xenopus, no differential regulation of 5S rRNA gene expression is required. The oocyte-type 5S genes of Xenopus amount to extra copies which are needed only during oogenesis; since the Euplotes life cycle contains no analogous developmental stage, presumably two types of 5S genes are not required.

Obviously, this inductive conclusion is subject to attack by a very simple and powerful criticism: if two forms of the 5S gene do in fact exist in Euplotes, and if they differ by only two base pairs within the internal control region as do the two forms of Xenopus 5S genes, then this subtle difference would not be detected in this assay. Since no heterogeneity was observed in the putative isolates of the Euplotes macronuclear 5S rRNA gene by investigation at the DNA level, the

question of sequence variants within this organism would probably best be pursued by sequencing native 5S rRNA isolated from Euplotes cells and looking for ambiguities at specific positions (a sequencing experiment performed on a mixture of molecules differing at a particular base will yield an ambiguous signal at that base).

M13 Subcloning of the p320 Insert

Ligation Theory

To generate single-stranded DNA templates for sequence analysis, the gel-purified Pst I-insert of p320 was subcloned into the Pst I site of bacteriophage M13 (RF) by cohesive-end ligation. Optimum ligation reaction conditions were predicted using the formulas given in Maniatis et al. (154, pp. 286-288). The two most important parameters in a ligation reaction are the relative abundance (molar ratio) of DNA reactants and the absolute concentration of DNA termini. Two sorts of joining reactions are possible: intermolecular interactions form linear concatamers and intramolecular interactions form circles. The ratio of these two sorts of reaction products is determined by two parameters: i and j. i is the total concentration of DNA termini in the reaction (in ends/ml). j is the effective concentration of one end of a DNA molecule in the neighborhood of the other end of the same molecule (in ends/ml). When $i < j$, circles are favored; when $i > j$, concatamers are favored; and when $i = j$ the two products accumulate at equal rates. Intuitively, this simply means that in a reaction containing a high DNA concentration the terminus of one DNA molecule has a relatively high probability

of colliding with the terminus of another DNA molecule, whereas in a dilute reaction one end of a DNA molecule will find the opposite end of the same molecule before encountering the terminus of a different DNA molecule, and at some intermediate DNA concentration these reaction rates are equal. Operationally, j is a constant for a DNA molecule of given length in a given buffer (and is a result of the physical properties of that DNA molecule), and i (the concentration of DNA termini in the reaction) is a parameter which can be adjusted by the investigator. One calculates j (using a theoretical equation) for the DNA species at hand and then adjusts i to be equal to, less than, or greater than j depending on the desired result (as per above). A molar ratio of insert:vector equal to 1:1 (or 2:1) is usually selected.

The calculation of j is based on the assumption that a linear duplex DNA molecule in solution behaves as a random coil. Since the longer a DNA molecule is the less frequently its ends will collide (the lower is the effective concentration of one end of the molecule in the neighborhood of the other end), j is inversely proportional to the length of the DNA molecule. Thus j is a function of molecular weight and expresses the concentration of one end of a DNA molecule as seen by the opposite end of the same molecule; j can be thought of as a measure of the frequency at which the termini of a given DNA molecule will collide. j is given by:

$$j = \left(\frac{3}{2\pi l\, b}\right)^{3/2} \qquad \text{(dimensions ends/ml)}$$

where

l = the length of the DNA in cm, and

b = the length of a randomly coiled segment of DNA in cm.

The value of b is dependent on the ionic strength of the buffer, which affects the rigidity of the DNA (154). For DNA in ligation buffer the above equation simplifies to:

$$j = \frac{5.504 \times 10^{22}}{MW^{3/2}} \qquad \text{(dimensions ends/ml)}$$

Thus j is a constant for a DNA molecule of given length in a given buffer.

The calculation of i, the total concentration of complementary DNA termini in the solution, is straightforward as expected. For DNA with self-complementary, cohesive ends:

$$i = 2\ AM \times 10^{-3} \qquad \text{(dimensions ends/ml)}$$

where

A = Avogadro's number, and

M = molarity.

Axiomatically, this is simply two times the concentration of molecules/ml, since each molecule has two ends. For DNA with noncohesive termini:

$$i = AM \times 10^{-3} \qquad \text{(dimensions ends/ml).}$$

This represents the concentration of unique termini and is simply the concentration of molecules/ml.

Thus to assemble a ligation reaction, one calculates j from theory, selects a molar ratio of insert:vector (typically 1:1), and then determines the required concentrations of DNA reactants which will result in $i > j$, $i < j$, or $i = j$ depending on if the desired product is concatamers, circles, or both, respectively. In this application, the intact p320 insert was subcloned into the Pst I site of M13 (RF) by cohesive-end ligation (site-directed insertion mutagenesis). A successful construct could only be formed via one pathway: intermolecular ligation of an insert molecule and a vector molecule (concatamerization) followed by intramolecular ligation of this chimeric linear molecule (circularization). Therefore, to maximize the yield of recombinant circles, the insertion ligation reactions were assembled such that $i = j$. In this situation one end of a particular DNA molecule is equally likely to contact the opposite end of the same molecule or an end of a different DNA molecule, so concatamers and circles (intermolecular interactions and intramolecular interactions) should form at equal rates. In the subsequent application, generation of small subfragments of the p320 insert for use as templates in the DNA sequencing reactions (site-directed deletion mutagenesis as described below), a successful construct could only be formed by intramolecular circularization, so the deletion ligation reactions were assembled such that $i << j$. In this situation the DNA is so dilute that an end of one DNA molecule will collide with the opposite end of the same molecule before it encounters a terminus of a different DNA molecule, so circles are favored. [This discussion of ligation theory was closely adapted from Maniatis et al. (154, pp. 286-288).]

Construction of M13 Subclones

Two different bacteriophage M13 vectors were used: M13mp18 and M13mp19. These differ only in the orientation of the polylinker. The primer binding site is located just 3' to the polylinker in both DNAs (3' in reference to the single-stranded viral DNA structure, the (+) strand which is packaged into phage particles and which serves as template for the sequencing reactions). The use of both vectors provides great versatility in manipulation of insert sequences and potentiates a broad spectrum of deletion mutagenesis, thereby allowing many routes of access to the interior of the insert and conveniencing the complete sequence determination of both strands. (Refer to Experimental Procedures for additional information on the M13 cloning and sequencing system.)

The insert of p320 was excised with Pst I and purified in the usual manner by extraction with phenol, preparative gel electrophoresis, column chromatography, and precipitation from ethanol. The M13 vectors were linearized with Pst I and purified by extraction with phenol and precipitation from ethanol. Ligation reactions were performed as described in Experimental Procedures and the products were introduced, without subsequent purification, into E. coli JM107 by RbCl-mediated transfection. Colorless plaques were used to infect liquid cultures of JM107 from which were isolated viral RF DNAs. The presence of unit-size p320 insert was verified by digestion of the recombinant phage DNA with Pst I and analysis of the reaction products by agarose gel electrophoresis. The orientation of inserts within the

chimeric constructs was determined by digestion of the RFs with Eco RI, an enzyme which cleaves the insert asymmetrically and which cuts the vector only once (in the polylinker). Since the restriction map of the insert was known, the sizes of the Eco RI fragments could be predicted for each of the two possible orientations of the insert. Four parental M13 subclones were isolated: the intact p320 insert was cloned in both orientations (A and B, as previously defined) into the Pst I site of both M13 vectors (mp18 and mp19). These four parental M13 constructs may be referred to as 18A, 18B, 19A, and 19B for brevity: 18A is M13mp18 with the p320 insert in orientation A, and so on (Fig. 13).

Since the limit of resolution of current polyacrylamide gel electrophoresis technology is about 350-400 bases, deletion mutagenesis was performed to generate subfragments of the insert which would allow initiation of the sequencing reactions within the interior of the insert. In this way each strand could be sequenced entirely. From the four parental M13 subclones the following deletions were made: 18A-Hind, 18B-Hind, 19A-Bam, 19B-Bam, 19A-Eco, 19B-Eco, 18A-Asu, and 19B-Asu. Deletion mutants were constructed by cleaving the appropriate recombinant viral RF with a restriction enzyme which cuts once in the insert and once in the polylinker at a position between the Pst I site and the primer binding site of the vector. This generates two fragments: a large fragment containing the vector joined to part of the insert and a small fragment containing the rest of the insert and a few bases ($\sim$10) from the polylinker. After restriction enzyme digestion,

Figure 13: M13 subcloning strategy.

The multiple cloning sites (MCS) of M13mp18 (above) and M13mp19 (below) are shown. The p320 insert (triangles) was cloned in both orientations (A and B) into the Pst I site of both vectors. Subclone 18A is M13mp18 with the insert in orientation A, and so on. Deletion mutants were made from these four parental M13 subclones using the enzymes indicated along the insert, as described in the text. DNA sequencing reactions initiate at the primer and proceed leftward.

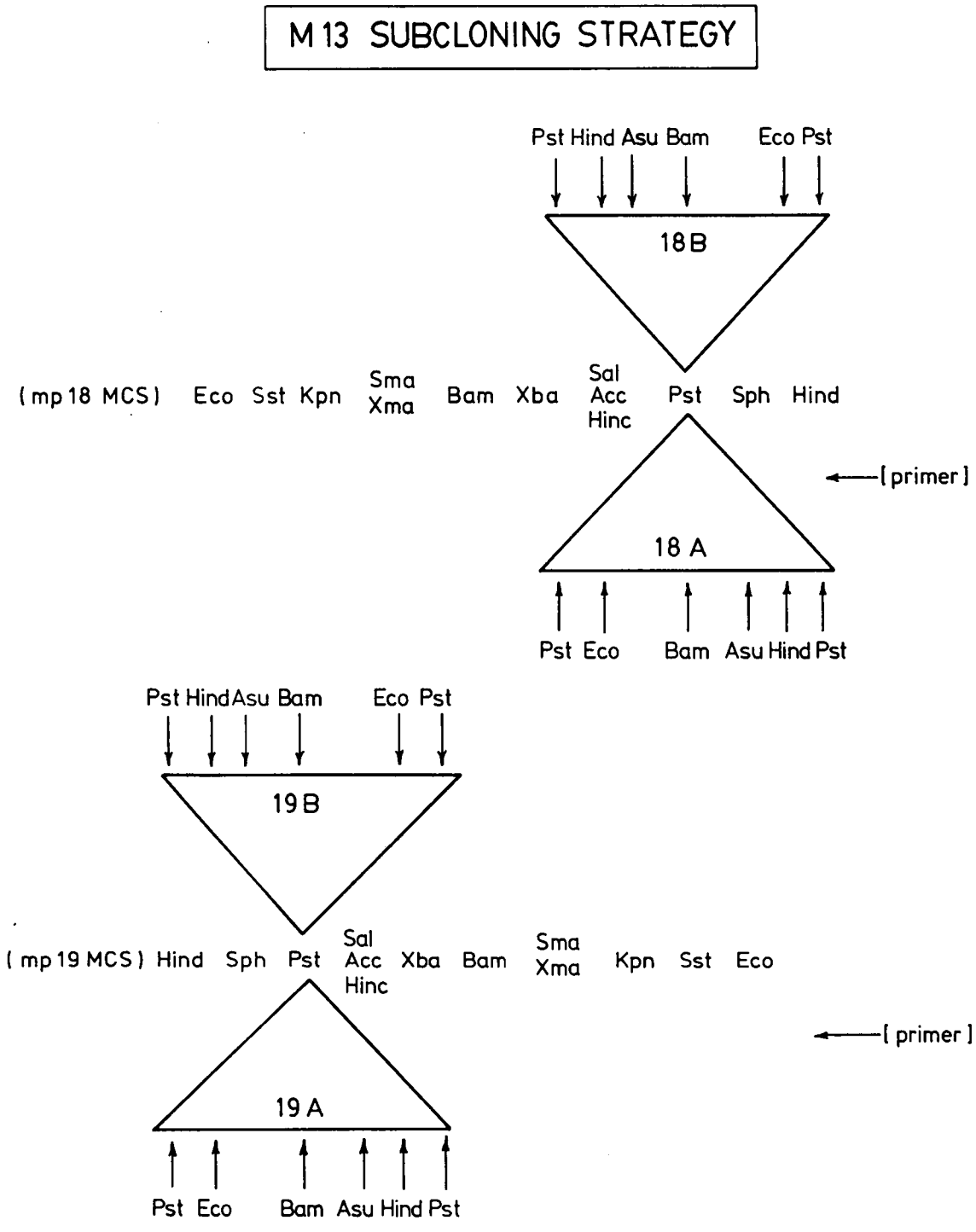

M13 SUBCLONING STRATEGY
Pst Hind Asu Bam
Eco Pst
18B
(mp18 MCS) Eco Sst Kpn Sma Xma Bam Xba Sal Acc Hinc Pst Sph Hind
[primer]
18 A
Pst Eco Bam Asu Hind Pst
Pst Hind Asu Bam
Eco Pst
19B
(mp19 MCS) Hind Sph Pst Sal Acc Hinc Xba Bam Sma Xma Kpn Sst Eco
[primer]
19 A
Pst Eco Bam Asu Hind Pst

Figure 13

the DNA was purified by extraction with phenol and precipitation from ethanol. The linear DNA fragments were then rejoined by cohesive-end ligation under reaction conditions were i << j so that formation of unimolecular circles would be highly favored. These mutant recombinants were introduced into bacteria and the RFs were screened by restriction enzyme digestion as before. Phage supernatants from successful constructs were then used in the preparation of viral genomic DNA (a single-stranded DNA structure) for use a template for the sequencing reactions.

As a specific example, consider the construction of the Bam HI deletion of clone 19A (refer to Fig. 13). Parental 19A was cleaved to completion with Bam HI, which cuts the insert once (near the middle) and the polylinker (MCS) once (right of the Pst I site). This results in two fragments: a large fragment containing the vector with the left half of the insert still attached to its Pst I site and a small fragment containing the right half of the insert plus the portion of the MCS between its Pst I and Bam HI sites. Upon religation in very dilute solution, these two linear fragments will form two unimolecular circles: a large circle containing the vector plus the left half of the insert, and a small circle containing the right half of the insert plus a few bases from the MCS. Thus the short Bam-Bam fragment, containing the right half of the insert, has been deleted. This mutant is called 19A-Bam and allows sequencing to initiate at the Bam site within 19A.

Since M13 contains no Asu II sites, construction of deletion mutants allowing initiation of sequencing at the Asu II site within the

insert was a bit more complicated. For generation of the Asu II deletion of 18A, the 18A-Hind mutant was double-digested with Asu II and Hind III. The resulting DNA termini were made flush using the Klenow fragment of E. coli DNA polymerase I and the structure was sealed by blunt-end ligation under conditions of i << j (favoring circles). Thus the fragment spanning the Asu II and Hind III sites within the insert was deleted, and the Asu II site of the insert was joined to the Hind III site of the MCS (allowing sequencing to initiate at the Asu II site of 18A and proceed leftward through the Bam HI site). For the Asu II deletion of 19B, the 19B-Bam mutant was double-digested with Asu II and Bam HI. The resulting DNA termini were made flush using the Klenow enzyme and the structure was circularized by blunt-end ligation as before. Thus the fragment spanning the Asu II and Bam HI sites within the insert was deleted, and the Asu II site of the insert was joined to the Bam HI site of the MCS (allowing sequencing to initiate at the Asu II site of 19B and proceed leftward through the Hind III site). (It may be worthy of notice that the above subcloning strategy involving site-directed deletion mutagenesis is much simpler and easier than the alternative approach of cloning each of the required fragments individually.)

DNA Sequence Analysis of the Euplotes Macronuclear 5S rRNA Gene

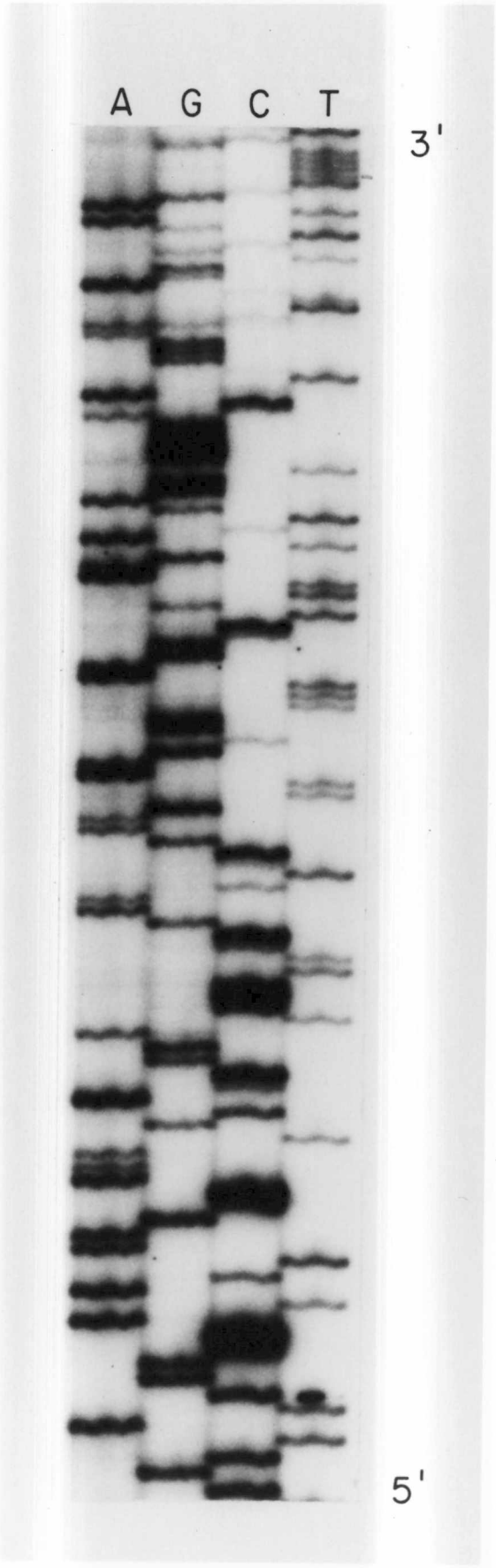
The collection of M13 subclones harboring the p320 insert allowed both strands of the Euplotes macronuclear DNA molecule to be

sequenced entirely with overlap between adjacent start sites; the deletion mutants allowed sequencing to initiate at sites internal to the insert (at the Eco RI, Bam HI, Asu II, and Hind III sites) and proceed outwardly from those sites in both directions. (Refer to Figures 11 and 13.) Constructs harboring the insert in orientation A permitted sequence determination of the (-) strand (as previously defined) while constructs harboring the insert in orientation B permitted sequence determination of the (+) strand.

Two independent isolates of each template were prepared; DNA sequencing reactions, high resolution polyacrylamide gel electrophoresis, and autoradiography were performed as described (Fig. 14). Both strands of the entire Euplotes macronuclear DNA insert were sequenced at least twice (using two independent template isolates) with overlap between adjacent start sites resulting in a minimum four-fold redundancy of data acquisition. [There are two minor exceptions to the previous generalizations: only one isolate of the Asu II deletion of 18A was prepared, so a small stretch of sequence between the Asu II and Bam HI sites was determined with only three-fold redundancy; also, it proved to be impossible to sequence through the poly(dG) tracts flanking the 5' ends of the insert (presumably due to secondary structure formation by the nascent transcript), so although the short segments flanking the Eco RI and Hind III sites were determined with four-fold redundancy this information was derived by sequencing the same strand four times (from four independent template isolates).] In all cases the data were internally consistent and reproducible; in no case were

Figure 14: A typical DNA sequencing gel.

Autoradiogram of a DNA sequencing gel showing the sense strand of the 5S rRNA coding region. Lanes from left to right are: A, G, C, T.


A
G
C
T
3'
5'

unambiguous but contradictory data obtained (in a few nucleotide positions the sequence information from one strand was ambiguous but the sequence from the opposite strand was clear). Up to 396 bases were determined from a single template, although 300-350 bases was more typical. The p320 insert was found to be 930 bp in length and to contain a single coding region for 5S rRNA (Fig. 15). Several independent lines of evidence support the conclusion that the p320 insert does in fact accurately represent an intact Euplotes macronuclear 5S rRNA gene:

1. The p320 insert hybridizes specifically to the 5S rRNA gene of Tetrahymena.
2. Nucleotide sequence analysis of the p320 insert clearly demonstrates that it contains a 5S rRNA gene coding function (see below).
3. The p320 insert hybridizes specifically to a Euplotes cellular RNA 120 nucleotides in length (see below).
4. The p320 insert accurately and specifically directs transcription of a 120 base RNA in vitro (see below).
5. That the p320 insert represents an intact isolate of a macronuclear DNA molecule is demonstrated by the presence of intact telomeric sequences flanking its ends (and by 6 and 7 below).
6. That the p320 insert represents an intact, unit size clone of the native macronuclear gene is also indicated by the observation that the insert specifically hybridizes to a single band in

Figure 15: Complete nucleotide sequence of the Euplotes macronuclear 5S rRNA gene.

The macronuclear 5S rRNA gene of Euplotes eurystomus was found to be 930 bp in length and to contain a single coding region of 120 bases. The coding region is underlined (positions 252-371); the strand homologous to the RNA is listed. The telomeres and presumptive replication origin are indicated. Fifteen bp of the G•C tracts added to each end of the DNA molecule during the cloning procedure are included. The sequence is listed in orientation A. (See text for additional discussion.)

```
          10         20         30         40         50         60         70         80
5' GGGGGGGGGGGGGGGCCAAAACCCCAAAACCCCAAAACCCCAAAACCCCAAAACCCCAAGTGAGAACTAAACACTTGAAT

          90        100        110        120        130        140        150        160
   TCAGATATAAGTATTATAATTCAGGAATCTGAGATTTACGGAATTATAAATCCAGATGGACTAGAGCATGCCCAAGCGTA

         170        180        190        200        210        220        230        240
   AGTAAACGGTTTTTCAAAGACTGGAGAGATGTTTTGGGTGAGTTCGATTTTAGGTGTTGAGTATATAAAGAGTGGCCTTA

         250        260        270        280        290        300        310        320
   CTATAATTATCGCTATCGGCCATACTAAGCCAAATGCACCGGATCCCTTCCGAACTCCGAAGTTAAGCGGTTTAAGGCCT

         330        340        350        360        370        380        390        400
   GTTAAGTACTGAGGTGGGCGACCACTCGGGAACTTCAGGTGCTGATAGCTTTTTGCTCCTGAAGCTATCTTTTTGCACTC

         410        420        430        440        450        460        470        480
   TTTCTTTTTATCTATCCACCCTTCAGTACTTCTCACACTATCCAGGTTGGCGAGTTTTTGTTTCGTCGTGCTTGGCATAA

         490        500        510        520        530        540        550        560
   TGGAAAATCATCCTTCTTTGTGATAAAATAGAAAACAAAATGGCTAGTACAATCTTGAGACACCCGATGCTGGGTTTTGG

         570        580        590        600        610        620        630        640
   GTACAAAGACTGCCAATACGTCCATCCCCTATCGAATTACCAAATATTGTTTTCTAATTTGGGTATTAAGTTCGAATCTC

         650        660        670        680        690        700        710        720
   GCCAATATTTAAGTTTTTATTTCGCCATTAATCACAAATATGGCTAAATCTCAAATAAAAATCAGATATTAACTACCCAT

         730        740        750        760        770        780        790        800
   TTGGATTTAATATCCGATCCAAGACTCTATACCTGATGGTTGCAAGTTCTAAAACAATTTATTTTGTTGGGGAAGACAAA

         810        820        830        840        850        860        870        880
   GAGTCGAATAAAACCCTAATTAATTTTATTGTACATCTTGGCTTCAGTGCCATTATTAATTTTCAAGCTTTAAAGTTTTA

         890        900        910        920        930        940        950        960
   GTTCAAGTCTGGAATCGTAGCACGGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTTGGCCCCCCCCCCCCCCC  3'
```

Figure 15

native Euplotes total macronuclear genomic DNA which is the same size as the insert.

7. That the p320 insert represents a faithful copy of the native Euplotes macronuclear 5S rRNA gene is indicated by the fact that multiple isolates of the gene are identical. (If the p320 insert had been the victim of some sort of cloning artifact such as degradation, insertion, deletion, rearrangement, recombination, or any other type of mutation, or if the p320 insert happened to be an atypical version of the macronuclear 5S rRNA gene or a pseudogene, then not all of the other isolates of the gene would have been the same. The probability of the same accident happening to all 19 isolates is essentially zero.)

Sequence analysis confirmed the positions of restriction enzyme sites mapped earlier; a fine-structure map was generated by computer (Fig. 16).

The structure of the Euplotes macronuclear 5S rRNA gene lends support to the dogma of macronuclear genome structure and development. The linear macronuclear DNA fragment contains a single coding region without introns (although this is not especially enlightening since no 5S rRNA genes contain introns) and is bounded by telomeres. The sequence appears to contain all of the elements necessary for its transcription, replication, and stability. The sequence represents a discrete, autonomous, monocistronic genetic function as per the canon of macronuclear gene structure (compare Figures 4 , p. 36, and 17). This work represents the first report of a DNA sequence from Euplotes

Figure 16: Restriction enzyme sites within the Euplotes macronuclear 5S rRNA gene.

A high resolution restriction enzyme map of the Euplotes macronuclear 5S rRNA gene was generated from the nucleotide sequence by computer. The abscissa represents the gene in orientation A; restriction enzymes are listed along the ordinate; stars indicate cutting sites.

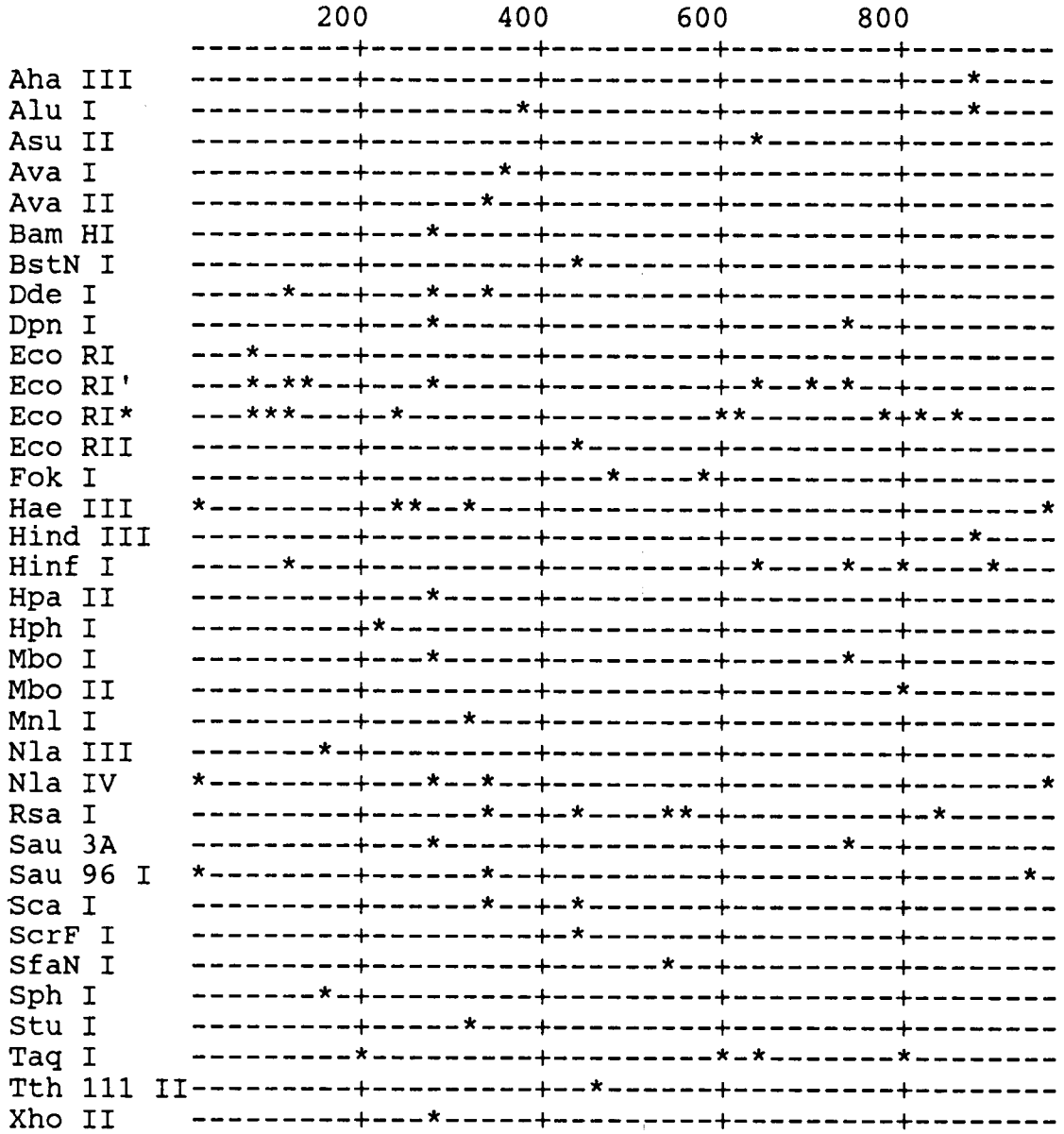
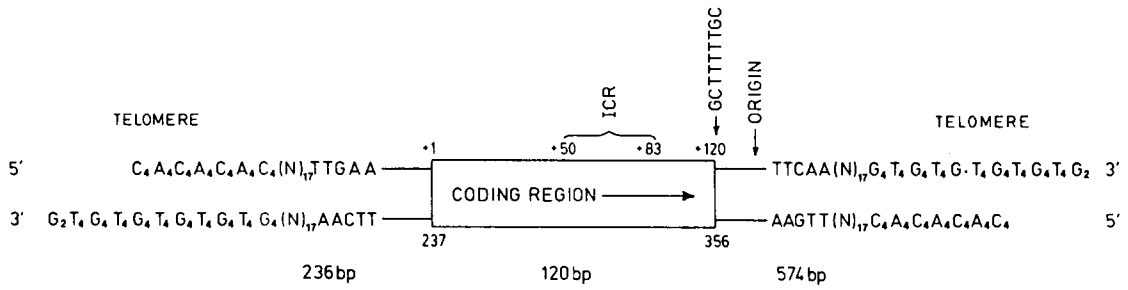

200
400
600
800
Aha III
Alu I
Asu II
Ava I
Ava II
Bam HI
BstN I
Dde I
Dpn I
Eco RI
Eco RI'
Eco RI*
Eco RII
Fok I
Hae III
Hind III
Hinf I
Hpa II
Hph I
Mbo I
Mbo II
Mnl I
Nla III
Nla IV
Rsa I
Sau 3A
Sau 96 I
Sca I
ScrF I
SfaN I
Sph I
Stu I
Taq I
Tth 111 II
Xho II

Figure 16

Figure 17: Schema of the Euplotes macronuclear 5S rRNA gene.

A diagram indicating the salient structural features and sequence elements of the Euplotes macronuclear 5S rRNA gene.



STRUCTURE OF THE EUPLOTES MACRONUCLEAR 5S rRNA GENE (930bp)

Figure 17

eurystomus, the first report of a 5S rRNA gene sequence from any species of hypotrichous ciliate (the 5S rRNA of Euplotes woodruffi has been sequenced, but not its gene), the first report of a complete nucleotide sequence of any Euplotes macronuclear DNA molecule (the large rRNA gene of Euplotes aediculatus has been partially sequenced), and only the fourth report of a complete nucleotide sequence of any macronuclear DNA molecule from any species of hypotrichous ciliate (several others have been sequenced in part) (refer to Table 3, p. 39).

Sequence Elements Within the Euplotes Macronuclear 5S rRNA Gene

The telomeric sequence reported here for Euplotes eurystomus (see Figs. 15 and 17) is identical to that previously published for Euplotes aediculatus (121), although here no effort was made to demonstrate the presence of single-stranded 3' overhangs. Also conserved is the pentanucleotide sequence TTGAA located 17 bp inward from the telomeres at each end; the function of this sequence, which is unique to Euplotes, and the reason for its conservation are unknown. The fact that the same terminal structure is conserved among all of the macronuclear DNA fragments within an individual organism as well as among the various species of hypotrichous ciliates strongly implies that its sequence is essential to its function. It has been suggested that perhaps these sequences serve as recognition sites for the excision of genes from the micronucleus during macronuclear development. This possibility was

discounted when it was found that the telomeric sequences are not associated with macronuclear genes as they reside in the micronucleus (116-118). The actual function of the telomeric sequence may be to confer stability to the macronuclear DNA fragments. This is by provision of a mechanism of replication of the linear DNA structure. (Recall that ciliate telomeres allow linear plasmids to be stably maintained in yeast; see 125 and 131-134.) The stringent conservation of the telomere sequence may also suggest that these DNA structures are recognized by specific proteins. Recently two such telomere-specific proteins have been identified in Oxytricha (135). These proteins, of 26 and 55 KDa, are not extracted by 2 M salt or by chloroform. They specifically recognize and bind to the unique single-stranded 3' tail of native macronuclear DNAs (plus the adjacent duplex region) and have been demonstrated to protect the DNA termini from digestion by BAL 31 exonuclease in vitro. These telomeric proteins may function to protect the linear DNAs from exonucleolytic attack in vivo (and thus enhance their stability), they may be involved in replication, they may play a role in chromatin structure, they may constitute anchors against which the linear DNAs can be torsionally strained, they may bind the DNA termini to prevent nonspecific end effects (to be discussed in more detail later), or they may be involved in some other function(s).

At the 3' end of the coding region (position 368) of the Euplotes macronuclear 5S rRNA gene was found the canonical termination signal for all class III genes: GCTTTTTGC (some genes contain only 4 T's) (1). Termination is after the second T.

The Euplotes macronuclear 5S rRNA gene was found to contain sequences similar to those believed to be origins of replication in other organisms. Referring to the (+) strand in orientation A, at position 771 is found the A•T rich sequence AAAACAATTTATTTT which is capable of forming a self-complementary cruciform structure and nearby at position 817 is the inverse repeat TAATTAAT. Also, at position 852 is found another inverse repeat ATTATTA and adjacent at position 861 is the sequence TTTCAAGCTTTAAA which is capable of forming a self-complementary cruciform structure. These structures are very similar to the presumptive replication origins reported for Oxytricha (148, 149) and Stylonychia (150, 151) and also resemble the replication origin of the polyoma virus chromosome (208) and autonomously replicating sequences (ARS) from yeast (209, 210). ARS elements are typically A•T rich and contain a hairpin (or cruciform) structure near an inverse repeat. ARS elements confer ability to replicate to bacterial plasmids in yeast. In the few examples which are available, such sequences have been found ∿50 bp inward from the telomeres at one or both ends of macronuclear DNA molecules from hypotrichous ciliates. By electron microscopy bubbles and forks have been observed to correlate with these positions (151). Delineation of which of the above pairs of hairpins and repeats serves as the replication origin of the Euplotes macronuclear 5S rRNA gene (if either) awaits further investigation. The second pair, between positions 852 and 874, is a likely candidate since in this structure the hairpin is separated from the repeat by only 2 bp (as is the case for the actin gene of Oxytricha; 148, 149), since

this sequence is located ∿50 bp inward from one of the telomeres, and since this position coincides with a G•C minimum of the macronuclear DNA molecule (see below).

The G•C content along the Euplotes macronuclear 5S rRNA gene was calculated as a sliding local average; the value of % G•C was determined at intervals of 5 bp over a range of 20 bp on either side of each position (Fig. 18). Interestingly, the coding region (positions 252-371) coincides with the G•C maximum of the macronuclear DNA fragment (∿50% average G•C) while the presumptive replication origin (positions 852-874) resides at a G•C minimum (∿25% average G•C). This was not unexpected as coding regions in general are enriched in G•C and, as previously mentioned, ARS elements are A•T rich. Since A•T base pairs are less stable than G•C base pairs, replication origins may be A•T rich to allow easier local denaturation (melting) of the duplex thus providing enzymes and single-strand DNA binding proteins access to the interior of the structure and assisting in unwinding of the double helix during the initiation of replication. The telomeric sequences represent the second-most G•C rich structures of the macronuclear DNA molecule. Also provocative is the observation that within the 5S coding region the local G•C minimum coincides with the location of the internal control region. Destabilization of this sequence may be related to its function as a promoter by enhancing the accessibility of the sequence information to transcription factors. Upstream from the transcription start site are two depressions in local G•C content (between positions 70 to 120 and 205 to 240); one of these may represent an entry site for

Figure 18: G•C content of the Euplotes macronuclear 5S rRNA gene.

The G•C content along the Euplotes macronuclear 5S rRNA gene is displayed as a sliding local average; the value of % G•C was calculated at intervals of 5 bp over a range of 20 bp on either side of each position. Numbers along the abscissa refer to nucleotide positions of the macronuclear 5S gene and % G•C is indicated along the ordinate. The coding region (positions 252-371) coincides with the G•C maximum while the presumptive replication origin (positions 852-874) resides at a G•C minimum.

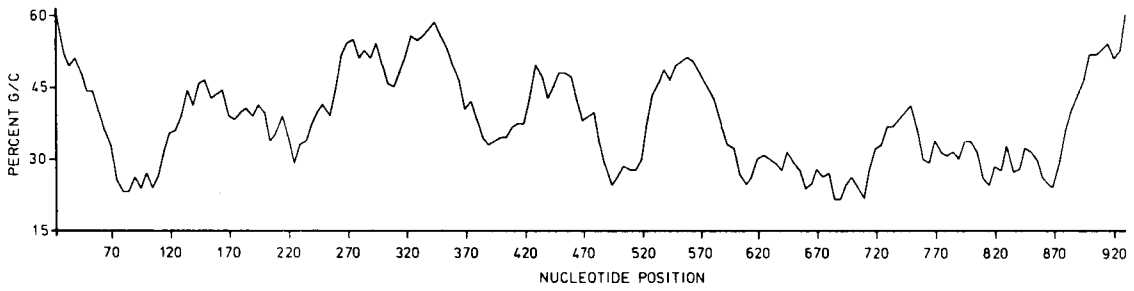

60
45
30
15
PERCENT G/C
70
120
170
220
270
320
370
420
470
520
570
620
670
720
770
820
870
920
NUCLEOTIDE POSITION

Figure 18

RNA polymerase III. In general, the local G•C content along the macronuclear 5S rRNA gene is not random nor does it vary in a uniform or regular way; various functional and regulatory sequence elements demonstrate particular characteristics which suggest that local thermal stability of the DNA structures has been adapted for the particular functions of those structures. A comparison of the G•C content of the Euplotes macronuclear 5S rRNA gene to other macronuclear DNAs from hypotrichous ciliated protozoa is presented in Table 8. The enrichment of G•C base pairs in the coding region of the Euplotes macronuclear 5S rRNA gene appears to be an occurrence common to macronuclear genes in general. The most significant difference between the Euplotes 5S gene and the other DNAs is in the noncoding portion of the molecule; as mentioned previously, Euplotes has the least similarity to the other hypotrichous ciliates. A plot was constructed wherein is displayed the ratio of purines/pyrimidines as a sliding local average along the Euplotes macronuclear 5S rRNA gene (data available upon request). The local A+G/C+T ratio is not random along the length of the macronuclear DNA molecule nor does its distribution vary in a uniform or regular way. The physical meaning and biological significance of this observation are unclear.

Comparative Sequence Analysis

As expected, the 5S rRNA coding sequence of the Euplotes macronuclear DNA fragment is highly homologous to 5S rRNA sequences from closely related organisms and is less homologous to sequences from more

Table 8: Comparative G•C Content of Hypotrichous Ciliate Macronuclear DNAs

DNA	%G•C	Reference
Total Euplotes MAC 5S gene	38.6[1]	here
Coding region Euplotes MAC 5S gene	52.5	here
Noncoding region Euplotes MAC 5S gene	36.5[1]	here
Total Euplotes MAC DNA	37	
Total Oxytricha MAC DNA	32.4[2]	101
Total Stylonychia MAC DNA	32	
Coding region hypotrich MAC DNA (ave.)	49	97
Noncoding region hypotrich MAC DNA (ave.)	25	97

[1]These values include the telomeric sequences but not the G•C tracts added during the cloning procedure.

[2]Reports in the literature for this value include 26%, 32.4%, and 42%.

distantly related organisms (Fig. 19). The DNA sequence presented here for E. eurystomus differs by only one nucleotide from the RNA sequence previously published for E. woodruffi (Genebank); the T at position +37 (in reference to the transcription start site) in the E. eurystomus sequence is replaced by an A in E. woodruffi. The Tetrahymena and Paramecium sequences were both found to be 79% homologous to the Euplotes sequence while the Euglena sequence is 72% homologous. The nontranscribed spacer sequence found in the 5S rRNA gene repeat structure of Tetrahymena is not significantly homologous to the DNA flanking the coding region of the Euplotes macronuclear 5S rRNA gene, thus the nontranscribed sequences must evolve much more rapidly (are less conserved) than do the coding sequences (assuming this is a valid comparison). This has been reported previously for the 5S rRNA genes of Xenopus (6) and the large rRNA genes of Stylonychia, Euplotes, and Oxytricha (145). Other sorts of organisms have 5S rRNA sequences less similar to the Euplotes sequence than do these protozoa. The primary nucleotide sequence of 5S rRNA has obviously been carefully conserved during evolution with the human sequence being 67% homologous to the Euplotes sequence. The sequences of man, mouse, rat, and most if not all other mammals are identical. The sequence is highly conserved among the vertebrates with the Xenopus somatic sequence differing from the mammalian sequence in only eight positions. The mammalian sequence is 60% homologous to that of yeast and as much as 50% of the sequence is conserved between eukaryotes and prokaryotes. In comparing the 13

Figure 19: Alignment of 5S rRNA gene sequences.

Note: The nucleotide sequences of twelve 5S rRNA genes are juxtaposed with that of Euplotes eurystomus. Sequences are listed in order of most to least homologous to the E. eurystomus sequence. Capital letters indicate homology to the E. eurystomus sequence, small letters indicate differences. Structural domains denoted A-E refer to helical regions within the RNA secondary structure (see p. 205). Stars indicate conserved residues. The consensus sequence was generated by assigning nucleotides to specific positions if they appeared there at >90% frequency (R = purine, Y = pyrimidine, N = nucleotide). Sequence data was obtained from Genebank.

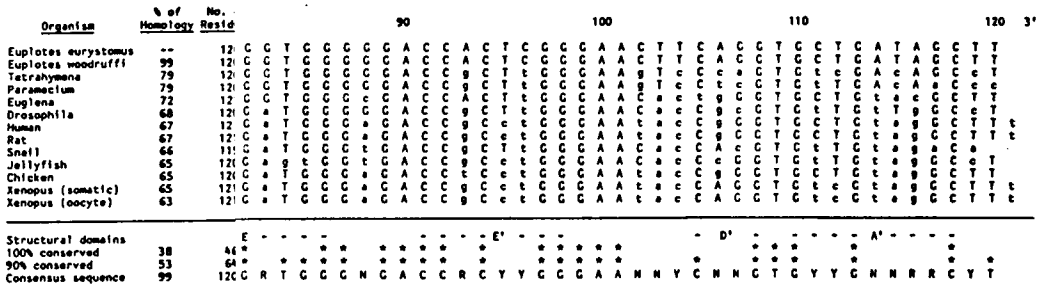

Organism
% of Homology
No. Resid
90
100
110
120
3'
Euplotes eurystomus
Euplotes woodruffi
Tetrahymena
Paramecium
Euglena
Drosophila
Human
Rat
Snail
Jellyfish
Chicken
Xenopus (somatic)
Xenopus (oocyte)
Structural domains
100% conserved
90% conserved
Consensus sequence

Figure 19

sequences listed in Fig. 19 it was observed that specific nucleotides are invariably conserved at 38% of the positions and at 53% of the positions specific nucleotides are conserved at >90% frequency. Within the ICR 56% of the positions are conserved at >90% frequency, so this structure is not conserved significantly more stringently than is the rest of the sequence. More than 90% sequence conservation was observed in 47% of the helical residues and in 63% of the nonhelical residues. Thus the single-stranded RNA structures, which are believed to be involved in protein binding, constitute the most highly conserved part of the molecule. The Euplotes sequence is more similar to the somatic-type sequence of Xenopus than to the oocyte-type sequence (but not by much) both with respect to the overall sequence and to the ICR sequence. The length of the molecule has also been highly conserved with all sequences listed containing 120 ± 1 bases. (The numbers reported above will of course be reduced when a larger number of sequences from a more diverse group of organisms is considered. In general, when purines and pyrimidines are considered as invariant residues, approximately 35% of the positions are universally conserved (85).)

Phylogenic Relationships Inferred From Comparison of 5S rRNA Sequences

When a comparative alignment of sequences, similar to that in Fig. 19, is expanded to include a large number of species, phylogenic relationships may be inferred. The 5S rRNA sequence has been widely applied to this end because it is ubiquitous to all organisms, it is abundant and easily isolated, it is relatively small and easily

determined, its function has been highly conserved, and most importantly its primary and secondary structure has been highly conserved (that is, it evolves slowly). A liability of the 5S rRNA sequence as a phylogenic marker is that since it is so short it contains a small number of signatures (diagnostic group-specific characteristics) and this limits its usefulness in comparison of closely related organisms.

Two general numerical methods have been applied to the alignment of sets of sequences and the generation of phylogenic trees from sequence data; within these two sorts of approaches there are several variations (85, 88, 91-93a). One method involves the comparison of the entire sequence and the generation of similarity matricies. The other method involves emphasis of unique signatures, such as group-specific homologies, insertions, deletions, and structures and positions of loops and helicies. Qualitative information about the relatedness of organisms is often most evident from signature markers while quantitative information is best attained by comparison of the complete sequences. "Limitations in tree analyses reside in problems of determining interrelatedness quantitatively within a genus or between genera because of too few differences in the sequences. Also, estimates of times of divergence between organisms are subject to errors arising from differences in rates of mutations" (from 85). The reader should be aware that the alignment of sequences and the inference of evolutionary relationships is not a trivial matter. Different numerical methods yield different dendrograms, thus various phylogenic schemes based on comparison of 5S rRNA sequences are not in perfect agreement. Furthermore, evolutionary trees based on 5S rRNA sequences are not in perfect agreement

with trees calculated from sequences of other informational macromolecules (other DNA, RNA or protein sequences). The ultimate phylogenic tree will be derived from correlation of many sorts of diverse informational sequences plus other biological criteria. Two phylogenic trees from the literature, one encompassing a diversity of life forms and the other emphasizing protists (including many protozoa), are presented (Fig. 20). (Refer also to 83-89 for additional information.)

Secondary Structure of 5S rRNA

The secondary structure of 5S rRNA has been predicted by methods based on the maximization of base pairing (maximization of thermodynamic stability). The greatest proof of this secondary structure (the "minimal model"; refer to Fig. 1, p. 8) is that all 5S rRNA sequences conform to it. (Note that the same secondary structure depicted in the minimal model can be drawn in several different but equivalent two dimensional representations.) Although no crystallographic data are available to confirm the proposed structure, various enzymatic, chemical, and physical studies provide support for it (see 85). Partial digestion of 5S rRNA with single-strand-specific ribonucleases A, T_1, and T_2 and double-strand-specific ribonuclease V_1 from cobra venom and subsequent analysis of the cleavage fragments by sequencing and other techniques support the proposed model. The interaction of a chemical or enzymatic structural probe with a macromolecule depends on the reactivity of the probe, the accessibility of the target sites, and the reaction conditions (85). The structure of 5S rRNA has been probed with various chemical and enzymatic reagents

Figure 20: Phylogenic trees based on 5S rRNA sequences.

20A: Dendrogram illustrating the evolutionary relationship of various eukaryotic, archaebacterial, eubacterial, and organellular species as determined by comparison of their 5S rRNA sequences [from Kuntzel, H., Piechulla, B., and Hahn, U. (1983) Consensus structure and evolution of 5S rRNA. Nucl. Acids Res. 11, 893-900].

20B: Dendrogram illustrating the evolutionary relationship of various protist species as determined by comparison of their 5S rRNA sequences [from Walker, W. F. (1985) 5S and 5.8S ribosomal RNA sequences and protist phylogenetics. BioSystems 18, 269-278].

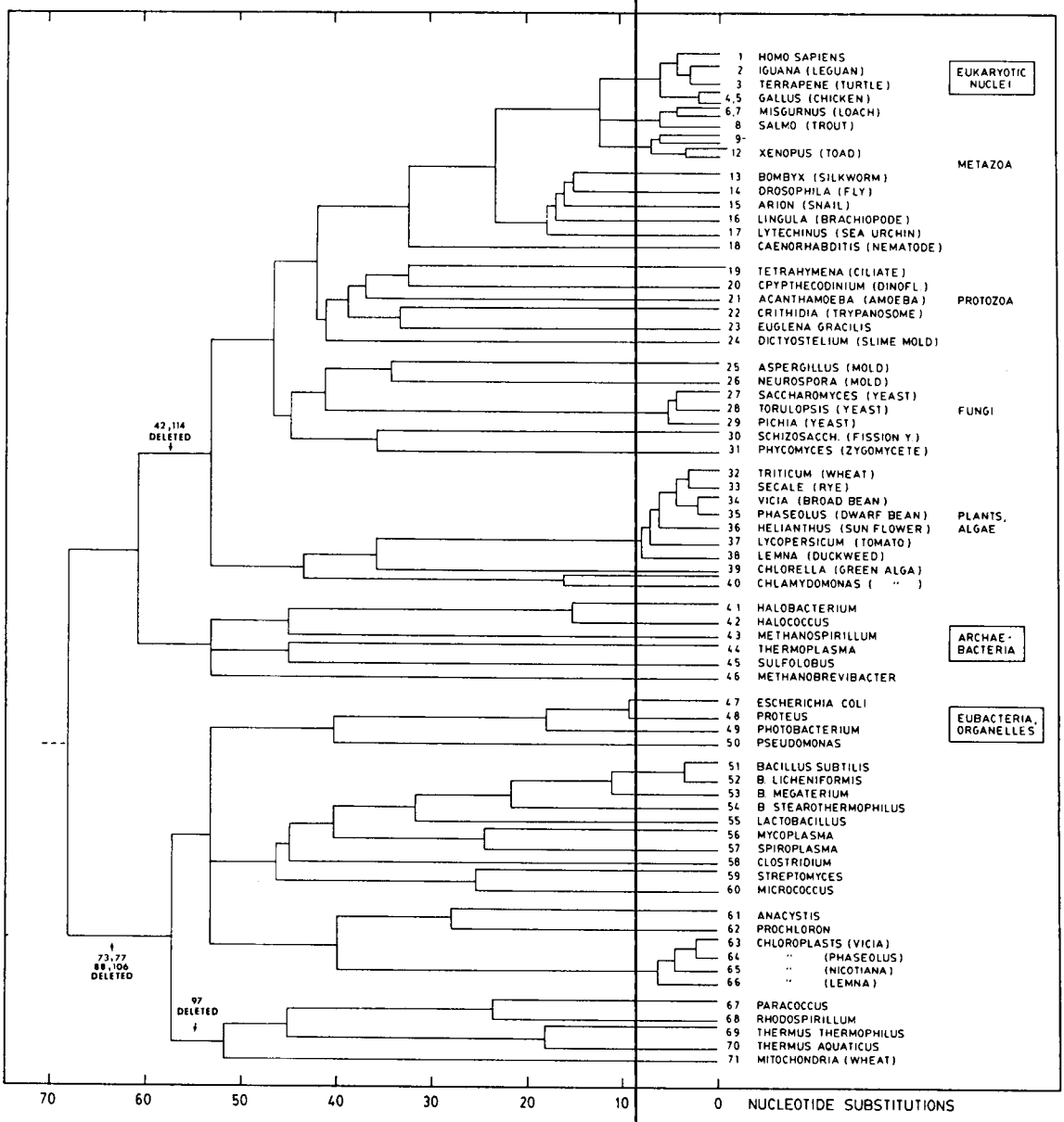

1 HOMO SAPIENS
2 IGUANA (LEGUAN)
3 TERRAPENE (TURTLE)
4,5 GALLUS (CHICKEN)
6,7 MISGURNUS (LOACH)
8 SALMO (TROUT)
9-12 XENOPUS (TOAD)
EUKARYOTIC NUCLEI
METAZOA
13 BOMBYX (SILKWORM)
14 DROSOPHILA (FLY)
15 ARION (SNAIL)
16 LINGULA (BRACHIOPODE)
17 LYTECHINUS (SEA URCHIN)
18 CAENORHABDITIS (NEMATODE)
19 TETRAHYMENA (CILIATE)
20 CPYPTHECODINIUM (DINOFL.)
21 ACANTHAMOEBA (AMOEBA)
22 CRITHIDIA (TRYPANOSOME)
23 EUGLENA GRACILIS
24 DICTYOSTELIUM (SLIME MOLD)
PROTOZOA
25 ASPERGILLUS (MOLD)
26 NEUROSPORA (MOLD)
27 SACCHAROMYCES (YEAST)
28 TORULOPSIS (YEAST)
29 PICHIA (YEAST)
30 SCHIZOSACCH. (FISSION Y.)
31 PHYCOMYCES (ZYGOMYCETE)
FUNGI
32 TRITICUM (WHEAT)
33 SECALE (RYE)
34 VICIA (BROAD BEAN)
35 PHASEOLUS (DWARF BEAN)
36 HELIANTHUS (SUN FLOWER)
37 LYCOPERSICUM (TOMATO)
38 LEMNA (DUCKWEED)
39 CHLORELLA (GREEN ALGA)
40 CHLAMYDOMONAS (")
PLANTS, ALGAE
41 HALOBACTERIUM
42 HALOCOCCUS
43 METHANOSPIRILLUM
44 THERMOPLASMA
45 SULFOLOBUS
46 METHANOBREVIBACTER
ARCHAE-BACTERIA
47 ESCHERICHIA COLI
48 PROTEUS
49 PHOTOBACTERIUM
50 PSEUDOMONAS
EUBACTERIA, ORGANELLES
51 BACILLUS SUBTILIS
52 B. LICHENIFORMIS
53 B. MEGATERIUM
54 B. STEAROTHERMOPHILUS
55 LACTOBACILLUS
56 MYCOPLASMA
57 SPIROPLASMA
58 CLOSTRIDIUM
59 STREPTOMYCES
60 MICROCOCCUS
61 ANACYSTIS
62 PROCHLORON
63 CHLOROPLASTS (VICIA)
64 " (PHASEOLUS)
65 " (NICOTIANA)
66 " (LEMNA)
67 PARACOCCUS
68 RHODOSPIRILLUM
69 THERMUS THERMOPHILUS
70 THERMUS AQUATICUS
71 MITOCHONDRIA (WHEAT)
42,114 DELETED
73,77 88,106 DELETED
97 DELETED
70 60 50 40 30 20 10 0 NUCLEOTIDE SUBSTITUTIONS

Figure 20A

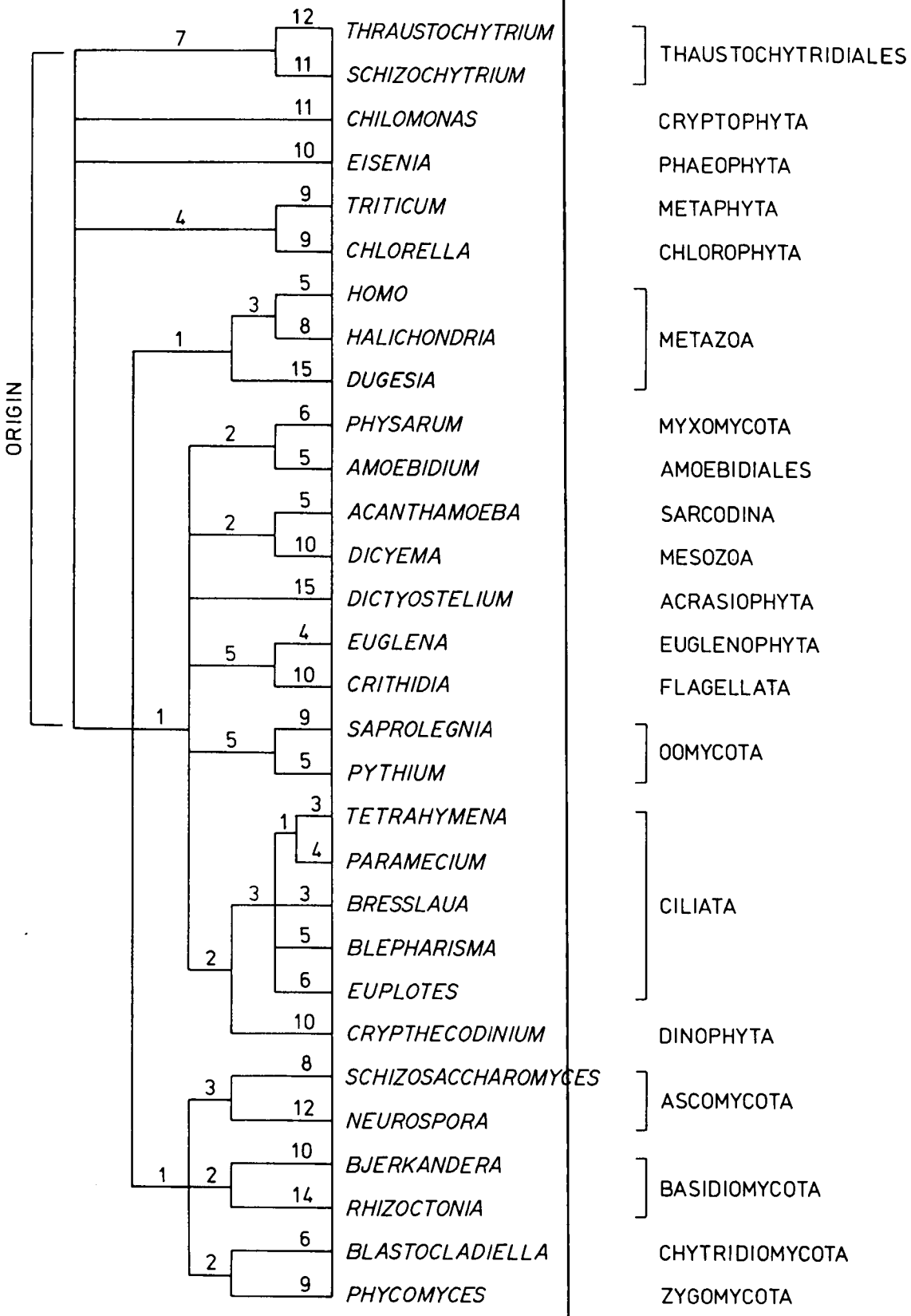

ORIGIN
12
7
11
THRAUSTOCHYTRIUM
SCHIZOCHYTRIUM
THAUSTOCHYTRIDIALES
11
CHILOMONAS
CRYPTOPHYTA
10
EISENIA
PHAEOPHYTA
9
4
TRITICUM
METAPHYTA
9
CHLORELLA
CHLOROPHYTA
5
3
HOMO
8
1
HALICHONDRIA
METAZOA
15
DUGESIA
6
2
PHYSARUM
MYXOMYCOTA
5
AMOEBIDIUM
AMOEBIDIALES
5
2
ACANTHAMOEBA
SARCODINA
10
DICYEMA
MESOZOA
15
DICTYOSTELIUM
ACRASIOPHYTA
4
5
EUGLENA
EUGLENOPHYTA
10
CRITHIDIA
FLAGELLATA
1
9
SAPROLEGNIA
5
OOMYCOTA
5
PYTHIUM
3
1
TETRAHYMENA
4
PARAMECIUM
3
3
BRESSLAUA
CILIATA
5
BLEPHARISMA
2
6
EUPLOTES
10
CRYPTHECODINIUM
DINOPHYTA
8
3
SCHIZOSACCHAROMYCES
12
NEUROSPORA
ASCOMYCOTA
10
1
2
BJERKANDERA
14
RHIZOCTONIA
BASIDIOMYCOTA
6
2
BLASTOCLADIELLA
CHYTRIDIOMYCOTA
9
PHYCOMYCES
ZYGOMYCOTA

Figure 20B

and by various physical techniques under varying conditions of temperature, salt, and Mg^{+2}. The intact RNA and various of its cleavage subfragments have been investigated by laser light scattering, optical absorbance, NMR, NOE (nuclear Overhauser measurement of the downfield portion of the NMR spectrum), and thermal denaturation. In addition to these physical and enzymatic techniques, the RNA structure has also been investigated by protein and oligonucleotide binding and diethyl pyrocarbonate reactivity. All of these lines of evidence taken together, plus the comparative sequence analysis, largely substantiate the minimal model (see 85 and 211 for additional information). The Euplotes macronuclear 5S rRNA sequence was observed to conform readily to the minimal model for eukaryotic 5S rRNA secondary structure represented by Erdmann and Wolters (86, 88) (Fig. 21).

Although essentially nothing factual is known about the tertiary structure of 5S rRNA, two models have been advanced anyway. In one, the arm containing helix III (or C) folds to make base pair contacts with the arm containing helicies IV and V (or D and E) (212). These regions include some conserved positions. In the other model, based on chemical crosslinking studies with the bifunctional phenyldiglyoxal reagent (213), a tertiary interaction between residues 37-40 and 73-76 has bee proposed. Since purified 5S rRNA exhibits no enzymatic or biological activity (except specific binding of certain proteins) and since the secondary structural conformation of the molecule appears quite dynamic and flexible, perhaps purified native 5S rRNA has no specific tertiary structure by itself. That is, the function of 5S rRNA

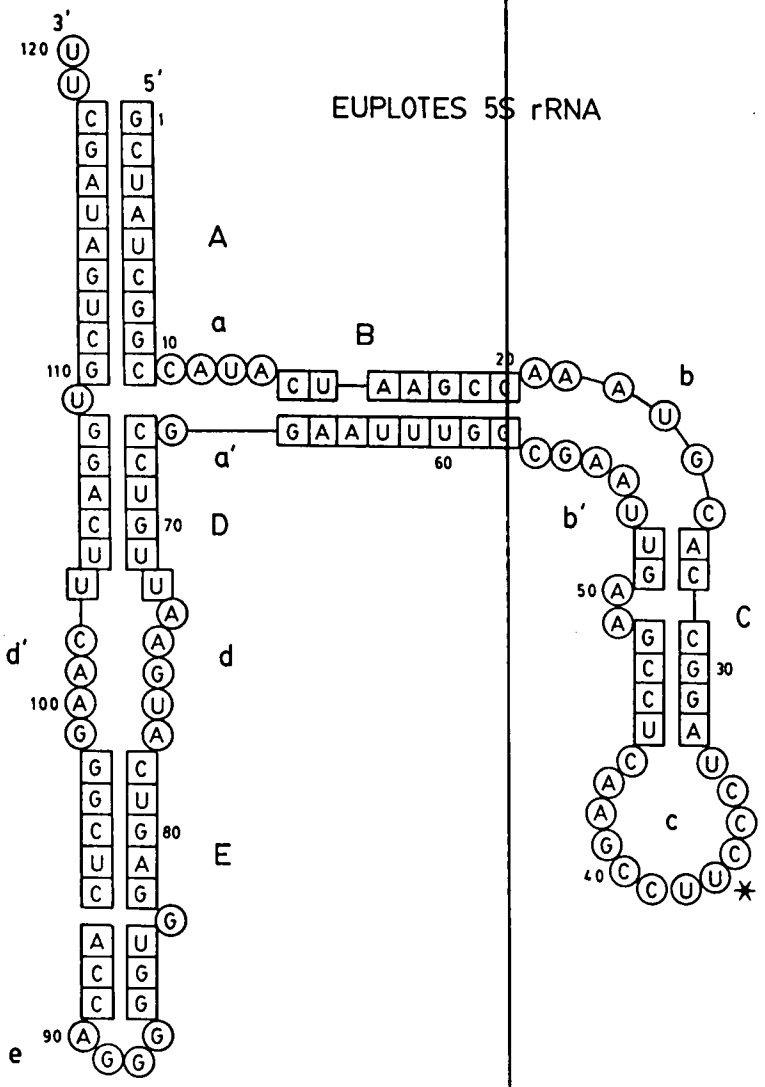


Figure 21: Secondary structure of Euplotes 5S rRNA.

The macronuclear 5S rRNA sequence of Euplotes eurystomus is fitted to the minimal model of eukaryotic 5S rRNA secondary structure by Erdmann and Wolters [from Erdmann, V. A., and Wolters, J. (1986) Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. Nucl. Acids Res. 14 (suppl.), r1-r59; Wolters, J., and Erdmann, V. A. (1986) Cladistic analysis of 5S rRNA and 16S rRNA secondary and primary structure--the evolution of eukaryotes and their relation to archaebacteria. J. Mol. Evol. 24, 152-166]. Squares indicate base pairing, circles indicate unpaired nucleotides. The star indicates that E. woodruffi has an A at this position. Letters A-E indicate structural domains of the molecule (as similarly denoted in Figs. 1, p. 8, and 19, p. 196).

is in ribosomes and as a complex with TFIIIA in the 7S storage particle and there its three dimensional structure may be dictated by interaction with other ribosomal components or TFIIIA, respectively. The tertiary structure of purified 5S rRNA may be biologically meaningless and irrelevant if it exists at all.

Prediction of Nucleosome Positions Along the Euplotes Macronuclear 5S rRNA Gene

The in vivo positions of nucleosomes in the native chromatin form of the linear Euplotes macronuclear 5S rRNA gene were predicted using a numerical algorithm shown to produce maps in agreement with experimental data (214). The method is based on two properties of DNA which has been observed to bind nucleosomes at nonrandom positions (it is believed that histone octamers actively select particular positions along a given DNA segment where their binding is most favorable.) First, a two-fold symmetry in the pattern of purines and pyrimidines is often observed about the nucleosomal dyad axis, so the program identifies 146 bp regions which display internal symmetry of dinucleotide pair types. Second, the DNA tends toward a specific 5'→3' arrangement of purines and pyrimidines, so the program locates DNA regions which display this sequence pattern. Two-fold symmetry of dinucleotide distribution is recognized by calculation of a dinucleotide symmetry coefficient. The specific 5'→3' pattern of dinucleotide pairs is recognized by calculation of a dinucleotide probability matrix correlation coefficient. These two indicies are simultaneously plotted as a function of position along a given DNA sequence and locations where their maxima coincide

indicate probable sites of nucleosome formation (such predictions have been confirmed by experiment). More specifically, the dinucleotide probability matrix locates general areas with the sequence properties appropriate for nucleosome formation while the dinucleotide symmetry constraints determine the precise nucleosome centers within those areas. That is, the matrix probe measures the local probability of nucleosome formation and the symmetry probe specifies the precise locations of nucleosomes. When this technique was applied to the Euplotes macronuclear 5S rRNA gene sequence a simple, unique, and unambiquous nonrandom distribution of nucleosomes was not predicted (Fig. 22). This can be interpreted to mean either that nucleosomes are randomly distributed along this sequence or else that, if nucleosomes are nonrandomly positioned, there is not a simple, unique set of nucleosome locations. Operationally this means that the results of the nucleosome mapping experiment are expected to be complicated if not ambiguous. Interestingly, the region of the macronuclear DNA molecule displaying the highest local probability density of nucleosome formation is centered around the 3' end of the coding region. If this prediction is accurate, this may represent a mechanistically relevant structural feature.

The Length of the Euplotes Macronuclear 5S rRNA Gene

Finally, why is the macronuclear DNA fragment harboring the 5S rRNA gene in Euplotes so long? In Oxytricha fallax the 5S gene is reported to reside on a macronuclear DNA of 690 bp (101) and in Stylonychia mytilus on a macronuclear DNA of 200-300 bp (142). The 930 bp

Figure 22: Predicted nucleosome positions along the Euplotes macronuclear 5S rRNA gene.

Positions of nucleosomes along the Euplotes macronuclear DNA molecule harboring the 5S rRNA gene were predicted using a numerical aligorithm as described in the text. The symmetry coefficient (top) and the dinucleotide probability matrix correlation coefficient (bottom) are plotted as a function of position along the DNA molecule.

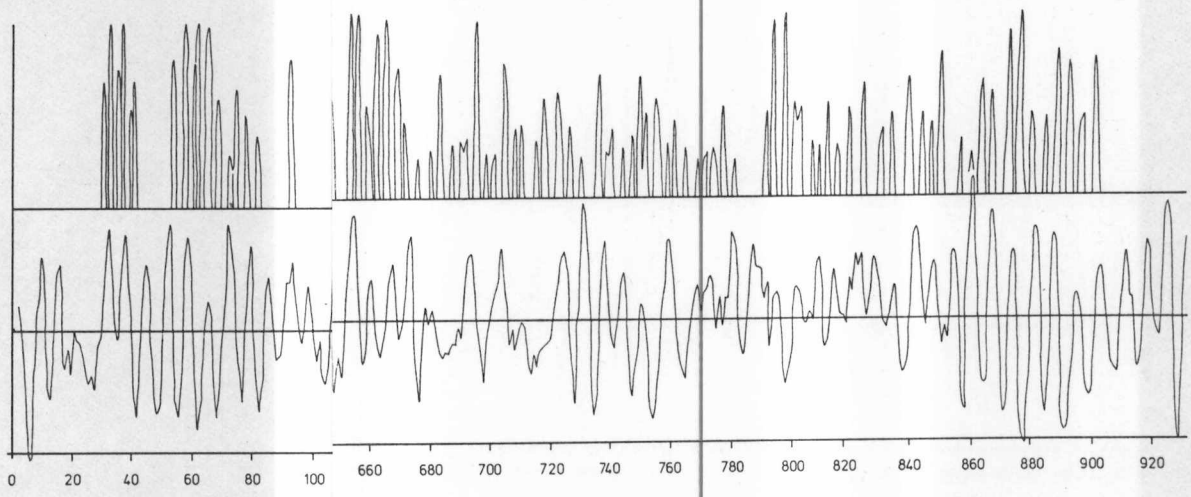

0
20
40
60
80
100
660
680
700
720
740
760
780
800
820
840
860
880
900
920

Figure 22

gene of Euplotes eurystomus is long enough to contain seven coding regions, although only one is found. Several explanations are possible. Perhaps the nontranscribed DNA serves to maintain or promote certain chromatin structural domains relevant to transcription. This seems unlikely since if such a large amount of nontranscribed DNA was necessary to establish the required chromatin structure(s) it would be difficult to explain the functionality of the smaller versions of the gene in Oxytricha and Stylonychia. Perhaps the DNA flanking the coding region contains a functional or regulatory sequence element(s) as yet unidentified. The most straightforward and reasonable explanation may be that the length of this mature macronuclear DNA molecule is simply a consequence of the pattern of fragmentation of the micronuclear genome during macronuclear development. If 5S rRNA genes are arranged in clusters of tandem repeats in the micronucleus, then the repeat length of these units is apparently fairly large. The repeat length of tandem 5S gene reiterations varies in Tetrahymena from 250-290 bp (167) and in Xenopus from 200-1400 bp (1). Perhaps the periodicity of the putative micronuclear 5S gene tandem repeats in Euplotes is on the order of 846 bp (that is, the length of the mature macronuclear fragment minus the telomeres). Since no IESs need to be removed from the 5S genes, maybe they are excised from the micronuclear genome simply by fragmentation of the putative tandem repeat cluster and modified during macronuclear development only by telomere addition. (This hypothesis is testable by Southern blot analysis of the restriction enzyme cleavage products of micronuclear genomic DNA.)

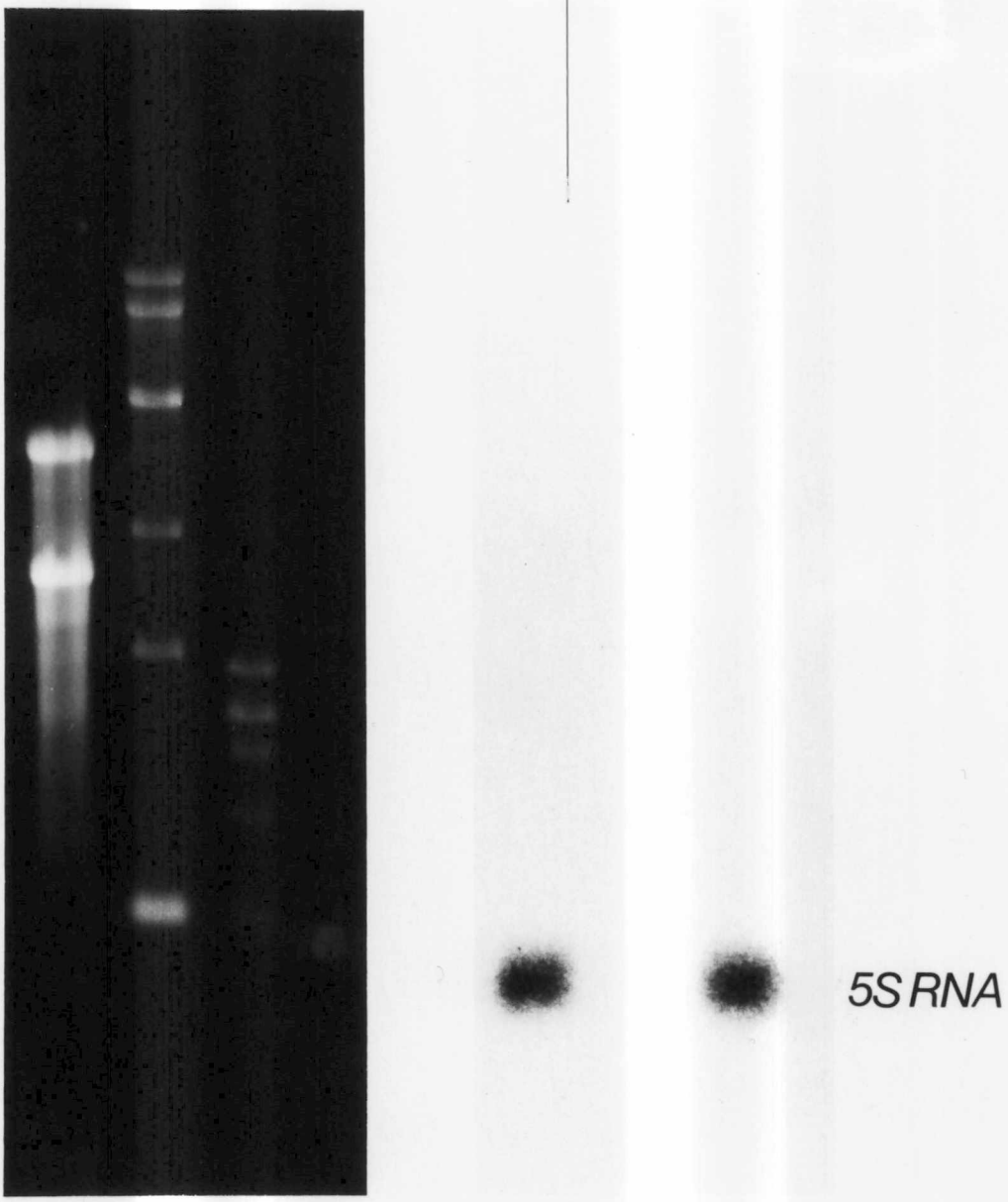
Northern Hybridization Analysis

Total cellular RNA was isolated from Euplotes eurystomus by the guanidinium thiocyanate/cesium chloride method, fractionated by agarose gel electrophoresis (in the presence of formaldehyde), and transferred to nitrocellulose for hybridization analysis using the gel-purified p320 insert as probe (Fig. 23). Denatured Euplotes total cellular RNA was observed to migrate as a continuum of bands ranging in size from ∿100 bases to ∿3700 bases superimposed on which were several species of enhanced abundance (most notably 19S rRNA at 2100 bases and 25S rRNA at 3700 bases). The p320 insert was observed to specifically hybridize to a single RNA species of ∿120 bases. This experiment was designed to demonstrate only a few specific points: (1) the p320 insert (the Euplotes macronuclear 5S rRNA gene) is transcribed in vivo; (2) only a single transcript is synthesized from this template; and (3) this transcript is ∿120 bases in length, the size of 5S rRNA. (Precise and reliable determination of the size of the p320-cognate RNA is not possible from this experiment as the hybridization signal falls outside the range of sizes covered by the RNA molecular weight standards. The exact size of the p320 transcription product is better deduced from the DNA sequence data and from the in vitro transcription experiment (see below) in which the p320 transcript was observed to comigrate with Xenopus 5S rRNA.) Two lanes of Euplotes total cellular RNA were included on the northern blot to demonstrate that the observed hybridization signal is not a random artifact.

Figure 23: Northern hybridization analysis.

Left: Acridine orange staining of an agarose gel loaded left to right as follows: Euplotes total cellular RNA, RNA molecular weight standards, ØX174 x Hae III DNA molecular weight standards.

Right: Autoradiogram of a northern blot probed with the p320 insert. The lanes here (which were adjacent to those shown in the left panel) were loaded left to right as follows: Euplotes total cellular RNA, RNA molecular weight standards, Euplotes total cellular RNA.


5S RNA

Relative Abundance of the 5S rRNA Gene in the Macronucleus of Euplotes Eurystomus

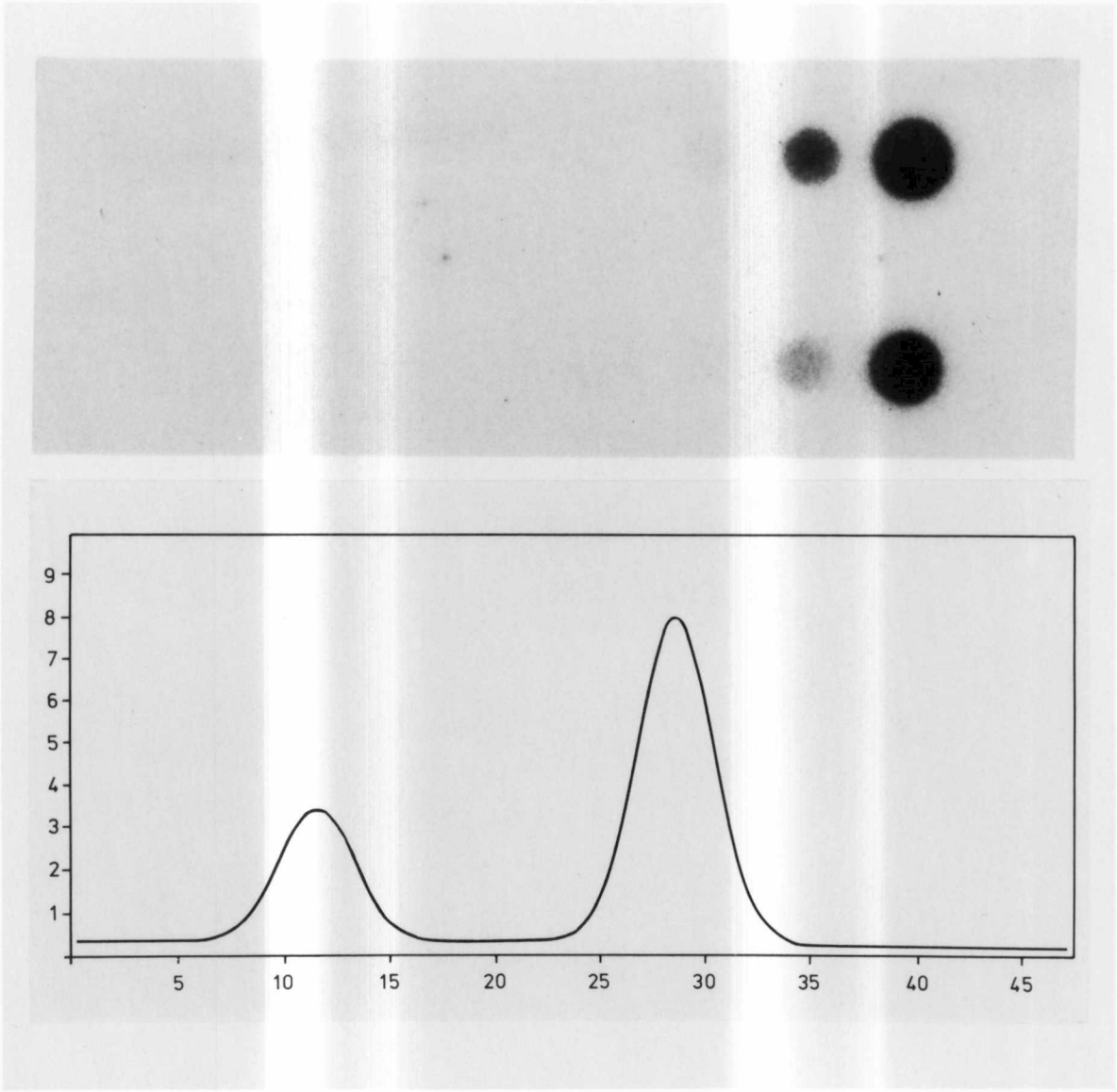
The relative abundance of the 5S rRNA gene in native Euplotes total macronuclear genomic DNA was investigated by hybridization analysis. Two dilution series of known amounts (spanning seven orders of magnitude) of macronuclear DNA and the gel-purified p320 insert were applied to nitrocellulose using a dot blot vacuum manifold. The dot blot was probed by hybridization with the gel-purified p320 insert under conditions where only identical matches would be detected. Thus in both series of dilutions the hybridization signal is generated by the p320 insert hybridizing to itself: in one case with its purified self (the cloned gene) and in the other case with itself as it resides in total macronuclear genomic DNA (the native gene). The validity of such a determination is based on the implicit assumption that the cloned gene is identical to the native gene. This is true except for the presence of the G•C tracts artificially added to the ends of the cloned gene during the cloning procedure. These sequences amount to (36/(930 + 36)) = 3.7% of the length of the cloned gene (one of the G•C tracts is 16 bp long and the other is 20 bp long). Thus an error of up to ~4% may be introduced by this assumption (although the true error which actually results is probably much less than this due to the stability of the 930 bp hybrid). After hybridization and washing, the bound probe was detected by autoradiography. Hybridization intensity was quantified by laser scanning densitometry of the autoradiogram (Fig. 24). This measurement technique introduces an additional error of about 5% due to

Figure 24: Relative abundance of the 5S rRNA gene in the macronucleus of Euplotes eurystomus.

The relative abundance of the 5S gene in the macronucleus was determined by dot blot hybridization analysis.

Top: Autoradiogram of a dot blot probed with the purified 5S gene insert. The upper row of dots is a dilution series of total macronuclear genomic DNA [(signals originating from 0.1 µg and 1.0 µg DNA are visible) (signals originating from 0.1 ng and 1.0 ng DNA are visible)] and the lower row is a dilution series of the purified 5S gene insert (signals originating from 0.1 ng and 1.0 ng DNA are visible).

Bottom: Laser scanning densitometry of a portion of the above autoradiogram to quantify hybridization intensities: optical density is plotted as a function of position. The area under the left peak represents the intensity of hybridization of the probe to 0.1 ng of the purified 5s gene insert and that under the right peak represents the intensity of hybridization to 0.1 µg of total macronuclear genomic DNA. Areas were calculated by numerical integration. (Full scale equals 1.0 OD.)


9
8
7
6
5
4
3
2
1
5
10
15
20
25
30
35
40
45

inaccuracy in the numerical method of integration used to calculate peak intensity. Thus embodied in the final result will be an inherent uncertainty on the order of 10% (ignoring additional contributions from pipetting errors and the like).

The hybridization signals with optical densities best suited for measurement (that is, dots which were dark enough to measure accurately but not so dark as to saturate the film; an OD of about 1) corresponded to those containing 0.1 μg of total macronuclear genomic DNA and 0.1 ng of the pure p320 insert, so the calculation was performed using these values. As the laser scanned across a dot the optical density of the signal was plotted as a function of position (refer to Fig. 24). The area under the resulting curve represents the intensity of the hybridization signal (and allows direct comparison of two signals) and this was calculated by numerical integration using a computer. These raw intensity values were adjusted to correct for equal amounts (masses) of DNA per dot and for equal numbers of DNA molecules per dot. The relative abundance of 5S rRNA genes in the macronucleus of Euplotes eurystomus was thus determined to be 1 per 200 DNA molecules. Since the macronucleus of this organism contains ∿23,500 different DNA molecules (refer to Table 2, p. 30) a gene of typical abundance is present at a frequency of 1/23,500 (by definition). The 5S rRNA gene is therefore enriched in the Euplotes macronucleus ((1/200)/(1/23,500)) = 118-fold relative to a gene of typical abundance. From these values the number of 5S rRNA gene copies per macronuclear genome can be calculated in two ways (refer to Table 2):

1. The total number of macronuclear DNA fragments times the relative abundance of the 5S gene = (192×10^6 MAC DNA molecules/MAC genome) (1 5S gene/200 MAC DNA molecules) = 960,000 5S gene copies per macronuclear genome.
2. The average copy number of a macronuclear gene times the fold enrichment of the 5S gene = (8170 copies/genome) (118 X) = 960,000 5S gene copies per macronuclear genome.

The ∿120-fold hyperamplification of 5S rRNA gene sequences observed in the Euplotes macronucleus correlates well with the hyperamplification of 19S + 25S rRNA gene sequences (rDNA) of ∿100-fold in the Oxytricha macronucleus (97, 109) and less well with that of ∿40-fold in the Stylonychia macronucleus (97). Presumably the 5S and the 19S + 25S rRNA gene sequences are coordinately amplified in a particular organism so as to better synchronize the rates of ribosomal component production. It would be interesting to determine the relative abundance of a macronuclear ribosomal protein gene.

Transcription of the Euplotes Macronuclear 5S rRNA Gene In Vitro

The functionality of the Euplotes macronuclear 5S rRNA gene as template for the specific and accurate in vitro transcription of 5S rRNA was tested in collaboration with Alan Wolffe in Don Brown's lab at the Carnegie Institute of Washington in Baltimore. A soluble cell-free extract of Xenopus laevis oocyte nuclei was prepared and supplemented with the appropriate buffer components including ATP, CTP, UTP, and

α^{32}P-GTP. Three forms of the Euplotes 5S gene were added as template: the intact covalently closed circular supercoiled recombinant plasmid p320, the gel-purified p320 insert, and native Euplotes total macronuclear genomic DNA. After 1 hour at 22°C the labelled nascent transcription products were purified from the reaction and analyzed by denaturing polyacrylamide gel electrophoresis followed by autoradiography. 5S rRNA synthesis was quantified by excising a gel fragment containing the 5S rRNA band and measuring Cerenkov radiation.

In the first experiment (Fig. 25) a negative control reaction was performed by combining all of the reaction components except exogenous DNA. No signal was observed. This demonstrates that the nuclear extract is free of Xenopus genomic DNA and that the reaction is completely dependent on addition of exogenous DNA. A positive control reaction was performed by addition of a plasmid containing the somatic 5S rRNA gene of Xenopus laevis, pX1s 11 (10). This demonstrates the fidelity, accuracy, and specificity of the transcription system and also provides a 5S rRNA size marker (120 nucleotides) for the gel. The experimental samples included the three structural forms of the Euplotes macronuclear 5S rRNA gene mentioned above; each template generated a specific RNA of 120 nucleotides. A signal of 400 cpm was generated by 1 µg of native Euplotes total macronuclear genomic DNA while 500 ng of intact p320 directed the incorporation of 1200 cpm into 5S RNA. Non-specific transcription products are also apparent on the gel.

In the second experiment (Fig. 26) the efficiency of 5S rRNA transcription from a linear DNA template was investigated. Under rate

Figure 25: In vitro transcription of the Euplotes macronuclear 5S rRNA gene by an extract of Xenopus oocyte nuclei.

Autoradiogram of a polyacrylamide gel containing in vitro transcription reaction products as follows: Lane 1, negative control reaction with no added DNA; Lane 2, positive control reaction with 200 ng pX1s 11 (a plasmid containing the Xenopus 5S gene); Lane 3, reaction with 1 μg native Euplotes total macronuclear genomic DNA; Lane 4, reaction with 500 ng p320; Lane 5, reaction with 100 ng p320 insert. Lanes 1-4 were exposed 30 minutes, lane 5 was exposed 6 hours.

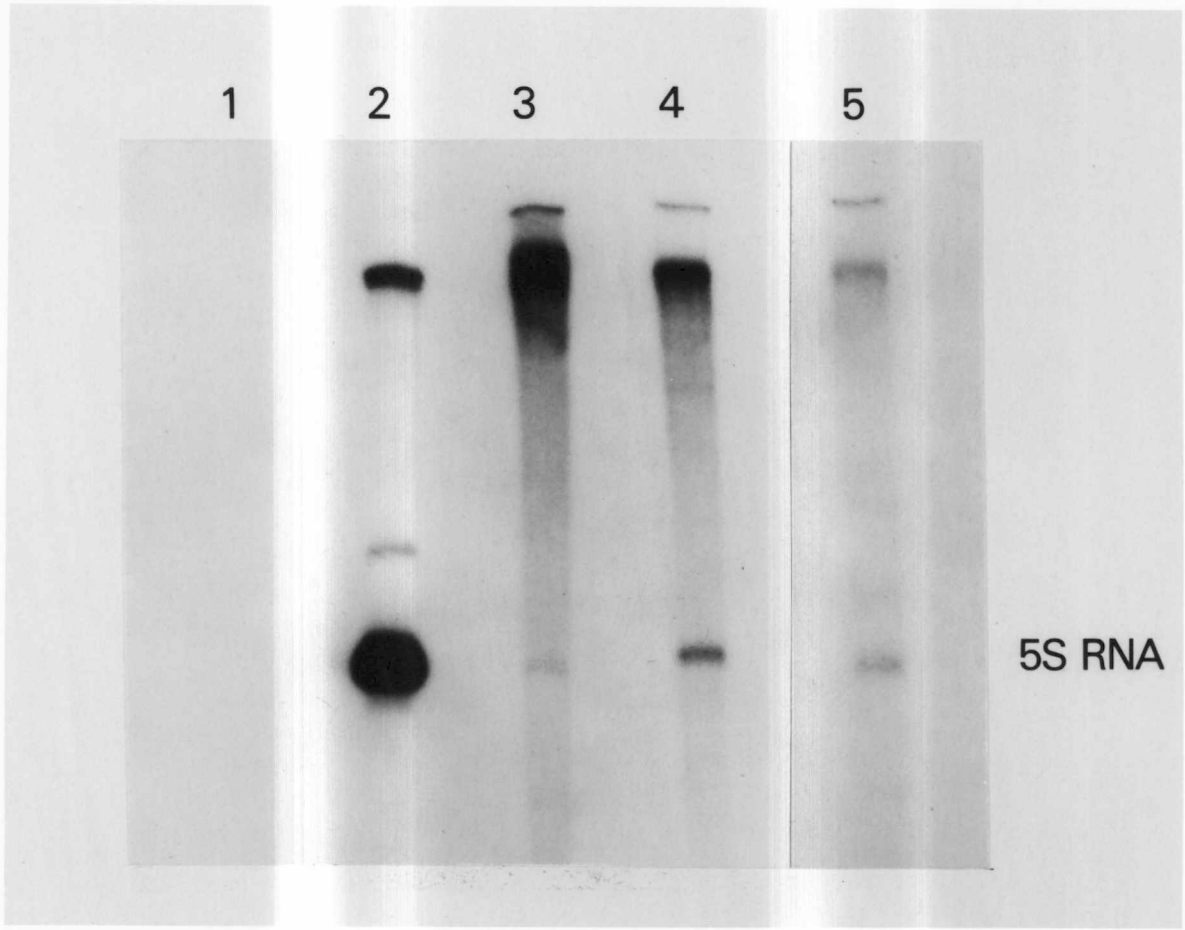
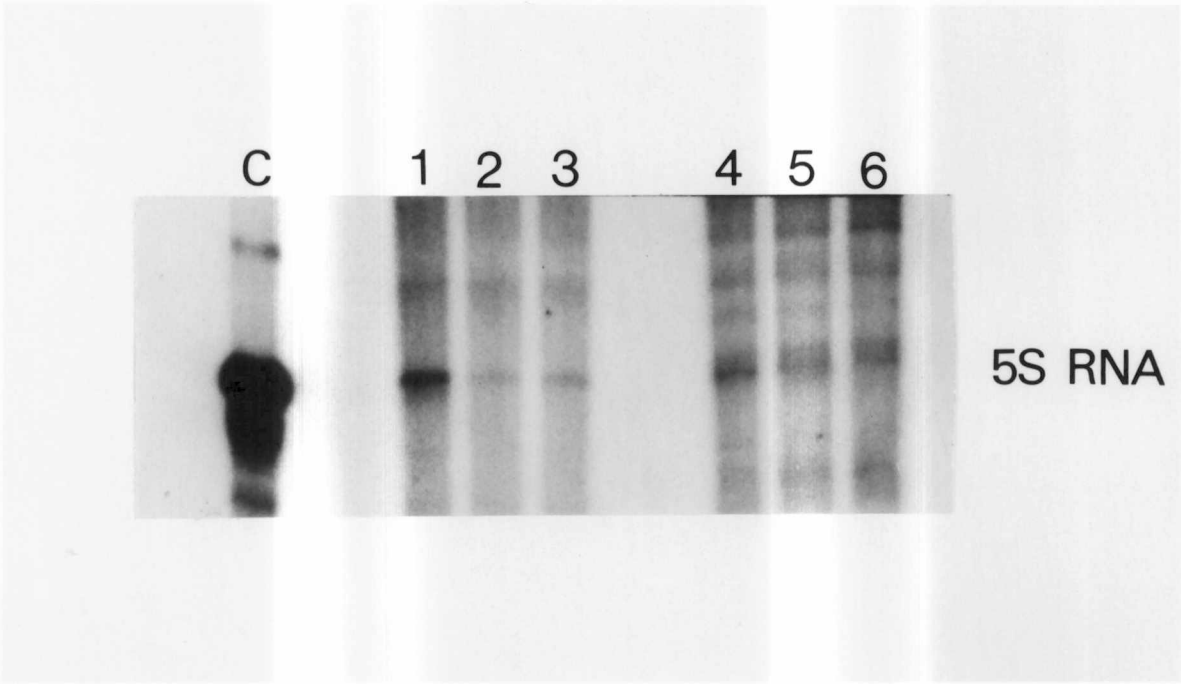

1
2
3
4
5
5S RNA

Figure 26: In vitro transcription of the linear Euplotes macronuclear 5S rRNA gene.

Autoradiogram of a polyacrylamide gel containing in vitro transcription reaction products as follows: Lane C, 200 ng pX1s 11 (positive control); Lane 1, 500 ng p320; Lane 2, 100 ng p320 + 400 ng pUC9; Lane 3, 20 ng p320 + 480 ng pUC9; Lane 4, 125 ng p320 insert + 375 ng pUC9; Lane 5, 25 ng p320 insert + 475 ng pUC9; Lane 6, 5 ng p320 insert + 495 ng pUC9. Note that lanes 1-3 contain a circular template while lanes 4-6 contain a linear template and that equal amounts of the coding sequence are included in lanes 1 & 4, 2 & 5, and 3 & 6; all reactions contain the same amount of total DNA.


C
1 2 3
4 5 6
5S RNA

enhanced conditions (discussed in the introduction) in vitro transcription efficiency may approach that observed in vivo. Similarly, under optimum conditions, the efficiency of transcription of linear DNA templates may approach that of circular templates. Nonspecific end effects may be reduced by decreasing the amount of specific linear template put into the reaction while keeping total DNA concentration constant by inclusion of nonspecific carrier (see below for elaboration). In this experiment decreasing amounts of circular p320 or the linear p320 insert were used while keeping the total DNA concentration constant by addition of carrier pUC9 (see legend Fig. 26). Again all template samples were observed to generate a labelled RNA of 120 bases and under optimum conditions the linear template functioned nearly as efficiently as the circular template.

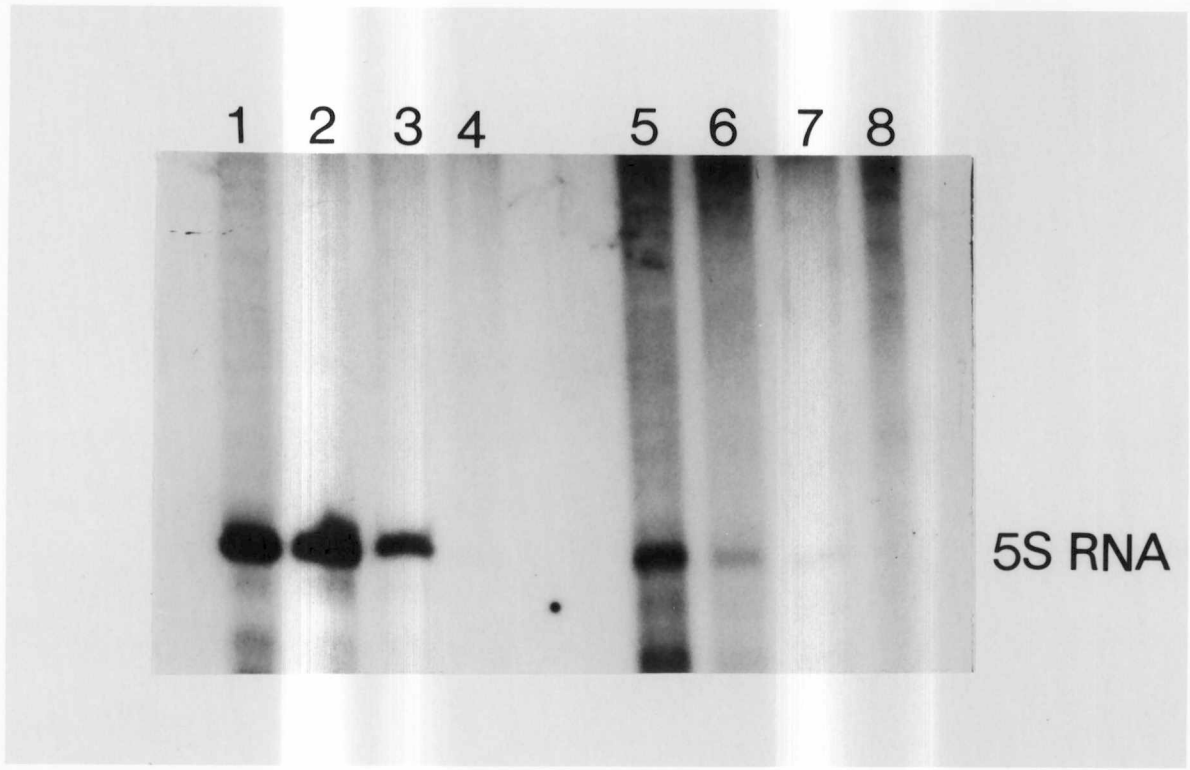
In the third experiment (Fig. 27) transcriptional activity of the Euplotes macronuclear 5S rRNA gene was assayed in the presence of increasing amounts of α-amanitin in order to identify which RNA polymerase is responsible for transcription of this gene. In parallel assays pX1s 11 and p320 were transcribed in reactions containing 0, 10, 50, and 100 μg/ml α-amanitin. Following electrophoresis and autoradiography, bands containing 5S rRNA were excised from the gel and the transcription was quantified by Cerenkov counting. Relative transcriptional activity (percent of control radioactivity) was plotted as a function of inhibitor concentration.

Figure 27: Sensitivity of Euplotes macronuclear 5S rRNA gene transcription to α-amanitin.

The identity of the RNA polymerase responsible for in vitro transcription of the Euplotes 5S gene was investigated by α-amanitin sensitivity.

27A: Autoradiogram of a polyacrylamide gel containing in vitro transcription reaction products as follows: Lanes 1-4, reactions with pX1s 11 in the presence of 0, 10, 50 and 100 μg/ml α-amanitin; Lanes 5-8, reactions with p320 in the presence of the same α-amanitin concentrations.

27B: 120 bp bands were excised from the above gel and radioquantitated. Relative transcriptional activity was plotted as a function of α-amanitin concentration.


1 2 3 4 5 6 7 8
5S RNA

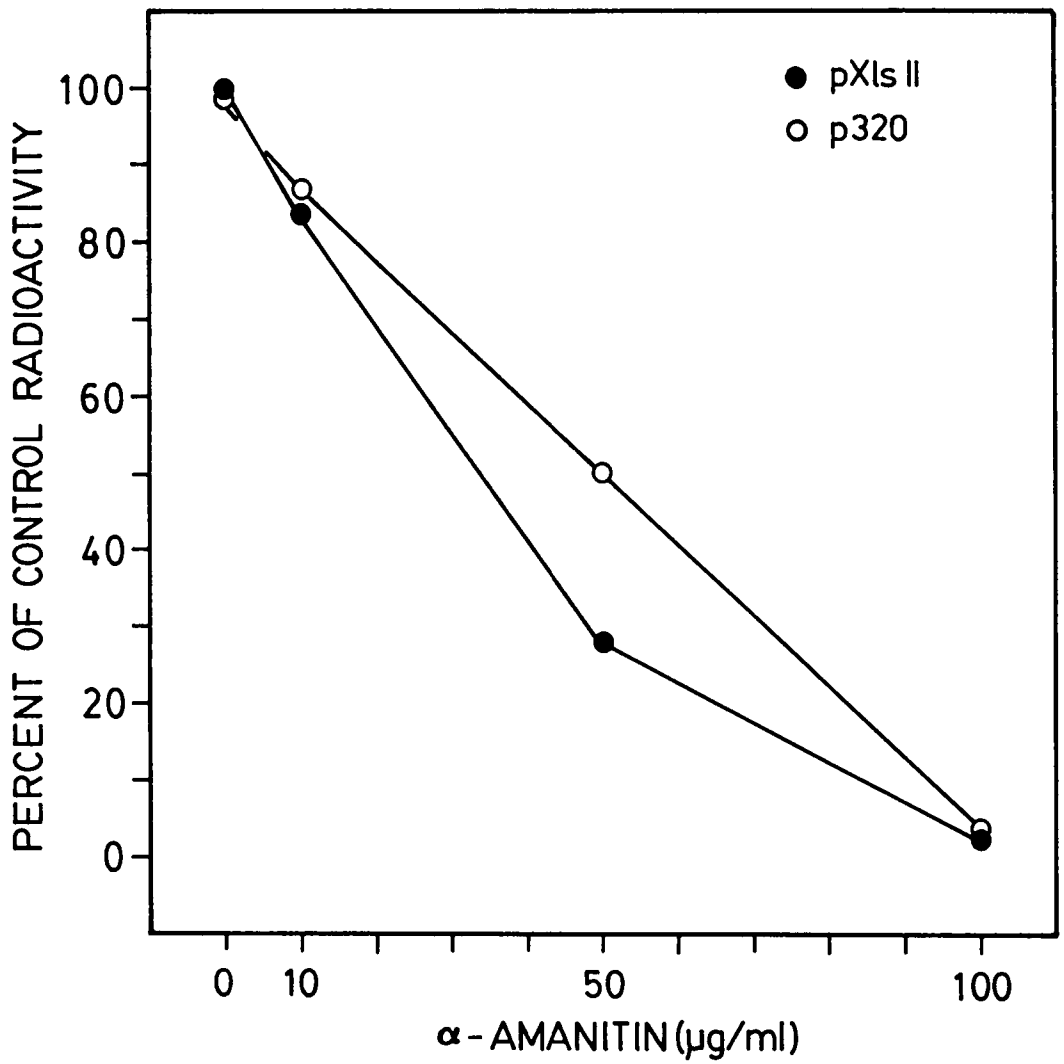

pXls II
p320
PERCENT OF CONTROL RADIOACTIVITY
100
80
60
40
20
0
0
10
50
100
α - AMANITIN (μg/ml)

Figure 27B

These experiments are interpreted to yield several biological conclusions:

1. The cloned Euplotes 5S rRNA gene while residing in circular plasmid p320 is competent in 5S rRNA synthesis in vitro. Thus the p320 clone represents a real, functional 5S gene (not a pseudogene) with the transcriptional coding and regulatory information intact (faithfully represented).
2. The linear p320 insert is competent in 5S rRNA synthesis in vitro.
3. Native (linear) Euplotes total macronuclear genomic DNA (containing the noncloned 5S gene) is competent in 5S rRNA synthesis in vitro.
4. The size of the transcript generated from each of these templates is the same as Xenopus 5S rRNA: 120 nucleotides.
5. Transcription from each of these templates is accurate and reasonably specific (although this conclusion is more about the extract than about the Euplotes templates per se).
6. Since the Euplotes 5S gene is transcribed by an extract of Xenopus oocyte nuclei the transcription mechanism and machinery of this gene have obviously been highly conserved over evolution. The ICR of the Euplotes gene as well as additional regulatory sequence elements are recognizable by the Xenopus transcription factors and RNA polymerase.
7. Although the Euplotes gene is active in the Xenopus system, it is transcribed less efficiently than the homologous template

(compare lanes 2 and 4 of Figure 25 which contain circular plasmids harboring the Xenopus and Euplotes 5S genes respectively). Transcription of the Euplotes 5S gene in this system is on the order of 5% as efficient as the Xenopus 5S gene.

8. Under optimum conditions the linear Euplotes 5S DNA template is transcribed as efficiently as the circular template (compare lanes 1 & 4, 2 & 5, and 3 & 6 in Figure 26: these pairs of reactions were assembled so as to contain equal amounts of the 5S coding sequence in either circular or linear form). (Also see comments below.)
9. Like the Xenopus gene, the Euplotes 5S gene is transcribed in vitro (and presumably in vivo) by RNA polymerase III.

This work is relevant to an ongoing discussion in the literature about transcription from linear versus circular templates. Abe Worcel contends that supercoiling and torsional stress (presumably only achievable on circular DNA templates or on linear template whose ends are anchored so as to prohibit revolution of one strand about the other) are essential for transcription (as discussed in the Introduction) while Don Brown disagrees. Worcel's model of gyrotory transcriptional activation is subject to the three criticisms levied in the Introduction plus the fact that no one can reproduce his results. [At the annual meeting of the American Society of Biological Chemists in June 1987, he explained that other labs are less skilled in the techniques of microanalysis than is his. We shall see.] The results presented here for the Euplotes system demonstrate with abundant clarity that not only are linear templates capable of transcription but that under appropriate conditions the

efficiency of transcription from linear templates approaches that of circular templates. While upon casual examination circular templates are in fact observed to be much better substrates for transcription, upon closer scrutiny the reasons for this are seen to be artifactual. If total DNA concentration is kept constant and the quantity of specific template DNA is reduced then the transcription rates of short, linear templates approach those of circles (Alan Wolffe, personal communication; reference 47; and Fig. 26 of this manuscript). The reason for the lower transcription rate of linear templates at high concentration is the presence of free DNA ends, especially single-stranded DNA at the ends, which sequestor both the transcription factors and the RNA polymerase (effectively removing them from solution). This may represent another function of the telomeric proteins of the hypotrich macronucleus: to block nonspecific end effects of the DNA (that is, nonspecific protein binding by the DNA termini). In vitro, end effects may be reduced by decreasing the amount of linear template DNA put into the reaction. In the interest of fairness (and accuracy), the discrepancy between Worcel's results and everyone else's is probably best explained by assuming that there must be some subtle but important difference in the way the experiments are executed in different labs. (A possible but unlikely explanation is that the Xenopus oocyte nuclear extract as prepared in Brown's lab contains proteins capable of anchoring DNA termini and thus allowing gyration of linear DNA structures, in which case Worcel's model may be correct.) That

Worcel has developed an in vitro system capable of inducing DNA gyration is clear, that this is related to transcriptional activation is not (refer to Introduction).

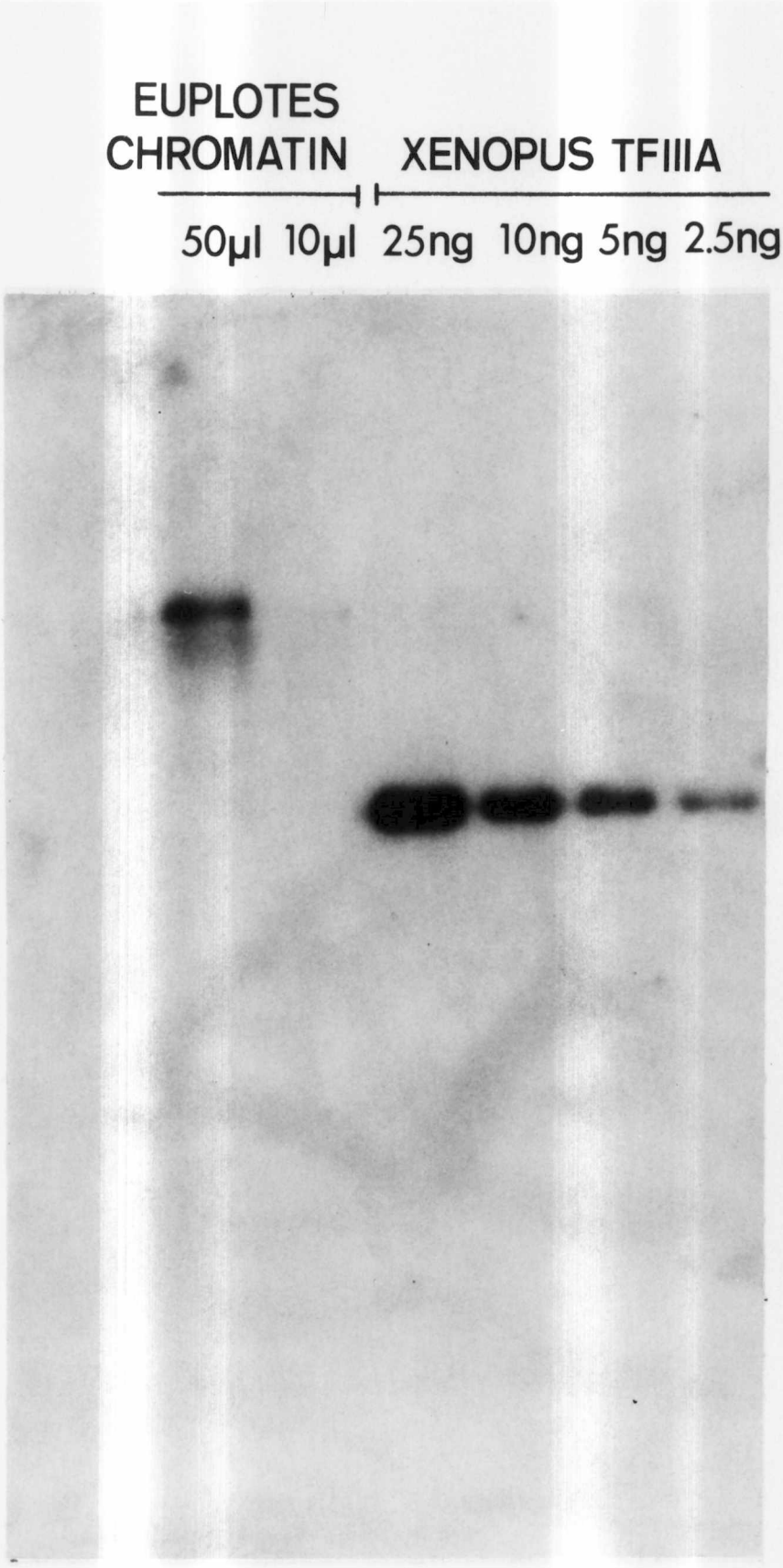
This work represents the first report of in vitro transcription of any hypotrich gene and one of the few reports about transcription in the hypotrichs in general.

Western Blot Analysis of Euplotes Macronuclear Chromosomal Proteins With Anti-TFIIIA Antibodies

Soluble Euplotes macronuclear chromatin was suspended in SDS gel sample loading buffer and the chromatin-associated proteins were resolved by SDS-polyacrylamide gel electrophoresis. The proteins were transferred from the gel to nitrocellulose by electroblotting and were probed with polyclonal anti-Xenopus TFIIIA antibodies in collaboration with Matt Andrews in Don Brown's lab at the Carnegie Institute of Washington in Baltimore (Fig. 28). While Xenopus TFIIIA has a molecular weight of 38-39 KDa, a Euplotes protein of 78-85 KDa was observed to crossreact. It must be emphasized that the detected Euplotes protein represents a species antigenically related to the Xenopus TFIIIA and is not necessarily the true Euplotes TFIIIA. This result must be interpreted with extreme caution for several reasons. First, a complete set of control experiments (which are absolutely required to unambiguously interpret immunodetection analyses) was not performed. Second, TFIIIA is believed to be fairly well conserved (the transcription machinery of the 5S gene in general is thought to be well conserved) so detection of a protein with molecular weight so deviant from that of the Xenopus

Figure 28: Western blot of Euplotes macronuclear chromosomal proteins probed with anti-TFIIIA.

Autoradiogram of a western blot of soluble Euplotes macronuclear chromatin-associated proteins (left) and purified Xenopus TFIIIA (right) probed with anti-Xenopus TFIIIA antibodies.


EUPLOTES
CHROMATIN
XENOPUS TFIIIA
50µl
10µl
25ng
10ng
5ng
2.5ng

TFIIIA was quite unexpected. Third, sequence analysis has demonstrated that the Euplotes 5S ICR conforms to the consensus structure, so a highly aberrant TFIIIA was unexpected (but not inconceivable). Fourth, even more compelling and disturbing is the fact that the Euplotes 5S gene is transcribed in the Xenopus system, which means that Xenopus TFIIIA faithfully recognizes and properly functions with the Euplotes 5S gene, implying that the Xenopus and Euplotes TFIIIAs are similar. Fifth, there is no corroborating evidence to support that this protein is in fact the Euplotes TFIIIA (such as demonstration of its association with 5S DNA or RNA). Therefore, although the anti-TFIIIA antibodies react specifically with a single band in Euplotes proteins, the reader is encouraged to consider this result with extreme skepticism.

Enrichment of Specific Genes as Native Chromatin Structures

A number of methods have been used to enrich specific gene sequences and actively transcribing or replicating gene sequences from populations of chromatin fragments which have been solubilized by mild nucleolysis. These methods include: HMG affinity chromatography; ectham-cellulose chromatography; immunoprecipitation with anti-HMG monoclonal antibodies; preparative agarose gel electrophoresis; sucrose gradient ultracentrifugation; digestion of nuclei followed by sedimentation to release solubilized chromatin; nuclease digestion and concomitant ultrafiltration of liberated chromatin fragments; biotin incorporation into nascent transcripts; and, of course, rRNA gene chromatin has been enriched by isolation of nucleoli. Perhaps the most

promising of these is the method of Herman (194-196) involving the incorporation of biotin-substituted ribonucleotides or deoxyribonucleotides into nascent transcripts of RNA or DNA. In this technique, biotin is attached to nucleotides via a linker containing a disulfide bond rendering it cleavable by reducing agents. The biotinylated nucleotide analog is presented to cells or in vitro nucleotide polymerizing systems and is incorporated into RNA or DNA. Afterwards, nucleic acid is purified and fractionated by streptavidin-biotin-cellulose (or avidin-biotin-cellulose) affinity column chromatography; any biotinylated RNA or DNA will adhere to the column while unmodified material freely passes through. The labelled sample is then recovered by washing the column with a reducing agent (dithiothreitol or β-mercaptoethanol) to sever the linker arm and release the nucleic acid, resulting in a dramatic enrichment of nascent polymers.

Most of the chromatin fractionation schemes mentioned above have been primarily applied toward the enrichment of actively transcribing chromatin. The results have generally been more confusing than enlightening. This is a consequence of the fact that different genes are regulated in different ways and thus a bulk population of molecules, however enriched in transcribing sequences, is still comprised of a myriad of heterogeneous chromatin forms. Thus little detailed mechanistic information can be learned from analysis of transcriptionally-active chromatin preparations containing a large number of different genes. One generalization which may be drawn from such studies is the observation that actively transcribing chromatin structures seem to be

depleted in histone H1. This is not surprising since this protein is causally related to chromatin condensation (196a), and retrieval of genetic information by the transcription apparatus requires a somewhat relaxed chromatin template (66, 67). In addition, active nucleosomes are found to contain a relatively high proportion of the high mobility group proteins (the HMGs). The meaning of this correlation is unknown. Also, several postsynthetic modifications of histones such as phosphorylation, acetylation, methylation, poly-ADP-ribosylation, and ubiquitination are believed to be involved in the modulation of chromatin architecture, and thus in the regulation of gene expression [this sentence adapted from Bryan and Destree (67)]. Finally, methylation of DNA appears to correlate with suppression of transcriptional activity.

Because of the complexity and diversity of chromatin forms, and because different genes are regulated in different ways, detailed molecular mechanistic information about the causal relationship between the modulation of chromatin structure and the regulation of gene expression is best attained by investigation of specific genes. The hypotrichous ciliated protozoan cell system is strategic for this purpose because the organism, as a normal part of its life cycle, has enriched the active sequences, solubilized them by precise and specific fragmentation into discrete functional genetic units, and compartmentalized them in an easily isolated organelle--the macronucleus. What remains is to fractionate this population of active chromatin molecules to enrich for specific gene sequences.

In collaboration with Carmen Cadilla, soluble Euplotes macronuclear chromatin was fractionated by isokinetic sucrose density gradient ultracentrifugation. Individual fractions collected from the gradient were analyzed by DNA and protein gel electrophoresis, Southern and dot blot hybridization, and electron microscopy. By this technique specific gene sequences were enriched as native chromatin structures. [Since this work has been previously presented, a rigorous treatment of the data does not appear here. Only the salient points will be mentioned. The reader may refer to 138, 140, and 197 for additional information.]

This experiment underscores the utility of the hypotrichous ciliated protozoan system in the investigation of chromatin structure (also refer to Introduction). Most chromatin must be treated with nucleases or mechanically sheared to be solubilized. (Solubility is prerequisite to most analytical techniques.) Thus in most cases the chromatin structure to be investigated must be perturbed (if not destroyed) to render it amenable to examination. Furthermore, not only does nuclease treatment alter the native chromatin structure but it also fails to respect the underlying gene structure (that is, genes may be cleaved within the coding region and so on).

Soluble Euplotes macronuclear chromatin was fractionated by ultracentrifugation according to sedimentation coefficient. It was empirically ascertained that the S value of a chromatin structure (a chromatin molecule) correlates linearly with both the DNA length of the fragment and the number of nucleosomes per fragment (data available upon request; see 197). (This is a bulk property and may not hold true for chromatin species displaying aberrant nucleosome packing.) Therefore genes

of different sizes may be readily resolved by sedimentation of their chromatin forms.

Native, soluble Euplotes total macronuclear chromatin was isolated and applied to a 10 ml 5-27% (w/v) isokinetic sucrose density gradient. Following ultracentrifugation at 28,000 rpm for 8 hours in an SW41 rotor at 4°C, 0.4 ml fractions of the gradient were collected and their A_{260} was determined. As expected based on the ethidium bromide staining pattern of DNA gels, bulk chromatin was found to have a roughly Gaussian distribution throughout the gradient (that is, more chromatin was at the middle than at the extremes because most DNA molecules are mid-sized; see below). Individual fractions from the gradient were analyzed by DNA and protein gel electrophoresis (Fig. 29). Sedimentation rate of chromatin as measured by isokinetic sucrose gradient centrifugation was observed to correlate with DNA size (as was electrophoretic mobility). Some proteins were found in all of the fractions (most notably the histones) while other proteins were enriched or depleted in certain fractions. Southern blot hybridization of the gradient fractions revealed that DNA molecules remained intact during the isolation and fractionation scheme (Fig. 30). The rRNA gene (7.5 kbp) was found to be enriched in fractions at the bottom of the gradient and its DNA was unit-length (the same length as in purified DNA) in all fractions which contained it. Electron microscopy demonstrated that the nucleosomal structure of chromatin molecules within individual fractions was maintained and that sedimentation of the chromatin was a function of DNA length and nucleosome number (Fig. 31). That is, fractions

Figure 29: Analysis of chromatin fractions by DNA and protein gel electrophoresis.

Top: Ethidium bromide staining of DNA in an agarose-sarcosyl gel loaded as follows: Lanes M, DNA molecular weight standards ØX and λ; Lane E, Euplotes total macronuclear DNA; Lanes 9-24, aliquots of gradient fractions 9-24.

Bottom: Staining of proteins in an SDS-polyacrylamide gel loaded as follows: Lane M, protein molecular weight standards; Lane E, Euplotes total macronuclear chromatin; Lanes 9-24, aliquots of gradient fractions 9-24.

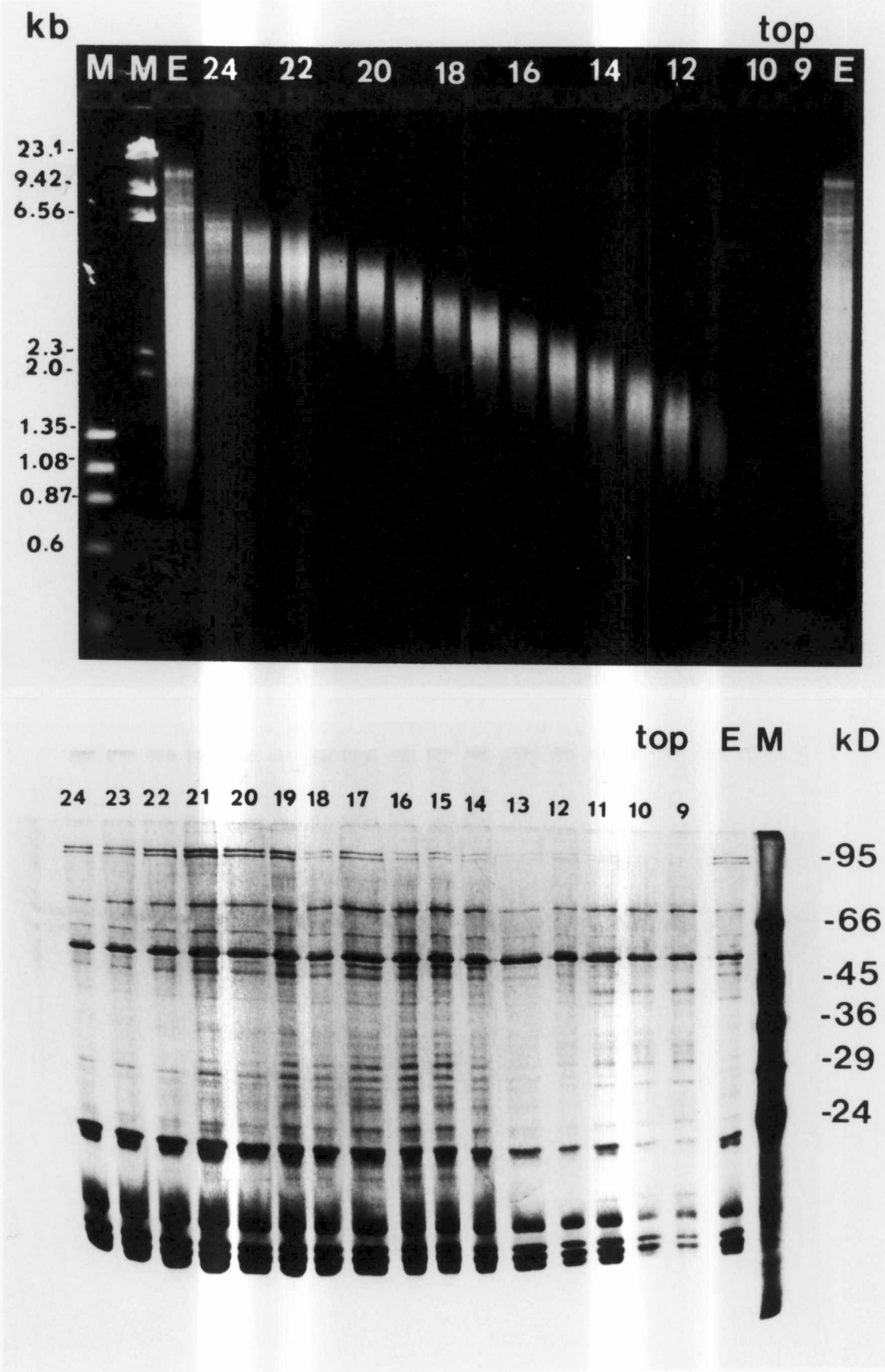
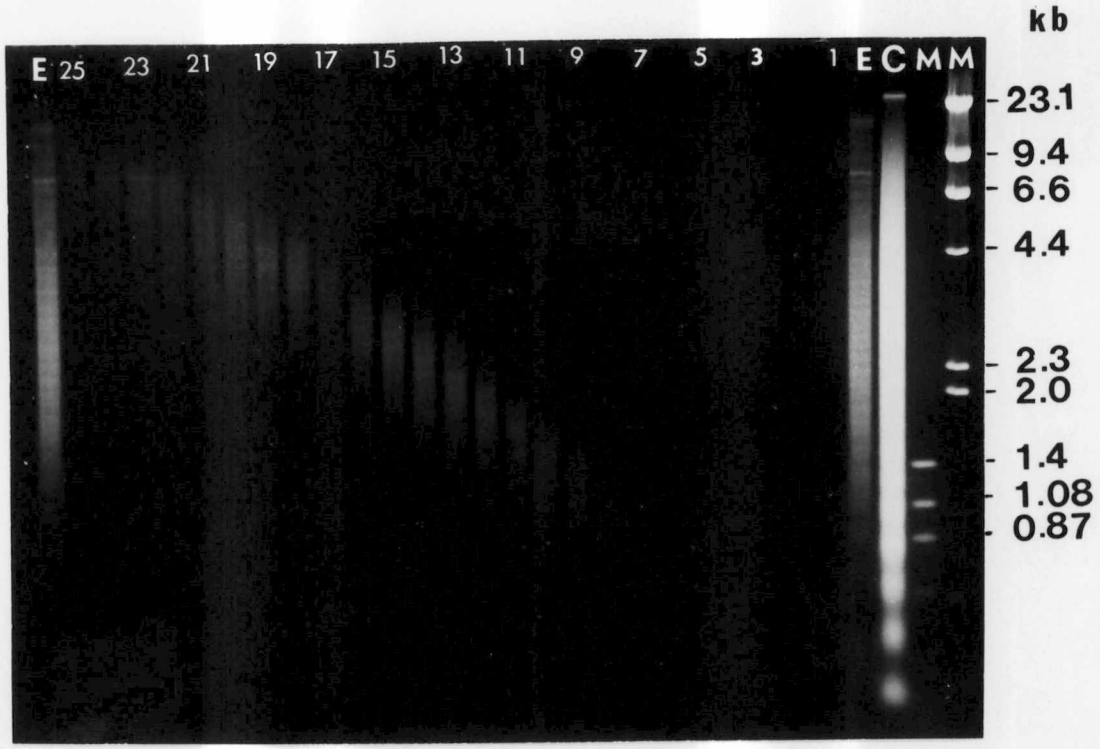

kb
top
M M E 24 22 20 18 16 14 12 10 9 E
23.1-
9.42-
6.56-
2.3-
2.0-
1.35-
1.08-
0.87-
0.6
top E M kD
24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9
-95
-66
-45
-36
-29
-24

Figure 30: Analysis of chromatin fractions by Southern blot hybridization.

Top: Ethidium bromide staining of DNA in an agarose gel loaded as follows: Lanes E, Euplotes total macronuclear DNA; Lane C, micrococcal nuclease digestion products of chicken erythrocyte nuclei; Lanes M, molecular weight standards ØX and λ; Lanes 1-25, aliquots of gradient fractions 1-25.

Bottom: Autoradiogram of a Southern blot of the above gel probed with the rRNA gene of Euplotes aediculatus.


kb
E 25 23 21 19 17 15 13 11 9 7 5 3 1 E C M M
23.1
9.4
6.6
4.4
2.3
2.0
1.4
1.08
0.87

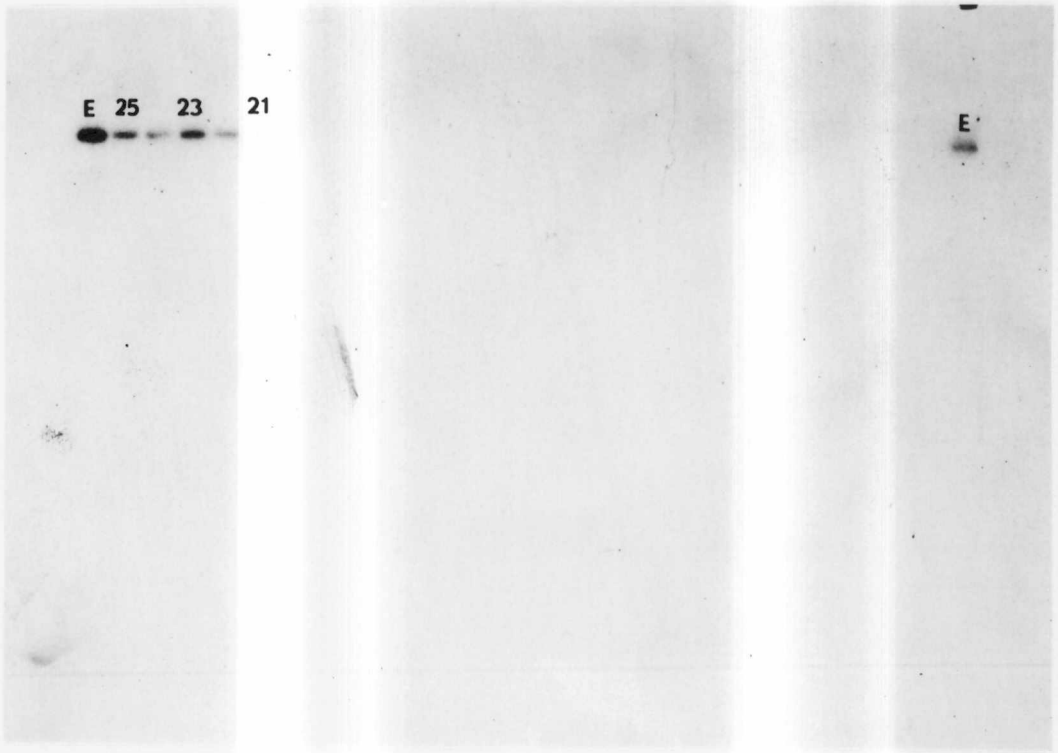
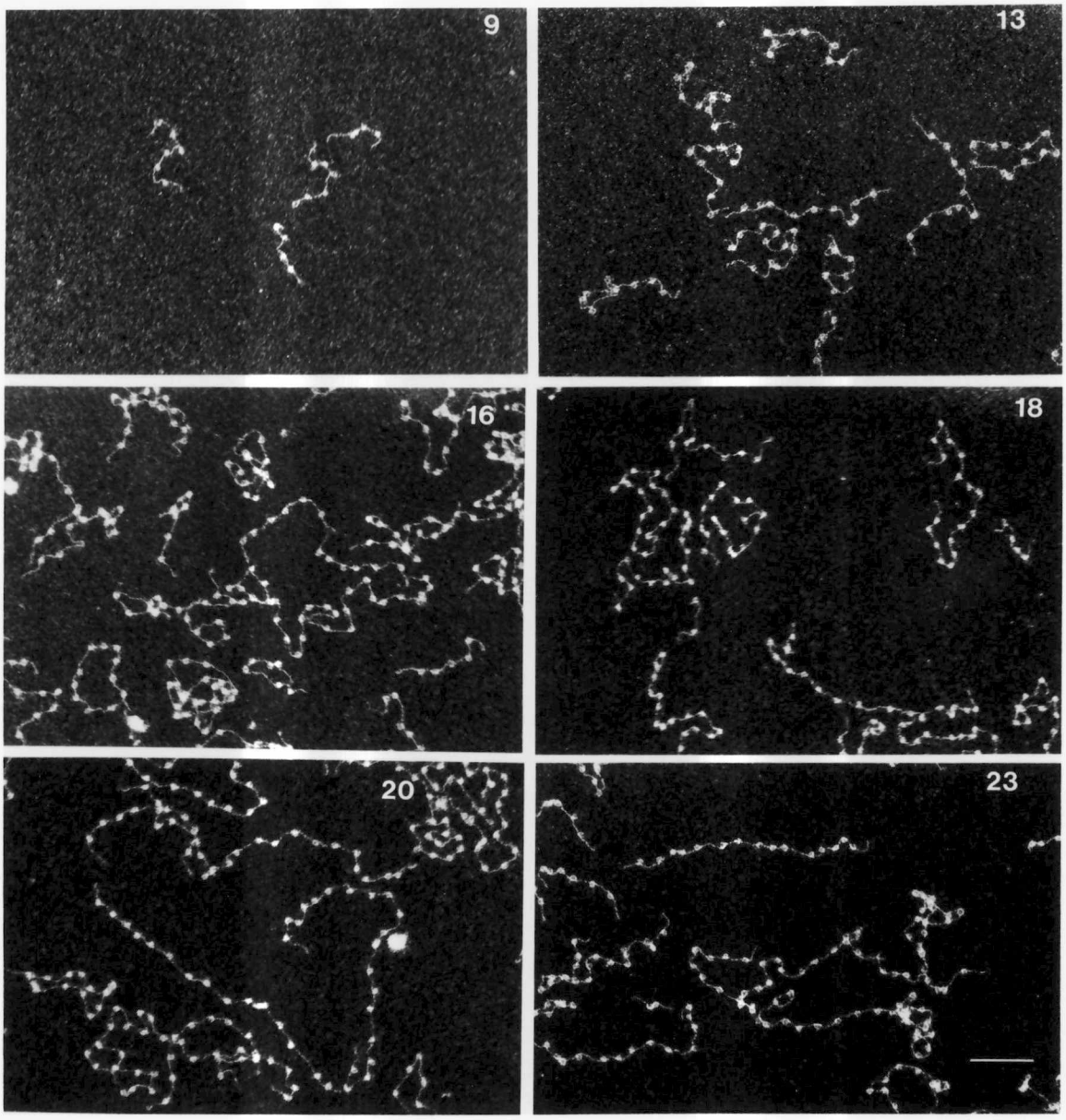

E 25 23 21
E'
7.5

Figure 31: Analysis of chromatin fractions by electron microscopy.

Darkfield electron micrographs of chromatin from various gradient fractions as indicated on each panel. (Bar = 0.1 μm.)


9
13
16
18
20
23

proceeding down the gradient were seen to contain longer DNA molecules and these were seen to contain more nucleosomes. Dot blot hybridization analysis of aliquots of the gradient fractions clearly showed that specific gene sequences were enriched in certain fractions and depleted in others (Fig. 32). Generally only a few fractions were observed to contain a particular sequence. Dot blot hybridization intensity was quantified by laser scanning densitometry of the autoradiograms as previously described. From these values enrichment of a specific sequence in a particular fraction was calculated using:

$$E_f = \frac{I_f}{\left(\sum_{f=1}^{N} I_f\right)/N} \qquad \text{(dimensionless)}$$

where E_f = enrichment of a sequence in fraction f (typically called "fold enrichment"),

I_f = normalized hybridization intensity of fraction f (determined by numerical integration),

N = number of fractions in the gradient, and

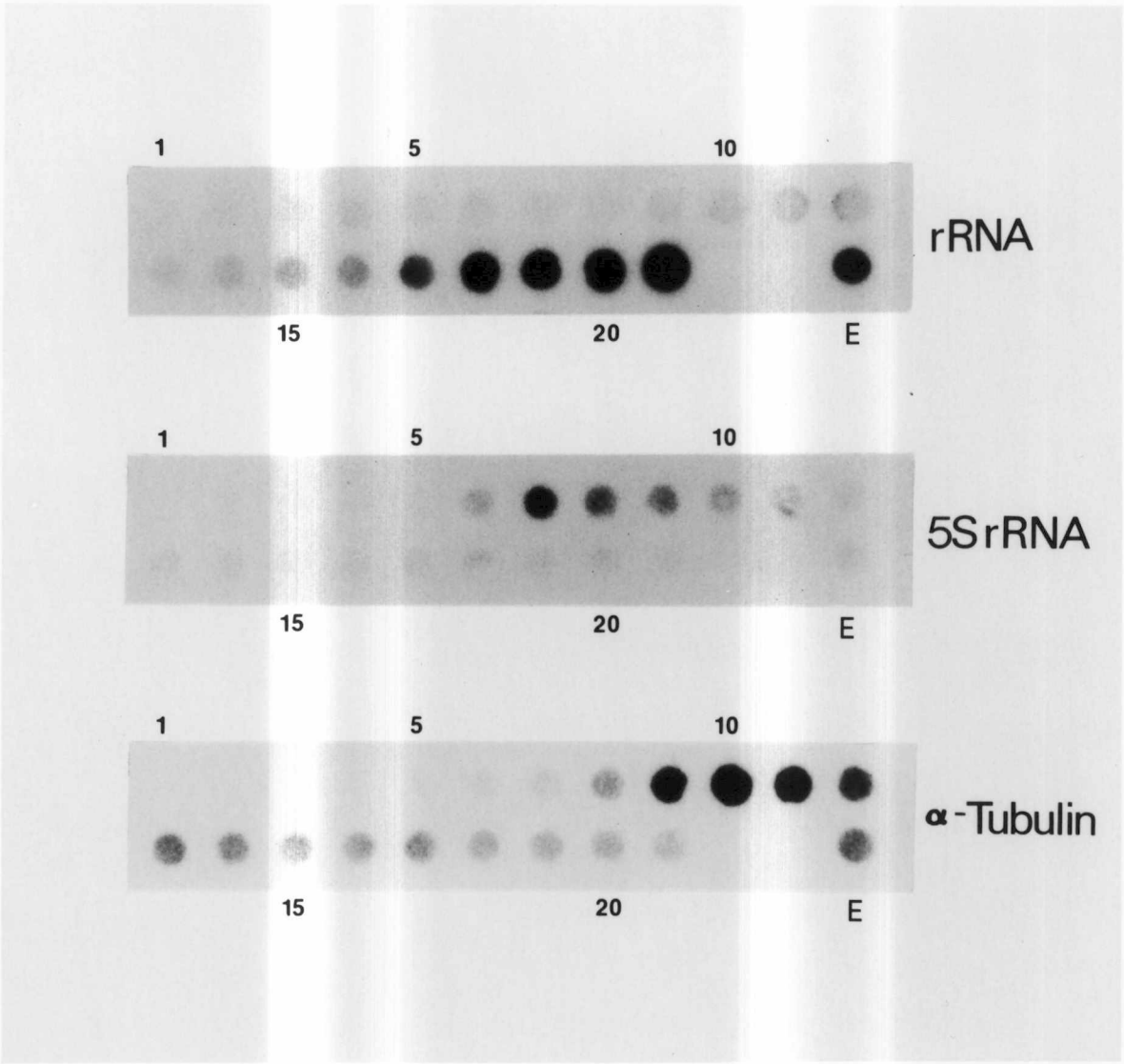
f = fraction number.

This simply says that the enrichment of a sequence in a particular fraction is the hybridization intensity of that fraction divided by the average hybridization intensity of all fractions across the gradient. (The application of this equation is a bit dangerous since the results will always be internally consistent even if they are wrong. A method including an external standard would be preferable.) Fold enrichment

Figure 32: Analysis of chromatin fractions by dot blot hybridization.

Autoradiograms of dot blots of gradient fractions as indicated by number; E indicates 1 μg Euplotes total macronuclear DNA. Specific DNA sequences were detected by hybridization with various gene probes as follows:

Top: The rRNA gene of Euplotes.
Middle: The 5S rRNA gene of Tetrahymena.
Bottom: The α-tubulin gene of Stylonychia.


1
5
10
rRNA
15
20
E
1
5
10
5S rRNA
15
20
E
1
5
10
α-Tubulin
15
20
E

is a measure of the increase in relative abundance of a species compared to its relative abundance in the starting material. Implicit to this calculation are the usual assumptions that hybridization signal intensity is proportional to relative abundance of the probe-cognate sequence and that autoradiographic signals fall within the linear range of film detection. The bulk distribution of chromatin (as measured by A_{260}) and the fold enrichment of specific gene sequences as native chromatin structures were plotted as a function of fraction number for a typical gradient (Fig. 33). A seven-fold enrichment of the Euplotes macronuclear 5S rRNA gene in its native chromatin form was obtained. Since the relative abundance of the 5S gene in the macronucleus is 1 in 200, in the peak fraction about 1 in 30 chromatin molecules (3.5%) harbors a 5S rRNA gene.

Figure 33: Enrichment of specific genes as native chromatin structures.

Enrichment of specific sequences was calculated from dot blot hybridization intensities of the gradient fractions. Fold enrichment (left ordinate) and A_{260} (right ordinate) are plotted as a function of fraction number (abscissa) as follows:

△ : A_{260} of each fraction.
□ : fold enrichment of 5S rRNA gene chromatin.
▽ : fold enrichment of α-tubulin gene chromatin.
◇ : fold enrichment of rRNA gene chromatin.

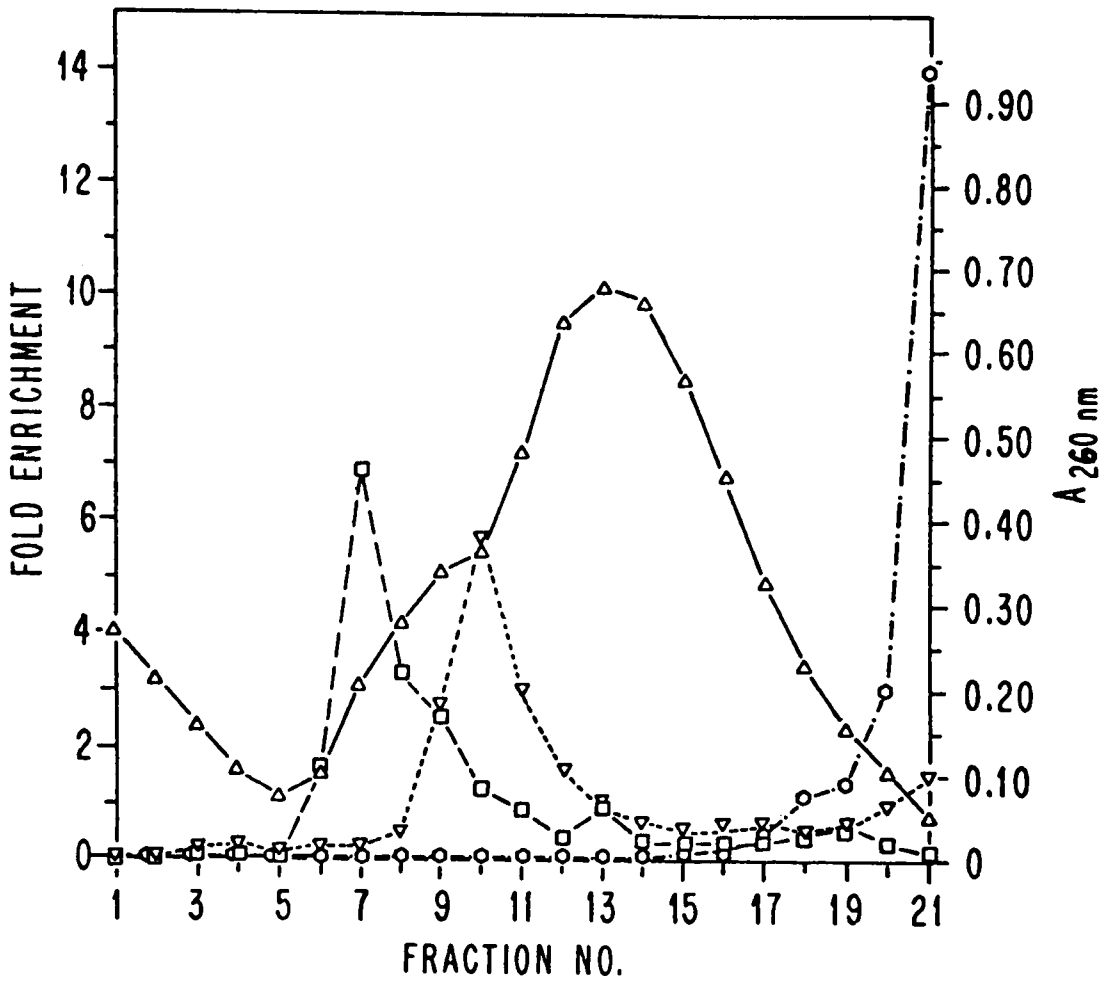

FOLD ENRICHMENT
A 260 nm
FRACTION NO.
0
2
4
6
8
10
12
14
0
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1
3
5
7
9
11
13
15
17
19
21

Figure 33

CHAPTER IV

SUMMARY AND PERSPECTIVE

For the convenience of the reader the major findings of this project are summarized:

1. The macronuclear 5S rRNA gene of Euplotes eurystomus resides on a molecule of 930 bp.
2. The macronuclear 19S + 25S rRNA gene of Euplotes eurystomus resides on a molecule of 7.5 kbp.
3. The macronuclear α-tubulin gene of Euplotes eurystomus resides on a molecule of 1.6 kbp.
4. The Euplotes macronucleus seems to contain only one version of 5S rRNA gene.
5. The Euplotes macronuclear 5S rRNA gene harbors a single coding region.
6. The Euplotes 5S rRNA coding region is 120 nucleotides in length, contains an internal promoter, and is not interrupted by introns.
7. The telomeric sequence of Euplotes eurystomus macronuclear DNA is identical to that of Euplotes aediculatus.
8. The Euplotes macronuclear 5S rRNA gene contains a sequence structurally similar to replication origins of other organisms.
9. The 3' end of the Euplotes 5S coding region contains the canonical RNA polymerase III termination signal.

10. The local G•C maximum of the macronuclear DNA molecule coincides with the coding region while a local G•C minimum coincides with the putative replication origin.
11. The 5S rRNA sequence of Euplotes eurystomus differs in only one position from that of Euplotes woodruffi. This sequence is ∿80% homologous to that of other protozoa such as Tetrahymena, Paramecium, and Euglena, and ∿65% homologous to the Xenopus sequence.
12. The Euplotes 5S rRNA sequence conforms readily to the minimal model of eukaryotic 5S rRNA secondary structure.
13. In the Euplotes macronucleus about 1 in 200 DNA molecules is a 5S rRNA gene. This represents a 120-fold amplification relative to a gene of typical abundance, or about 960,000 copies per macronucleus.
14. The Euplotes macronuclear 5S rRNA gene is a template for a single transcript of ∿120 nucleotides in vivo.
15. Both native linear and cloned linear or circular Euplotes macronuclear 5S rRNA genes are accurately transcribed in vitro by a Xenopus oocyte nuclear extract. The transcript is 120 nucleotides in length and is synthesized by RNA polymerase III.
16. Linear DNA structures may serve as efficient templates for transcription under appropriate conditions.
17. Anti-Xenopus TFIIIA antibodies recognize a Euplotes macronuclear chromatin-associated protein of ∿80 kDa.

18. Specific Euplotes macronuclear gene sequences can be enriched as native, intact chromatin structures.
19. The overall structure of the Euplotes macronuclear 5S rRNA gene supports the dogma of hypotrich macronuclear genome structure in that it contains a single genetic function, has no introns, is bounded by telomeres, and contains all of the sequence information necessary for regulation, transcription, and replication. (This point bears mentioning only because so few hypotrich genes have been characterized in this much detail.)

This work represents the first identification of specific genes in Euplotes eurystomus, the first detailed analysis of any hypotrich 5S rRNA gene, the first gene sequenced from Euplotes eurystomus, the first complete macronuclear gene sequence reported from all Euplotes species, the first in vitro transcription of any hypotrich gene (in collaboration with Alan Wolffe), the first enrichment of specific gene sequences as native, intact chromatin structures (in collaboration with Carmen Cadilla), and only the fourth complete macronuclear gene sequence reported from all hypotrich species. The results summarized above meet, and in fact exceed, the proposed research objectives of this project.

This report represents a relevant and significant contribution to biology in the following ways:

1. The dogma of hypotrichous ciliate macronuclear genome structure and development is supported by a sparcity of detailed

information on specific genes, as only three complete macronuclear gene sequences have been previously reported. The Euplotes macronuclear 5S rRNA gene sequence thus represents a significant contribution to the molecular biology of hypotrichous ciliated protozoa.

2. This work provides strong evidence relating to the ongoing controversy regarding the dependency of transcription on torsional stress. Here it is made clear that linear DNA structures can serve as efficient templates for transcription. Thus this work represents a significant contribution to the molecular biology of the 5S gene in general and transcription in particular.

3. The Euplotes eurystomus 5S rRNA sequence represents a small but real contribution to the field of molecular evolution and phylogenetics.

4. This work represents a contribution to the field of chromatin structure in that the detailed characterization of the Euplotes macronuclear 5S rRNA gene at the DNA level constitutes the foundation for further studies on the chromatin structure of this gene. In addition, powerful biological reagents (specific gene probes) are now available for the investigation of this gene as chromatin. Furthermore, various analytical and recombinant DNA techniques (such as hybridization, in vitro transcription, and chromatin fractionation) have been successfully applied to this system.

This report describes the DNA structure of the Euplotes macronuclear 5S rRNA gene in molecular detail as well as a scheme for its enrichment as a native chromatin structure so that the causal mechanisms relating chromatin structure and gene function may be better understood.

REFERENCES

REFERENCES

1. Brown, D. (1982) How a simple animal gene works. The Harvey Lectures 76, 27-44.

2. Wormington, W. M., and Brown, D. D. (1983) Onset of 5S RNA gene regulation during Xenopus embryogenesis. Dev. Biol. 99, 248-257.

3. Wormington, W. M., Schlissel, M., and Brown, D. D. (1983) Developmental regulation of the Xenopus 5S RNA genes. Cold Spring Harbor Symp. Quant. Biol. 47 879-884.

4. Brown, D. (1984) The role of stable complexes that repress and activate eucaryotic genes. Cell 37, 359-365.

5. Brown, D. D., and Sugimoto, K. (1973) 5S DNAs of Xenopus laevis and Xenopus mulleri: evolution of a gene family. J. Mol. Biol. 78, 397-415.

6. Ford, P. J., and Brown, D. D. (1976) Sequences of 5S ribosomal RNA from Xenopus mulleri and the evolution of 5S gene-coding sequences. Cell 8, 485-493.

7. Brown, D. D. (1981) Gene expression in eukaryotes. Science 211, 667-674.

8. Wegnez, M., Monier, R., and Denis, H. (1972) Sequence heterogeneity of 5S RNA in Xenopus laevis. FEBS Lett. 25, 13-20.

9. Ford, P. J., and Southern, E. M. (1973) Different sequences for 5S RNA in kidney cells and ovaries of Xenopus laevis. Nature New Biol. 241, 7-12.

10. Peterson, R. C., Doering, J. L., and Brown, D. D. (1980) Characterization of two Xenopus somatic 5S DNAs and one minor oocyte-specific 5S DNA. Cell 20, 131-141.

11. Roeder, R. G. (1976) Eukaryotic nuclear RNA polymerases. In Losick, R. and Chamberlin, M. (eds.), RNA Polymerase (p. 285). Cold Spring Harbor Lab.

12. Pelham, H. R. B., and Brown, D. D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Natl. Acad. Sci. USA 77, 4170-4174.

13. Birkenmeier, E. H., Brown, D. D., and Jordan, E. (1978) A nuclear extract of Xenopus laevis oocytes that accurately transcribes 5S RNA genes. Cell 15, 1077-1086.

14. Weil, P. A., Segall, J., Harris, B., Ng, S. Y., and Roeder, R. G. (1979) Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. 254, 6163-6173.

15. Shastry, B. S., Ng, S. Y., and Roeder, R. G. (1982) Multiple factors involved in the transcription of class III genes in Xenopus laevis. J. Biol. Chem. 257, 12979-12986.

16. Parker, C. S., and Roeder, R. G. (1977) Selective and accurate transcription of the Xenopus laevis 5S RNA genes in isolated chromatin by purified RNA polymerase III. Proc. Natl. Acad. Sci. USA 74, 44-48.

17. Segall, J., Matsui, T., and Roeder, R. G. (1980) Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem. 255, 11986-11991.

18. Emerson, B. M., and Roeder, R. G. (1984) DNA sequences and transcription factor interactions of active and inactive forms of mammalian 5S RNA genes. J. Biol. Chem. 258, 7926-7935.

19. Sakonju, S., Bogenhagen, D. F., and Brown, D. D. (1980) A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region. Cell 19, 13-25.

20. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3' border of the region. Cell 19, 27-35.

21. Bogenhagen, D. F. (1985) The intragenic control region of the Xenopus 5S RNA gene contains two factor A binding domains that must be aligned properly for efficient transcription initiation. J. Biol. Chem. 260, 6466-6471.

22. Fairall, L., Rhodes, D. ,and Klug, A. (1986) Mapping of the sites of protection on a 5S RNA gene by the Xenopus transcription factor IIIA--A model for the interaction. J. Mol. Biol. 192, 577-591.

23. Engelke, D. R., Ng, S. Y., Shastry, B. S., and Roeder, R. G. (1980) Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell 19, 717-728.

24. Sakonju, S., Brown, D. D., Engelke, D. R., Ng, S. Y., Shastry, B. S., and Roeder, R. G. (1981) The binding of a transcription factor to deletion mutants of a 5S ribosomal RNA gene. Cell 23, 665-669.

25. Bogenhagen, D. F., Wormington, W. M., and Brown, D. D. (1982) Stable transcription complexes of Xenopus 5S RNA genes: a means to maintain the differentiated state. Cell 28, 413-421.

26. Sakonju, S., and Brown, D. D. (1982) Contact points between a positive transcription factor and the Xenopus 5S RNA gene. Cell 31, 395-405.

27. Brown, D. D., and Schlissel, M. S. (1985) A positive transcription factor controls the differential expression of two 5S RNA genes. Cell 42, 759-767.

28. Ginsberg, A. M., King, B. O., and Roeder, R. G. (1984) Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell 39, 479-489.

29. Smith, D. R., Jackson, I., and Brown, D. D. (1984) Domains of the positive transcription factor specific for the Xenopus 5S RNA gene. Cell 37, 645-652.

30. Bieker, J., and Roeder, R. G. (1984) Physical properties and DNA-binding stoichiometry of a 5S gene-specific Transcription Factor. J. Biol. Chem. 259, 6158-6164.

31. Böhm, S., and Drescher, B. (1985) Multiple internal repeats within the structure of the 5S RNA/DNA binding transcription factor TF-IIIA from Xenopus laevis. Studia Biophysica 107, 237-247.

32. Tso, J. Y., Van Den Berg, D. J., and Korn, L. J. (1986) Structure of the gene for Xenopus transcription factor TFIIIA. Nucleic Acids Res. 14, 2187-2200.

33. Vincent, A. (1986) TFIIIA and homologous genes. The "finger" proteins. Nucleic Acids Res. 14, 4385-4391.

34. McConkey, G. A., and Bogenhagen, D. F. (1987) Transition mutations within the Xenopus borealis somatic 5S RNA gene can have independent effects on transcription and TFIIIA binding. Mol. and Cell. Biol. 7, 486-494.

35. Lasser, A. B., Martin, P. L., and Roeder, R. G. (1983) Transcription of class III genes: formation of preinitiation complexes. Science 222, 740-743.

36. Mary, J. L. (1984) The riddle of development. Science 226, 1406-1408.

37. Gottesfeld, J., and Bloomer, L. S. (1982) Assembly of transcriptionally active 5S RNA gene chromatin in vitro. Cell 28, 781-791.

38. Schlissel, M. S., and Brown, D. D. (1984) The transcriptional regulation of Xenopus 5S RNA genes in chromatin: the roles of active stable transcription complexes and histone H1. Cell 37, 903-913.

39. Rhodes, D. (1985) Structural analysis of a triple complex between the histone octamer, a Xenopus gene for 5S RNA and transcription factor IIIA. EMBO Journal 4, 3473-3482.

40. Cozzarell, N. R., Gerrard, S. P.,Schlissel, M., Brown, D. D., and Bogenhagen, D. F. (1983) Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell 34, 829-835.

41. Lewis, M. K., and Burgess, R. R. (1980) Transcription of simian virus 40 DNA by wheat germ RNA polymerase II. J. Biol. Chem. 255, 4928-4936.

42. Rosenberg, M., and Court, D. A. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13, 319-353.

43. Lin, S. Y., and Riggs, A. D. (1975) The general affinity of lac repressor for E. Coli DNA: implications for gene regulation in procaryotes and eucaryotes. Cell 4, 107-111.

44. Setzer, D. R., and Brown, D. D. (1985) Formation and stability of the 5S RNA transcription complex. J. Biol. Chem. 260, 2483-2492.

45. Segall, J. (1986) Assembly of a yeast 5S RNA gene transcription complex. J. Biol. Chem. 261, 11578-11584.

46. Bieker, J. J., Martin, P. L., and Roeder, R. G. (1985) Formation of a rate-limiting intermediate in 5S RNA gene transcription. Cell 40, 119-127.

47. Wolffe, A. P., Jordan, E., and Brown, D. D. (1986) A bacteriophage DNA polymerase transcribes through a Xenopus 5S RNA gene transcription complex without disrupting it. Cell 44, 381-389.

48. Wolffe, A. P., and Brown, D. D. (1986) DNA replication in vitro erases a Xenopus 5S RNA gene transcription complex. Cell 47, 217-227.

49. Ryoji, M., and Worcel, A. (1984) Chromatin assembly in Xenopus oocytes: in vivo studies. Cell 37, 21-32.

50. Glikin, G. C., Ruberti, I., and Worcel, A. (1984) Chromatin assembly in Xenopus oocytes: in vitro studies. Cell 37, 33-41.

51. Kmiec, E. B., and Worcel, A. (1985) The positive transcription factor of the 5S RNA gene induces a 5S DNA-specific gyration in Xenopus oocyte extracts. Cell 41, 945-953.

52. Kmiec, E. B., Ryoji, M., and Worcel, A. (1986) Gyration is required for 5S RNA transcription from a chromatin template. Proc. Natl. Acad. Sci. USA 83, 1305-1309.

53. Kmiec, E. B., Razvi, F., and Worcel, A. (1986) The role of DNA-mediated transfer of TFIIIA in the concerted gyration and differential activation of the Xenopus 5S RNA genes. Cell 45, 209-218.

54. Kmiec, E. B., and Worcel, A. (1986) The positive transcription factor of the 5S RNA gene proteolyses during direct exchange between 5S RNA sites. J. Cell. Biol. 103, 673-681.

55. Liu, L. F. (1984) DNA topoisomerases-enzymes that catalyse the breaking and rejoining of DNA. CRC Critical Reviews in Biochemistry 15, 1-24.

56. Sander, M., and Hsieh, T. (1985) Drosophila topoisomerase II double-strand DNA cleavage: analysis of DNA sequence hemology at the cleavage site. Nucl. Acids Res. 13, 1057-1072.

57. Gottesfeld, J. M. (1986) Novobiocin inhibits RNA polymerase III transcription in vitro by a mechanism distinct from DNA topoisomerase II. Nucleic Acids Res. 14, 2075-2088.

58. Roberson, A. E., Wolffe, A., and Olins, D. E. (1987, June) Analysis of a 5S rRNA gene cloned from Euplotes eurystomus. Abstract for the annual meeting of the American Society of Biological Chemists.

59. Olins, A. L., and Olins, D. E. (1973) Spheroid chromatin units (ν bodies). J. Cell. Biol. 59, 252a.

60. Woodcock, C. L. F. (1973) Ultrastructure of inactive chromatin. J. Cell. Biol. 59, 368a.

61. Olins, A. L., and Olins, D. E. (1974) Spheroid chromatin units (ν bodies). Science 183, 330-332.

62. Olins, D. E., and Olins, A. L. (1978) Nucleosomes: the structural quantum in chromosomes. American Scientist 66, 704-711.

63. Lilley, D. M. J., and Pardon, J. F. (1979) Structure and function of chromatin. Ann. Rev. Genet. 13, 197-233.

64. Mathis, D., Oudet, P., and Chambon, P. (1980) Structure of transcribing chromatin. Prog. in Nucl. Acid Res. and Mol. Biol. 24, 1-55.

65. Cartwright, I. L., Abmayr, S. M., Fleischmann, G., Lowenhaupt, K., Elgin, S. C. R., Keene, M. A., and Howard, G. C. (1982) Chromatin structure and gene activity: the role of nonhistone chromosomal proteins. CRC Critical Reviews in Biochemistry 13, 1-86.

66. Weisbrod, S. (1982) Active chromatin. Nature 297, 289-295.

67. Bryan, P. N., and Destree, O. H. J. (1982) Activation and function of chromatin in "Horizons in Biochemistry and Biophysics."

68. Igo-Kemenes, T., Hörz, W., and Zachau, H. G. (1982) Chromatin. Ann. Rev. Biochem. 51, 89-121.

69. Tsanev, R. (1983) Transcriptionally active chromatin. Molec. Biol. Rep. 9, 9-17.

70. Elgin, S. C. R. (1983) Chromatin structure and gene expression in Drosophila. Curr. Topics in Cell. Anat., 103-143.

71. Butler, P. J. G. (1984) The folding of chromatin. CRC Critical Reviews in Biochemistry 15, 57-91.

72. Eissenberg, J. C., Cartwright, I. L., Thomas, G. H., and Elgin, S. C. R. (1985) Selected topics in chromatin structure. Ann. Rev. Genet. 19, 485-536.

73. Yaniv, M., and Cereghini, S. (1986) Structure of transcriptionally active chromatin. CRC Critical Reviews in Biochemistry 21, 1-26.

74. Gottesfeld, J. M., and Bloomer, L. S. (1980) Nonrandom alignment of nucleosomes on 5S RNA genes of X. laevis. Cell 21, 751-760.

75. Young, D., and Carroll, D. (1983) Regular arrangement of nucleosomes on 5S rRNA genes in Xenopus laevis. Mol. and Cell. Biol. 3, 720-730.

76. Reynolds, W. F., and Gottesfeld, J. M. (1985) Torsional stress induces as S1 nuclease hypersensitive site within the promoter of the Xenopus laevis oocyte-type 5S RNA gene. Proc. Natl. Acad. Sci. USA 82, 4018-4022.

77. Louis, C., Schedl, P., Samal, B., and Worcel, A. (1980) Chromatin structure of the 5S RNA genes of D. melanogaster. Cell 22, 387-392.

78. Cartwright, I. L., and Elgin, S. C. R. (1984) Chemical footprinting of 5S RNA chromatin in embryos of Drosophila melanogaster. EMBO Journal 3, 3101-3108.

79. Fitzgerald, P. C., and Simpson, R. T. (1985) Effects of sequence alterations in a DNA segment containing the 5S RNA gene from Lytechinus variegatus on positioning of a nucleosome core particle in vitro. J. Biol. Chem. 260, 15318-15324.

80. Nazar, R. N., Yaguchi, M., and Willick, G. E. (1981) The 5S RNA-protein complex from yeast: a model for the evolution and structure of the eukaryotic ribosome. Can. J. Biochem. 60, 490-496.

81. Nazar, R. N., Yaguchi, M., Willick, G. E., Rollin, C. F., and Roy, C. (1979) The 5S RNA binding protein from yeast (Saccharomyces cerevisiae) ribosomes--evolution of the eukaryotic 5S RNA binding protein. Eur. J. Biochem. 102, 573-582.

82. Nazar, R. N. (1979) The ribosomal protein binding site in Saccharomyces cerevisiae ribosomal 5S RNA--a conserved binding site in 5S RNA. J. Biol. Chem. 254, 7724-7729.

83. Küntzel, H., Piechulla, B., and Hahn, U. (1983) Consensus structure and evolution of 5S rRNA. Nucl. Acids Res. 11, 893-900.

84. Kumazaki, T., Hori, H., and Osawa, S. (1983) Phylogeny of protozoa deduced from 5S rRNA sequences. J. Mol. Evol. 19, 411-419.

85. Delihas, N., Andersen, J., and Singhal, R. P. (1984) Structure, function, and evolution of 5-S ribosomal RNAs. Prog. in Nucl. Acid Res. and Mol. Biol. 31, 161-190.

86. Erdmann, V. A., and Wolters, J. (1986) Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. Nucl. Acids Res. 14 (suppl.), r1-r59.

87. Hori, H., and Osawa, S. (1986) Evolutionary change in 5S rRNA secondary structure and a phylogenic tree of 352 5S rRNA species. BioSystems 19, 163-172.

88. Wolters, J., and Erdmann, V. A. (1986) Cladistic analysis of 5S rRNA and 16S rRNA secondary and primary structure--the evolution of eukaryotes and their relation to archaebacteria. J. Mol. Evol. 24, 152-166.

89. Walker, W. F. (1985) 5S and 5.8S ribosomal RNA sequences and protist phylogenetics. BioSystems 18, 269-278.

90. Wilson, A. C., Carlson, S. S., and White, T. J. (1977) Biochemical evolution. Ann. Rev. Biochem. 46, 573-639.

91. Klotz, L. C., Komar, N., Blanken, R. L., and Mitchell, R. M. (1979) Calculation of evolutionary trees from sequence data. Proc. Natl. Acad. Sci. USA 76, 4516-4520.

92. Klotz, L. C., and Blanken, R. L. (1981) A practical method for calculating evolutionary trees from sequence data. J. Theor. Biol. 91, 261-272.

93. Hogeweg, P., and Hesper, B. (1984) The alignment of sets of sequences and the construction of phyletic trees: an integrated method. J. Mol. Evol. 20, 175-186.

94. Lawn, R. M., Heumann, J. M., Herrick, G., and Prescott, D. M. (1978) The gene-sized DNA molecules in Oxytricha. Cold Spring Harbor Symp. Quant. Biol. 42, 483-492.

95. Prescott, D. M., Heumann, J. M., Swanton, M., and Boswell, R. E. (1979) The genome of hypotrichous ciliates. In Engberg, J., Klenow, H., and Leick, V. (eds.), Specific Eukaryotic Genes, published by Munksgaard, Copenhagen.

96. Prescott, D. M. (1983) The C-value paradox and genes in ciliated protozoa. Modern Cell Biol. 2, 329-352.

97. Kraut, H., Lipps, H. J., and Prescott, D. M. (1986) The genome of hypotrichous ciliates. International Review of Cytology 99, 1-28.

98. Allen, R. L., and Olins, D. E. (1984) Cytochemistry of the chromatin replication band in hypotrichous ciliated protozoa staining with silver and thiol-specific coumarin maleimide. Chromosoma 91, 82-86.

99. Allen, R. L., Kennel, S. J., Cacheiro, L., Olins, A. L., and Olins, D. E. (1986) Examination of the macronuclear replication band in Euplotes eurystomus with monoclonal antibodies. J. Cell Biol. 102, 131-136.

100. Prescott, D. M., Bostock, C. J., Murti, K. G., Lauth, M. R., and Gamow, E. (1971) DNA of ciliated protozoa. I. Electron microscopic and sedimentation analyses of macronuclear and micronuclear DNA of Stylonychia mytilus. Chromosoma 34, 355-366.

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

102. Nock, A. (1981) RNA and macronuclear tarnscription in the ciliate Stylonychia mytilus. Chromosoma 83, 209-220.

103. Pluta, A. F., and Spear, B. B. (1981) Localization of genes for ribosomal RNA in the nuclei of Oxytricha fallax. Exp. Cell Res. 135, 387-392.

104. Olins, A. L., Olins, D. E., Franke, W. W., Lipps, H. J., and Prescott, D. M. (1981) Stereo-electron microscopy of nuclear structure and replication in ciliated protozoa (Hypotrichia). Eur. J. Cell Biol. 25, 120-130.

105. Dawson, D., Stetler, D. J., Swanton, M. T., and Herrick, G. (1983) Tandemly repeated sequence families in micronuclear DNA of the ciliate Stylonychia pustulata. J. Protozool. 30, 629-635.

106. Cartinhour, S. W., and Herrick, G. A. (1984) Three different macronuclear DNAs in Oxytricha fallox share a common sequence block. Mol. Cell. Biol. 4, 931-938.

107. Steinbuck, G. (1983) Overemplification of genes in macronuclei of hypotrichous ciliates. Chromosoma 88, 156-163.

108. Ammermann, D., and Muenz, A. (1982) DNA and protein content of different hypotrich ciliates. Eur. J. Cell Biol. 27, 22-24.

109. Swanton, M. T., Heumann, J. M., and Prescott, D. M. (1980) Gene-sized DNA molecules of the macronuclei in three species of hypotrichs: size distribution and absence of nicks. Chromosoma 77, 217-227.

110. Steinbrück, G., Hass, I., Hellmer, K. H., and Ammermann, D. (1981) Characterization of macronuclear DNA in five species of ciliates. Chromosoma 83, 199-208.

111. Ammermann, D., Steinbrück, G., vonBerger, L., and Hennig, W. (1974) The development of the macronucleus in the ciliated protozoan Stylonychia mytilus. Chromosoma 45, 401-429.

112. Lauth, M. R., Spear, B. B., Heumann, I., and Prescott, D. M. (1976) DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of Oxytricha. Cell 7, 67-74.

113. Spear, B. B., and Lauth, M. R. (1976) Polytene chromosmes of Oxytricha: Biochemical and morphological changes during macronuclear development in a ciliated protozoan. Chromosoma 54, 1-13.

114. Kloetzel, J. A. (1981) Nuclear roles in the postconjugant development of the ciliated protozoan Euplotes aediculatus. I. Evidence for sequential roles of the differentiating macronucleus in exconjugant development. Dev. Biol. 83, 20-32.

115. Kloetzel, J. A. (1981) Nuclear roles in the postconjugant development of the ciliate Euplotes aediculatus. II. Experimentally induced regneration of old macronnuclear fragments. J. Protozool. 28, 108-116.

116. Boswell, R. E., Klobutcher, L. A., and Prescott, D. M. (1982) Inverted terminal repeats are added to genes during macronuclear development in Oxytricha nova. Proc. Natl. Acad. Sci. USA 79, 3255-3259.

117. Klobutcher, L. A., Jahn, C. L., and Prescott, D. M. (1984) Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan Oxytricha nova. Cell 36, 1045-1055.

118. Roth, M., and Prescott, D. M. (1985) DNA intermediates and telomere addition during genome reorganization in Euplotes crassus. Cell 41, 411-417.

119. Lipps, H. J. (1985) A reverse transcriptase like enzyme in the developing macronucleus of the hypotrichous ciliate Stylonychia. Curr. Genet. 10, 239-243.

120. Wesley, R. D. (1975) Inverted repetitious sequences in the macronuclear DNA of hypotrichous ciliates. Proc. Natl. Acad. Sci. USA 72, 678-682.

121. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc. Natl. Acad. Sci. USA 78, 3015-3019.

122. Lipps, H. J., and Erhardt, P. (1981) DNase I hypersensitivity of the terminal inverted repeat DNA sequences in the macronucleus of the ciliate Stylonychia mytilus. FEBS Lett. 126, 219-222.

123. Pluta, A. F., Kaine, B. P., and Spear, B. B. (1982) The terminal organization of macronuclear DNA in Oxytrich fallax. Nucl. Acids Res. 10, 8145-8154.

124. Blackburn, E. H., Budarf, M. L., Challoner, P. B., Cherry, J. M., Howard, E. A., Katzen, A. L., Pan, W. C., and Ryan, T. (1983) DNA termini in ciliate macronuclei. Cold Spring Harbor Symposia on Quantitative Biology 47, 1195-1207.

125. Blackburn, E. H. (1984) Telomeres: do the ends justify the means? Cell 37, 7-8.

126. Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

127. Dawson, D., and Herrick, G. (1982) Micronuclear DNA sequences of Oxytricha fallax homologous to the macronuclear inverted terminal repeat. Nucl. Acids Res. 10, 2911-2924.

128. Dawson, D., and Herrick, G. (1984) Telomeric properties of C_4A_4-homologous sequences in micronuclear DNA of Oxytricha fallax. Cell 36, 171-177.

129. Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A., and Williams, K. (1985) Mobile elements bounded by C_4A_4 telomeric repeats in Oxytricha fallax. Cell 43, 759-768.

130. Cherry, J. M., and Blackburn, E. H. (1985) The internally located telomeric sequences in the germ-line chromosomes of Tetrahymena are at the ends of transposon-like elements. Cell 43, 747-758.

131. Szostak, J. W., and Blackburn, E. H. (1982) Cloning yeast telomeres on a yeast linear plasmid. Cell 29, 245-255.

132. Dani, G. M., and Zakian, V. A. (1983) Mitotic and meiotic stability of linear plasmids in yeast. Proc. Natl. Acad. Sci. USA 80, 3406-3410.

133. Murray, A. W., and Szostak, J. W. (1983) Construction of artificial chromosomes in yeast. Nature (London) 305, 189-193.

134. Pluta, A. F., Dani, G. M., Spear, B. B., and Zakian, V. A. (1984) Elaboration of telomeres in yeast: recognition and modification of termini from Oxytricha macronuclear DNA. Proc. Natl. Acad. Sci. USA 81, 1475-1479.

135. Gottschling, D. E., and Zakian, V. A. (1986) Telomere proteins that recognize the molecular ends of Oxytricha macronuclear DNA. Abstract for annual meeting of the American Society for Cell Biology. J. Cell Biol. 103, 289A.

136. Lipps, H. J., Gruissem, W., and Prescott, D. M. (1982) Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate Oxytricha nova. Proc. Natl. Acad. Sci. USA 79, 2495-2499.

137. Butler, A. P., Laughlin, T. J., Cadilla, C. L., Henry, J. M., and Olins, D. E. (1984) Physical structure of gene-sized chromatin form the protozoan Oxytricha. Nucl. Acids Res. 12, 3201-3217.

138. Cadilla, C. L., Harp, J., Flanagan, J. M., Olins, A. L., and Olins, D. E. (1986) Preparation and characterization of soluble macronuclear chromatin from the hypotrich Euplotes eurystomus. Nuc. Acid Res. 14, 823-841.

139. Laughlin, T. J., Herrmann, A. L., and Olins, D. E. (1984) Fractionation of the gene-size macronuclear chromatin fragments of the binucleated eukaryote Oxytricha. Molecular and Cellular Biochemistry, 62, 157-163.

140. Cadilla, C. L., Roberson, A. E., Harp, J., Olins, A. L., and Olins, D. E. (1986) Subfractionation of soluble macronuclear chromatin and enrichment of specific genes as chromatin from Euplotes eurystomus. Nucl. Acids Res. 14, 8501-8512.

141. Gottschling, D. E., and Cech, T. R. (1984) Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex. Cell 38, 501-510.

142. Lipps, H. J., and Steinbrück, G. (1978) Free genes for rRNAs in the macronuclear genome of the ciliate Stylonychia mytilus. Chromosoma 69, 21-26.

143. Swanton, M. T., Geslin, A. F., and Prescott, D. M. (1980) Arrangement of coding and noncoding sequences in the DNA molecules coding for rRNAs in Oxytricha sp. Chromosoma 77, 203-215.

144. Spear, B. B. (1980) Isolation and mapping of the rRNA genes in the macronucleus of Oxytricha fallax. Chromosoma 77, 193-202.

145. Swanton, M. T., McCarroll, R. M., and Spear, B. B. (1982) The organization of macronuclear rDNA molecules of four hypotrichous ciliated protozoans. Chromosoma 85, 1-9.

146. Sogin, M. L., Swanton, M. T., Gunderson, J. H., and Elwood, H. J. (1986) Sequence of the small subunit ribosomal RNA gene from the hypotrichous ciliate Euplotes aediculatus. J. Protozool. 33, 26-29.

147. Roberson, A. E., and Olins, D. E. (1986) Cloning of 5S rRNA genes from Euplotes eurystomus. J. Cell Biol. 103, 308a.

148. Kaine, B. P., and Spear, B. B. (1980) Putative actin genes in the macronucleus of Oxytricha fallax. Proc. Natl. Acad. Sci. USA 77, 5336-5340.

149. Kaine, B. P., and Spear, B. B. (1982) Nucleotide sequence of a macronuclear gene for actin in Oxytricha fallax. Nature 295, 430-432.

150. Helftenbein, E. (1984) Nucleotide sequence of a macronuclear gene sized DNA molecule coding for α-tubulin in Stylonychia lemnae. Abstract from the Cold Spring Harbor meeting on ciliate molecular genetics.

151. Helftenbein, E. (1985) Nucleotide sequence of a macronuclear DNA molecule coding for α-tubulin from the ciliate Stylonychia lemnae. Special codon usage: TAA is not a translation termination codon. Nucl. Acids Res. 13, 415-433.

152. Elsevier, S. M., Lipps, H. J., and Steinbrück, G. (1978) Histone genes in macronuclear DNA of the ciliate Stylonychia mytilus. Chromosoma 69, 291-306.

153. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.

154. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular cloning--A laboratory manual. Cold Spring Harbor Laboratory.

155. Clewell, D. B. (1972) Nature of ColE$_1$ plasmid replication in Escherichia coli in the presence of chloramphenicol. J. Bacteriol. 110 667-676.

156. Carmichael and McMaster. (1980) Methods in Enzymology 65, 47.

157. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227, 680-685.

158. Thomas, J. O., and Kornberg, R. D. (1975) An octamer of histones in chromatin and free in solution. Proc. Natl. Acad. Sci. USA 72, 2626-2630.

159. Dessev, G. N., and Grancharov, K. (1973) Precipitation of RNA, DNA, and nucleoprotein particles from very dilute solutions. Anal. Biochem. 53, 269-271.

160. Wray, W., Boulikas, T., Wray, V. P., and Hancock, R. (1981) Silver staining of proteins in polyacrylamide gels. Analytical Biochemistry 118, 197-203.

161. Bethesda Research Laboratories Catalogue and Reference Guide. (1985)

162. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separatd by gel electrophoresis. J. Mol. Biol. 98, 503-517.

163. Lehrach, Diamond, D., Wozey, J. M., and Boedtker, H. (1977) RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical review. Biochem. 16, 4743-4751.

164. Thomas, P. S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77, 5201-5205.

165. Thomas, P. S. (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods in Enzymology 100, 255-266.

166. Komaiko, W., and Felsenfeld, G. (1985) Solubility and structure of domains of chicken erythrocyte chromatin containing transcriptionally competent and inactive genes. Biochem. 24, 1186-1193.

167. Pederson, D. S., Yao, M.-C., Kimmel, A. R., and Gorovsky, M. A. (1984) Sequence organization within and flanking clusters of 5S ribosomal RNA genes in Tetrahymena. Nucl. Acids Res. 12, 3003-3021.

168. Laskey, R. A., and Mills, A. D. (1975) Quantitative film detection of ^{3}H and ^{14}C in Polyacrylamide gels by fluorography. Eur. J. Biochem. 56, 335-341.

169. Mees, C. E. K., and James, T. H. (eds.). (1966) The theory of the photographic process. 3rd ed. Macmillan, New York.

170. Lüthi, V., and Waser, P. G. (1965) Nature (London) 205, 1190-1192.

171. Gatti, R. A., Cancannon, P., and Salser, W. (1984) Multiple use of Southern blots. Biotechniques 2, 148-155.

172. Kato, K., Goncalves, J. M., Houts, G. E., and Bollum, F. J. (1967) Deoxynucleotide-polymerizing enzymes of calf thymus gland II: Properties of the terminal deoxynucleotidyltransferase. J. Biol. Chem. 242, 2780-2789.

173. Roychoudhury, R., and Wu, R. (1980) Terminal transferase-catalyzed addition of nucleotides to the 3' termini of DNA. Methods in Enzymology 65, 43-62.

174. Deng, G., and Wu, R. (1981) An improved procedure for utilizing terminal transferase to add homopolymers to the 3' termini of DNA. Nucl Acids Res. 9, 4173-4188.

175. Vieira, J., and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

176. Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557-580.

177. Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P. H. (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a Hind II fragment of the lac regultory region in the M13 replicative form in vitro. Proc. Natl. Acad. Sci. USA 75, 3642.

178. Heidecker, G., Messing, J., and Gronenborn, B. (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene 10, 68.

179. Messing, J., Crea, R., and Seeburg, P. H. (1981) A system of shotgun DNA sequencing. Nucl. Acids Res. 9, 309.

180. Messing, J., and Vieira, J. (1982) A new pair of M13 vectors for selecting either strand of a double-digest restriction fragment. Gene 19, 269.

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

182. Norrander, J., Kempe, T., and Messing, J. (1983) Construction of imporved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26, 101.

183. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

184. M13 Cloning/Dideoxy Sequencing Instruction Manual. Bethesda Research Laboratories.

185. Birnboim, H. C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1513.

186. Marko, H. (1983) Cybernetics, Information theory and communication theory. In Hoppe, W., Lohmann, W., Markl, H., and Ziegler, H. (eds.), Biophysics (pp. 788-794). Springer-Verlag.

187. Morowitz, H. J. (1970) Probability and information. Entropy, information, and energy. In Entropy for Biologists--An Introduction to Thermodynamics (pp. 61-71; 105-112). Academic Press.

188. Maxam, A. M., and Gilbert, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560.

189. Maxam, A. M., and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology 65, 499-560.

190. Sanger, F., and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94, 441.

191. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463.

192. Smith, A. J. H. (1980) DNA sequence analysis by primed synthesis. Methods in Enzymology 65, 560-581.

193. Sanger, F. (1981) Determination of nucleotide sequences in DNA. Science 214, 1205.

194. Shimkus, M. L., Guaglianone, P., and Herman, T. M. (1986) Synthesis and characterization of Biotin-labelled nucleotide analogs. DNA 5, 247-255.

195. Herman, T. M., Lefever, E. and Shimkus, M. (1986) Affinity chromatography of DNA labeled with chemically cleavable biotinylated nucleotide analogs. Anal. Biochem. 156, 48-55.

196. Roseman, B., Lough, J., Houkom, E., and Herman, T. (1986) Affinity isolation of transcriptionally active DNA. Biochem. Biophys. Res. Comm. 137, 474-479.

196a. Allan, J., Cowling, G. J., Harborne, N., Cattini, P., Craigie, R., and Gould, H. (1981) Regulation of the higher-order structure of chromatin by histones H1 and H5. J. Cell Biol. 90, 279-288.

197. Cadilla, C. L. (1986) Characterization and subfractionation of macronuclear chromatin from the hypothrich Euplotes eurystomus. Ph.D. Dissertation, The University of Tennessee, Oak Ridge.

198. Noll, M. (1974) Subunit structure of chromatin. Nature (London) 251, 249-251.

199. McCarty, K. S., Jr., Vollmer, R. T., and McCarty, K. S. (1974) Improved computer program data for the resolution and fractionation of macromolecules by isokenetic sucrose density gradient sedimentation. Anal. Biochem. 61, 165-183.

200. Dubochet, J., Ducommun, M., Zollinger, M., and Kellenberger, E. (1971) A new preparation method for dark field electron microscopy of macromolecules. J. Ultrastructural Res. 35, 147-267.

201. Olins, A. L. (1979) Eletron microscopy: a tool for visualizing chromatin. In: Chromatin Structure and Function (C. A. Nicolini, ed.), NATO Advanced Study Institutes Series: Series A, vol. 21, pt. A, Plenum Press, New York, pp. 31-40.

202. Weil, P. A., Luse, D. S., Segall, J., and Roeder, R. G. (1979) Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18, 469-484.

203. Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77, 3855-3859.

204. Tsai, S. Y., Tsai, M.-J., Kops, L. E., Minghetti, P. P., and O'Malley, B. W. (1981) Transcription factors from oviduct and HeLa cells are similar. J. Biol. Chem. 256, 13055-13059.

205. Grummt, I. (1981) Specific transcription of mouse ribosomal DNA in a cell-free system that mimics control in vivo. Proc. Natl. Acad. Sci. USA 78, 727-731.

206. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acid Res. 11, 1475-1489.

207. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193, 265-275.

208. Soeda, E., Miura, K., Nakaso, A., and Kimura, G. (1977) Nucleotide sequence around the replication origin of polyoma virus DNA. FEBS Lett. 79, 383-389.

209. Stinchcomb, D. T., Thomas, M., Kelly, J., Selker, E., and Davis, R. W. (1980) Eukaryotic DNA segments capable of autonomous replication of yeast. Proc. Natl. Acad. Sci. USA 77, 4559-4563.

210. Montiel, J. F., Norbury, C. J., Tuite, M. F., Dobson, M. J., Mills, J. S., Kingsman, A. J., and Kingsman, S. M. (1984) Characterization of human chromosomal DNA sequences which replicate autonomously in Saccharomyces cerevisiae. Nucl. Acids Res. 12, 1049-1068.

211. Singhal, R. P., and Shaw, J. K. (1983) Prokaryotic and Eukaryotic 5S RNAs: Primary sequences and proposed secondary structures. Progress in Nucleic Acid Research and Molecular Biology 28, 177-209.

212. Pieler, T., and Erdmann, V. A. (1982) Three-dimensional structural model of eubacterial 5S RNA that has functional implications. Proc. Natl. Acad. Sci. USA 79, 4599-4603.

213. Hancock, J., and Wagner, R. (1982) Nucl. Acids Res. 10, 1257.

214. Uberbacher, E. C., Wilkinson-Singley, E., Harp, J. M., and Bunick, G. J. (1988) DNA sequence effects on nucleosome positioning and DNA bending. In Olson, W. K., Sarma, M. H., Sarma, R. H., and Sundaralingam, M. (eds.), Structure and Expression DNA Bending and Curvature. Adenine Press.

215. Mannella, C. A., Frank, J., and Delihas, N. (1987) Interrelatedness of 5S RNA sequences investigated by correspondence analysis. J. Mol. Evol. 24, 228-235.

216. Klobutcher, L. A., and Prescott, D. M. (1986) The special case of the hypotrichs. In Gall, J. G. (ed.), The Molecular Biology of Ciliated Protozoa (p. 111-154). Academic Press.

VITA

Arthur E. Roberson was born to James and Greta Roberson on October 15, 1959, in Los Angeles, California. He graduated from Beavercreek High School in Dayton, Ohio, in 1977. He was awarded the Bachelor of Arts degree with a major in chemistry in 1981 and the Doctor of Philosophy degree with a major in biomedical sciences in 1988, both from The University of Tennessee. His doctoral research was under the supervision of Donald Olins at the Oak Ridge National Laboratory. Mr. Roberson has a wife, Twilah, and a daughter, Sarah.