Studies on the Structure, Expression, Genome Organization, and Evolution of U3 snoRNA Genes in *Euglena gracilis*

by

Joseph Michael Claude Charette

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

at

Dalhousie University
Halifax, Nova Scotia
January 2007

Your file Votre référence
ISBN: 978-0-494-27176-6
Our file Notre référence
ISBN: 978-0-494-27176-6

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages
- Examiners Signature Page (pii)
- Dalhousie Library Copyright Agreement (piii)

Appendices
- Copyright Releases (if applicable)

DEDICATION

This thesis is dedicated to my family.

To my loving wife Ellen, for her presence, her bountiful love, and unending support.

To my mother, for her love, and her years of patience, sacrifice, and support.

To my father's memory...He would be so proud.

"The complexity inherent in the production of ribosomal RNAs (rRNAs), the key components of the ribosome, is interesting in itself. Add to this the complex nature of the production and regulation of the myriad other components involved in ribosome biogenesis, and the topic becomes fascinating."

Brown, J.W.S. and Shaw, P.J. 1998. Small nucleolar RNAs and pre-rRNA processing in Plants. Plant Cell 10: 649-657.

"Evolution, as truth, insofar as we can comprehend it at the moment; as a realistic assessment of our position in the universe; and as a joyous celebration of our potential future."

John Sulston, co-winner of the 2002 Nobel Prize in Medicine or Physiology in answer to the question: "If you could teach the world one thing about science, what would it be?" posed by Alom Shaha. The Scientist, May 23, 2005, pg 12.

"It is our choices, Harry, that show what we truly are, far more than our abilities."

Harry Potter and the Chamber of Secrets; Chapter 18, Dobby's Reward. In a conversation between Dumbledore and Harry Potter after having emerged, victorious, from the Chamber of Secrets.

"Though faith is above reason, there can never be any real discrepancy between faith and reason. Since the same God who reveals mysteries and infuses faith has bestowed the light of reason on the human mind, God cannot deny himself, nor can truth ever contradict truth. Consequently, methodical research in all branches of knowledge, provided it is carried out in a truly scientific manner and does not override moral laws, can never conflict with the faith, because the things of the world and the things of faith derive from the same God. The humble and persevering investigator of the secrets of nature is being led, as it were, by the hand of God in spite of himself, for it is God, the conserver of all things, who made them what they are."

Catechism of the Catholic Church, item number 159: *Faith and science*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Small nucleolar RNAs (snoRNAs) are small RNA molecules present in eukaryotic nucleoli. They guide nucleoside modification (formation of $O^{2'}$-methylribose and pseudouridine) and endonucleolytic cleavage events necessary in the processing, folding, and assembly of the pre-rRNA transcript into mature rRNAs. In all known cases, the site of modification or cleavage in the pre-rRNA/rRNA is specified by short nucleotide sequences present in the snoRNA, complementary to the target rRNA sequence. This specificity, provided by the snoRNA through base pairing interactions with the target rRNA, localizes a specific snoRNP to the rRNA target where the relevant protein-catalyzed modification or cleavage reaction occurs. U3 snoRNA plays an essential role in many of the endonucleolytic processing events that produce the mature 18S rRNA.

This thesis examines the organization of U3 snoRNA genes in the genome of *Euglena gracilis.* Southern analysis revealed at least 13 U3-hybridizing bands. Both PCR and genomic λ library screening were used to isolate U3 sequences. Analysis of the recovered sequences identified 14 U3 gene variants. Sequence heterogeneities identified in the U3 gene variants were located in the 3′-stem-loop domain. An examination of the genomic neighborhood of *Euglena* U3 snoRNA genes revealed the presence of at least three different genomic organizations: *i*) stand-alone, *ii*) linked to $tRNA^{Arg}_{UCG}$ genes, and *iii*) linked to a U5 snRNA gene. In the U3 snoRNA-$tRNA^{Arg}$ linkage, the U3 snoRNA gene is linked to two identical, downstream, convergently oriented (relative to the U3 gene) $tRNA^{Arg}$ genes. This scenario is reminiscent of, but different in a number of key ways from, the U3 snoRNA-tRNA gene linkage in trypanosomatids. Twelve different U3 snoRNA-U5 snRNA gene linkages were identified. In each case, the U3 gene is linked to a downstream and convergently oriented U5 snRNA gene. The multiple U3 snoRNA-U5 snRNA gene linkages cluster into distinct families based on sequence similarities within the intergenic spacer. The multiple *Euglena* U3 gene copies may have arisen by genome, chromosome, and/or locus duplications. Finally, the unexpected variability in the signal intensities of the multiple Southern hybridizing bands raises the possibility that *Euglena* might contain a naturally aneuploid chromosome complement.

LIST OF ABBREVIATIONS AND SYMBOLS USED

A	adenosine
A_0	A_0 processing site in pre-rRNA
A_1	A_1 processing site in pre-rRNA
A_2	A_2 processing site in pre-rRNA
A_3	A_3 processing site in pre-rRNA
A_{260}	absorbance at a light wave length of 260 nm
Arg	arginine
BLAST	basic local alignment search tool
BLASTn	nucleotide query vs. nucleotide database BLAST
bp	base pair(s)
BSA	bovine serum albumen
°C	degree(s) Celsius
C	cytidine
cDNA	complementary DNA
Ci	Curie
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxyribonucleoside triphosphate
dTTP	deoxythymidine triphosphate
DNA	deoxyribonucleic acid

DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylene-diamine-tetra-acetic acid
est	expressed sequence tag
ETS	external transcribed spacer (of the pre-rRNA; 5′-ETS and 3′-ETS)
EWB	extraction wash buffer
γ	gamma
g	gram
g	gravity units (in relative centrifugal force)
G	guanosine
gRNA	guide RNA
hr	hour(s)
IGS	intergenic spacer
indel	insertion/deletion (of nucleotide or protein sequence)
ITS	internal transcribed spacer (of the pre-rRNA; ITS1, ITS2)
J	β-D-glucosyl-hydroxymethyluracil
kbp	kilobase pair(s)
KOAc	potassium acetate
K-turn	kink-turn
λ	lambda
l	liter
LB	Luria Bertani media

LSU	large subunit (of the ribosome)
5-MeC	5-methylcytosine
μCi	microCurie
μg	microgram
μl	microliter
μM	micromolar
M	molar
MDa	megadalton
mg	milligram
MgOAc	magnesium acetate
min	minute(s)
ml	milliliter
mM	millimolar
mMol	millimole
NaOAc	sodium acetate
ND-PAGE	non-denaturing PAGE
ng	nanogram
NH_4OAc	ammonium acetate
nm	nanometer
Nm	methylation of the oxygen at ribose position 2 (2′-*O*-methylnucleoside or $O^{2'}$-methylnucleoside)
NMF	*N*-methylformamide

nt	nucleotide
OD_{600}	optical density at 600 nm light wave length
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque-forming unit
pg	picogram
pH	potential of hydrogen
pmol	picomole
pre-rRNA	precursor ribosomal RNA
Ψ	pseudouridine (in RNA) pseudogene (in gene nomenclature)
RACE	rapid amplification of cDNA ends
rDNA	ribosomal DNA (gene)
RNA	ribonucleic acid
RNAP I	RNA polymerase I (responsible for pre-rRNA transcription)
RNAP II	RNA polymerase II (responsible for mRNA transcription)
RNAP III	RNA polymerase III (responsible for tRNA, snRNA, and 5S rRNA transcription)
RNase A	ribonuclease A
rpm	revolutions per minute
rRNA	ribosomal RNA

S	Svedberg sedimentation unit
SDS	sodium dodecyl sulphate
sec	second
SL RNA	spliced leader RNA
sRNA	small RNAs (archaeal snoRNA)
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SSU	small subunit (of the ribosome)
TMG	trimethylguanosine ($m_3^{2,2,7}G$)
T	thymidine
Taq	*Thermus aquaticus* (DNA polymerase)
TFIIIC	transcription factor IIIC
T_m	melting temperature
Tris	tris (hydroxymethyl) aminomethane
tRNA	transfer RNA
U	unit (of enzyme activity) uridine
Utp	U-three-associated protein
V	volt
v/v	volume/volume
w/v	weight/volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

ACKNOWLEDGEMENTS

I must first acknowledge and thank my loving wife Ellen for her help and support throughout the many years of graduate school and during the writing of this thesis. She was my support, my source of motivation, and the only person capable of calming my frequently frayed nerves. Without her, this thesis would never have been completed; I am forever in her debt. You, Ellen, are the best thing that I got out of graduate school.

Many members of the Gray lab deserve special thanks and recognition. First and foremost, I wish to thank my supervisor, Dr. Michael W. Gray, for offering me the golden opportunity of working in his lab. It has been an intellectually rigorous and challenging experience that is bearing fruit and opening countless doors. Thank you, BBM, for your time, endless patience, supervision, tough love, mentorship, and financial support.

Special thanks must also be given to Drs. Murray N. Schnare and David F. Spencer for their constant willingness and keen enthusiasm to help me though the many technical and biological difficulties experienced during this project. More than anyone, they have taught me about the importance of understanding the technical aspects of experimental protocols. Also, they have taught me to be a more critical and discerning scientist, not only in evaluating the validity and accuracy of my work, but also of the published literature.

Many thanks must be given to Dr. Amanda Lohan. Amanda kept me laughing, kept my eyes on the prize, was there for me during the many difficult parts of graduate school, and most importantly, was my biggest fan. Thank you, Amanda, for your unending support and encouragement, academically, experimentally, professionally, and personally. I would not have gotten through the difficult parts of this thesis without you! Similarly, without your encouragement, I would not be embarking on this next promising step in my career. Thank you Amanda!!!

So many other people are deserving of thanks. To Dr. John M. Archibald for opening my eyes to the wonders of evolution. I fondly remember our many kitchen talks during which you pushed the boundaries of my knowledge and opened my eyes to so many evolutionary theories. Most importantly, thank you, John, for serving as my personal reminder that a successful scientific career can be accomplished without losing a sense of self and while maintaining normal human interactions with others and with family. You have served, and continue to serve, as my role model!

Thank you to the many true friends that I have made during graduate school. While too many to list here, each and everyone one of you has served as a source of support, be it academic, social, or intellectual. Special thanks must be given to Dr. Melanie Dobson for her continued support and encouragement throughout the year. I have learnt more about doing science from her than from anyone else. She too has served as my role model and as the person I aspire to become. Many thanks to all the coaches and fellow swimmers of the

Dalhousie Masters Swim team for the friendship over the years, and for keeping my body and (more importantly) my mind working in top shape.

Lastly, I wish to acknowledge the support and family and friends. Thank you mom for you patience over the many years, for your willingness to put so many plans on hold while I finish graduate school, and for your financial and emotional support. My father would be so proud. Special thanks must also be given to my wife's family. To my mother- and father-in-law (Elaine and Leonard), grandmother (Rose) brothers-in-law (the best a guy could get...Robert and John), and sister-in-law (Randy Lynn) thank you for your support and encouragement. You, more than anyone, know how difficult this process has been. Thank you for saving me from myself, for forcing me to think clearly and rationally, and for preventing me from acting rashly and making decisions that I would have later regretted. This thesis belongs to you.

Financial support from the Canadian Institutes of Health Research (CIHR) and interaction support from the Program in Evolutionary Biology of the Canadian Institute for Advanced Research (CIAR) is gratefully acknowledged.

CHAPTER 1: INTRODUCTION

1.1: Perspective

In the past 15 years, the field of ribosome research has been revolutionized by the discovery of the small nucleolar RNAs (snoRNAs). When originally identified, the precise role of snoRNAs was not obvious. They were initially found to localize to the nucleolus, the site of ribosome biosynthesis, then to associate with ribosomes, and finally to possess sequence elements that are complementary, or anti-sense, to rRNA. Subsequently, these small, versatile, *trans*-acting molecules have been implicated in nearly every aspect of ribosome biosynthesis, from the endonucleolytic cleavage of the pre-rRNA transcript to the chemical modification of ribonucleotides, and the folding and assembly of mature ribosomes. More than 100 different snoRNAs and countless associated proteins have been identified in all examined organisms. Every aspect of snoRNA biology has been fascinating, with discoveries about their structure, mechanism of action, numerous protein partners, unusual mode of gene expression, and unexpected genomic organization all challenging many well-established dogmas.

This thesis explores the gene organization of one particular snoRNA, U3, in the genome of the protist *Euglena gracilis*. U3 snoRNA is the most extensively studied snoRNA. It plays a critical role in pre-rRNA processing and maturation. Most of our knowledge on snoRNAs and on ribosome biosynthesis stems from work done in yeast and vertebrates. Thus, our understanding of these phenomena may not be truly representative of the many, potentially different strategies used by phylogenetically disparate organisms. The many well known idiosyncratic features of the RNA biology of *Euglena*, a distant but specific relative of trypanosomatids, represent such an example. Thus, an examination of the organization of U3 snoRNA genes in the genome of *Euglena* may provide a valuable insight into the diversity of biological systems and into the evolution of snoRNAs and of the ribosome itself.

1.2: An Introduction to rRNA Processing and Maturation

In actively growing cells, a high proportion of the cellular metabolism is devoted to the transcription, processing, modification, folding, and assembly of pre-rRNAs into mature rRNAs and functional ribosomes. Indeed, in the yeast *Saccharomyces cerevisiae*, as many as 2000 ribosomes may be synthesized every minute [1].

In most eukaryotic organisms, hundreds of rDNA gene operons are organized in multi-copy arrays. Each operon encodes one copy each of the 18S, 5.8S, and 28S rDNA genes. Transcription by RNA Polymerase I (RNAP I) produces a pre-rRNA transcript that must be processed to release the mature rRNA species (see Figure 1.1) [2]. The pre-rRNA is cleaved at multiple locations, initially at A_0 in the 5′-external transcribed spacer (5′-ETS) upstream of the 18S rRNA. Subsequently, cleavage at site A_1 produces the mature 5′-end of the 18S rRNA. Further cleavage events, at sites A_2 and A_3 in the internal transcribed spacer 1 (ITS1), liberate the 3′-region of the 18S rRNA (in addition to subsequent processing that generates the mature 3′-end) from the remainder of the pre-rRNA transcript. Subsequent processing steps liberate the mature 5.8S and 28S rRNAs. During this time, many nucleosides in the rRNA are modified by the formation of pseudouridine (Ψ) and $O^{2'}$-methylribose (Nm) residues. The rRNAs also fold and associate with ribosomal proteins in the formation of mature ribosomes [3,4].

1.3: An Introduction to snoRNAs

The small nucleolar RNAs (snoRNAs) constitute a very large assortment of small RNA molecules associated with the nucleolus, the site of ribosome biosynthesis. SnoRNAs are ubiquitously found in eukaryotes and in Archaea (where they are called snoRNA-like or small RNAs (sRNAs) since Archaea do not have nucleoli) [5-7]. They are not present in Bacteria, though bacterial homologues of select snoRNA-interacting proteins have been identified [8]. Two families of snoRNAs have been defined based on sequence features, functional roles, and protein components: the box C/D and the box H/ACA snoRNAs.

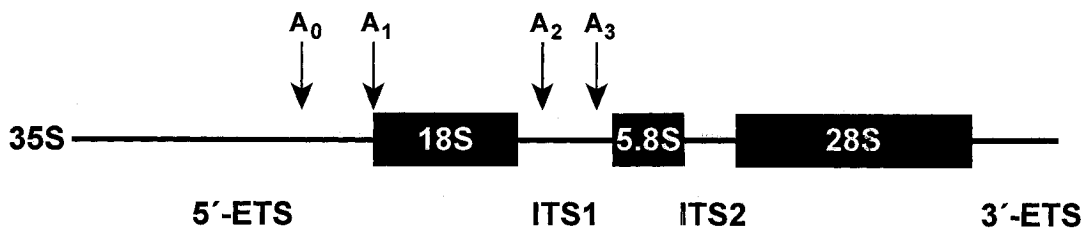


Figure 1.1 Gene Organization, Transcription, and Processing of rRNAs. A single rDNA gene operon is shown, corresponding to a 35S pre-rRNA transcript (in the yeast *Saccharomyces cerevisiae* [2]). The transcript is endonucleolytically processed by cleavage *i*) within the 5′-ETS at site A_0, *ii*) at site A_1 to generate the mature 5′-end of the 18S rRNA, and *iii*) within the ITS1 at sites A_2 and A_3 (additional processing events will generate the mature 3′-end of the 18S rRNA). These multiple processing events liberate a mature 18S rRNA from the pre-rRNA transcript. Additional processing events, not shown, liberate mature 5.8S and 28S rRNAs from the pre-rRNA transcript.

The box C/D snoRNA family and its role in rRNA nucleotide modification was first described in 1996 by the laboratory of Jean-Pierre Bachellerie [9,10]. Box C/D snoRNAs are defined by their short sequence elements, boxes C (RUGAUGA) and D (CUGA)(see Figure 1.2-A1). Members of this snoRNA family display minimal secondary structure features, other than a short terminal stem structure formed by base pairing interactions of the 5′- and 3′-ends of the snoRNA. (However, this structure is absent from *Euglena* [11], *Trypanosoma brucei* [12], and *Dictyostelium discoideum* [13] box C/D snoRNAs.) The box C and D sequence elements are positioned near the terminal stem structure; degenerate internal versions of boxes C and D, called C′ and D′ are frequently found. Sequences within the juxtaposed box C and D (or box C′/D′) elements [14,15] participate in the formation of an RNA secondary structure motif called a kink-turn (K-turn; see Figure 1.3) [16,17]. This motif was first described in the crystal structure of U4 snRNA bound to the sn/snoRNP protein 15.5-kD/Snu13p [17] and in the crystal structure of the SSU and LSU rRNAs of *Haloarcula marismortui* [16]. The K-turn is an ~15 nt two-stranded RNA motif consisting of a stem I-bulge-stem II structure [16]. Stem I consists exclusively of canonical base-pairs and ends with a 3-nt internal bulge. Stem II follows the internal loop and starts with two noncanonical sheared G•A and A•G pairs then continues with additional canonical base-pairs. The internal bulge between the helical stems is always asymmetrical and typically consists of 3 unpaired nucleotides on one strand and none on the other. These base-pairing interactions result in the stacking of the nucleotides comprising stem I and stem II which induces a characteristic 120° kink in the phosphodiester backbone between stem I and stem II [16].

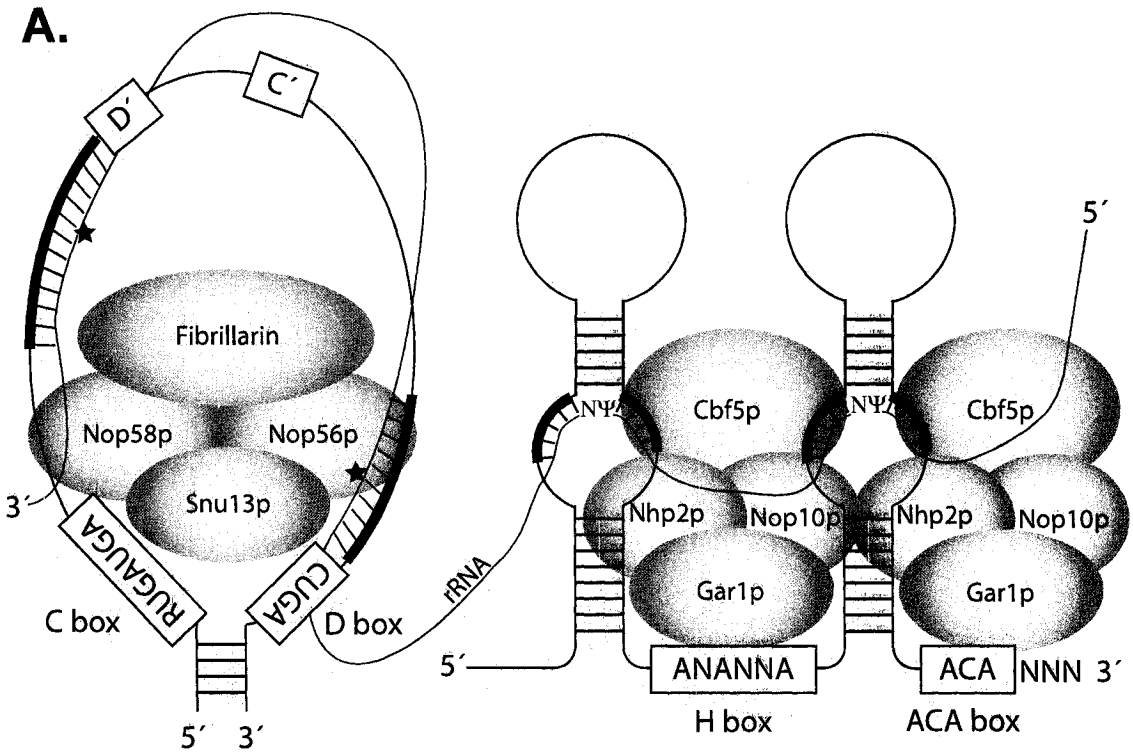
A short (10- to 20-nt) region of complementarity to an RNA target (most frequently rRNA, but also snRNAs [18,19]) is located immediately upstream of the D (or alternative D′) box sequence. The snoRNA guide sequence base-pairs to the rRNA/snRNA target. The fifth nucleotide upstream of the D (or D′) box in the target sequence is modified by methylation of the $O^{2'}$ position of the sugar (2′-OH to a 2′-*O*-methyl), thus forming an Nm residue (see Figure 1.2-B1). Some box C/D snoRNAs, such as U3 and U8 are not known to guide Nm formation, but are instead involved in endonucleolytic processing of the pre-rRNA

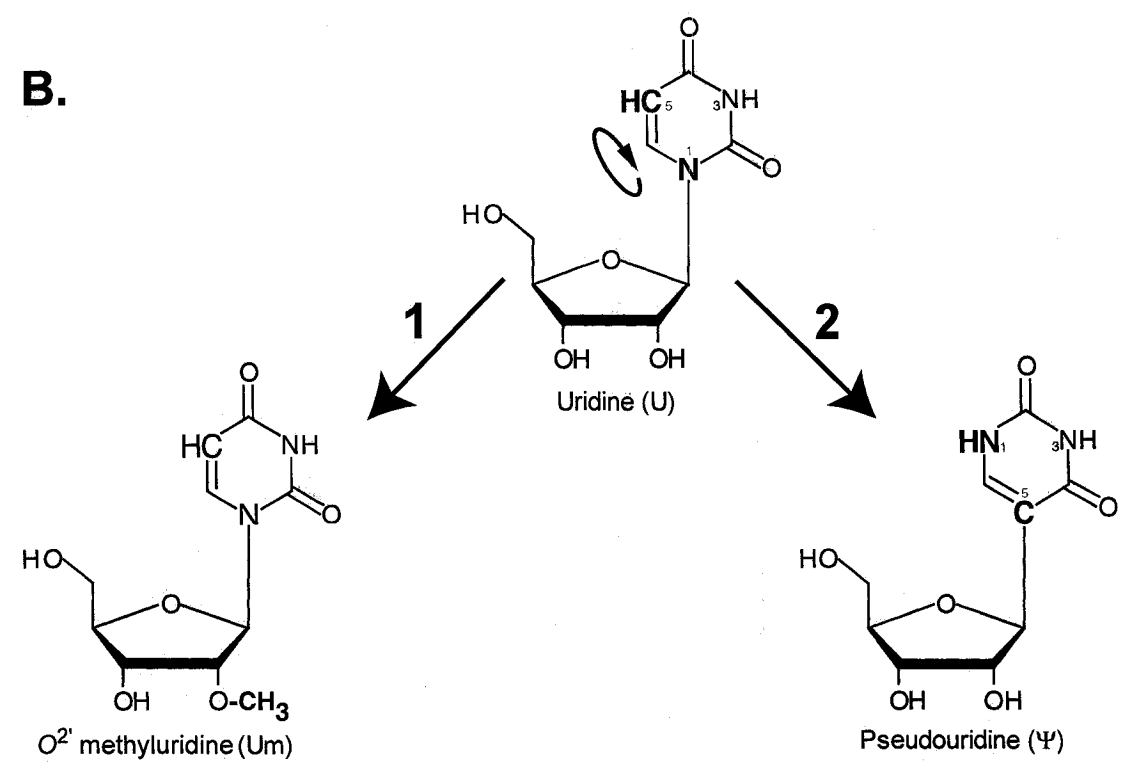
transcript. Box C/D snoRNAs are known to associate with four core proteins, Snu13p, Nop56, Nop58, and Nop1 (the latter is also known as fibrillarin, the methyltransferase that mediates Nm formation [20]) to form small nucleolar ribonucleoprotein (snoRNP) complexes.

Nm residues are found in tRNAs, snRNAs, snoRNAs, and rRNAs (though eukaryotic box C/D snoRNAs only guide Nm formation in rRNAs and at select positions in snRNAs). In all cases, Nm residues are found in functionally important regions of the molecule, such as those involved in intermolecular RNA-RNA contacts in the ribosome [21-23] and in snRNAs [24]. The molecular role played by Nm residues remains speculative. Based on nuclear magnetic resonance studies, the large methyl group can sterically clash with the base. Thus, in an Nm residue, the ribose will preferentially adopt a 3′-endo conformation resulting in the base adopting an axial position to minimize unfavorable steric interactions [25]. Any nucleoside modification that stabilizes the 3′-endo ribose conformation promotes stacking of the RNA bases which results in the stabilization of the local RNA structure and increased structural rigidity of the region [25,26].

The box H/ACA snoRNA family and its role in rRNA nucleotide modification was first described in 1996 by the laboratory of Maurille Fournier [29,30]. Box H/ACA snoRNAs frequently consist of a two stem-loop structure with box H (ANANNA) located in the hinge region between the two stem-loops and box ACA located downstream of the second stem-loop structure, 3 nucleotides from the 3′-end of the RNA (see Figure 1.2-A2). Bulge-loop regions within the stem-loop structures contain sequences that are complementary, or anti-sense, to rRNA and snRNA targets [19,31,32]. Base-pairing interactions occur on either side of the bulge-loop structure, leaving an unpaired residue in the target sequence. This nucleoside, a uridine, is isomerized to a Ψ residue (see Figure 1.2-B2). While most box H/ACA snoRNAs guide Ψ formation, one box H/ACA snoRNA, U17/snR30, participates in the endonucleolytic processing of the pre-rRNA transcript [33-35]. As with the box C/D snoRNAs, box H/ACA snoRNAs also associate with proteins to form a snoRNP complex. These proteins include Nhp2, Nop10, Gar1, and the pseudouridine synthase [36], Cbf5.

Figure 1.2 Secondary Structure and Protein Components of Box C/D and Box H/ACA snoRNAs and their Resulting Modified Nucleosides. A. Typical (yeast or vertebrate) box C/D (**A1**) and box H/ACA (**A2**) snoRNAs and their associated core proteins [27]. The snoRNAs are represented by black lines whereas the rRNA is illustrated in blue. The sequence of the conserved box elements, boxes C and D in the C/D snoRNA along with boxes H and ACA in the box H/ACA snoRNA, is shown. Regions of the snoRNAs that are complementary, or anti-sense, to the rRNA target, are represented by thick black lines, with snoRNA/rRNA base-pairing interactions shown by dashes. For box C/D snoRNAs, the fifth base-paired nucleotide (indicated by red dashes) upstream of the D (or D′) box in the target rRNA is modified by methylation (represented by a red star) at the $O^{2'}$ position of the ribose. **B.** An unmodified ribonucleoside (uridine in the example shown) and its corresponding $O^{2'}$-methylated constituent (indicated in red) are show below (**B1**). For box H/ACA snoRNAs, sequences on both sides of each bulge-loop structure base-pair to a rRNA target, leaving two unpaired residues, 3′-NU-5′, in the target sequence. The uridine nucleoside is isomerized to a Ψ residue (shown in red). The uracil base in uridine (shown below, **B**) is linked through its N-1 position (in green) to the C-1′ position of the ribose. Pseudouridine (**B2**) is formed by cleavage of the N1-C1′ glycosyl bond separating the base from its ribose partner. The freed uracil base is then rotated 180° through a diagonal N3-C6 axis (circular red arrow in **B**) and reattached to the ribose by a C5-C1′ glycosyl bond (C5 is in blue)[28].


A.
D´
C´
Fibrillarin
Nop58p
Nop56p
Snu13p
3´
RUGAUGA
CUGA
C box
D box
5´ 3´
rRNA
5´
NΨA
Cbf5p
Nhp2p
Nop10p
Gar1p
ANANNA
H box
ACA
NNN 3´
ACA box
1. C/D snoRNP
2. H/ACA snoRNP


B.
HC
NH
N
O
HO
OH
OH
Uridine (U)
1
2
O-CH3
O2' methyluridine (Um)
HN
C
Pseudouridine (Ψ)

The distribution and role of Ψ residues has been reviewed by Charette and Gray [28]. As with Nm residues, Ψ residues are also distributed in functionally important regions in tRNAs, snRNAs [37], snoRNAs [38], and rRNAs [22] (though eukaryotic box H/ACA snoRNAs only guide Ψ formation in rRNAs and at select positions in snRNAs). Loss of any one Ψ residue results in minimal (if any) phenotypic effect. However, loss of certain rRNA Ψ residue(s) appears to perturb ribosome structure and activity [23,39,40]. Unlike uridine, Ψ residues can coordinate a structural water molecule between the additional hydrogen bond donor at N1-H and the phosphate backbone [28,41]. This results in a subtle but significant rigidifying influence [42] on the nearby sugar-phosphate backbone and also enhances base stacking. Pseudouridine residues often act in concert to exert a cooperative effect, thus stabilizing the structure of the RNA beyond the site of modification.

SnoRNA genes have been found in a variety of unusual genomic organizations. Some snoRNA genes, such as those encoding U3 snoRNA, exist as conventional stand-alone (independent) genes with typical promoter and termination elements. Unexpectedly, in animals and plants (and to a lesser extent yeast), snoRNA genes are frequently encoded within intron sequences, either as a single snoRNA or as a snoRNA cluster [27]. In such cases, the intron is removed from the pre-mRNA transcript and following lariat debranching, the snoRNA(s) is/are liberated from the intron sequence. Other snoRNAs, as in trypanosomatid protozoa, are encoded in snoRNA gene clusters and expressed as polycistronic transcripts [43-45]. One example of a snoRNA gene overlapping a protein-coding gene has also been described in yeast [46].

1.4: An Introduction to U3 snoRNA

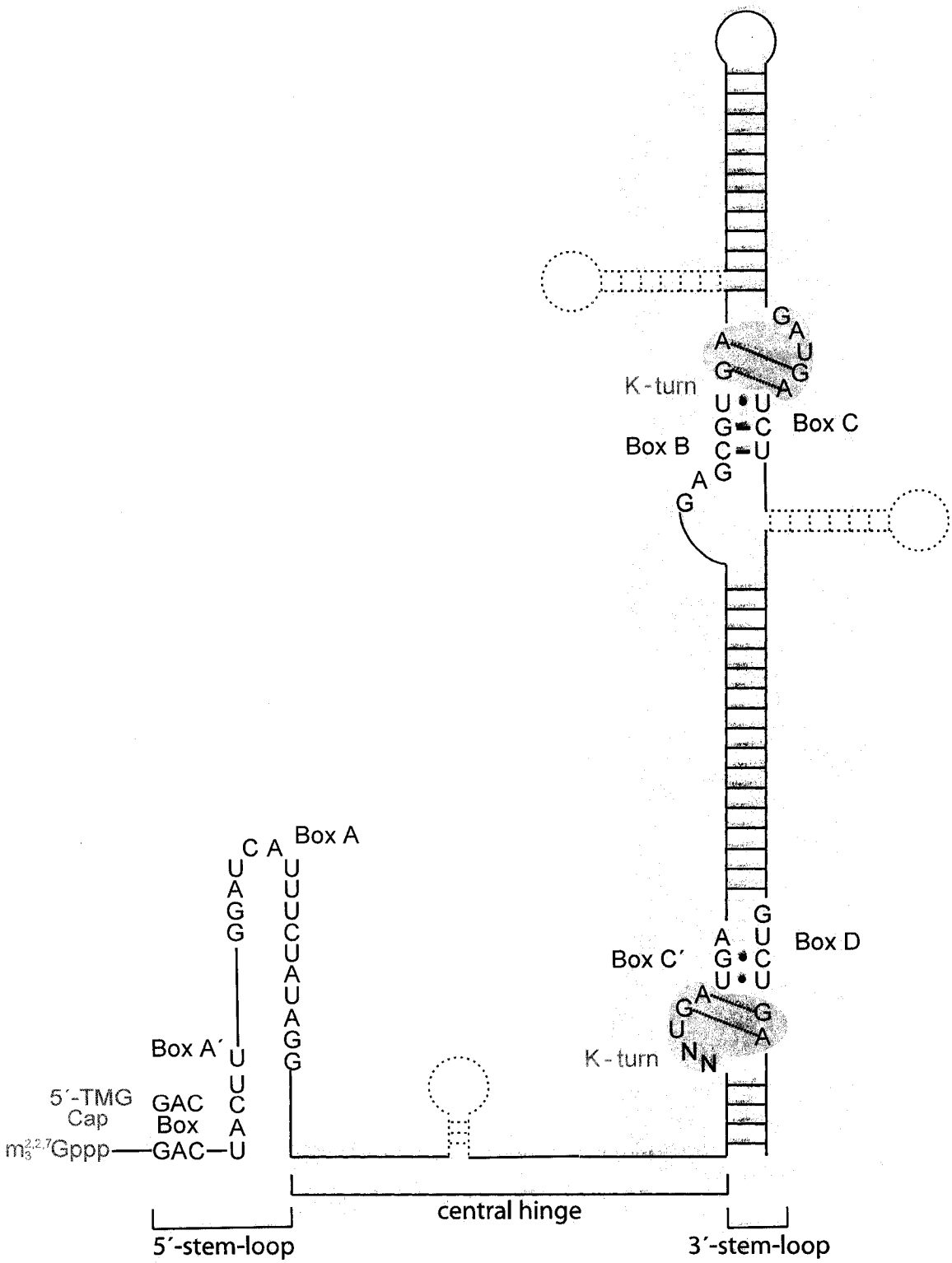
U3 was the first snoRNA identified, and has since become the most extensively studied snoRNA. It was discovered in 1968 by James L. Hodnett and Harris Busch [47] in the course of investigations on the small RNAs of animal cells. Initially, these investigators identified three uridine-rich small RNAs. Based on their electrophoretic mobility, the smallest was named U1, the next largest U2, and the largest U3. The nucleoside compositions of U1 and U2 exhibited a slightly larger proportion of uridine residues, 27% to 29%, compared to that of other small RNAs, such as tRNAs (typically ~20% U). It is now

known that this elevated frequency is partly attributable to the uridine-rich Sm-binding sequence (AAUUUUUGG). This sequence, present in U1, U2, U4, and U5 snRNAs, serves as a binding site for Sm proteins, a group of structural snRNA-specific proteins. It was also noted that the nucleoside composition of U3 consisted of a nearly equal distribution of all four ribonucleosides, consistent with our current-day knowledge that U3 snoRNA does not possess an Sm-binding sequence. Nonetheless, by virtue of its experimental association with the U-rich U1 and U2 RNAs, U3 RNA maintained its "U" designation. Shortly thereafter, U1, U2, and U3 were found to be predominantly associated with the nucleus and thus the term small nuclear RNA (snRNA) was coined. However, U3 was subsequently found to be enriched in the nucleolus (whereas U1 and U2 were not) and to co-purify with ribosomes. Thus, it was proposed that U1, U2, and U3 participated in various aspects of RNA processing and maturation. However, their roles in these pathways remained enigmatic for some years after their discovery.

In the early 1970s, U3 RNA was implicated in pre-rRNA processing by virtue of its conclusive localization in the nucleolus by cell fractionation studies and later by immunolocalization microscopy. Subsequently, regions of complementarity between U3 snoRNA and the pre-rRNA/rRNA transcript were identified and confirmed by chemical cross-linking. Knockout of U3 snoRNA genes was found to impede pre-rRNA processing and result in the accumulation of unprocessed transcripts and concomitant cell lethality.

U3 snoRNA consists of 5′- and 3′-domains separated by a hinge region (see Figure 1.3) [48]. The 5′-domain contains the conserved box GAC, A′, and A sequence elements. The sequence of these conserved box elements is complementary to regions of the 5′-ETS of the pre-rRNA and to the 5′-end of the 18S rRNA [49-52]. Base-pairing interactions between the pre-rRNA and its complementary regions in U3 snoRNA guide, by a complex and poorly understood mechanism that includes the participation of U14 [53-55] and U17/snR30 [33-35] snoRNAs, the multiple sequence-specific pre-rRNA cleavage events that eventually liberate the mature 5′-end of the 18S rRNA [49-51,56-59]. This region of U3 snoRNA can be modeled as a stable stem-loop structure. However, in an active U3 snoRNA

Figure 1.3 Conserved Sequence Elements and Secondary Structure Features of U3 snoRNA. U3 snoRNA consists of three domains: a 5′-stem-loop (shown here unpaired, in an "open" conformation), a central hinge, and an extended 3′-stem-loop. The conserved backbone structure is represented by a solid line, whereas non-conserved hairpins are represented by dashed lines. Conventional Watson-Crick base-pairing interactions are depicted as dashes (–) and G/U pairings are identified by dots (•). The 5′-TMG cap structure ($m_3^{2,2,7}G$) is shown in orange. The conserved box elements, boxes GAC, A′, A, C′, B, C, and D are shown in green, based on the *Saccharomyces cerevisiae* sequence. Two K-turns, formed by boxes C′/D and B/C in the 3′-stem-loop, are highlighted in grey, with the two noncanonical sheared G•A and A•G pairs indicated by small dashed lines. Many proteins, shown in beige, are known to bind to sequence elements in U3 snoRNA, particularly in the 3′-stem-loop domain [64]. Multiple protein-protein interactions are also known to occur [65,66].


K-turn
Box C
Box B
Box D
Box C′
K-turn
Box A
Box A′
5′-TMG
Cap
GAC
Box
m3 2,2,7Gppp—GAC—U
central hinge
5′-stem-loop
3′-stem-loop

molecule, this region is thought to adopt an open conformation in which U3 snoRNA-pre-rRNA/rRNA intermolecular interactions displace the intramolecular base-pairing interactions of a closed, inactive U3 snoRNA molecule. Experimental results from *in vivo* chemical and enzymatic structure probing experiments [51,60] support this model.

The central hinge region of U3 snoRNA consists of an unstructured region with a small, centrally located stem-loop structure (in some organisms) (see Figure 1.3). This region is thought to provide proper spacing between the RNA-binding 5′-domain and the protein-complexed 3′-domain of U3 snoRNA [48,52]. In addition, base-pairing interactions also occur between the central hinge region of U3 snoRNA and the 5′-ETS of the pre-rRNA [52,61-63].

The 3′-domain of U3 snoRNA, important in protein binding, RNA stability, and nuclear retention, contains the box C/D sequence elements associated with this snoRNA family (see Figure 1.3) [48,51]. The domain consists of an extended stem-loop structure punctuated by a number of bulge-loop elements. Although they are not linearly adjacent in sequence, the box C′ and D elements, as with the box B and C elements, are juxtaposed in the secondary structure context. Furthermore, the box elements occupy the single-stranded regions of the bulge-loop structures. As with other box C/D snoRNAs, this region of U3 also contains two kink-turn (K-turn) RNA motifs within the juxtaposed box C′/D and B/C elements [67]. Thus, the 3′-domain of U3 snoRNA could be regarded as a (relatively) structured box C/D snoRNA. In some organisms, additional non-conserved and non-essential stem-loop structures are found in this domain [48].

U3 snoRNA associates with at least 40 proteins (the four core box C/D-binding proteins plus the U-three-associated proteins, Utps) to form a large ribonucleoprotein complex [68-71]. This massive 80S complex of ~2.2 MDa has been named the SSU processome, a term coined by Dr. Maurille Fournier [48] and popularized by Dr. Susan Baserga [68]. Electron microscopy studies suggest that the SSU processome corresponds to the terminal knobs observed on the 5′-ends of growing pre-rRNAs in Miller-spread "Christmas trees" [68,72].

A co-transcriptional rRNA processing and assembly model has been proposed based on previous knowledge of U3 snoRNA function and the recent identification of the protein components of the SSU processome [71,73-77]. In this model, rDNA transcription is coupled to pre-rRNA processing. After RNAP I initiates transcription, a large number of processing factors, such as snoRNPs and a subset of the Utps, are recruited to the pre-rRNA. Nucleoside modifications are thought to occur very early in pre-rRNA processing [78-80]. The SSU processome is recruited to the growing pre-rRNA and endonucleolytic cleavages occur before the completion of transcription. Frequently, the multiple base-pairing interactions between U3 snoRNA and the pre-rRNA/rRNA transcript cannot occur simultaneously. Multiple intra- and inter-molecular rearrangements and displacements must occur in the formation of the mature 5′-end of the 18S rRNA [81]. Thus, U3 snoRNA has been postulated to exert a chaperone-like activity in the co-transcriptional folding of the 18S rRNA [56,81] by preventing incorrect, short-range, intra-domain base-pairings, while favoring correct, long-range, inter-domain interactions [82].

Unlike the varied and unusual gene organizations described for other snoRNAs, all known U3 snoRNA genes are encoded as standard genes, with upstream promoter and downstream termination sequences. The genomic organization of U3 snoRNA genes has mainly been investigated in animals, yeast, plants, and a limited number of protists (see Table 1.1 and Figure 1.4). This limitation is perhaps attributable to low overall sequence conservation (especially across broad phylogenetic distances) among U3 snoRNAs and the resultant technical difficulties when using similarity-searching algorithms such as BLAST.

U3 snoRNA is encoded as a multi-copy gene in most vertebrate genomes. In humans, U3 snoRNA genes are solitary, found on large, ~45-kbp, nearly identical inverted repeats clustered within a ~200 kbp chromosomal locus [83] (see Table 1.1). In mouse, most U3 snoRNA genes are solitary, and arranged in a similar organization to that seen in humans [84]. In *Xenopus laevis*, U3 snoRNA genes are multi-copy, with both clustered (within 18.7 kbp) and dispersed genomic organizations [85]. U3 snoRNA pseudogenes have been identified in the humans [86-88], rat [89,90], and mouse [91,92]. They are postulated to have arisen by both DNA- and RNA-mediated pseudogene formation events.

In plants, U3 snoRNA is found as a multi-copy gene, and both clustered and dispersed organizations have been identified. In *Lycopersicon* (tomato), a U3 snoRNA gene is clustered with four neighboring U3 snoRNA pseudogenes [93]. Similarly, in *Solanum* (potato), U3 snoRNA is encoded with other snRNA genes and pseudogenes and with U3 snoRNA pseudogenes (my re-annalysis of [94]) (see Figure 1.4).

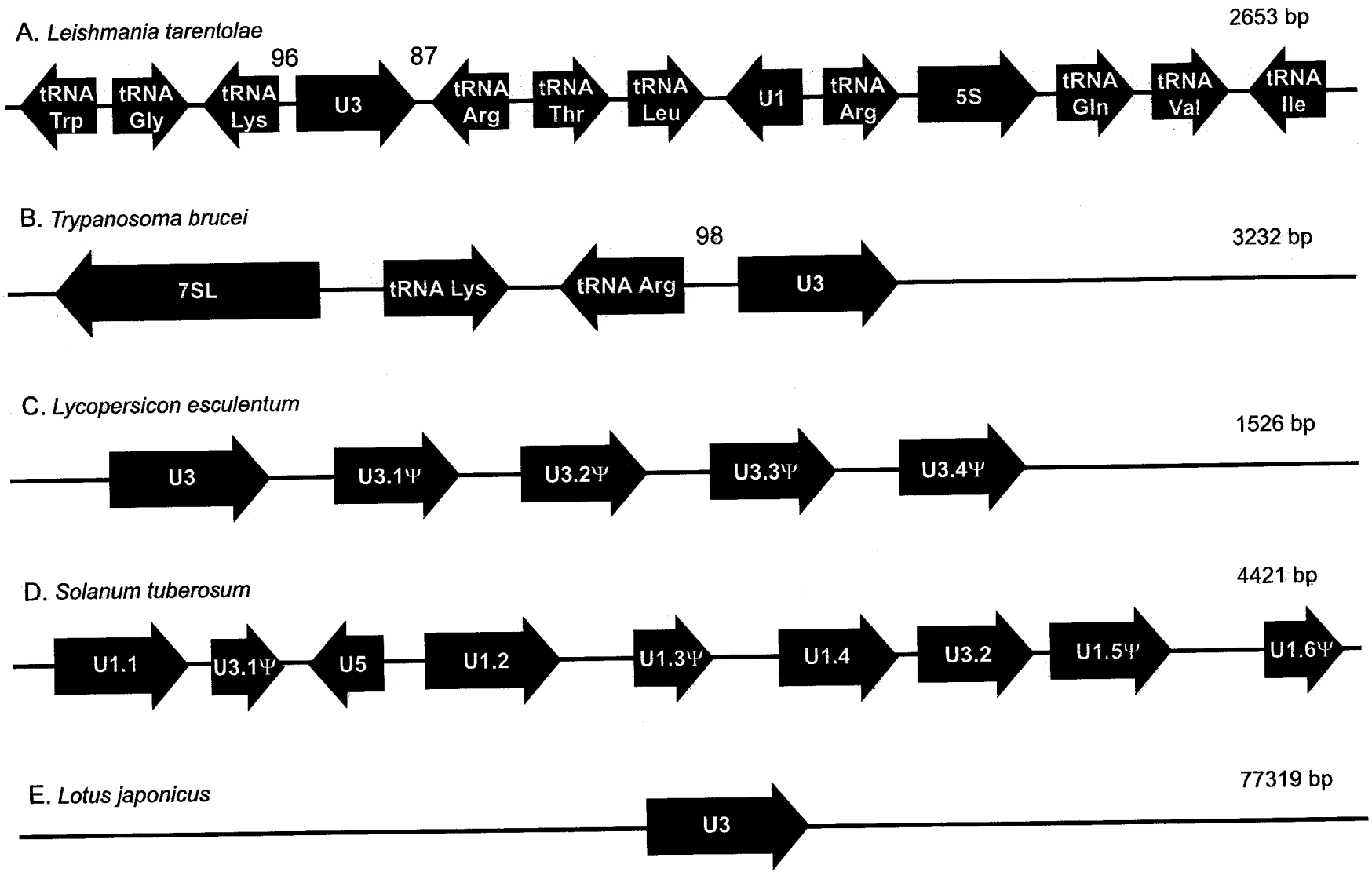
In protists, the *Dictyostelium* genome encodes five solitary and dispersed U3 snoRNA genes [95]. Similarly, four solitary and dispersed U3 snoRNA genes have been identified in *Tetrahymena* [96,97].

The U3 snoRNA genes of trypanosomatids have been extensively studied. In all known cases, U3 snoRNA is encoded by a single-copy gene, with many other small RNA genes (tRNAs, snRNAs, and 5S rRNA) in the genomic neighborhood (see Figure 1.4) [98-100].

Table 1.1 Organization and Gene Copy Number of U3 snoRNA Genes in Representative Organisms.

Organism	U3 gene copy #	Notes	Ref
Human	5-10 per haploid genome	Solitary, in ~45-kbp nearly identical inverted repeats, clustered within a ~200-kbp region of chromosome 17p11.2	[83]
Mouse	4	Solitary, tandemly clustered	[84]
Xenopus	14 - 20	Solitary, clustered and dispersed	[85]
Dictyostelium	5	Solitary and dispersed	[95]
Tetrahymena	4	Solitary and dispersed	[96,97]

Figure 1.4 Organization of U3 snoRNA Genes in Representative Organisms. The U3 snoRNA genes are shown in red and all other genes are shown in black. Pseudo-genes are indicated by Ψ. Numbers shown between the various genes correspond to the size of the intergenic spacer separating the two genes. Numbers on the right hand of the figure indicate the size, in nucleotides, of the relevant genomic region. Figure not drawn to scale. *Lycopersicon esculentus* (tomato); *Solanum tuberosum* (potato).


A. Leishmania tarentolae
2653 bp
96
87
tRNA Trp
tRNA Gly
tRNA Lys
U3
tRNA Arg
tRNA Thr
tRNA Leu
U1
tRNA Arg
5S
tRNA Gln
tRNA Val
tRNA Ile
B. Trypanosoma brucei
98
3232 bp
7SL
tRNA Lys
tRNA Arg
U3
C. Lycopersicon esculentum
1526 bp
U3
U3.1Ψ
U3.2Ψ
U3.3Ψ
U3.4Ψ
D. Solanum tuberosum
4421 bp
U1.1
U3.1Ψ
U5
U1.2
U1.3Ψ
U1.4
U3.2
U1.5Ψ
U1.6Ψ
E. Lotus japonicus
77319 bp
U3

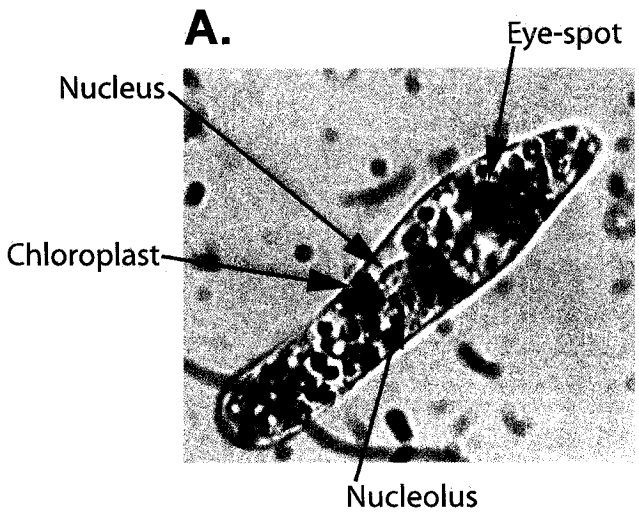
1.5: *Euglena gracilis*: an Interesting Organism with Many Unusual RNAs

Euglena gracilis is a free-living, flagellated, photosynthetic protist first described in 1830 by Ehrenberg. *Euglena* and its close relatives are found worldwide in stagnant ponds and brackish waters, where they frequently contribute to algal blooms. The cells are large and cigar-shaped, and contain an orange-red carotenoid-based [101] light-sensitive eye-spot from which *Euglena* derives its name: *eus* and *glêne* or "good" and "eye", respectively (see Figure 1.5A). *Euglena* and its 800 to 1000 close relatives form a phylogenetic grouping called the euglenids. *Euglena* and other euglenids are distant but specific relatives of the kinetoplastids (trypanosomes). Together, the euglenids and kinetoplastids form a larger phylogenetic grouping called Euglenozoa, which is a member of the Excavata (excavates) supergroup of protists [102] (see Figure 1.6).

The chloroplasts of *Euglena* have been the subject of intensive investigation. Terrestrial plants and most green algae are believed to have acquired their chloroplasts by a primary endosymbiotic event between a eukaryote and a photosynthetic bacterium [103,104]. In such organisms, the chloroplasts are surrounded by two membranes. However, *Euglena* (and select other protists) is thought to have acquired its chloroplasts through a secondary endosymbiotic event in which a non-photosynthetic kinetoplastid-like organism engulfed a green photosynthetic unicellular eukaryote [103-108]. In such circumstances, the host cell has been known to keep the engulfed photosynthetic eukaryote, presumably by virtue of a beneficial symbiotic relationship that eventually becomes fixed and irreversible [104,107,109,110]. However, in *Euglena*, only the chloroplasts of the engulfed photosynthetic eukaryote remain; in consequence, the chloroplasts are surrounded by three membranes [105]. *Euglena*'s secondary endosymbiotic acquisition of the chloroplast has been further substantiated by phylogenetic analysis of chloroplast genes [111-114].

In addition to being a staple of introductory biology classes, certain aspects of *Euglena* biology have been the focus of intense research. In the early part of the last century, *Euglena* was considered a model organism for biochemical research on respiratory pathways, amino acid biosynthesis, and lipid metabolism. In the 1950's, the development of techniques to manipulate the chloroplasts of *Euglena* made it an organism of choice for photosynthesis research (see Materials and Methods, section 2.1). More recently, *Euglena* has been used in

Figure 1.5 Micrographs of *Euglena gracilis*. A. Light micrograph of *Euglena*. The eye-spot, chloroplast, nucleus, and nucleolus are each indicated by an arrow. The eye-spot is seen in the anterior region of the cell, whereas the chloroplasts are uniformly distributed throughout the cell. Unlike most organisms, the nucleolus of *Euglena* is visible by light microscopy, even in the absence of nucleolar staining. The structure of the nucleus/nucleolus frequently resembles that of a woman's wide rimed hat, with the nucleolus adopting a raised position corresponding to the hat's head position, surrounded by the nucleus corresponding to the hat's rim. Although not seen in this light micrograph, structures seen at higher magnification in the nucleus of unstained *Euglena* cells are believed to correspond to permanently condensed chromosomes. Photo, 1000X magnification. **B.** Transmission electron micrograph of a thin section through the nucleus of *Euglena*. The nuclear membrane and the nucleolus are each indicated by an arrow, as are the very prominent and permanently condensed chromosomes. As with other organisms, the nucleolus is near the periphery of the nuclear membrane. Unlike what is seen in other organisms, the *Euglena* nucleolus is a multi-lobed structure with deep invaginations between the multiple lobes (seen in this plane of view as three discrete structures that are assumed to be part of the same nucleolus). Photo, 24,000X magnification, is courtesy of Dr. David F. Spencer (Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada).


A.
Eye-spot
Nucleus
Chloroplast
Nucleolus

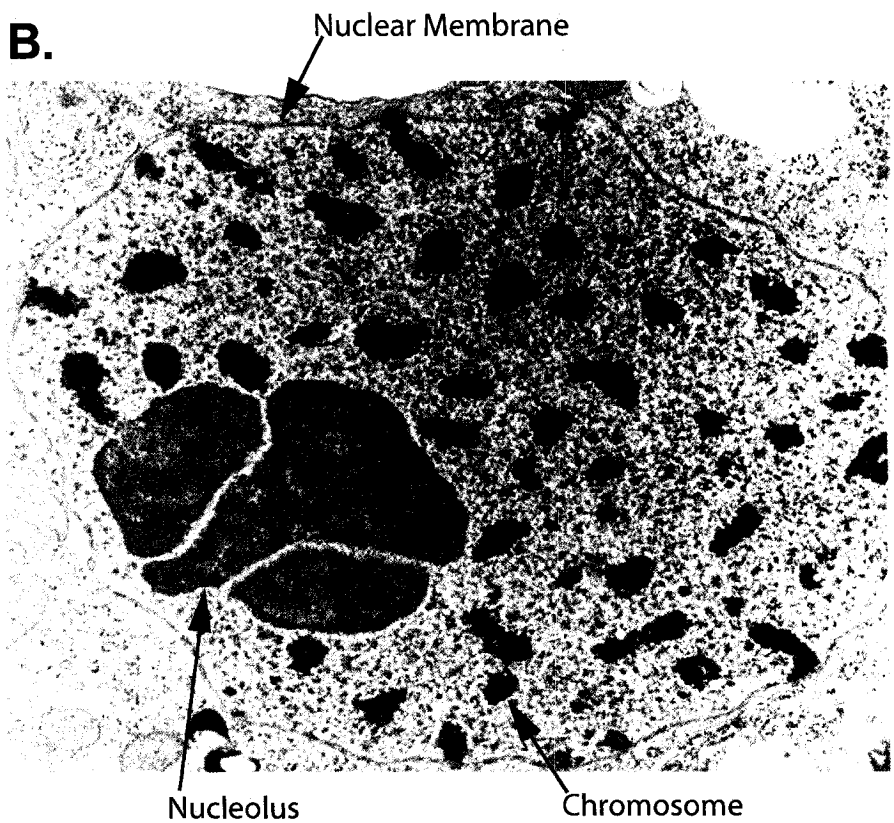
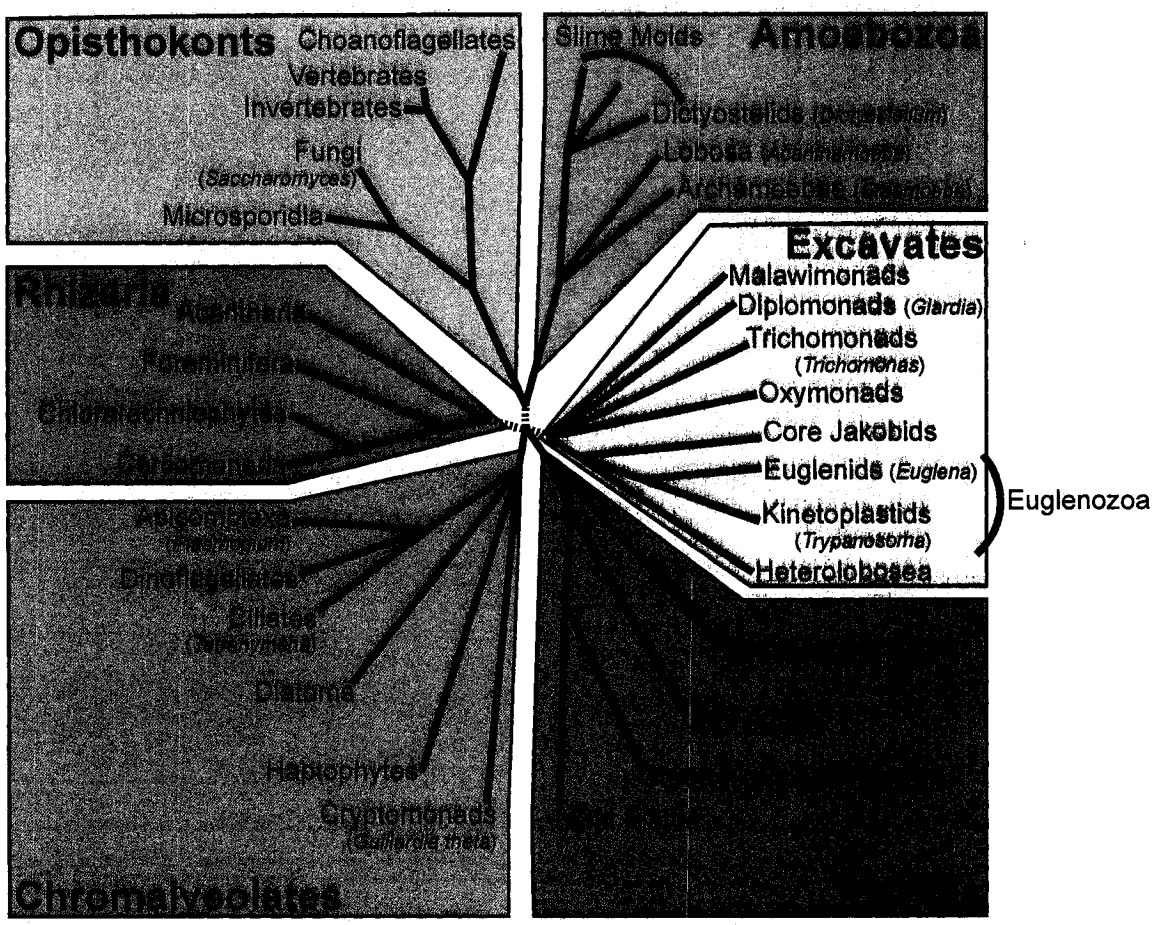

B.
Nuclear Membrane
Nucleolus
Chromosome

Figure 1.6 Consensus Phylogenetic Tree of Eukaryotic Organisms. This tree, modified from Keeling *et al.* [102], is derived from various types of phylogenetic and evolutionary data. The six main phylogenetic "supergroups" are shown: *i*) the Opisthokonts (animals and fungi), *ii*) the Rhizaria (an assemblage of protists, frequently with hard shells, found in fossils), *iii*) the Chromalveolates (an assemblage of algae, including kelps, diatoms, and dinoflagellates, and protozoa, including apicomplexans and ciliates), *iv*) the Plantae (plants and red and green algae), *v*) the Excavates (an assemblage of protists including kinetoplastids (trypanosomes and relatives), euglenids (including *Euglena*), trichomonads, and diplomonads), and *vi*) the Amoebozoa (an assemblage of protists believed by some to be specifically related to the Opisthokonts (and thus forming with the latter the Unikonts grouping), consisting of amoeboid protozoa such as *Dictyostelium*). Select phylogenetic groupings are shown along with representative organisms. Well-supported phylogenetic groupings are shown by a solid line. Uncertain phylogenetic relationships are left unresolved (*i.e.*, unclear branching order). Dashed lines at the base of the eukaryotic tree indicate tentative relationships between the supergroups.


Opisthokonts
Choanoflagellates
Vertebrates
Invertebrates
Fungi
(Saccharomyces)
Microsporidia
Slime Molds
Amoebozoa
Dictyostelids (Dictyostelium)
Lobosa (Acanthamoeba)
Archamoebae (Entamoeba)
Excavates
Malawimonads
Diplomonads (Giardia)
Trichomonads
(Trichomonas)
Oxymonads
Core Jakobids
Euglenids (Euglena)
Kinetoplastids
(Trypanosoma)
Heterolobosea
Euglenozoa
Rhizaria
Acantharia
Foraminifera
Chlorarachniophytes
Cercozoa
Radiolaria
Alveolates
(Plasmodium)
Dinoflagellates
Ciliates
(Paramecium)
Diatoms
Haptophytes
Cryptomonads
(Guillardia theta)
Chromalveolates

microgravity research [115] and its tolerance of heavy metals (as exemplified by its nutritional requirements [116], see Materials and Methods, section 2.2) has made it a focus of research on heavy metal bioaccumulation and environmental decontamination [117-119].

More germane to this study, *Euglena*, other euglenids and kinetoplastids have garnered considerable interest due to the many, very unusual aspects of their RNA molecular biology. Each of these aspects will be briefly introduced.

The chloroplast genome of *Euglena* was one of the first plastid genomes to be completely sequenced [120]. Analysis of this circular 143-kbp genome revealed the presence of 155 Group II and Group III introns [120,121], significantly more than in any other known plastid genome. Group III introns are believed to represent streamlined Group II introns. Among these introns are 15 twintrons, in which one or more Group II introns are inserted within a Group II intron. The presence, phylogenetic distribution, and mechanism of excision of the chloroplast introns of *Euglena* has been the focus of much research, the results of which have been used to argue in support of the late-origin-of-introns theory [121].

The *Euglena*-trypanosomatid phylogenetic grouping (Euglenozoa) is best known for its unusual intron processing requirements: *trans*-splicing [122-125] in addition to conventional *cis*-splicing events are both present [125-127]. In *trans*-splicing, a short spliced leader RNA (SL RNA), consisting of an exon at its 5′-end and intron at its 3′-end, undergoes *trans*-splicing with intronic sequence at the 5′-end of a pre-mRNA, in a spliceosome-dependent mechanism [123-125,128]. Consequently, the 5′-exon sequence ("spliced leader") of the SL RNA becomes the first exon of the mature mRNA, whereas intronic sequences at the 3′-end of the SL RNA and 5′-end of the pre-mRNA are excised. In *Euglena*, an additional, non-conventional form of *cis*-splicing is also known to take place [129]. In this case, non-canonical introns adopt extensive, stable secondary structures that bring the 5′- and 3′-ends of the intron into close proximity. The splicing mechanism for these non-conventional introns is unknown.

The ribosomes of *Euglena* and of trypanosomes are also highly unusual. In most organisms, hundreds to thousands of repeating copies of the rDNA operon are tandemly arrayed on one or more chromosomes in a head-to-tail configuration. In *Euglena*, a single copy of the rDNA operon is encoded by an extrachromosomal circular plasmid (see Figure

1.7), present in high copy number and autonomously replicating [63,130-135]. Few, if any, integrated chromosomal rDNA genes are evident [132,135]. While this scenario is unusual, it is not unique to *Euglena*. Similar arrangements have also been identified in other protists such as *Paramecium tetraurelia* [136], *Dictyostelium discoideum* [137], *Physarum polycephalum* [138], *Tetrahymena pyriformis* [139], *Entamoeba histolytica* [140-142], and *Naegleria gruberi* [143,144]. Furthermore, a number of additional, novel internal transcribed spacer (ITS) sequences in the *Euglena* 28S rDNA are excised during pre-rRNA transcription and processing (see Figure 1.7) [145]. Thus, a total of 13 ITS sequences (ITS2 plus 12 novel ITSs) are excised from the pre-rRNA transcript of the *Euglena* large subunit (LSU) rRNA gene [145]. Consequently, the mature *Euglena* LSU rRNA (usually consisting of two stable species, 5.8S plus 28S) is composed of 14 rRNA pieces (5.8S plus 13 "fragments") [146]. The *Euglena* 18S rRNA is not fragmented. A similar, but less extensive fragmentation in trypanosomes yields seven fragments (5.8S plus six fragments) [147-149]. Other unrelated organisms also encode fragmented LSU rRNAs, including *Acanthamoeba castellanii* (3 pieces) [150], *Tetrahymena thermophila* and *T. pyriformis* (3 pieces) [151], *Prorocentrum micans* and other dinoflagellates (3 pieces) [152,153], and *Drosophila melanogaster* (4 pieces) [154,155]. Early work on *Euglena* cytoplasmic ribosomes had come to the conclusion that the fragmented LSU rRNAs were degradation products [156,157]. However, the presence of a 5′-phosphate and 3′-hydroxyl on the ends of the LSU rRNA fragments confirms, along with other types of experimental verification, that they represent mature rRNA fragments and not degradation products [146]. (Most degradative nucleases produce products with 5′-hydroxyl and 3′-phosphate termini.) Despite the fragmentation of the *Euglena* LSU rRNA, the most extensive known to date, the rRNA pieces associate in *trans* [158] to form fully functional ribosomes [156,158-161]. Furthermore, ongoing mapping of Nm and Ψ positions in the *Euglena* ribosome (unpublished data, Dr. Murray N. Schnare, Department of Biochemistry and Molecular Biology, Dalhousie University) suggests that it is the most highly modified rRNA of any organism examined to date [11].

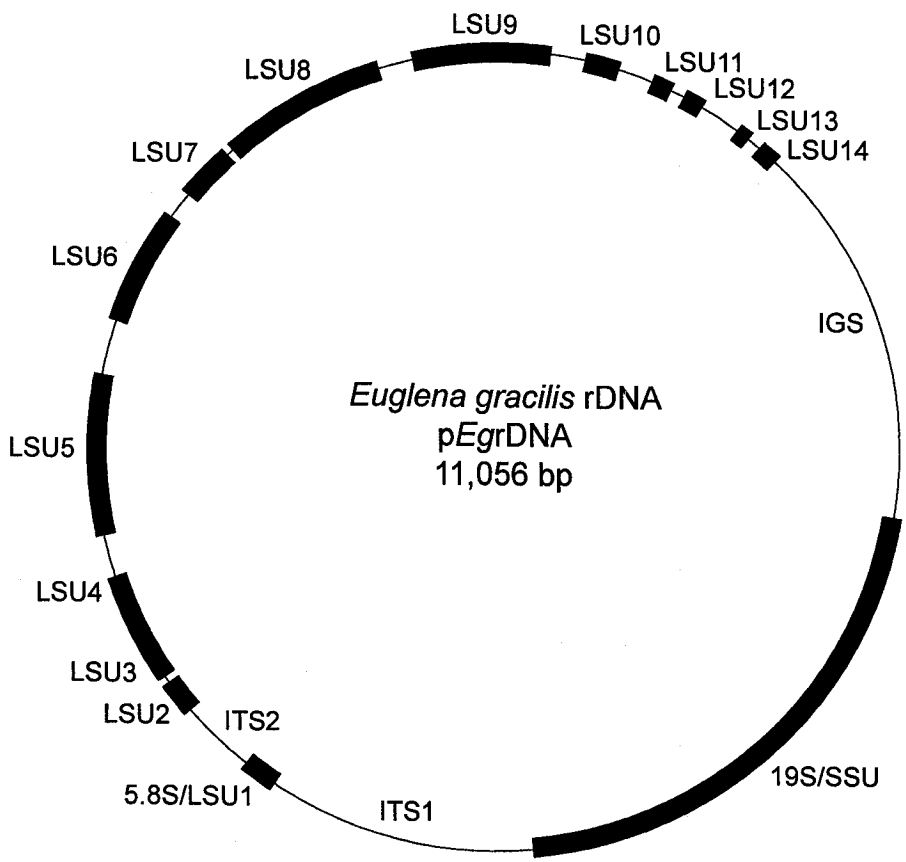


Figure 1.7 The Extrachromosomal Circular rDNA Plasmid of *Euglena*, p*Eg*rDNA. The 11,056-bp plasmid encodes one complete copy of the rDNA operon. The mature rRNA components are indicated by boxes. The first fragment of the ribosomal LSU of *Euglena*, LSU1, corresponds to the 5.8S rRNA. The LSU species 2 to 14 correspond to fragmented components of the 28S rRNA. The intergenic spacer (IGS), ITS1, ITS2, and the 12 novel *Euglena* ITSs, are represented by the thin line of the circle. The small 10-nt ITS separating LSU species 3 and 4 is too short to be represented here.

Trypanosomes are also known for RNA editing by uridine (U) insertion/deletion in mitochondrial mRNA transcripts. In this system, the DNA sequence of mitochondrially-encoded protein-coding genes is missing (or has additional) thymidine (T) residues, such that the protein-coding frame is disrupted. Using a system analogous to the snoRNA-guided modification of rRNA, complementary guide RNAs (gRNA), acting in *trans* by base pairing to the target region, guide enzyme complexes to the site of U insertion or deletion in the mRNA transcript [162,163]. Editing of the mRNA transcript by U insertion or deletion restores its proper protein-coding frame. This system is known to occur in many different kinetoplastids. A very limited exploration of *Euglena* mitochondrial genes has not identified a need for RNA editing by U insertion/deletion (personal communication, Dr. David F. Spencer, Dalhousie University) [164,165]. Thus, it appears that this form of RNA editing may be a kinetoplastid-specific phenomenon.

Information is scarce about the genetic features of *Euglena*, such as its genome size, chromosome number, and gene content. The *Euglena* haploid genome size has been estimated at ~1.36 X 10^9 bp [166,167]. This estimate is based on $C_0t_{½}$ values for non-repetitive DNA that are consistent with *Euglena* being a diploid organism [167]. However, this assumption may not be valid. Evidence is mounting for aneuploidy in many fungi and protists [168-170], including in the distant but specific relatives of *Euglena*, the kinetoplastids [170-175]. In such circumstances, no true and absolute ploidy value exists, as each chromosome may be present at a different copy number in each cell. Thus, it may be preferable to simply report that the *Euglena* nucleus contains 3 pg of DNA [176], which corresponds to 2.72 X 10^9 bp per cell [167]. The nuclear guanine plus cytosine (G+C) content has been estimated at ~48% [166]. *Euglena* is known to reproduce by division of the cell along its longitudinal plane. Since mating has never been observed, *Euglena* is assumed to be asexual.

The number of chromosomes in *Euglena* is also uncertain, but counts ranging from four [177] to 45 [101,178,179] have been reported. Furthermore, the chromosomes of *Euglena* have been described as permanently condensed (see Figure 1.5B): they have never been observed uncondensed. Indeed, structures believed to be chromosomes can be seen by light microscopy of unstained *Euglena* cells. The chromosomes of other protists, such as

those of many different dinoflagellate species, have also been described as permanently condensed (personal communication, Dr. Charles Delwiche, Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA)[180,181]. Thus, one must assume that this cytological description is not analogous to chromosome condensation (heterochromatin) in other organisms, which typically refers to a chromosomal conformation in which gene expression is repressed. Therefore, the state of permanently condensed chromosomes in these organisms in probably not a state of permanent heterochromatin, which would obviously result in cell inviability (personal communication, Dr. Charles Delwiche, University of Maryland). In these organisms, alternative chromosome features, such as a different protein composition of chromatin and/or different histone constituents or characteristics, may explain the observed cytological form.

The nuclear DNA of *Euglena* contains base J, β-D-glucosyl-hydroxymethyluracil, an unusual minor derivative of uracil. Base J occurs in place of select thymines and accounts for ~0.2% of the nucleotides in the *Euglena* genome [178]. The genomes of trypanosomatids also contain base J, as does the genome of the marine flagellate *Diplonema* (another Euglenozoa). However, this minor constituent is not present in the nuclear DNA of animals, plants, fungi, or other examined protists (unrelated to *Euglena*). Base J is thought to be analogous in function to 5-methylcytosine (5-MeC), which in animals and plants is involved in CpG island formation and transcriptional regulation.

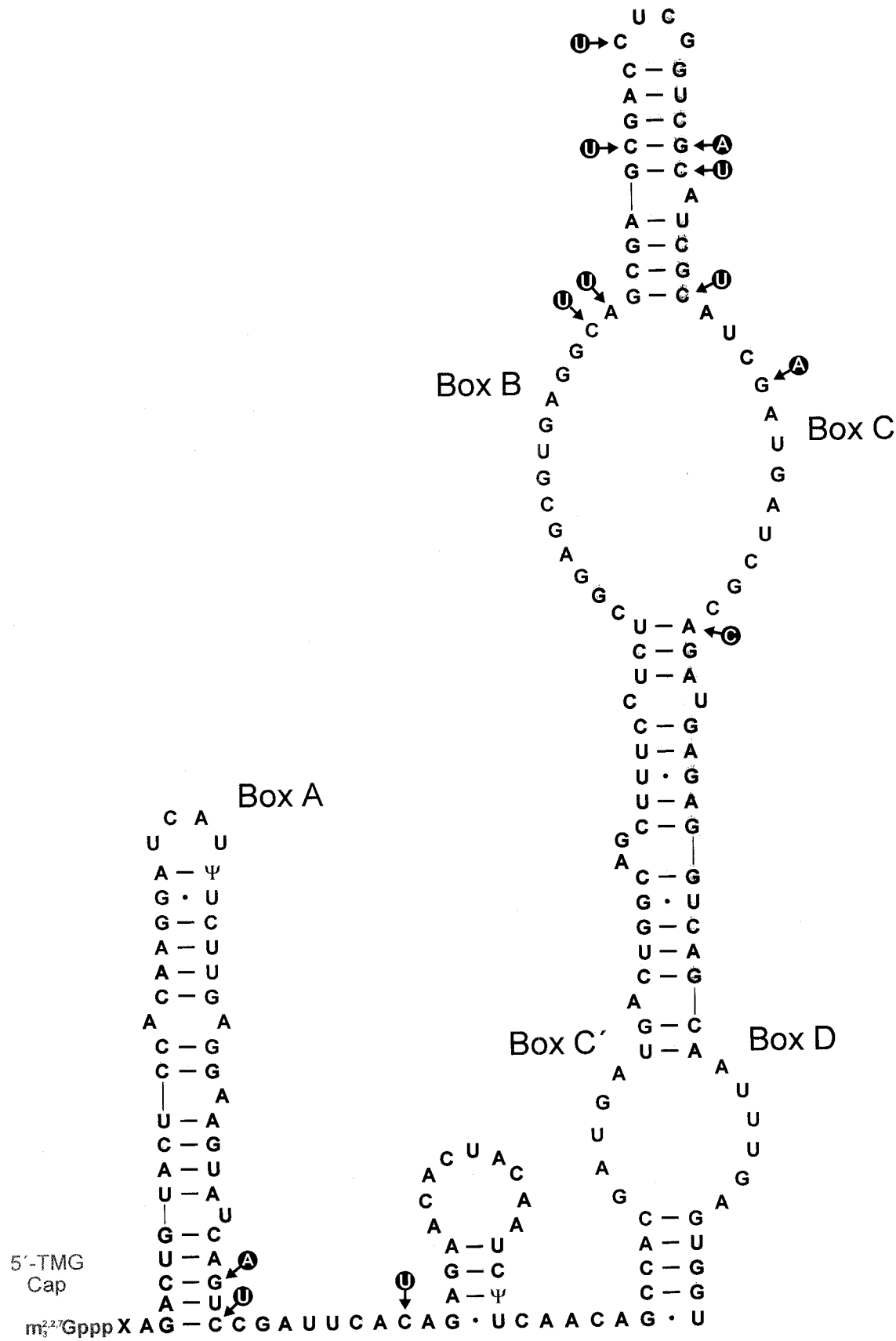
Unlike in trypanosomatids, the nuclear genome of *Euglena* is currently not amenable to genetic manipulation. Furthermore, the genome's many unusual features, such as its possible aneuploid composition, do not make it an attractive model organism for genetic research. No accounts have been published of attempts to genetically manipulate the nuclear genome of *Euglena*, leaving open the possibility that such work could be technically difficult. In contrast, a biolistic transformation method has been used successfully to genetically manipulate *Euglena* chloroplasts [182].

1.6: *Euglena* U3 snoRNA: What is Known and Unanswered Questions

The *Euglena* U3 snoRNA was identified by Greenwood *et al.* [38]. When total *Euglena* RNA is resolved by denaturing polyacrylamide gel electrophoresis and visualized by ethidium bromide staining (see Results, Figure 3.11), RNA species corresponding to the 18S rRNA and to the 14 fragments of the LSU rRNA are seen along with the 5S rRNA and tRNAs. Additional lower-abundance bands are also seen. RNA from three such bands was recovered from the 164- to 234-nt region of the gel, radioactively labeled, and re-electrophoresed in a higher-resolution denaturing polyacrylamide gel. The top and bottom bands resolved into separate homogeneous species. However, the central band resolved into four different bands of varying intensities. Enzymatic and chemical RNA sequencing of the central band revealed its identity as the 180-nt U3 snoRNA of *Euglena.* SnoRNAs are typically low-abundance transcripts; however, in all organisms, U3 snoRNA is known to be 100 to 1000 times more abundant than other snoRNAs, thus explaining its successful visualization by ethidium bromide staining of electrophoretically resolved total *Euglena* RNA, which allowed isolation of sufficient amounts of transcript for RNA sequencing.

Sequence analysis of *Euglena* U3 snoRNA identified all of the hallmark U3 snoRNA sequence features known at the time (see Figure 1.8). As with other U3 snoRNAs, the *Euglena* U3 snoRNA is believed to possess a post-transcriptionally added 5′-trimethylguanosine cap structure (5′-TMG;$m_3^{2,2,7}$G) based on the ability to immunoprecipitate the molecule with anti-TMG antibodies [38]. The 5′-region of U3 snoRNA, known to interact with the 5′-ETS of the pre-rRNA and with the 5′-region of the 18S rRNA, contains a GAC box (not identified at the time of discovery of the *Euglena* U3 snoRNA) followed by A′ (not identified at the time) and A boxes [48]. A short hinge region separates the 5′-rRNA-interacting and 3′-protein-binding domains of U3 snoRNA. The 3′-domain of U3 snoRNA contains the conserved sequence boxes C′, B, C, and D, known to interact with RNA-binding proteins. Two Ψ residues were identified in the *Euglena* U3 snoRNA: one is located in the conserved A box sequence whereas the other is found in the central hinge region. While modification status was not examined over the entirety of the molecule, no Nm residues were identified.

Figure 1.8 Sequence and Secondary Structure of *Euglena* U3 snoRNA as Proposed by Greenwood *et al.* In this original model of the *Euglena* U3 snoRNA [38], the 5′-TMG cap (in orange), along with boxes A, C′, B, C, and D (in green) are shown. The 5′-terminal nucleotide, shown as X had not been determined. Two Ψ residues, in box A and the hinge region, are shown in blue. Nucleotide heterogeneities, identified at the RNA level, are shown in red. Heterogeneities in the 3′-region of the 5′-stem-loop domain and in the 5′-region of the central hinge domain were tentatively identified.


Box A
Box B
Box C
Box C´
Box D
5´-TMG
Cap
m3 2,2,7Gppp X A G — C C G A U U C A C A G · U C A A C A G · U

The original secondary structure model proposed for *Euglena* U3 snoRNA [38] incorporated the structural features known at that time. The 5′-region, containing boxes GAC, A′, and A, was folded into a near perfect 17-bp stem-loop structure. A short 4-bp stem-loop structure was identified in the middle of the central hinge domain. The 3′-terminal domain containing boxes C′, B, C, and D was modeled into an extended stem-loop structure punctuated by bulge-loops. The conserved C′/D and B/C box motifs, though not close by in the linear sequence, were adjacently localized in the secondary structure, forming single-stranded regions corresponding to the bulge-loop structures.

Sequencing of the four bands obtained upon higher electrophoretic resolution of the U3 snoRNA band identified additional U3 snoRNA transcripts exhibiting a number of nucleotide sequence heterogeneities. The heterogeneities in the different U3 snoRNA transcripts slightly altered the structure of the molecules during electrophoresis, thus resulting in the resolution of the different species. In the secondary structure model of *Euglena* U3 snoRNA, two sequence heterogeneities were located at the 3′-position of the 5′-stem-loop structure, near but not involving sequences known to interact with the pre-rRNA. One heterogeneity was found in the 5′-region of the central hinge domain. The majority of heterogeneities (nine) were located in the extended 3′-stem-loop domain where they potentially result in minimal biologically relevant changes to the secondary structure of the region. Most heterogeneities occurred in single-stranded regions or resulted in transitions from canonical (G-C) to non-canonical (G•U) base pairings.

The identified sequence heterogeneities in the *Euglena* U3 snoRNA suggested that multiple U3 snoRNA genes were present in the *Euglena* genome. This scenario was unexpected since U3 snoRNA is a single-copy gene in all examined trypanosome species (see Figure 1.4) [98-100]. Multiple bands were obtained in Southern analysis of *Euglena* DNA hybridized with a *Euglena* U3 snoRNA gene probe. This result further supported the inference that U3 snoRNA may be encoded as a multi-copy gene family in *Euglena gracilis* [38]. However, in the Greenwood *et al.* study [38], *Euglena* U3 snoRNA genes were not examined. Thus, neither the number of U3 snoRNA genes in *Euglena* nor their genomic organization were elucidated.

In trypanosomes, the genomic neighborhood of the U3 snoRNA locus is particularly rich in genes for other small RNAs, such as tRNAs and snRNAs [99,100]. Furthermore, in the trypanosomatid genomic context, the proximity of the U3 snoRNA gene to a closely linked, upstream and divergently oriented tRNA gene is critical to the expression of the U3 snoRNA gene [183,184]. For these reasons, the intriguing and contrasting initial observations on *Euglena* U3 snoRNA genes seemed worthy of further investigation.

1.7: Project Rationale

Our current understanding of the many players, such as U3 snoRNA and its protein components, and the multiple intricate and highly coordinated events in rRNA processing and ribosome biogenesis stems largely from work done in the narrow phylogenetic grouping of animals and yeast. Therefore, our current knowledge of ribosome biogenesis may not be truly representative of the many potentially different strategies used by, nor of the universal commonalities that may exist in, phylogenetically disparate organisms. Thus, the well-known idiosyncratic features of the RNA biology of *Euglena gracilis*, in addition to its interesting phylogenetic placement, make this organism an ideal model system for the study of rRNA processing, ribosome biogenesis, and the many key players that participate in this highly complex assembly process. Moreover, this investigation may further our understanding of the evolution of the ribosome and of the many snoRNA and protein factors responsible for rRNA transcription, processing, modification, and assembly. Finally, such study will undoubtedly increase our appreciation of the many particularities of the *Euglena* genome.

In this thesis, *Euglena gracilis* U3 snoRNA genes and their genomic contexts are described. Unlike the situation in trypanosomatids, where U3 snoRNA is the product of a single copy gene, *Euglena* U3 snoRNA is encoded by a multi-copy gene family with at least 14 members. As found in trypanosomes, most *Euglena* U3 snoRNA genes are encoded near genes for other small RNAs, such as tRNAs and U5 snRNA. However, many significant differences between the *Euglena* and the trypanosomatid arrangements are present, and the genomic and gene expression consequences of these differences will be discussed.

CHAPTER 2: MATERIALS AND METHODS

2.1: Origin of the Bleached Variant of *Euglena gracilis* Strain Z

A streptomycin-bleached, aplastidic variant of *Euglena gracilis* strain Z [133] was used in this study. It was obtained from Dr. James R. Cook, formerly of the Department of Zoology at the University of Maine (Orono, Maine, USA). This strain was derived from the UCLA variety of *Euglena gracilis* strain Z by taking advantage of a phenomenon, first described in the late 1940s, known as "bleaching" [185]. Streptomycin interferes with translation in Bacteria by irreversibly binding to protein S12 of the 30S ribosomal SSU. In *Euglena*, the chloroplast-encoded ribosomes are inhibited by streptomycin, presumably owing to their bacterial ancestry and properties. Thus, in the presence of streptomycin, the translation of chloroplast-encoded mRNAs is inhibited. Over several generations, a gradual "bleaching" is observed as the organelle becomes unable to synthesize components of its replication, transcription, translation and photosynthesis machinery. Such a chloroplast-deficient, or aplastidic, strain is advantageous in that its vestigial chloroplasts contain a highly reduced genome present at a very low copy number [186], thus resulting in less contaminating chloroplast DNA in total DNA preparations.

2.2: Laboratory Cultivation of the Bleached Variant of *Euglena gracilis* Strain Z

The bleached variant of *Euglena gracilis* strain Z was grown at room temperature in liquid cultures of modified Cramer and Myers *Euglena gracilis* salt medium [116]. The solution was modified in the following ways: addition of a 30 mM ethanol carbon source [133,187,188] (owing to the bleached cell's inability to photosynthesize), $CoCl_2$ (1.3 mg/l) and Na_2MoO_4 (0.2 mg/l) in place of $Co(NO_3)_2$ and H_2MoO_4, respectively, and adjustment to pH 6.5 with phosphoric acid [189].

Euglena stock cultures were maintained by periodic passaging of 500 µl of culture into 5 ml of fresh medium. One liter (1 l) cultures were grown in large Fernbach flasks (2800 ml size) by inoculation with 2.5 ml of fresh, exponentially growing culture, and kept aerated by gentle agitation on an orbital shaker. Nucleic acids were isolated after 4 to 5 days of growth, once cells reached mid- to late-log phase. At this growth stage, the culture is

white and creamy with a slight yellow-green hue, has an OD_{600} between 0.8 and 1.0 units, and yields an approximately 4.5-g cell pellet. Under the microscope, nearly all cells are active, mobile, and cigar-shaped and not stationary, rounded, or bloated.

2.3: Extraction of Total DNA from *Euglena gracilis*

Total *Euglena* DNA was prepared using the classic Marmur detergent/chloroform/ phenol extraction procedure [190] modified for use with *Euglena* by Dr. David F. Spencer (Dalhousie University).

Euglena cultures were harvested by subjecting cells to a cold shock prior to centrifugation. Since *Euglena* is a flagellated protist, this procedure decreases cell swimming and metaboly, a *Euglena*-specific movement, and results in a more compact cell pellet. Accordingly, the *Euglena* culture was decanted into six pre-chilled 250-ml centrifuge bottles and kept on ice for approximately 30 min. The cells were then harvested in an IEC B-22M centrifuge at 3,520*g* (5,000 rpm in a Thermo IEC 877 fixed-angle rotor) for 20 min at 4°C. They were subsequently washed in a total of 150 ml of cold Extraction Wash Buffer [EWB = 25 mM EDTA-Tris (pH 8.5): an EDTA solution titrated to pH 8.5 with solid Tris base]. Pooled cells were recovered by centrifugation at 3,520*g* for 10 min and resuspended to a final volume of 10 ml with room temperature EWB.

Cells were lysed by the addition of 2.5 ml of 25% SDS. The solution was mixed by repeated gentle inversion of the tube until homogeneous (~2 to 3 min), adjusted to 25 ml by the addition of 12.5 ml of room temperature EWB and mixed again until homogeneous.

Deproteination of nucleic acids was initiated by adding 4 ml of 8 M sodium perchlorate to the cell lysate. The solution was mixed gently until homogeneous. Nucleic acids were extracted by adding 20 ml of chloroform:isoamyl alcohol (24:1) to the cell lysate. The solution was mixed gently until emulsified and transferred to 30-ml Corex® centrifuge tubes. The aqueous phase was resolved by room temperature centrifugation at 11,180*g* (10,000 rpm in a Thermo IEC 875 fixed-angle rotor) for 10 min, removed and re-extracted. Nucleic acids were precipitated with an equal volume (~14.5 ml) of room temperature isopropanol. High-molecular-weight DNA was preferentially enriched by spooling onto a

glass rod. The DNA was washed in 10 ml of 80% ethanol, collected by centrifugation at 11,180*g* for 5 min at 4°C, briefly dried and dissolved in 10 ml of TE [10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA].

The DNA was further deproteinized by phenol extraction. Briefly, 0.1 volumes of 3 M NaOAc (1 ml) and an equal volume (10 ml) of Tris-buffered phenol-cresol [500 g of phenol, 55 ml of dH_2O, 70 ml of *m*-cresol, and 0.5 g of 8-hydroxyquinoline [191], equilibrated successively with 500 mM, 200 mM, and 10 mM Tris-HCl (pH 8.0)] were added to the aqueous DNA solution. The mixture was gently agitated until emulsified and the aqueous and phenol phases were resolved by centrifugation at 11,180*g* for 10 min. The aqueous phase was removed and the phenol extraction was repeated (~5 to 6 times) until no material was left at the interface.

DNA was precipitated from the aqueous phase with 2.5 volumes of 95% ethanol (at -20°C) and collected by centrifugation at 11,180*g* for 15 min. The DNA was washed with 10 ml of 80% ethanol, recovered by centrifugation at 11,180*g* for 5 min, briefly dried and dissolved in 2.5 ml of TE. The DNA was then re-precipitated and high-molecular-weight DNA was preferentially enriched by spooling onto a glass rod (when possible). The DNA was washed with 80% ethanol, briefly dried, and redissolved in 1.0 ml of TE.

Contaminating RNA in the DNA preparation was removed by RNase treatment followed by PEG precipitation [192]. Briefly, 200 µl (120 U) of heat-treated RNase A was added to the DNA sample and the reaction mixture was incubated at 37°C for 30 min. High-molecular-weight DNA was precipitated by adjusting the solution to final concentrations of 0.5 M NaCl (172 µl of 5 M NaCl) and 10% PEG (344 µl of 50% PEG), and incubating on ice for 30 min. The PEG-precipitated DNA was recovered by centrifugation at 11,180*g* for 15 min. The DNA was washed twice with 10 ml of 80% ethanol, collected by centrifugation at 11,180*g* for 10 min, briefly dried, and redissolved in 2 ml of TE. Residual PEG was removed from the aqueous DNA by additional (3) phenol extractions and (2) ethanol precipitations. The DNA was redissolved in 1.0 ml TE, quantified by measurement of A_{260}, and assessed for suitability in subsequent steps by A_{260}/A_{280} and by gel electrophoresis. Routinely, 400 to 800 µg of high-molecular-weight DNA was obtained from a 1 l culture of *Euglena*.

2.4: Isolation of *Euglena gracilis* Total RNA

Total RNA was extracted from *Euglena* using a protocol modified from Schnare and Gray [146], with all steps done on ice. *Euglena* cells were grown as described above and the culture was harvested by first cold-shocking the cells. Briefly, the culture was decanted into six 250-ml centrifuge bottles and maintained on ice for ~30 min. The cells were recovered by centrifugation at 2,260*g* (4,000 rpm in a Thermo IEC 877 fixed-angle rotor) for 20 min at 4°C. The cells were pooled, washed in cold 0.85% NaCl in a total volume of 40 ml, and transferred to a 50-ml conical centrifuge tube. Pooled cells were recovered by centrifugation in an IEC CR-6000 model centrifuge at 1,900*g* (3,000 rpm in a swinging bucket rotor # 219) for 10 min at 4°C and re-washed as described above. The cells were finally washed in 40 ml of 50 mM Tris-HCl (pH 7.6) and harvested by centrifugation, as described above.

The cells were resuspended in a homogenizing tube with a Teflon® pestle in cold 50 mM Tris-HCl (pH 7.6) to a final volume of 10 ml. The cells were lysed by passage through a French pressure cell at 15,000 lb in^{-2}, and the cell lysate transferred to a 50-ml conical tube. RNA was extracted using the detergent/phenol-cresol method [193] by the addition of 10 ml of 2X detergent mix [2% sodium sarkosyl, 100 mM NaCl] and 20 ml of Tris-buffered phenol-cresol. The solution was mixed at 4°C for 10 min by agitation on a platform shaker and transferred to 30-ml Corex® centrifuge tubes. The aqueous phase was resolved by centrifugation at 1,000*g* (3,000 rpm in a Thermo IEC 875 fixed-angle rotor) for 10 min at 4°C. The aqueous phase was transferred to a clean 30-ml Corex® centrifuge tube, and 0.6 g of NaCl was added. Once the NaCl had dissolved, the aqueous phase was mixed with 20 ml of Tris-buffered phenol-cresol, extracted at 4°C for 5 min by agitation on a platform shaker, and recovered by centrifugation at 1,000*g* (3,000 rpm) for 10 min at 4°C. The phenol-cresol extraction was repeated (3 to 4 times) until no material was left at the interface.

The RNA was precipitated from the aqueous phase with 2 volumes of 95% ethanol at room temperature. The solution was agitated using a vortex mixer and any DNA that precipitated was removed with a glass rod and discarded. The RNA was precipitated overnight at -20°C and collected by centrifugation at 1,000*g* (3,000 rpm) for 10 min at 4°C. The RNA was washed with 5 ml of 80% ethanol, recovered by centrifugation at 1,000*g* (3,000 rpm) for 5 min at 4°C, dried, and dissolved in 1 ml of dH_2O. The aqueous RNA was

transferred to 1.5-ml microcentrifuge tubes. The RNA was extracted with 0.1 volumes of 3 M NaOAc and 1 volume of phenol-cresol and mixed by vigorous agitation with a vortex mixer. The aqueous phase was resolved by centrifugation at 23,880*g* (14,000 rpm) for 5 min at 4°C in a microcentrifuge, and re-extracted. The RNA was precipitated from the aqueous phase with 2 volumes of 95% ethanol, at room temperature, and incubated at -70°C for 30 min. The RNA was sedimented by centrifugation at 23,880*g* for 5 min at 4°C in a microcentrifuge, washed in 75% ethanol, dried, and re-dissolved in a total of 400 µl of dH_2O. The recovered RNA was re-precipitated, as described above, and dissolved in 400 µl of dH_2O. The final product was quantified by A_{260} and assessed for suitability in subsequent steps by A_{260}/A_{280} and by gel electrophoresis. For long-term storage, 1 volume of 95% ethanol was added to the purified RNA, and the solution was maintained -20°C as a 50% ethanol/50% dH_2O mixture.

2.5: Growth Conditions for Laboratory Strains of *E. coli*

E. coli strains were grown for 16 to 20 hr at 37°C on LB [1.0% tryptone, 0.5% yeast extract, 0.5% NaCl, adjusted to pH 7.0 with NaOH] agar (2.0%) plates supplemented with 100 µg/ml of ampicillin (unless stated otherwise) and 40 µg/ml of X-Gal (when used). Liquid cultures were grown to a high cell density (16 to 20 hr) at 37°C with gentle agitation in a roller drum in 3 ml of LB supplemented with 100 µg/ml of ampicillin (unless stated otherwise) and 100 µg/ml of X-Gal (when used [194]). *E. coli* strains and clones were maintained in long-term storage at -70°C as 25% glycerol slurries [500 µl of cell culture in 500 µl of 50% glycerol].

2.6A: *E. coli* Plasmid DNA Extraction - Alkaline-Lysis Miniprep

E. coli cells from a 3-ml overnight culture were decanted into a 1.5-ml tube and harvested by centrifugation in a microcentrifuge for 30 sec at maximum speed. The cell pellet was washed in 1 ml of cold TE-NaCl [100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)] by vigorous agitation with a vortex mixer and collected by centrifugation. The cell pellet was resuspended in 100 µl of cold GTE-lysis buffer [50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA] by vigorous agitation with a vortex

mixer and incubated for 5 min at room temperature. The cells were lysed by the addition of 100 µl of NaOH/SDS lysis solution [200 mM NaOH, 1% SDS] and mixed by gentle inversion (~10 times) of the tube. The solution was immediately neutralized with 100 µl of cold 3 M KOAc (pH 5.5), mixed gently by inverting the tube ~30 times, and incubated for 5 min on ice. Cell debris was removed by centrifugation at maximum speed for 5 min in a cold microcentrifuge. The supernatant, ~275 µl, was transferred to a clean 1.5-ml tube. Plasmid DNAs were precipitated with 0.6 volumes of isopropanol (~170 µl), vigorously agitated with a vortex mixer, incubated for 10 min at room temperature, and recovered by centrifugation at 23,880*g* (14,000 rpm) for 15 min in a 4°C microcentrifuge. The DNA pellet was washed with 1 ml of 70% ethanol, dried, and redissolved in 50 µl of TE. Contaminating RNAs were removed by the addition of 3 U of RNase A and incubated at 37°C for 30 min. The plasmid DNAs were extracted with phenol (once or twice), precipitated with ethanol, washed, dried and redissolved in 30 µl of TE.

2.6B: *E. coli* Plasmid DNA Extraction - QIAprep® Spin Miniprep

Plasmid DNA templates used in DNA sequencing were prepared using the alkaline-lysis-based QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. For plasmid-cloned PCR products, no deviations were made from the protocol. However, for large plasmids encoding genomic λ inserts, the QIAprep Spin Column was washed with 0.5 ml of Buffer PB and the DNA was eluted from the column with 50 µl of Buffer EB (at 70°C), both as recommended by the manufacturer.

Manufacturer-supplied and in-house buffers were both used. The latter included: Buffer P1 [50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 600 U/ml RNase A]; Buffer P2 [200 mM NaOH, 1% SDS]; Buffer N3 [4.09 M guanidine-HCl, 759 mM KOAc, 1.2% (v/v) acetic acid]; Buffer PE [80% ethanol, 17.2 mM Tris-HCl (pH 7.6), 104 mM KOAc, 2 mM EDTA]; and Buffer EB [10 mM Tris-HCl (pH 8.5)].

Used columns were regenerated by the following 750-µl washes with a 30- to 60-sec centrifugation between each wash: *i*) two washes with dH_2O; *ii*) two washes with Renew

Buffer [3 M NaCl, 0.15% Triton X-100]; *iii*) one wash with 1 M NaOH; and *iv*) two washes with 80% ethanol. Residual ethanol was removed by centrifugation and the columns were air dried.

2.7: DNA Amplification by Polymerase Chain Reaction (PCR)

Many different DNA templates were PCR-amplified for a variety of experimental purposes. All PCR reactions were typically done in 50 µl volumes using standard amplification conditions.

PCR amplifications of plasmid DNA templates were done using a small volume (~1 µl or less) of DNA solution (estimated at ~100-500 ng/µl). Amplifications of genomic DNA templates were typically done using 100 to 500 ng of total *Euglena* DNA. Certain genomic DNA amplifications were improved by the use of sheared DNA templates (50 passes through a 29-gauge syringe). This beneficial effect may be attributed to an increase in the accessibility of the oligonucleotide primer to the target DNA template, presumably owing to the sheared DNA's greater ease of denaturation. Alternatively, single-stranded DNA, when sheared, may hybridize preferentially to the oligonucleotide primer rather than to the complementary DNA strand.

Appropriate DNA oligonucleotide primer combinations (see Table 2.1) were used depending on the desired amplification product. Typically, each 50 µl PCR amplification reaction contained 20 pmol of each oligonucleotide primer.

The PCR reaction also contained 200 µM of each dNTP, 1X PCR buffer (New England BioLabs (NEB); ThermoPol Reaction Buffer; contains 2 mM $MgSO_4$) and 2.5 U *Taq* DNA polymerase (NEB).

PCR thermal cycling was done in a Perkin Elmer GeneAmp PCR System 2400. Cycling conditions typically consisted of an initial denaturation at 96°C for 5 min. This was followed by 35 cycles of *i*) denaturation at 95°C for 40 sec, *ii*) annealing at 55°C for 40 sec, and *iii*) extension at 72°C for 50 sec. The thermal cycling was completed with a final extension at 72°C for 7 min. Cycling parameters for PCR products intended for cloning in the TOPO TA Cloning® system (Invitrogen™) consisted of 30 instead of 35 amplification

cycles, to minimize *Taq*-induced errors, and a final extension at 72°C for 30 min instead of 7 min, to ensure that all PCR products were full length and 3′ adenylylated, as recommended by the cloning kit manufacturer.

2.8A: DNA Purification by Non-Denaturing Polyacrylamide Gel Electrophoresis

Short DNA fragments were resolved by non-denaturing polyacrylamide gel electrophoresis (ND-PAGE) and recovered from the gel using the "crush and soak method" [195]. Briefly, DNAs were resolved by electrophoresis in a 7.5% (w/v) polyacrylamide/1x TBE [89 mM Tris-HCl, 89 mM boric acid, 2 mM EDTA] gel. The electrophoresed gel was stained in a light-proof container in 100 ml of ethidium bromide (1 μg/ml) for 5 min, washed in dH_2O for 5 min, and visualized using the Gene Genius Bio Imaging System (Syngene) gel documentation system. Gel slices containing the DNA fragments of interest were excised from the gel with a scalpel and transferred to 1.5-ml microcentrifuge tubes. The polyacrylamide gel slices were crushed and the DNA was eluted by incubation in 1 ml of elution buffer [500 mM NH_4OAc, 10 mM MgOAc, 1 mM EDTA (pH 8.0), 0.1% SDS] for 16 - 18 hr at 37°C with gentle agitation in a roller drum. The crushed polyacrylamide was sedimented by centrifugation at 23,880*g* (14,000 rpm) for 1 min at 4°C. The supernatant was removed and kept. Residual DNA was eluted by re-washing the crushed polyacrylamide with 50 μl of fresh elution buffer, vigorously agitating the mixture with a vortex mixer, and sedimenting the polyacrylamide by centrifugation, as described above. The polyacrylamide was discarded and the supernatant was kept. The supernatants were pooled and the DNA was precipitated, facilitated by the addition of 10 μl of 0.25% linear polyacrylamide carrier [196], with 2 volumes of cold 95% ethanol and incubated at -20°C for 30 min. The DNA was recovered by centrifugation at 23,880*g* for 15 min in a cold microcentrifuge, washed in 75% ethanol, dried, and redissolved in 50 μl of TE. The DNA was then extracted with an equal volume of 24:1 chloroform:isoamyl alcohol. The aqueous DNA was precipitated with 0.1 volumes of 3 M NaOAc and 2.5 volumes of cold 95% ethanol, incubated at -20°C for ~1.5 hr, recovered by centrifugation as described above, washed with 75% ethanol, dried, and redissolved in 10 μl of TE.

2.8B: DNA Purification in Low-Melting-Point Agarose

DNA fragments were resolved by electrophoresis in 1 to 2.5% (w/v) low-melting-point agarose/1x TAE [100 mM Tris, 0.1 mM Na_2•EDTA, titrated to pH 8.0 with glacial acetic acid] gels submerged in 1x TAE buffer. Gels were stained in a light-proof container in 100 ml of ethidium bromide (1 μg/ml) for 5 min, washed in dH_2O for 5 min, and visualized. Gel slices containing the DNA fragments of interest were excised from the gel with a scalpel, subdivided when necessary, and 0.4 to 0.55 g was transferred to each 2-ml microcentrifuge tube. The gel slices were crushed by vigorous agitation with a vortex mixer and melted by incubation at 65°C for 30 min with occasional mixing. The melted gel was incubated with 0.5 volumes of hot (65°C) Tris-buffered phenol-cresol for 5 min at 65°C with frequent vigorous agitation with a vortex mixer. The supernatant was recovered by centrifugation at 23,880*g* (14,000 rpm) for 10 min in a room temperature microcentrifuge. The supernatant was phenol extracted two additional times, as described above. After the addition of 0.1 volumes of 3 M NaOAc to the supernatant, four additional room temperature phenol extractions were done, or until no material was left at the interface. DNA was precipitated, facilitated by the addition of 5 μl of 0.25% linear polyacrylamide carrier [196], with 2.5 volumes of cold 95% ethanol and incubated at -20°C for ~2 hr. The DNA was recovered by centrifugation at 23,880*g* for 20 min in a cold microcentrifuge, washed in 75% ethanol, dried, and redissolved in 5 to 10 μl of TE.

2.8C: DNA Purification Using the Sephaglas™ BandPrep Kit

DNA fragments were resolved by electrophoresis in a 2.5% (w/v) agarose/1x TBE gel submerged in 1x TBE buffer. Gels were stained in a light-proof container in 100 ml of ethidium bromide (1 μg/ml) for 5 min, washed in dH_2O for 5 min, and visualized. Gel slices containing the DNA fragments of interest were excised from the gel with a scalpel. DNA was recovered from the gel slices using the NaI/glass milk-based Sephaglas™ BandPrep Kit (Amersham Biosciences), following the manufacturer's protocol. Briefly, gel bands were crushed and up to 1000 mg was transferred to each 2-ml microcentrifuge tube. The gel was liquified by the addition of 1 μl of fresh 6 M NaI for each 1 mg of agarose and 5 μl of 50% acetic acid for each 250 μl of 6M NaI, vigorously agitated with a vortex mixer, and incubated

at 65°C for 10 min, or until the agarose was dissolved. The Sephaglas BP glass milk solution was vigorously mixed to form a uniform suspension and 10 µl was added to the dissolved agarose in each tube. The mixture was gently agitated and incubated at room temperature for 5 min, with occasional gentle mixing to keep the Sephaglas in suspension. The Sephaglas-bound DNA was recovered by centrifugation at 23,880*g* (14,000 rpm) for 30 sec in a room temperature microcentrifuge, washed three times with 160 µl of NEET wash buffer [50 mM NaCl, 50% ethanol, 1 mM EDTA (pH 8.0), 10 mM Tris-HCl (pH 7.6)], and air-dried for 20 min. DNA was eluted from each Sephaglas pellet by resuspension in 20 µl of elution buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA], gently mixed, and incubated for 5 min at room temperature with periodic mixing. The Sephaglas was sedimented by centrifugation at 23,880*g* for 1 min in a room temperature microcentrifuge and the aqueous DNA supernatant was transferred to a clean microcentrifuge tube.

2.9: Cloning of PCR Products

PCR products were amplified as previously described. The *Euglena* U3 snoRNA PCR product was resolved by ND-PAGE, eluted from the gel, ligated into the pT7Blue T-tailed vector (Novagen), and transformed into chemically competent *E. coli* DH5α cells, all as described in relevant sections of Materials and Methods. All other PCR products were purified, if necessary, in low-melting-point agarose gels or using the Sephaglas™ BandPrep Kit (Amersham Biosciences), cloned using the TOPO TA Cloning® kit (Invitrogen™), and transformed into commercial chemically competent cells.

2.9.1A: Plasmid Cloning with T4 DNA Ligase

DNA ligations were done in 10 µl volumes. A mixture of the insert (1 - 4 µl) and vector (50 ng) DNAs was incubated at 37°C for 5 min to melt cohesive termini, then cooled on ice. The remaining components, 1x T4 DNA ligase buffer (Invitrogen™) and 1 U of T4 DNA ligase (Invitrogen™) were then added. The ligation mixture was incubated at 16°C for 16 to 18 hr. The enzyme was heat-inactivated at 65°C for 15 min. The ligated products were transformed into chemically competent cells, as described below.

2.9.1B: Preparation and Transformation of Chemically Competent Cells

A fresh 1-ml overnight culture of *E. coli* DH5α cells grown in LB medium was used to inoculate a 50-ml volume of LB medium. The culture was grown for ~1.5 hr at 37°C with vigorous agitation in an orbital shaker until an OD_{550} of 0.3 to 0.5 was reached. The culture was put on ice and transferred to a sterile 50-ml conical tube. The cells were harvested by centrifugation at 4°C in an IEC CR-6000 model centrifuge for 10 min at 3,430*g* (4,000 rpm in a swinging-bucket #219 rotor). The cell pellet was gently washed in 50 ml of ice-cold 100 mM $MgCl_2$ and collected by centrifugation as described above. The cell pellet was then gently washed in 25 ml of ice-cold 100 mM $CaCl_2$. The resuspended cells were incubated on ice for a minimum of 20 min. The cells were harvested a final time by centrifugation, as described above, resuspended in 2 ml of ice-cold 100 mM $CaCl_2$, and incubated on ice for 1 to 18 hr. The cells, now competent, were used immediately or stored as 115-µl aliquots in 15% glycerol, flash frozen in liquid nitrogen, and kept at -70°C.

Chemically competent *E. coli* DH5α cells were thawed on ice. Typically, 0.5 to 4 µl of plasmid or ligation reaction was added to 115 µl of competent cell suspension and the mixture was incubated on ice for 5 to 20 min. The cells were heat-shocked at 42°C for 90 sec and placed on ice. Transformed cells were allowed to recover from heat shock by the addition of 250 µl of SOC medium [2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, adjusted to pH 7.0 with NaOH] and incubated for 1 hr at 37°C with gentle agitation in a roller drum. The cells were spread on LB agar plates supplemented with ampicillin and X-Gal.

2.9.2: TOPO TA Cloning® of PCR Products

PCR products were cloned, using the TOPO TA Cloning® kit (Invitrogen™), into the pCR®2.1-TOPO® vector and transformed into the One Shot® chemically competent *E. coli* TOP10 cells (Invitrogen™), both procedures as described by the kit manufacturer. TOPO TA Cloning® relies on the activity of topoisomerase I, covalently linked to the vector, to ligate the A-tailed PCR product into the T-tailed vector. Briefly, 1 to 3 µl of a solution of fresh PCR product, gel-purified if necessary, was mixed with 1 µl of salt solution [200 mM

NaCl, 10 mM $MgCl_2$] and 1 µl of TOPO® vector (10 ng) in a 6-µl volume. The topoisomerase I ligation reaction was gently mixed, incubated at room temperature for 5 to 30 min, and placed on ice.

To transform the cells, 2 µl of the TOPO® ligation mixture was added to a vial (50 µl) of chemically competent cells and mixed gently. The cells were incubated on ice for 30 min, heat-shocked at 42°C for 30 sec, and placed on ice. Transformed cells were allowed to recover from heat shock by the addition of 250 µl of SOC medium at room temperature and incubated for 1 hr at 37°C with gentle agitation in a roller drum. The cells were spread on LB agar plates containing ampicillin and X-Gal.

2.10.1: Hybridization Probes - 5′-End Labeling of Oligonucleotide Primers with [γ-^{32}P]ATP

Oligonucleotide primers were 5′-end labeled with [γ-^{32}P]ATP in a 25-µl reaction mixture prepared on ice. Briefly, the reaction mixture contained ~15 pmol (or ~100 ng) of oligonucleotide primer, 1x kinase reaction buffer [50 mM Tris (pH 7.6), 10 mM $MgCl_2$], 5 mM DTT, 1 mM spermidine-HCl, 1 to 2 µl (150 µCi) of [γ-^{32}P]ATP (6000 Ci/mMol), and 10 U of T4 polynucleotide kinase (Invitrogen™). Labeling reactions were incubated at 37°C for 1 hr.

2.10.2A: Hybridization Probes - Preparation of Southern Hybridization Probes

All Southern hybridization probes were derived from cloned PCR products. Plasmid DNAs were extracted by the alkaline-lysis miniprep procedure, as described above. Probes were generated from the plasmid DNAs by either of two methods: *i*) restriction digestion of the plasmid and gel purification of the cloned insert or *ii*) PCR amplification of the plasmid insert.

In the first method, the cloned insert, corresponding to the probe of interest, was excised from the plasmid backbone by *Eco*R1 digestion. The insert and vector DNAs were resolved by electrophoresis in low-melting-point agarose, and the insert was recovered from the gel, as described above.

Alternatively, the desired plasmid insert was PCR-amplified from the plasmid DNA template using insert-specific oligonucleotide primers (see Table 2.1). Typically, the PCR

product obtained was extracted with phenol-cresol (2X), precipitated with ethanol (3X), and redissolved in 20 µl of TE. (The multiple phenol extraction and ethanol precipitation steps were included to deplete the PCR products of the high salt content introduced by the PCR buffer (ThermoPol Reaction Buffer, NEB), as these salts may interfere with subsequent enzymatic manipulations.) When necessary, the recovered PCR product was purified in low-melting-point agarose, as described above.

2.10.2B: Hybridization Probes - Random Hexamer [α-^{32}P]dATP-Labeling of DNA

DNA probes were labeled with [α-^{32}P]dATP using the random hexamer-primed synthesis method of Feinberg and Vogelstein [197]. In a 10.5-µl initial volume, 100 to 300 ng of probe DNA was denatured in the presence of 5 µg of random hexamers by boiling for 3 min followed by rapid annealing on ice. The mixture was then incubated with 1X Klenow fragment buffer [50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 50 ng/µl BSA], 1 mM DTT, 3dNTP mix [50 µM each of dGTP, dCTP, and dTTP], 5.0 µl of 3000 Ci/mMol [α-^{32}P]dATP [50 µCi], and 8 U Klenow fragment [Invitrogen™; corresponding to the large fragment of DNA polymerase I]. The reaction (in a final volume of 25 µl) was conducted at room temperature for 4 to 8 hr.

The labeled probe was separated from the unincorporated [α-^{32}P]dATP by size-exclusion chromatography. Briefly, the reaction products, diluted in dH_2O to a final volume of 50 µl, were loaded into an S200 size-exclusion Sephacryl column (Pharmacia Corp/Pfizer) and eluted by centrifugation for 2 min at 920*g* (3,000 rpm) in a microcentrifuge. The efficiency of the labeling reaction was crudely determined using a Geiger counter to compare the number of counts from the purified probe to that of the unincorporated [α-^{32}P]dATP remaining in the S200 column.

2.11.1: Southern Hybridization Analysis - Restriction Digestion of Total *Euglena* DNA

Total *Euglena* DNA was digested with the specified restriction endonuclease in 100-µl reaction volumes for 5 hr at 37°C. Reaction conditions typically included 13 µg of total *Euglena* DNA, 1X of the appropriate REact buffer (Invitrogen™), 1 mM DTT, 100 ng/µl

BSA and 100 U of restriction endonuclease (Invitrogen™). The digested DNA fragments were extracted with phenol (2X), and then precipitated with ethanol (2X), washed with 75% ethanol, dried, and redissolved in TE.

2.11.2: Southern Hybridization Analysis - Gel Electrophoresis for Southern Transfer

Gel electrophoresis was performed in a 20 cm X 20 cm gel submerged in an electrophoresis chamber. Restricted *Euglena* DNA fragments, ~13 μg/7 mm lane, were resolved by electrophoresis in a 0.5% (w/v) agarose/1X TAE gel submerged in 1X TAE buffer. The DNA fragments were electrophoresed at ~1V/cm for ~20 hr. Gels were stained in 500 ml of ethidium bromide (1 μg/ml) for 30 min, washed in dH_2O for 5 min, and visualized.

2.11.3: Southern Hybridization Analysis - Capillary Southern Transfer of *Euglena* DNA

Capillary Southern transfer was performed as described [195,198,199]. Briefly, electrophoresed gels were soaked in depurination solution [0.25 M HCl] with gentle agitation for ~10 min. The gels were then rinsed with dH_2O and soaked in Southern denaturation/transfer solution [0.4 M NaOH] with gentle agitation for two 15-min incubations. The HCl treatment partially depurinates the DNA. Subsequent exposure to NaOH results in the formation of nicks at the depurinated sites, with the resulting smaller DNA fragments being more easily transferred out of the gel. The DNA fragments were transferred to a charged nylon membrane (GeneScreen Plus®, NEN® Life Science Products) by capillary Southern transfer according to the membrane manufacturer's alkaline transfer protocol. The nylon membrane was briefly wetted in dH_2O followed by a 15-min incubation in denaturation/transfer solution. The capillary Southern transfer was assembled [198] and allowed to proceed overnight. Denaturation/transfer solution was used as transfer agent, in place of the Southern high salt solution [198], owing to its ability to denature, transfer, and fix the DNA in a single step. Following the transfer, the membrane was neutralized in 0.2 M Tris-HCl (pH 7.6)/2X SSC [1X: 150 mM NaCl, 15 mM sodium citrate] for 5 min, washed in 2X SSC for 5 min, and baked at 80°C for ~8 hr. When necessary, the nylon membrane was cut into strips corresponding to groups of duplicated sample lanes.

2.11.4: Southern Hybridization Analysis - Southern Hybridization and Washing Conditions

Southern transfer membranes were wetted in 6X SSC for 5 min. The blots were sandwiched between two pieces of nylon mesh and placed in large Hybaid HB-OV-BL hybridization bottles. The membranes were washed for a minimum of 30 min in ~120 ml of 6X SSC at 42°C in a Mini Oven MK II (Hybaid) rotary hybridization oven. The blots were then pre-hybridized for 4 to 18 hr at 42°C in ~40 ml of Southern hybridization solution [5X Denhardt's solution (0.1% Ficoll (Type 400), 0.1% polyvinylpyrrolidone, 0.1% BSA (Fraction V)), 5X SSC, 25 mM Na_2HPO_4, 25 mM NaH_2PO_4, 180 µg/ml sheared and denatured herring testes DNA (boiled for 10 min and cooled on ice), 50% formamide (deionized), 1% SDS]. The pre-hybridization solution was discarded and replaced with 24 ml of pre-warmed hybridization solution. The random hexamer-radiolabeled probe was added to 1 ml of hybridization solution, denatured by boiling for 10 min, cooled on ice, and added to the hybridization bottle. Blots were hybridized at 42°C for 18- to 22-hr.

The hybridized membranes were washed in the hybridization bottles by five 10-min incubations at 42°C in 2X SSC/0.1% SDS in the rotary hybridization oven. The blots were transferred to a large Pyrex® baking dish and washed once or twice for 15 min each in 0.1X SSC/0.1% SDS with gentle agitation in a 42°C water bath. The membranes were then rinsed in 0.1X SSC at room temperature, placed in sealed bags, and autoradiographed at -70°C in the presence of an intensifying screen for 5 to 32 days, based on the intensity of the hybridization signal. Most autoradiographs were developed in a Kodak X-OMAT 1000A processor, with select films being manually developed (by Dr. Murray N. Schnare, Dalhousie University). Autoradiographs were digitized with an Epson Expression 1680 flat-bed scanner and Adobe® Photoshop® Elements version 2.0. When necessary, digital images were manipulated using Adobe® Photoshop® CS2 version 9.0.2. All image manipulations, such as background reduction, were uniformly applied to the entire digital image.

2.11.5: Southern Hybridization Analysis - Removal of Probes from Hybridized Southern Membranes

In some cases, hybridized Southern membranes were reused by removing the bound probe to permit subsequent hybridization with a different probe. Briefly, the hybridized

probe was removed, or "stripped off," by pouring ~500 ml of a boiling solution of 0.1% SDS onto the membrane and allowing the solution to slowly cool to room temperature [199]. The membrane was then briefly rinsed in 0.01X SSC at room temperature. Excess liquid was removed by blotting the hybridization membrane with dry Whatman 3MM filter paper and the stripped Southern membrane was placed in a sealed bag. The efficacy of probe removal was assessed by autoradiography, as described above. The stripped Southern membrane was subsequently hybridized as described above.

2.11.6: Southern Hybridization Analysis - Densitometry of Hybridizing Bands

The relative signal intensities of the multiple hybridizing bands were examined by densitometric analysis. Unsaturated autoradiographic film exposures of Southern hybridization membranes were digitized as described above. Digital images were analyzed using the image analysis program ImageJ version 1.36b [200]. For each lane, the intensity of the signal in each of the multiple hybridizing bands was transformed into peaks in a line graph. The area under each peak was determined and used as a relative quantification of the intensity of the hybridization signal.

2.12.1: Construction of a λ Library from *Euglena gracilis* Genomic DNA

A *Euglena gracilis* λ genomic DNA library [201] (prepared and kindly provided by Dr. Yoh-ichi Watanabe, formerly of the Department of Biochemistry and Molecular Biology, Dalhousie University) was constructed in the λBlueSTAR™ vector (Novagen, Inc.) according to the manufacturer's protocol. Briefly, *Euglena gracilis* genomic DNA (kindly provided by Dr. David F. Spencer, Dalhousie University) was partially hydrolyzed with *Sau*3AI. The two internal positions of the resulting restriction site overhang, 5′ GATC, were filled in using Klenow DNA polymerase and dATP and dGTP, thereby creating a 5′ GA overhang. Although recommended by the manufacturer, the partially filled-in restriction fragments were not size-fractionated prior to ligation onto the λ arms. The commercially prepared λ arms had been: *i*) hydrolyzed with *Xho*I to remove the stuffer fragment encoding the *E. coli lacZ* gene, *ii*) partially filled in using dCTP and dTTP, to be compatible with the partially filled-in *Sau*3AI genomic inserts, and to suppress insert-insert, vector-vector and

stuffer-vector ligations, and *iii*) dephosphorylated to further reduce the occurrence of non-productive ligations. Following the ligation of the insert and vector arms, recombinant molecules were packaged *in vitro* using the MaxPlax Packaging Extract (Epicentre Technology). The resulting *Euglena gracilis* λ genomic DNA library was amplified according to the manufacturer's protocol.

2.12.2: Screening of the *Euglena gracilis* λ Genomic DNA Library

The *Euglena gracilis* λ genomic DNA library, described above, was screened to isolate genomic inserts encoding U3 snoRNA genes, with the aim of exploring the wider genomic context of these genes. The library was screened by standard plaque lift and hybridization with a *Euglena* U3 snoRNA gene probe.

The number of recombinant phage plaques required to be screened in order to achieve a 99% probability of isolating a clone containing the desired genomic insert was determined. This standard statistical method of assessing library coverage [202] incorporated an estimated *Euglena* genome size of 2.72 X 10^9 bp [167,176] (but see discussion of the *Euglena* genome in section 1.5 of the Introduction) and an expected average λ library insert size of 12 kbp. Based on this model, a total of 1.04 X 10^6 screened λ plaques would be required. However, a total of 2.13 X 10^6 recombinant plaques was screened, representing twice the required number. Nevertheless, the final screening and sequencing results (see discussion in Results, Figure 3.3) revealed that deviations from the λ library construction protocol resulted in a library whose complexity could not accurately be estimated.

2.12.3A: λ Library Infection and Plating

The λ library was prepared for screening by plaque lift according to the manufacturer's protocol. Briefly, the λ host strain, *E. coli* ER1647, was maintained on LB agar plates containing 12.5 μg/ml of tetracycline. A 3-ml starter culture of liquid LB medium supplemented with 10 mM $MgSO_4$ and 0.2% maltose was inoculated with cells from a single colony of *E. coli* ER1647 and grown overnight. Fifteen to 30 μl of the overnight starter culture was used to inoculate new 3-ml cultures of identically supplemented LB medium and the cultures were grown to mid-log phase (OD_{600} of 0.5 to 0.6).

Primary screening of the λ library was done at high density (1 to 3 X 10^5 plaque-forming units (pfu)/plate) on 26 large 150-mm petri plates. A total of 330 µl of mid-log-phase λ host cells, *E. coli* ER1647, was mixed with 330 µl of λ phage suspension at an appropriate dilution in SM buffer [100 mM NaCl, 8.1 mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5), 0.01% gelatin]. The solution was gently mixed, then incubated at 37°C for 30 min to allow the phage to adsorb to the host cells. Ten ml of molten top agarose [1% tryptone, 0.5% NaCl, 0.6% agarose, supplemented with 200 µg/ml X-Gal] at 47°C was added to the λ phage/host cell mixture. The solution was gently mixed, immediately poured onto a slightly dry, 2- to 3-day-old, pre-warmed LB plate, and spread by gently swirling the plate. Once the top agarose had solidified, the plates were incubated at 37°C for 8 to 14 hr, depending on the desired plaque size and confluence.

2.12.3B: λ Library Plaque Lifts

Plaque lifts were done as recommended by the membrane manufacturer's protocol (Magna Lift, Osmonics Inc.). Phage plates were prepared for plaque lifts by cooling the plates at 4°C for 30 min. Precut, 137 mm (for large plates) or 82 mm (for small plates), circular nylon hybridization membranes were gently placed on the surface of the phage plates and allowed to adhere for 5 min. The nylon membrane and plate were then marked at three asymmetric locations with a 20-gauge syringe needle and a pen to allow for subsequent re-orientation of the membrane with respect to the phage plate. Nylon membranes were carefully lifted from the phage plates and floated, plaque side up, for 5 min each in small pools of *i*) high salt denaturation buffer [0.5 M NaOH, 1.5 M NaCl], *ii*) neutralization buffer [1.5 M NaCl, 0.5 M Tris-HCl, pH 8.0], and *iii*) wash buffer [2 X SSC]. The membranes were air dried, DNA side up, for 5 min on filter paper, sandwiched between fresh filter paper, wrapped in aluminum foil, and baked at 80°C for approximately 2 hr, or until dry.

2.12.3C: λ Library Plaque Hybridization and Washing Conditions

The *Euglena gracilis* λ genomic DNA library, plated and transferred to nylon membranes, was screened with a *Euglena* U3 snoRNA probe by plaque hybridization. The hybridization probe was prepared and random hexamer-labeled as described above. The

plaque membranes were hybridized and washed as described above for Southern hybridizations. However, hybridizations and washes were performed in round Tupperware® containers placed in a New Brunswick Scientific orbital shaker/water bath.

2.12.3D: Identification of Hybridizing Plaques, Elution of Phage, and Subsequent Plaque Screenings

The asymmetric markings on the hybridization membranes and phage plates were used to align the hybridization signals on the developed autoradiographs with the corresponding plaques on the screened plates. Putative hybridizing plaques were recovered from the phage plate by "coring" the identified regions with the wide end of a blue P1000 micropipetter tip. The phage were recovered from the cored regions by placing the cored media in 1 ml of SM buffer containing 2% chloroform, vigorously agitating the mixture with a vortex mixer, and eluting the phage for several hours at room temperature. The eluted phage were diluted in SM buffer to an appropriate pfu concentration and subsequent screenings were performed. Secondary (and in some cases, tertiary) screenings were performed as described above, but on small, 82 mm petri plates using a third of the volume of host cells, diluted λ phage, and top agarose. When single positive plaques could not be isolated, the region was cored with the wide end of a yellow P200 micropipetter tip, eluted and diluted, as described above, and re-screened. Single phage plaques, each corresponding to a positive hybridization signal, were cored with the narrow end of a Pasteur pipet and eluted in 50 µl of SM buffer containing 2% chloroform.

2.12.3E: *In vivo* Excision of λ Clones

Positive λ plaques were excised *in vivo* into plasmid clones by Cre-mediated recombination, as described in the λBlueSTAR™ kit protocol. The λ host strain used for *in vivo* excision, *E. coli* BM25.8, was maintained on LB plates containing 34 µg/ml of chloramphenicol and 50 µg/ml of kanamycin. Cells were grown in liquid LB medium supplemented with 10 mM $MgSO_4$ and 0.2% maltose to an OD_{600} of 1.0. The λ host cells used for *in vivo* excision were infected with an appropriate dilution of phage, as described

above, but omitting the molten top agarose. The infected cells were spread onto LB agar plates containing 100 μg/ml of ampicillin and 40 μg/ml of X-Gal, and grown at 37°C overnight.

According to the manufacturer's protocol, plasmid DNAs extracted from *E. coli* BM25.8, the λ host strain used for *in vivo* excision, are unsuitable for restriction digestion or DNA sequencing. Thus, plasmid DNAs were extracted from *E. coli* BM25.8 by alkaline-lysis plasmid miniprep and used to transform chemically competent *E. coli* DH5α cells, both as described above. All subsequent plasmid DNAs were extracted from *E. coli* DH5α cells.

2.13: 3′ Rapid Amplification of cDNA Ends (3′ RACE)

A 3′ RACE methodology [11] was used to map the mature 3′-ends of *Euglena* U3 snoRNA and U5 snRNA, and to identify which gene variants are expressed.

First, *Euglena* total RNA was poly(A)-tailed. In a 10-μl reaction volume, 2 μg of total RNA from *Euglena gracilis* strain Z (the parental strain from which the aplastidic variant was derived [133]; RNA kindly provided by Dr. Anthony G. Russell, Department of Biochemistry and Molecular Biology, Dalhousie University), 1x poly(A) polymerase buffer (Amersham BioSciences), and 500 μM of ATP were heated at 65°C for 5 min to disrupt RNA secondary structures, then chilled on ice. To the mixture, 1500 U of yeast poly(A) polymerase (Amersham BioSciences) was added and the reaction was incubated at 37°C for 30 min. The poly(A)-tailed RNA was phenol extracted (2x), ethanol precipitated, washed in 75% ethanol, dried, and redissolved in 11 μl of dH_2O.

In the cDNA synthesis step, 11 μl of the poly(A)-tailed RNA from the previous step was annealed to 1 pmol of the p94 oligonucleotide (see Table 2.1). The mixture was incubated at 65°C for 5 min, then slowly annealed at 47°C for 10 min and at room temperature for 30 min. To the annealed poly(A)-tailed RNA/p94 oligonucleotide mixture was added 1x First-Strand Synthesis buffer (Invitrogen™), 10 mM DTT, 40 U of RNaseOUT™ ribonuclease inhibitor (Invitrogen™), and 500 μM of each dNTP. The mixture (in a 21-μl volume) was heated at 47°C for 5 min and 200 U of Superscript II Reverse Transcriptase (Invitrogen™) was added. The reaction mixture was incubated at 47°C for 45 min. The enzyme was denatured by incubation at 90°C for 3 min, then cooled

to room temperature. The cDNAs were treated with 2 U of RNase H (Invitrogen™) at 37°C for 30 min to remove the RNA template strand. The resulting single-stranded cDNA solution was phenol extracted (2x), ethanol precipitated, washed with 75% ethanol, dried, and redissolved in 40 µl of dH_2O.

The PCR amplification step, performed in a 50-µl reaction volume as described above, contained 20 pmol each of oAR7 (nested to the reverse transcription-specific p94 (see Table 2.1)) and a gene-specific oligonucleotide primer designed to the 5′ region of the target RNA (see Table 2.1), and 2 µl of single-stranded cDNA product as template. Cycling conditions consisted of an initial denaturation at 94°C for 5 min. This step was followed by 30 cycles of *i*) denaturation at 94°C for 30 sec, *ii*) annealing at 55°C for 30 sec, and *iii*) extension at 72°C for 30 sec. The thermal cycling was completed with a final extension at 72°C for 7 min.

The various 3′ RACE-PCR products were resolved by electrophoresis. Bands of interest were gel purified if necessary, cloned by TOPO TA Cloning®, and sequenced.

2.14.1: Northern Hybridization Analysis – Gel Electrophoresis and Capillary Northern Transfer of *Euglena* Total RNA

Total *Euglena* RNA, 3.5 µg per lane in 10 µl of NMF/urea loading buffer [60% (v/v) *N*-methylformamide (deionized), 5 M urea, 10 mM Tris-HCl (pH 7.6), 1 mM EDTA, 0.1% (w/v) xylene cyanol [193]], was resolved by electrophoresis in a 20 cm X 20 cm X 0.15 cm vertical 6% (w/v) polyacrylamide/7M urea gel in 0.6x TBE. The gel was stained in 500 ml of ethidium bromide (1 µg/ml) for 30 min, washed in dH_2O for 5 min, and visualized.

The RNA was transferred from the gel to a charged nylon membrane (GeneScreen Plus®, NEN® Life Science Products) by capillary transfer, following modifications to the manufacturer's alkaline transfer protocol. The nylon membrane was briefly wetted in dH_2O followed by a 15-min incubation in alkaline northern transfer solution [5 mM NaOH]. The capillary transfer components were assembled as described for a capillary Southern transfer [198] and allowed to proceed overnight. The alkaline northern transfer solution offers the advantages of being able to denature, facilitate the transfer of (by random, infrequent nicking of large fragments via phosphodiester bond attack mediated by the 2′-OH of the RNA sugar at low concentrations of NaOH), and fix the RNA in a single step. Following the transfer,

the membrane was neutralized in 0.2 M Tris-HCl (pH 7.6)/2X SSC for 10 min, washed in 2X SSC for 10 min, and baked at 80°C for ~ 4 hr. The nylon membrane was cut into strips corresponding to duplicated sample lanes.

2.14.2: Northern Hybridization Analysis - Northern Hybridization and Washing Conditions

Northern hybridizations were performed as described above for the Southern hybridization method, with the following changes. The hybridization membranes were wetted in 6X SSC for 5 min, placed in small hybridization bottles, and washed for a minimum of 30 min in 30 ml of 6X SSC at 45°C in a rotary hybridization oven. The membranes were then pre-hybridized for 1.5 hr at 45°C in 30 ml of northern hybridization solution [10X Denhardt's solution, 5X SSC, 10 mM Na_2HPO_4, 10 mM NaH_2PO_4, 10% dextran sulfate, 7% SDS]. The pre-hybridization solution was discarded and replaced with 4 ml of pre-warmed hybridization solution. The 5′-end-labeled oligonucleotide probe was added to 1 ml of hybridization solution, denatured at 75°C for 10 min, cooled on ice, and added to the hybridization bottle. Blots were hybridized at 45°C (5-7°C lower than the T_m (melting temperature) of the oligonucleotide probes) for 19 hr.

The hybridized blots were washed in the hybridization bottles by five 10-min incubations at 40°C (7-12°C lower than the T_m of the oligonucleotide probes) in 2X SSC/0.1% SDS in the rotary hybridization oven. The blots were transferred to a large Pyrex® baking dish and washed three times for 15 min each in 0.1X SSC/0.1% SDS with gentle agitation in a 40°C water bath. The membranes were then rinsed in room-temperature 0.1X SSC, placed in sealed bags, and autoradiographed at -70°C in the presence of an intensifying screen for ~5 days, based on the intensity of the hybridization signal.

2.15: DNA Sequencing, Sequence Analysis, and Bioinformatics

All plasmid templates for DNA sequencing were prepared using the alkaline lysis-based QIAprep® Spin Miniprep Kit (Qiagen), as described above. Routinely, ~500 ng of plasmid DNA template (visually estimated by ethidium bromide fluorescence) and 6.4 pmol of oligonucleotide primer (see Table 2.1) were submitted (in a 15-µl volume) for sequencing to an in-house automated sequencing facility. Plasmid DNA templates were sequenced by

PCR-based cycle sequencing using the BigDye® Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on an ABI Prism™ 377 DNA sequencer (Applied Biosystems) by Ms. Marlena Dlutek (Department of Biochemistry and Molecular Biology, Dalhousie University). All sequence chromatograms were manually inspected and errors in automated base calling were rectified when possible. Sequence chromatograms were analyzed and contig assembly was done using the Staden Package version 2002.0 software suite [203].

All PCR and λ clones were initially sequenced using the appropriate vector-based M13R, T3, and/or T7 oligonucleotide primers (see Table 2.1). Short insert clones were completely sequenced on both strands. Multiple independent clones of PCR products were sequenced to eliminate *Taq*-generated errors. For large clones, such as λ genomic inserts, nucleotide regions encoding genes of interest were sequenced on both strands. Additional sequences were obtained using gene-specific oligonucleotide primers in combination with a primer walking strategy (see Table 2.1). Nucleotide regions recalcitrant to standard sequencing methodologies, such as regions of extreme nucleotide bias and simple repetitive elements, were sequenced, when possible, by the addition of 5% acetamide to the cycle sequencing reaction or by LiCor automated sequencing (NRC, Halifax, Nova Scotia).

GenBank (http://www.ncbi.nlm.nih.gov/blast/) and the Protist EST Program (TBestDB-http://amoebidia.bcm.umontreal.ca/pepdb/searches/login.php) sequence databases were searched using the Basic Local Search Alignment Tool (BLAST) suite of programs [204,205]. Other databases occasionally searched included: The Institute for Genomic Research (TIGR-http://www.tigr.org/), the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/), the US Department of Energy Joint Genome Institute (JGI-http://www.jgi.doe.gov/), and the Broad Institute of MIT and Harvard (http://www.broad.mit.edu/).

The pattern-searching algorithm PatScan [206] was used to search for small RNA genes, such as U3 snoRNA, in genomic sequences. The PatScan program searches for nucleotide sequences within user-defined secondary structures. Publicly available genome sequence databases were downloaded to a local SUN computer server, on which the PatScan

algorithm was run. Putative hits were further screened for their location in a non-coding region of the genome and for the potential to fold the sequence into the expected secondary structure model.

Sequence alignments were generated using the ClustalX version 1.81 algorithm [207] and manually edited with BioEdit Sequence Alignment Editor version 7.0.5.3 [208]. DNA sequences were manipulated, introns were excised, and open reading frames (ORFs) were translated using DNASIS version 2.5. RNA secondary structures were drawn by Dr. Murray N. Schnare (Dalhousie University) using the RNA secondary structure drawing program XRNA [209].

Table 2.1 DNA Oligonucleotide Primers Used for PCR Amplifications, DNA Sequencing, and 3′ RACE.

Oligonucleotide	Comments	Sequence (5′ to 3′)
M13R	Standard sequencing primer	CAGGAAACAGCTATGAC
T3	Standard sequencing primer	ATTAACCCTCACTAAAGGGA
T7	Standard sequencing primer	TAATACGACTCACTATAGGG
o*Eg*U3-F1	Near 5′ end of U3 snoRNA; Used for PCR, sequencing, and 3′ RACE; Designed by Greenwood *et al.* [38]	CTCCACAAGGATCATTTCTTGAGG
o*Eg*U3-F2	Near 3′ end of U3 snoRNA; Used for sequencing and 3′ RACE	GATGAGAGGTCAGCAATTTGAGTGG
o*Eg*U3-F3	Near 3′ end of U3 snoRNA; Used for sequencing and 3′ RACE	GATGAGAGGTCAGCAATTTGAGTGGTCTTTCC
o*Eg*U3-R1	Near 3′ end of U3 snoRNA; Used for PCR and sequencing; Designed by Greenwood *et al.* [38]	CCACTCAAATTGCTGACCTCTCATC
o*Eg*U3-R2	Near 5′ end of U3 snoRNA; Used for sequencing; Designed by Greenwood *et al.* [38]	CTCTGTGAATCGGACTGATACTTC
o*Eg*U3-R3	Near 3′ end of U3 snoRNA; Used for PCR	CCACTCAAATTGCTGACCTCTC
o*Eg*U4-F1	Near 5′ end of U4 snoRNA; Used for 3′ RACE	CGCTCGGGCAATCACTCAGAGC
o*Eg*U4-R1	Near 3′ end of U4 snRNA; Used for PCR	ACAGGGGGCACAGAAATTGATC
o*Eg*U5-F1	5′ end of U5 snRNA; Used for PCR, sequencing, and 3′ RACE	GCAACACAGCTCCGTGCTTACTCG
o*Eg*U5-F2	Middle of U5 snRNA; Used for 3′ RACE	CTAAAGATAGCCGTTGGCTACGGAGC
o*Eg*U5-R1	3′ end of U5 snRNA; Used for PCR and sequencing	GTTCCAAAAATTGATGTAACACATCG
o*Eg*U5-R2	3′ end of U5 snRNA; Used for PCR	GTTCCAAAAATTGATGTAACACATC

o*Eg*U5-R3	5′ end of U5 snRNA; Used for PCR and sequencing	GATAGATGCGAGTAAGCAC GGAGC
o*Eg*tRNA-Arg-F1	5′ end of $tRNA^{Arg}_{UCG}$; Used for PCR and sequencing	GTCGTGTGGCGCAATGGAT AG
o*Eg*tRNA-Arg-F2	Middle of $tRNA^{Arg}_{UCG}$; Used for PCR and sequencing	GCAATGGATAGCGCGTCGG GCTTCG
o*Eg*tRNA-Arg-R1	3′ end of $tRNA^{Arg}_{UCG}$; Used for PCR and sequencing	GATCGTGACAGGACTCGAA C
o*Eg*tRNA-Arg-R2	Middle of $tRNA^{Arg}_{UCG}$; Used for PCR and sequencing	CGTGACAGGACTCGAACCT GCAACCG
o*Eg*U3U5IGS A1-F1	5′ end of U3-U5 IGS linkage A1; Used for PCR and sequencing	GACCATAAACCATCACAAT CATC
o*Eg*U3U5IGS A1-R1	3′ end of U3-U5 IGS linkage A1; Used for PCR and sequencing	GTGCGATTGATTGGCAAGG TACAGC
o*Eg*U3U5IGS B1-F1	5′ end of U3-U5 IGS linkage B1; Used for PCR and sequencing	GCTTGCCAAAGTGATGTGA GG
o*Eg*U3U5IGS B1-R1	3′ end of U3-U5 IGS linkage B1; Used for PCR and sequencing	CTTGTGGTTCAAAAGTTGA GG
o*Eg*U3U5IGS B2-F1	5′ end of U3-U5 IGS linkage B2; Used for PCR and sequencing	GCTCGAAATCCACACAATT TGCC
o*Eg*U3U5IGS B2-R1	3′ end of U3-U5 IGS linkage B2; Used for PCR and sequencing	GGAACTTTCCTTTTCTTGT GG
o*Eg*U3U5IGS C1/2-F1	5′ end of U3-U5 IGS linkage C1 and C2; Used for PCR and sequencing	CTTGACGAAGTTCCAATTT CTC
o*Eg*U3U5IGS C1/2-R1	3′ end of U3-U5 IGS linkage C1 and C2; Used for PCR and sequencing	CCNTTTYTTTGCCAMCTCA AAGGTCCG
o*Eg*U3U5IGS D2-F1	5′ end of U3-U5 IGS linkage D2; Used for PCR and sequencing	CCTCAAGAAACAAAGATGG AGCGGG
o*Eg*U3U5IGS D2-R1	3′ end of U3-U5 IGS linkage D2; Used for PCR and sequencing	CCCTTCCTTTGTCAGTGCT TTG
o*Eg*U3U5IGS E1-F1	5′ end of U3-U5 IGS linkage E1; Used for PCR and sequencing	CACATGGTAAGTACGCCCA CAGG
o*Eg*U3U5IGS E2-F1	5′ end of U3-U5 IGS linkage E2; Used for PCR and sequencing	CCACATATGGTAAATACGC TGTC

o*Eg*U3U5IGS E1/2-R1	3′ end of U3-U5 IGS linkage E1 and E2; Used for PCR and sequencing	CCTTTTCCACACATTATRAACCTGG
o*Eg*λU3solitary-T3FW1	Sequence walking primer	GCCAACCTGTCAGGTGACAG
o*Eg*λU3solitary-T3FW2	Sequence walking primer	TGTCATATAGGAAGTTG
o*Eg*λU3solitary-T7FW1	Sequence walking primer	ATGAACTATGGACGGCCATG
o*Eg*λU3solitary-T7FW2	Sequence walking primer	CATGCTGAAATTGACTCAC
o*Eg*λU3solitary-T7FW3	Sequence walking primer	AACAGCTCGTGTGGAGCTCC
o*Eg*λU3solitary-T7FW4	Sequence walking primer	GCATATGTCAGTGCCATGCCATACCG
o*Eg*λU3solitary-T7FW5	Sequence walking primer	GGGTAGGAACCATGGGAATCACAGGG
o*Eg*λU3solitary-U3FW1	Sequence walking primer	GTTGGTTCCAATAACCTGCC
o*Eg*λU3solitary-U3FW2	Sequence walking primer	GGACTCGTGGGTATTCAAAGG
o*Eg*λU3solitary-U3FW3	Sequence walking primer	GATGCGCGATGGTCGCATTGAGC
o*Eg*λU3solitary-U3RW1	Sequence walking primer	TACTGGGAAGAGGTGGACACTC
o*Eg*λU3solitary-U3RW2	Sequence walking primer	CTCGCATTGATCTTCGCGTCTGGC
o*Eg*λU3solitary-U3RW3	Sequence walking primer	CTCCTTGATCCAACTTCATGCTGCC
o*Eg*λU3-tRNA-Arg-T3FW1	Sequence walking primer	GTTCCTTCATCTGGCTCACTTGC
o*Eg*λU3-tRNA-Arg-T3FW2	Sequence walking primer	GCGCAAAACCTGGTGCTGGAGGTCC
o*Eg*λU3-tRNA-Arg-T3FW3	Sequence walking primer	ACTTGACCAATGTGGGCTTGGCGCC

o*Eg*λU3-tRNA-Arg-T7FW1	Sequence walking primer	GAGGCTTCAACCAAAATCACCGGACTTC
o*Eg*λU3-tRNA-Arg-T7FW2	Sequence walking primer	GGTTGACCACTCTGGAGTCAAGGCC
o*Eg*λU3-tRNA-Arg-T7FW3	Sequence walking primer	CTGGTGCCAGGTCACTGCAACGG
o*Eg*λU3-tRNA-Arg-U3FW1	Sequence walking primer	GCTGAATCATGCAACACAGAGC
o*Eg*λU3-tRNA-Arg-U3FW2a	Sequence walking primer	ATGCGCAGATCTCGCGTTCG
o*Eg*λU3-tRNA-Arg-U3FW2b	Sequence walking primer	GGACTAGTTGAAGAGAGTGAGCG
o*Eg*λU3-tRNA-Arg-U3FW2c	Sequence walking primer	CCAGAATAAATCCATAAATTGCC
o*Eg*λU3-tRNA-Arg-U3FW3	Sequence walking primer	GCTGTTTCTGGGCCGAAAATGAGCG
o*Eg*λU3-tRNA-Arg-U3RW1	Sequence walking primer	ATATCTCACCCACCATGCACTTCG
o*Eg*λU3-tRNA-Arg-U3RW2	Sequence walking primer	GGCAATTTATGGATTTATTCTGG
o*Eg*λU3-tRNA-Arg-U3RW4	Sequence walking primer	GAACAACGATGCTTTACCAAACC
o*Eg*λU3U5A-T7FW1	Sequence walking primer	GGCTCATTGTGGGAAATTGGTGATGC
o*Eg*λU3U5A-T7FW2	Sequence walking primer	GCCTTCCAAAACTCCCATTCGG
o*Eg*λU3U5A-T7RW1	Sequence walking primer	ATCCTCGCCTTGCACACACTGC
o*Eg*λU3U5A-U3FW1	Sequence walking primer	CGACTGGGTTCATTAGGGAC
o*Eg*λU3U5A-U3FW2a	Sequence walking primer	GTATTCAGTGCTCATCAAGGCG
o*Eg*λU3U5A-U3FW2b	Sequence walking primer	AACATTCGGTCCATGGTGTTGGTGGC
o*Eg*λU3U5A-U3RW1	Sequence walking primer	ACCATGGACCGAATGTTGCCGTAACC

p94	Used for 3′ RACE [11]	AATAAAGCGGCCGCGGATC CAAT$_{17}$V
oAR7	Used for 3′ RACE; Designed by Russell *et al.* [11]	CCGGAATTCAATAAAGCGG CCGCGGATCCAA

Primer nomenclature

o*Eg*λ**U3**-FW1: oligonucleotide #1 designed to a *Euglena gracilis* λ genomic fragment encoding a **U3** snoRNA gene whose PCR amplification or DNA sequencing products are Walking Forward (*i.e.*: downstream or 3′ to the U3 snoRNA gene; alternatively, walking Reverse or upstream).

IUPEC nucleotide degeneracy code

R-puRine (A or G); Y-pYrimidine (C or T); M-aMino (A or C); K-Keto (G or T); S-Strong hydrogen bonding (C or G); W-Weak hydrogen bonding (A or T); B-not A (C, G or T); D-not C (A, G or T); H-not G (A, C or T); V-not T (A, C or G); N-aNy (A, C, G or T).

CHAPTER 3: RESULTS

Euglena gracilis U3 snoRNA was originally isolated at the RNA level [38]. Sequence heterogeneities observed in RNA sequencing and multiple hybridizing bands obtained in Southern analysis of *Euglena* DNA hybridized with a U3 probe suggested that U3 snoRNA is a multi-copy gene in the *Euglena* genome [38].

To further expand on this work, a detailed analysis of the copy number and genomic organization and context of *Euglena* U3 snoRNA genes was initiated. Results from this study will *i*) further our understanding of the genomic neighborhood of *Euglena* U3 snoRNA genes, *ii*) may provide a rationale for the unexpected multiplicity of U3 snoRNA genes, and *iii*) may increase our understanding of the many unusual features of the *Euglena* genome. Furthermore, considering *Euglena*'s phylogenetic placement and the well-known idiosyncratic features of its RNA biology, the rationale for this study was that it would provide a more phylogenetically balanced framework for our understanding of U3 snoRNA gene organization, function and evolution.

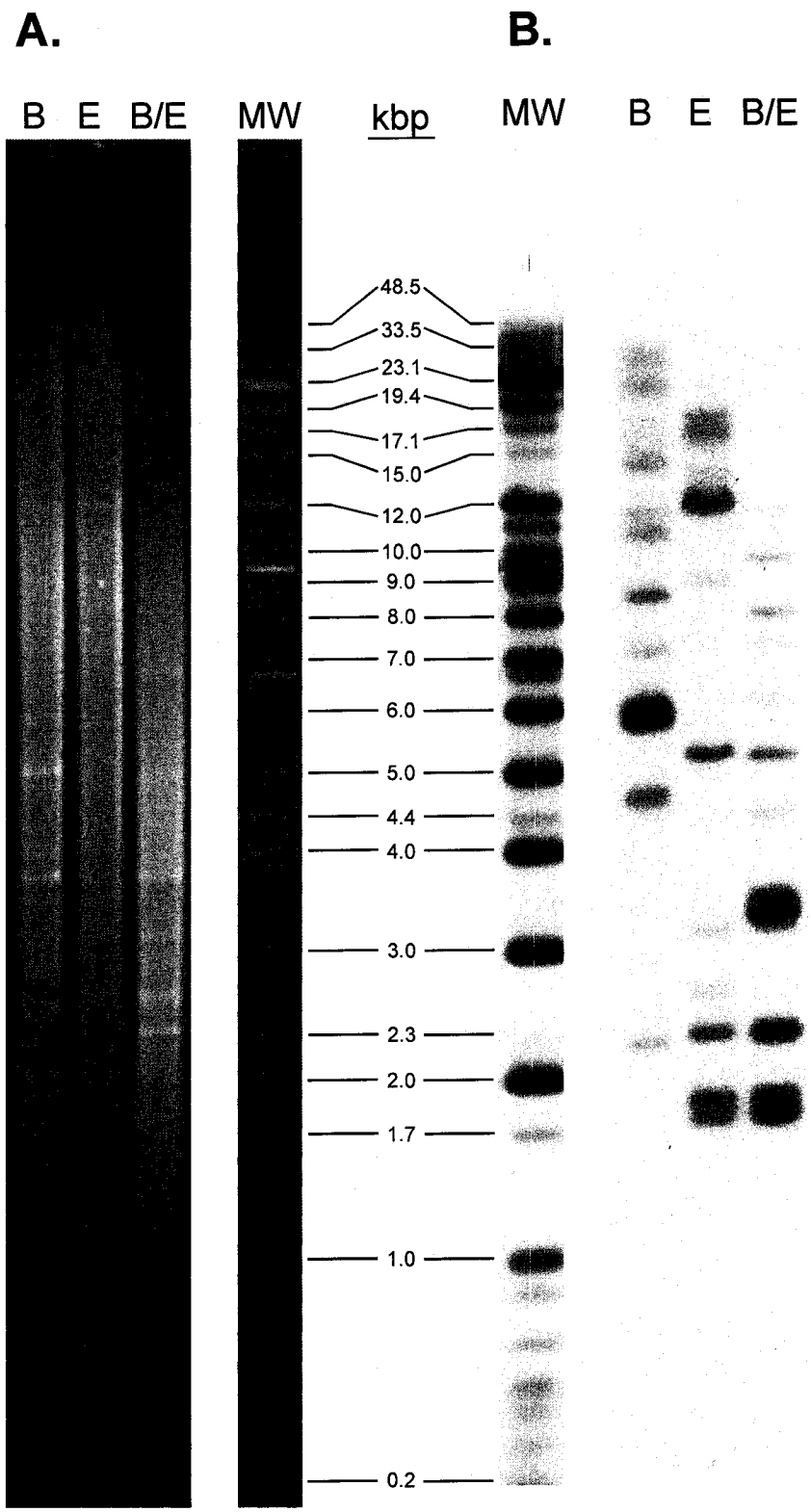
3.1: Multiple Bands Hybridizing with a U3 snoRNA Probe in Southern Analysis of *Euglena* DNA Imply that the U3 snoRNA Gene is Multi-Copy in the *Euglena* Genome

The multi-copy nature of U3 snoRNA genes in the *Euglena* genome was confirmed by replicating the Southern hybridization results obtained by Greenwood *et al.* [38].

Briefly, *Euglena* total DNA was hydrolyzed with *Bam*H1 and/or *Eco*R1 (see Figure 3.1A), two restriction endonucleases that do not have recognition sequences within the *Euglena* U3 snoRNA gene. In addition, preliminary work further justified the choice of restriction endonucleases, revealing that each enzyme yields an optimal distribution of restriction fragment sizes, thus allowing for maximal electrophoretic separation of DNA fragments.

Southern transfer and hybridization with a *Euglena* U3 probe reveals a complex and reproducible pattern of multiple hybridizing bands (see Figure 3.1B). This observation implies that U3 snoRNA is a multi-copy gene in the *Euglena* genome. The resulting hybridization patterns each consist of at least 13 hybridizing *Eco*R1, *Bam*H1, or

Figure 3.1 Multiple U3 snoRNA-Hybridizing Bands in Southern Analysis of *Euglena* DNA Imply that U3 snoRNA is a Multi-Copy Gene in the *Euglena* Genome. A. Total *Euglena* DNA, 13 μg per lane, was hydrolyzed with *Bam*H1 (B), *Eco*R1 (E), or with both (*Bam*H1/*Eco*R1; B/E) restriction enzymes, as described in relevant sections of Materials and Methods. Restriction fragments were resolved by electrophoresis in a 0.5% agarose/1X TAE gel and visualized by ethidium bromide staining. The majority of the DNA fragments range in size between 1.0 and 48.5 kbp, based on the DNA size ladder (MW; a mixture of Invitrogen™ λ DNA/*Hin*dIII, 1 Kb Plus, and λ DNA/High Molecular Weight Markers). The distinct banding pattern within each lane stems from restriction fragments of the *Euglena* rDNA circle, p*Eg*rDNA [63,133], and from the mitochondrial genome. **B.** Southern transfer and hybridization with a *Euglena* U3 snoRNA gene probe reveals a minimum of 13 hybridizing fragments, ranging in size from ~1800 bp to ~38.5 kbp, in each of the *Bam*H1, *Eco*R1, and *Bam*H1/*Eco*R1 lanes. The *Bam*H1 and *Eco*R1 lanes display a larger than expected proportion of large hybridizing fragments (~6 fragments each in the 10 to 38.5 kbp size range), possibly reflecting a lower than expected frequency of *Bam*H1 and *Eco*R1 restriction sites in the *Euglena* genome. Hybridizing fragments in the *Bam*H1/*Eco*R1 lane display a relatively uniform size distribution, with some hybridizing bands co-migrating with hybridizing fragments in either the *Bam*H1 or *Eco*R1 lanes. The exact number of hybridizing fragments in the 10 to 38.5 kbp size range is difficult to determine conclusively owing to the limited resolution of this size range on standard agarose gels. Furthermore, the faint hybridization intensity (discussed in later sections) of some of the larger hybridizing bands makes them difficult to detect, even when using maximal amounts of target DNA (13 μg of DNA per 7 mm lane approaches the electrophoretic capacity limit of standard agarose gels) and long autoradiographic exposures in the presence of intensifying screens (up to 32 days).

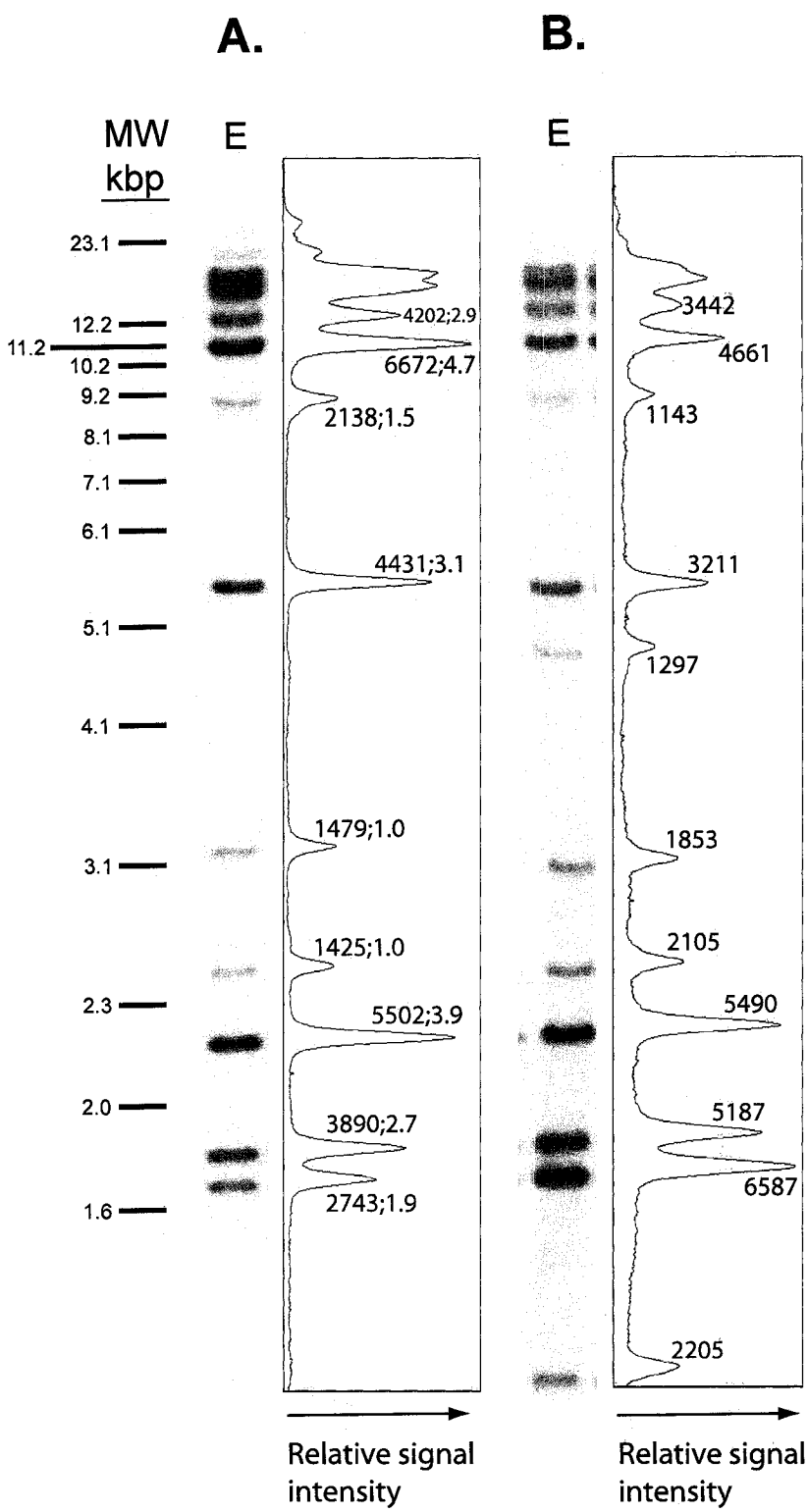

A.
B.
B E B/E
MW
kbp
MW
B E B/E
48.5
33.5
23.1
19.4
17.1
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7
1.0
0.2

*Eco*R1/*Bam*H1 fragments, ranging in size from ~1800 bp to ~38.5 kbp. The use of different restriction enzyme combinations and different electrophoresis conditions reveals the presence of up to 18 hybridizing bands (data not shown); many of these bands are in the 10 to 38.5 kbp size range and of faint hybridizing intensity, making them difficult to resolve under standard electrophoretic conditions.

The conclusion that the observed multiple hybridizing bands are legitimate is based on *i*) reproducible results from additional DNA preparations of separate *Euglena* cultures; *ii*) different restriction digestion conditions; *iii*) the use of reaction conditions favouring complete digestion, such as maximum amounts of enzyme, the addition of the stabilizing agents DTT and BSA, and extended (5-h) incubation times; and *iv*) stringent Southern hybridization washing conditions. Furthermore, multiple hybridizing bands were also obtained when *Euglena* DNA was hydrolyzed with *Bgl*II, *Cla*I, *Hin*dIII, *Pst*I, *Sal*I, and *Hae*III. Degraded and/or incompletely digested DNAs would be expected to yield additional hybridizing bands, unreproducible banding patterns and band intensities, fuzzy bands, and high background levels in Southern hybridization analysis. In addition, subsequent experiments, including genomic PCR amplifications and sequence data, further substantiate the authenticity of the observed multiple hybridization bands.

Unexpectedly, the multiple U3 snoRNA-hybridizing fragments show reproducible differences in hybridization signal intensities (see Figure 3.2), with this apparent non-stoichiometry being confirmed by densitometric analysis. Comparison of the signal intensities of the multiple hybridizing bands within each sample lane reveals a number of bands with very similar signal intensities. These include the ~2.5-, 3.1-, and 9.2-kbp hybridizing bands with relative signal intensities, as determined by the area under each peak, of 1425, 1479, and 2138, respectively (see Figure 3.2A). Similarly, hybridizing bands at ~1.8-, ~5.5-, and 12.2-kbp also display very similar relative signal intensities of 3890, 4431, and 4202, respectively (in Figure 3.2A). Thus, the relative signal intensities of the multiple hybridizing bands within a sample lane can also be very different. This is clearly illustrated by comparison of the ~1.7-, ~1.8-, ~2.2-, and ~2.5-kbp hybridizing bands, with relative signal intensities of 2743, 3890, 5502, and 1425, respectively, which are 1.9x, 2.7x, and 3.9x greater than that of the ~2.5-kbp band (1425). Similar results have also been obtained with

Figure 3.2 Densitometric Analysis Confirms that the Multiple U3 snoRNA-Hybridizing Bands in Southern Analysis of *Euglena* DNA Display Unequal Hybridization Signal Intensities. A. Total cellular DNA from *Euglena*, prepared as described in relevant sections of Materials and Methods. **B.** DNA preparation used by Greenwood *et al.* [38], enriched in nuclear DNA. Both DNA preparations were hydrolyzed with *Eco*R1 (E) prior to electrophoresis (on the same gel), Southern transfer, and hybridization with a *Euglena* U3 snoRNA gene probe, all as described above. Densitometric analysis of the hybridizing bands, done as described above using ImageJ [200], is shown to the right of each sample lane. The relative signal intensity of each hybridizing band corresponds to the area of the densitometry peak. For relevant hybridization bands, the area under the peak was determined and is indicated as the first numerical value near the peak. In **A.**, the second value represents the signal intensity of the specific hybridizing band relative (*i.e.*, 2X, 3X) to that of the hybridizing band having the lowest signal intensity in the lane, which corresponds to the 2.5-kbp hybridizing band with the relative signal intensity of 1425. Note that the slightly degraded DNA preparation (used in **B.**) may artifactually influence the densitometry values by contributing to higher background hybridization levels. MW; representative bands from a mixture of Invitrogen™ λ DNA/*Hin*dIII and 1 Kb DNA ladders.


A.
B.
MW
kbp
E
E
23.1
12.2
11.2
10.2
9.2
8.1
7.1
6.1
5.1
4.1
3.1
2.3
2.0
1.6
4202;2.9
6672;4.7
2138;1.5
4431;3.1
1479;1.0
1425;1.0
5502;3.9
3890;2.7
2743;1.9
3442
4661
1143
3211
1297
1853
2105
5490
5187
6587
2205
Relative signal intensity
Relative signal intensity

other hybridization probes (see Figures 3.6, 3.7, and 3.10). The reproducibility of the results obtained suggests that they are not attributable to incomplete restriction digestion of the DNA or to uneven transfer of restriction fragments from the gel to the membrane prior to Southern hybridization. Furthermore, minor sequence heterogeneities identified in the multiple U3 snoRNA gene variants (see section 3.6) are not believed to contribute significantly to the preferential hybridization of the U3 snoRNA gene probe to certain U3 snoRNA gene variants. Possible interpretations of this unexpected result will be presented in the Discussion.

Overall, the banding pattern obtained, including the different hybridization signal intensities, closely parallels the results obtained by Greenwood *et al.* [38]. However, two U3-hybridizing bands, at 1.1- and 4.8-kbp in the *Eco*R1 digest of the Southern hybridization pattern of Greenwood *et al.*, are absent from the analysis presented here (compare Figure 3.2A and B). Furthermore, densitometric comparisons of the relative signal intensities of the multiple co-hybridizing bands observed in the *Euglena* DNA samples of different origins (compare A and B in Figure 3.2) reveals changes in the relative hybridizing signal intensities of some, but not all, co-hybridizing bands. This is best exemplified by comparison of the ~1.7- and ~1.8-kbp hybridizing bands (in A and B of Figure 3.2). In the *Euglena* DNA used in this study (see Figure 3.2A), the ~1.7- and ~1.8-kbp hybridizing bands display relative signal intensities of 2743 and 3890, respectively. Thus, the ~1.7-kbp hybridizing band displays a lower signal intensity than that of the ~1.8-kbp band. This situation is reversed in the *Euglena* DNA used by Greenwood *et al.* [38] (see Figure 3.2B). In this circumstance, the ~1.7-kbp hybridizing band displays a higher relative signal intensity of 6587 in comparison to the ~1.8-kbp hybridizing band's lower relative signal intensity of 5187. Several factors may account for these minor discrepancies. Total *Euglena* DNA, prepared as described in the relevant section of Materials and Methods (see section 2.3), was used in the analysis presented here (Figure 3.2A). In contrast, the DNA used by Greenwood *et al.* [38] was enriched for nuclear DNA, as a by-product of a mitochondrial DNA preparation. However, the most likely explanation is that the observed differences may be attributable to clonal variations that may have accumulated in the two independently passaged cell populations from which the DNAs were harvested. Over the course of multiple independent

generations of growth, such clonal variations may have included changes in ploidy and chromosome number (see section 4.6 of the Discussion), chromosomal translocations and unequal crossing-over, and nucleotide changes (that could create or destroy restriction sites).

3.2: Screening of a λ Library of *Euglena* Genomic DNA: Isolation of Genomic Clones and Description of U3 snoRNA Gene Arrangements Identified

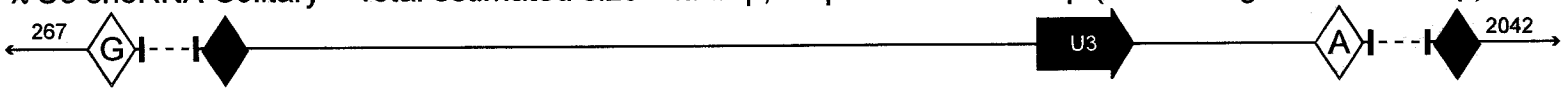
The large number of U3-hybridizing bands in Southern analysis of *Euglena* DNA prompted a more detailed study of the copy number and genomic organization of U3 snoRNA genes in the *Euglena* genome. Thus, a search for U3 snoRNA gene sequences was initiated by extensively screening a *Euglena* genomic λ library with a *Euglena* U3 snoRNA gene probe, which revealed multiple U3 snoRNA-hybridizing λ clones. Sequence walking in conjunction with BLAST analysis identified four unique U3 snoRNA gene variants in three different genomic contexts (see Figure 3.3). These U3 snoRNA genomic arrangements comprise *i*) a stand-alone gene, *ii*) linkage to tRNA genes, and *iii*) linkage to a U5 snRNA gene.

A solitary U3 snoRNA gene arrangement was identified in a large ~12-kbp genomic fragment (see Figure 3.3 - λ U3 snoRNA-Solitary). No additional, recognizable genes or nucleotide sequence elements were found by BLAST analysis of this genomic insert. This clone was sequenced by primer walking and by subcloning regions of the genomic fragment; however, a number of nucleotide regions proved difficult to sequence owing to the presence of repetitive sequence elements and to extreme biases in nucleotide composition, such as oligo-G stretches. Thus, substantial portions of this λ genomic clone remain unsequenced. Regardless, at least one U3 snoRNA gene appears to be encoded as a solitary gene in the *Euglena* genome. There appears to be no additional U3 snoRNA genes on this λ genomic insert.

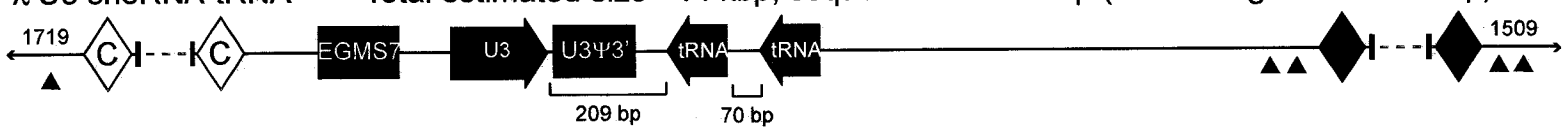
A U3-hybridizing λ clone was found to encode a U3 snoRNA gene neighbored downstream by two identical arginine tRNA genes ($tRNA^{Arg}_{UCG}$) encoded in the opposite transcriptional orientation (*i.e.*, convergently oriented) relative to the U3 snoRNA gene (see Figure 3.3 - λ U3 snoRNA-$tRNA^{Arg}$). The first $tRNA^{Arg}$ gene is located 209-nt downstream of the U3 snoRNA gene, whereas the second $tRNA^{Arg}$ gene is encoded 70-nt downstream of the first $tRNA^{Arg}$ gene. No additional, recognizable genes were found by BLAST analysis

Figure 3.3 Three Different U3 snoRNA Gene Arrangements in *Euglena* Genomic λ Clones: Solitary, Linked to $tRNA^{Arg}_{UCG}$ Genes, and Linked to U5 snRNA Genes. Clone maps of λ U3 snoRNA-Solitary, λ U3 snoRNA-$tRNA^{Arg}$, λ U3 snoRNA-U5 snRNA-A, and λ U3snoRNA-U5 snRNA-B are shown. Genes, and their orientation of transcription, are depicted by arrows: U3 snoRNA in red, $tRNA^{Arg}$ in blue, and U5 snRNA in green. Numbers shown between the U3 snoRNA-U5 snRNA and the U3 snoRNA-$tRNA^{Arg}$ genes corresponds to the size of the intergenic spacer separating the two genes. Various repetitive sequence elements are diagrammed. Homonucleotide stretches (*i.e.*, oligo-N regions) that prematurely terminate sequencing reactions, thus preventing the acquisition of downstream sequence, are illustrated by a diamond with the relevant nucleotide. Simple repeat regions that made the design of primer walking oligonucleotides impossible, and thus prevented the acquisition of additional sequence, are indicated by a filled diamond. Other repeat structures, such as the *Euglena EG*MS7 microsatellite repeat [210], repetitive sequences also found within certain *Euglena* γ-tubulin introns, and the U3 snoRNA pseudo-3′-end repeat (U3Ψ3′), are also indicated. Solid lines represent sequenced regions; dashed lines represent unsequenced regions. Note that sequence elements are not drawn to scale, and U3* in the λ U3snoRNA-U5 snRNA-B refers to a truncation in the 5′-region of the U3 snoRNA gene as a result of cloning. The large λ U3 snoRNA-Solitary and λ U3 snoRNA-$tRNA^{Arg}$ clones have not been sequenced in their entirety due to the technical challenges posed by the outlined repetitive sequence elements. The two U3 snoRNA-U5 snRNA-encoding genomic inserts are unexpectedly small owing to technical difficulties related to λ cloning. Extensive sequence analysis has revealed that these λ clones contain chimeric inserts comprising stuffer fragment and *Euglena* genomic DNA. As discussed in the relevant section of Materials and Methods (see section 2.12.1), the insert DNA was not size fractionated, as recommended by the manufacturer. Thus, it is believed that small genomic DNA fragments may have ligated with stuffer fragments and vector arms to form packagable λ phage molecules. While it cannot be definitively excluded, these small genomic λ inserts are not thought to be the result of aberrant recombination events. Regardless, the validity of these small genomic fragments has been established by subsequent Southern hybridization analysis and genomic PCR studies.

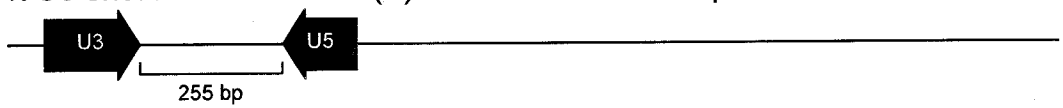
λ U3 snoRNA-Solitary Total estimated size ~12 kbp; sequenced - 4824 bp (centre fragment - 2515 bp)



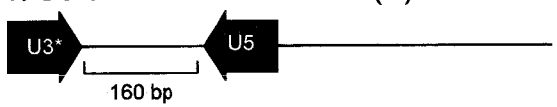
λ U3 snoRNA-tRNA^{Arg} Total estimated size ~14 kbp; sequenced - 6096 bp (centre fragment - 2868 bp)



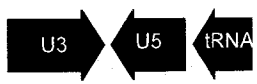
λ U3 snoRNA-U5 snRNA (A) Total size - 1772 bp



λ U3 snoRNA-U5 snRNA (B) Total size - 516 bp

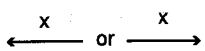


LEGEND



gene, indicating the orientation of transcription

repetitive elements



flanking sequence (containing no features of interest) of length x bp

oligo-N (A, G, or C) region that prematurely terminates sequencing reactions

simple heteronucleotide repeats that prevent effective primer design for additional primer walking

partial α-tubulin intron repeat sequence also found within one of the α-tubulin introns

unsequenced region of the λ clone

of this genomic insert. The U3 snoRNA-tRNA^{Arg} λ clone contains many of the short repetitive sequence elements identified in the other λ clones. In addition, this clone contains several other interesting repetitive sequence features, such as repeat sequences that are also found within the introns of the *Euglena* γ-tubulin gene paralogues [211]. The region upstream of the U3 snoRNA gene also contains a recently identified *Euglena* microsatellite sequence, *EG*MS7 [210], one of the first microsatellite sequences identified in the *Euglena* genome. Furthermore, a 52-nt region, consisting of the 3′-end and transcription termination sequence of this particular U3 snoRNA gene, appears to have been duplicated. The resulting sequence consists of a full-length U3 snoRNA gene and transcription termination signal followed by the 3′-terminal 22-nt of the U3 snoRNA gene (nt 159 to 180) and a nearly identical copy of the transcription termination signal. As with the solitary U3 snoRNA-encoding genomic fragment, this large ~14-kbp λ genomic clone contains large unsequenced regions owing to the presence of repetitive sequence elements and regions of extreme nucleotide bias that prove difficult to sequence.

The identified $tRNA^{Arg}_{UCG}$ gene (see Figure 3.4) is unremarkable in its sequence and secondary structure. It adopts a conventional clover-leaf secondary structure, displays the expected D and TΨC loops, and possesses a small loop in the variable region.

Two different U3-hybridizing λ clones were identified in which the U3 snoRNA gene is neighbored downstream by a convergently oriented U5 snRNA gene (see Figure 3.3 - λ U3 snoRNA-U5 snRNA-A and -B). No additional, recognizable genes were found by BLAST analysis of these genomic inserts. Only limited sequence is available upstream of the U3 snoRNA gene in the λ U3 snoRNA-U5 snRNA-A clone whereas sequence for the upstream and 5′-terminal regions of the U3 snoRNA gene are not available for the λ U3 snoRNA-U5 snRNA-B clone. While the U3 snoRNA and U5 snRNA gene sequences in the two genomic λ clones are highly similar, the size and sequence of the U3 snoRNA-U5 snRNA intergenic spacers are very different. In the λ U3 snoRNA-U5 snRNA-A and -B clones, the U5 snRNA gene is encoded 255- and 160-nt, respectively, downstream of the U3 snoRNA gene. Furthermore, the sequence of the regions downstream of the U5 snRNA genes is equally dissimilar in both λ genomic inserts. As with the other λ clones, both U3 snoRNA-U5 snRNA genomic λ fragments also contain many short repetitive sequence

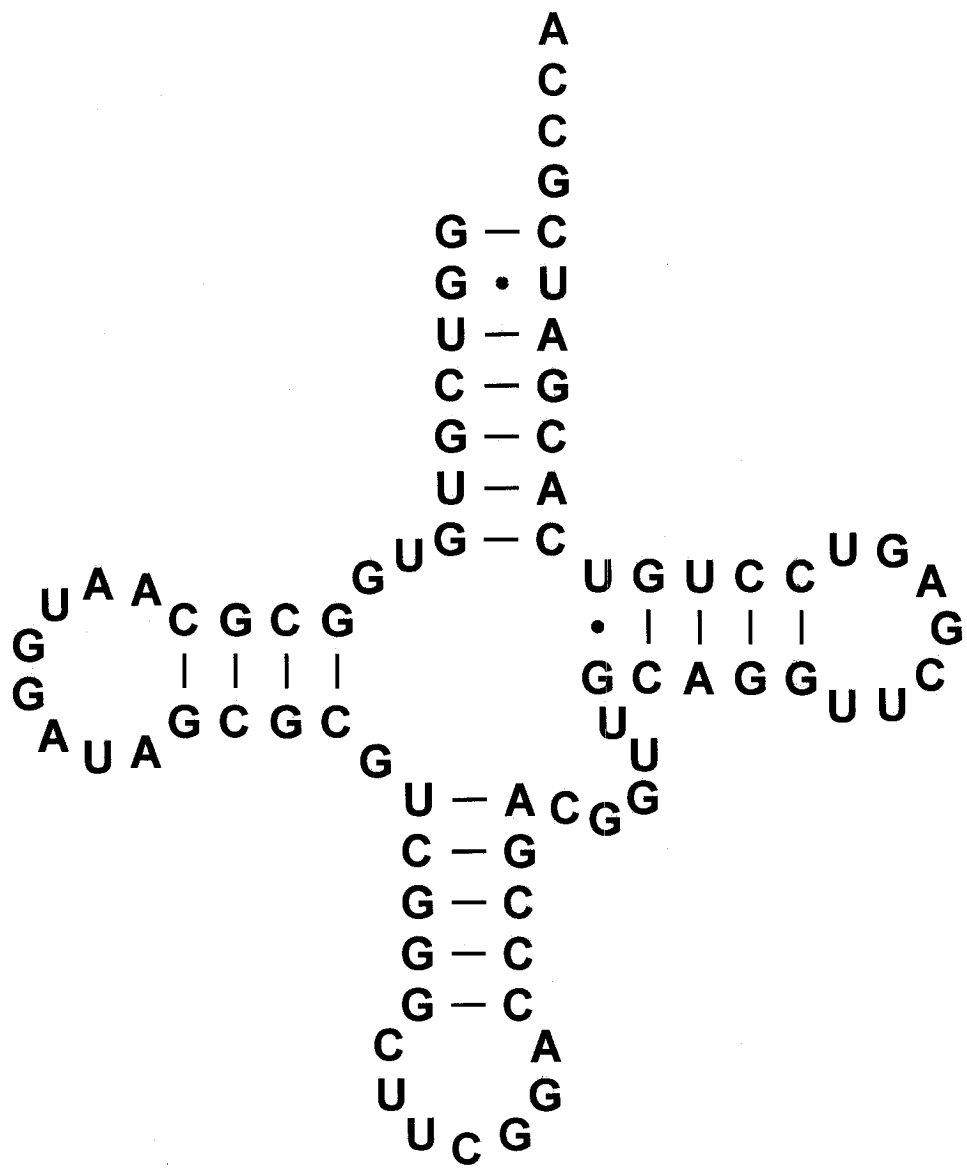


Figure 3.4 The *Euglena* $tRNA^{Arg}_{UCG}$ Adopts a Conventional Clover-Leaf Secondary Structure. In this proposed secondary structure model, conventional Watson-Crick base-pairing interactions are depicted as dashes (–) and G/U pairings with dots (•). The anti-codon nucleotides, UCG, are highlighted in red. As with other eukaryotic nucleus-encoded tRNAs, the 3′-terminal nucleotides (CCA, highlighted in blue) are not encoded in the genome. They are added post-transcriptionally during tRNA processing and maturation [212], though this has not been verified for this particular *Euglena* tRNA.

elements and regions of extreme nucleotide bias that proved difficult to sequence. However, the small size of these λ inserts (1772-bp for λ U3 snoRNA-U5 snRNA-A and 516-bp for λ U3 snoRNA-U5 snRNA-B) facilitated their complete sequence determination (see caption, Figure 3.3).

The sequence for the *Euglena* U5 snRNA gene was previously unknown. Its nucleotide sequence and secondary structure (see Figure 3.5) display features present in U5 snRNAs from other organisms. The *Euglena* U5 snRNA is 98-nt in length. It 5′-end has been inferred based on comparison with other U5 snRNA sequences. However, its precise 3′-end has been determined by 3′ RACE analysis and by chemical RNA sequencing (by Dr. Murray N. Schnare, Dalhousie University). The secondary structure of the *Euglena* U5 snRNA consists, in its 5′-region, of an extended stem-loop structure punctuated by a central bulge-loop. The 11-nt terminal loop I of the 5′-stem-loop structure contains the invariant 9-nt sequence (5′-GCCUUUUAC-3′) known to interact with exon sequences at the 5′- and 3′-splice sites [213]. In other organisms, this region is known to contain both 2′-*O*-methylated and pseudouridylated residues [214]; however, the modified nucleoside content of the *Euglena* U5 snRNA has not been examined. The 3′-region of *Euglena* U5 snRNA contains a conventional, unstructured, Sm binding site sequence. Notably, a small stem-loop structure, typically present near the 3′-end of U5 snRNAs, is not found in the *Euglena* U5 snRNA.

3.3: Southern Hybridization Analysis Suggests that U3 snoRNA Genes are Frequently Linked to U5 snRNA Genes in the *Euglena* Genome

Screening of the *Euglena* λ genomic DNA library identified four different U3 snoRNA genes in three different genomic contexts. Yet, Southern analysis of the *Euglena* genome reveals the presence of at least 13 (and up to 18) U3-hybridizing bands. Thus, many U3 snoRNA genes (and their genomic arrangements) remained unaccounted. Hence, Southern hybridization analysis was done to determine if additional variants of the linkages identified in the λ genomic fragments are present in the *Euglena* genome.

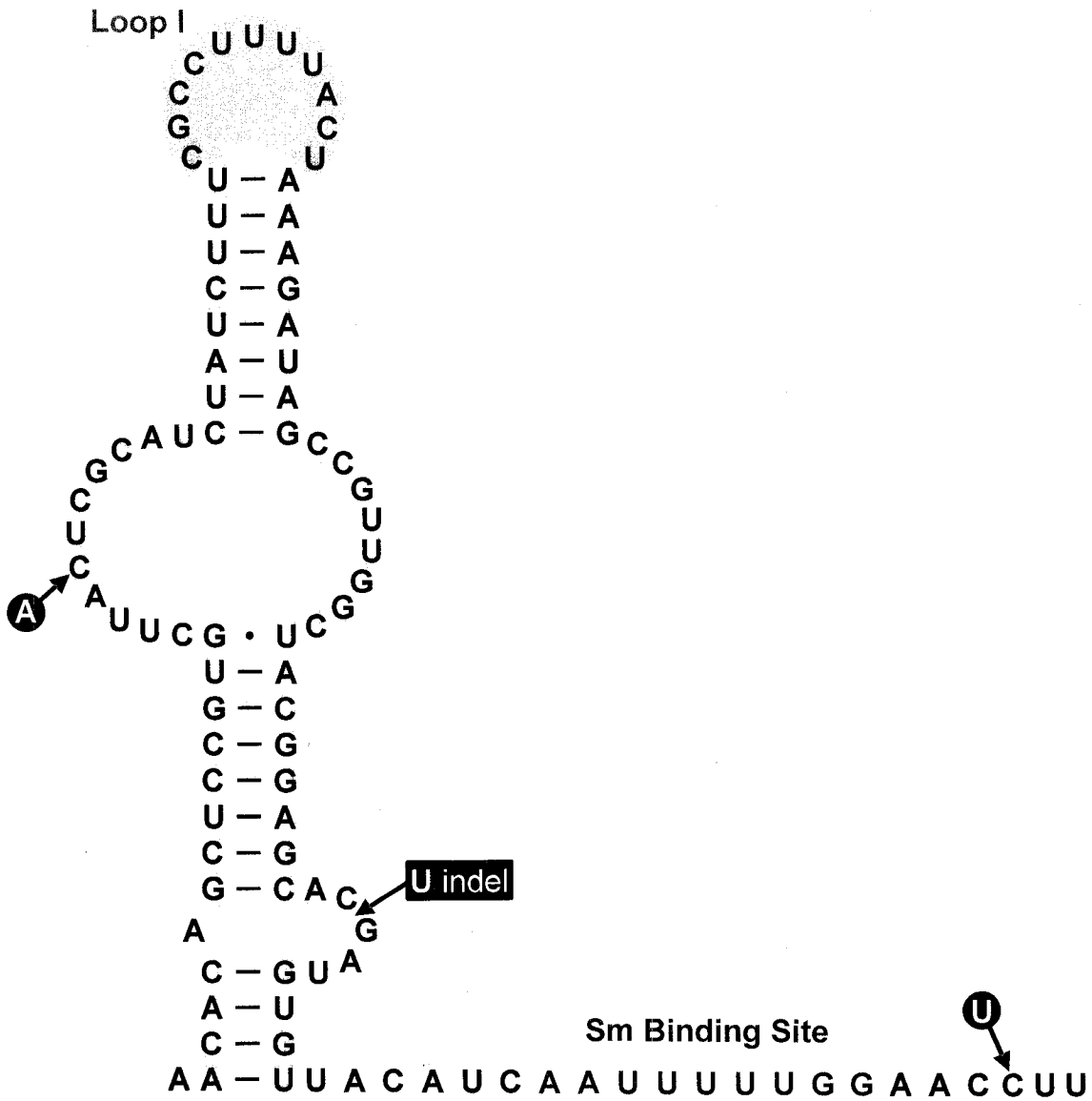


Figure 3.5 *Euglena* U5 snRNA Displays Conventional Secondary Structure Features. In this proposed *Euglena* U5 snRNA secondary structure model, the phylogenetically conserved and functionally essential exon-interacting nucleotides (in green) of Loop I (highlighted on a gray background), and the Sm binding site sequence (in blue) are highlighted. Nucleotide heterogeneities and an indel identified in the *Euglena* U5 snRNA gene variants are indicated in red. Conventional Watson-Crick base-pairing interactions are depicted as dashes (–) and G/U pairings are identified by dots (•).

Southern hybridization analysis of *Euglena* DNA with a $tRNA^{Arg}_{UCG}$ gene probe reveals the presence of multiple hybridizing bands (12 in *Bam*HI/*Eco*RI, see Figure 3.6), suggesting that $tRNA^{Arg}_{UCG}$ is a multi-copy gene in the *Euglena* genome. This is not unexpected, considering that tRNA genes frequently form large, multi-gene families.

When the Southern hybridization result using the $tRNA^{Arg}$ probe is compared to that using the U3 probe, very few co-hybridizing bands are identified (see Figure 3.6). Thus, while one example of a U3 snoRNA-$tRNA^{Arg}$ linkage is present in the *Euglena* genome, other members of the $tRNA^{Arg}$ gene family do not appear to be similarly linked to U3 snoRNA genes.

The authenticity of the U3-$tRNA^{Arg}$ co-hybridizing band is further substantiated by an additional Southern hybridization analysis. A probe, derived from the upstream region of the U3 snoRNA gene in the U3 snoRNA-$tRNA^{Arg}$ λ clone, predominantly hybridizes to the band that co-hybridizes with the U3 snoRNA and the $tRNA^{Arg}$ probes. However, this probe also contains a *Euglena* microsatellite sequence [210]. This may explain the high level of background hybridization obtained with this probe, despite extensive, stringent washing of the Southern nylon membrane.

Southern hybridization analysis of the *Euglena* genome with a U5 snRNA probe identifies at least 13 hybridizing bands ranging in size between 33.5 and 1.0 kbp (see Figure 3.7). Thus, U5 snRNA is also a multi-copy gene in the *Euglena* genome.

Comparison of these results with those obtained with a U3 probe unexpectedly reveals at least seven co-hybridizing bands. Thus, the majority, though not all, of U5 snRNA genes may be linked to U3 snoRNA genes in the *Euglena* genome.

In addition, as observed with the U3-hybridizing bands, the U5-hybridizing bands also showed reproducible differences in hybridization intensities. The relative hybridization signal intensities within the U5 pattern co-vary with those within the U3 pattern.

Figure 3.6 Southern Hybridization Analysis of *Euglena* DNA Reveals Few U3 snoRNA-tRNA^{Arg} Gene Linkages. Total *Euglena* DNA, 13 μg per lane, was hydrolyzed with *Bam*H1 (B), *Eco*R1 (E), or with both (*Bam*H1/*Eco*R1; B/E) restriction enzymes, as described in relevant sections of Materials and Methods. Restriction fragments were resolved by electrophoresis in a 0.5% agarose/1X TAE gel prior to Southern transfer and hybridization with probes corresponding to the *Euglena* U3 snoRNA gene, tRNA^{Arg} gene, and to a region upstream of the U3 snoRNA gene (UpStr U3) containing a *Euglena* microsatellite sequence. (MW; a mixture of Invitrogen™ λ DNA/*Hin*dIII, 1 Kb Plus, and λ DNA/High Molecular Weight Markers). Southern hybridization with a *Euglena* tRNA^{Arg} gene probe reveals 9-12 hybridizing fragments, ranging in size from ~2.5 kbp to ~33.5 kbp, in the *Bam*H1, *Eco*R1, or *Bam*H1/*Eco*R1 lanes. Southern hybridization with a probe corresponding to a region upstream of the U3 snoRNA gene containing a microsatellite sequence reveals one prominent band and a large number of faint bands. The prominent band is thought to correspond to the particular genomic sequence from which it is derived, while the faint bands are thought to represent heterologous hybridization of the probe's microsatellite sequence to other microsatellite sequences in the *Euglena* genome. Bands co-hybridizing with different probes are indicated by red asterisks.

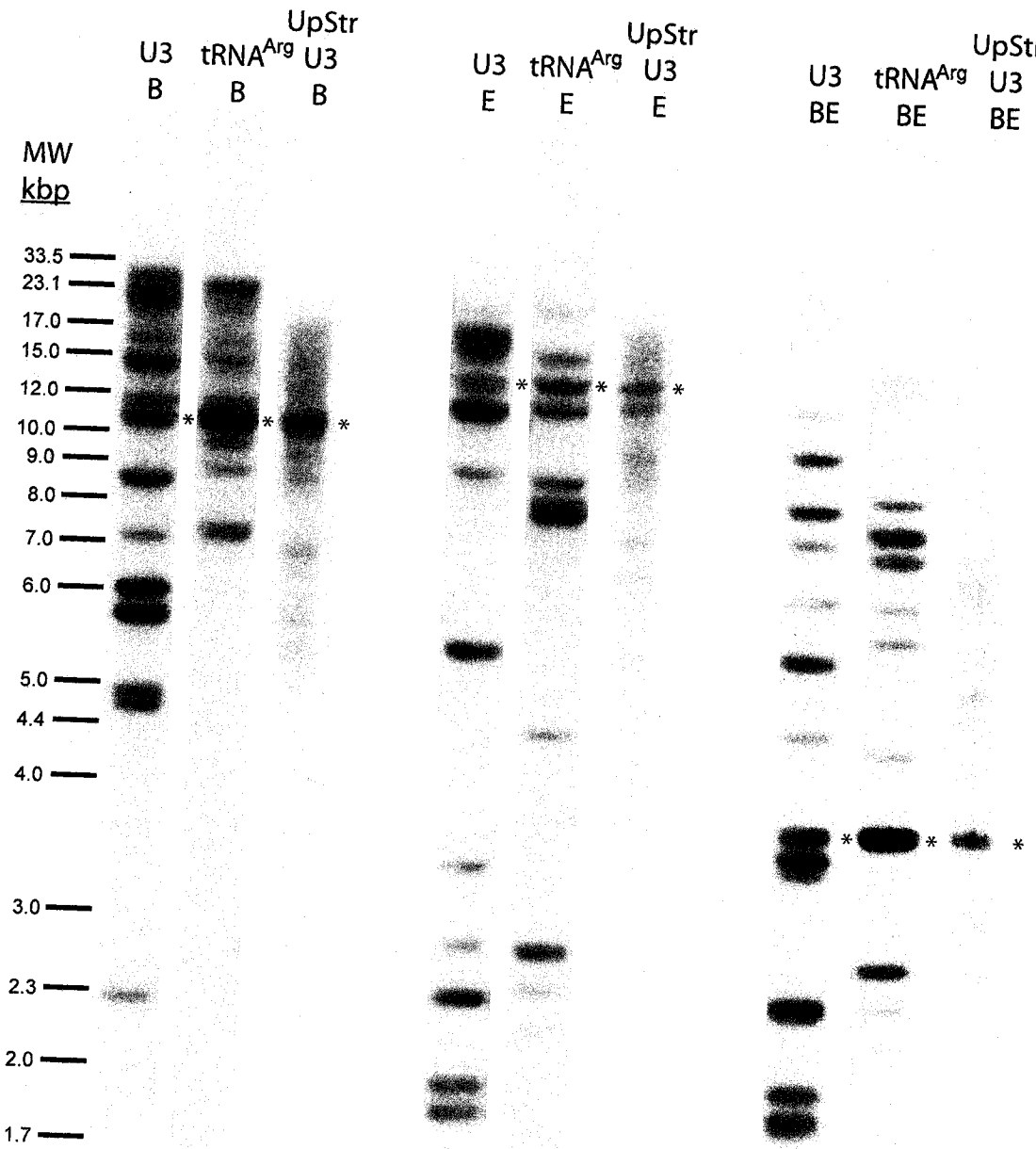
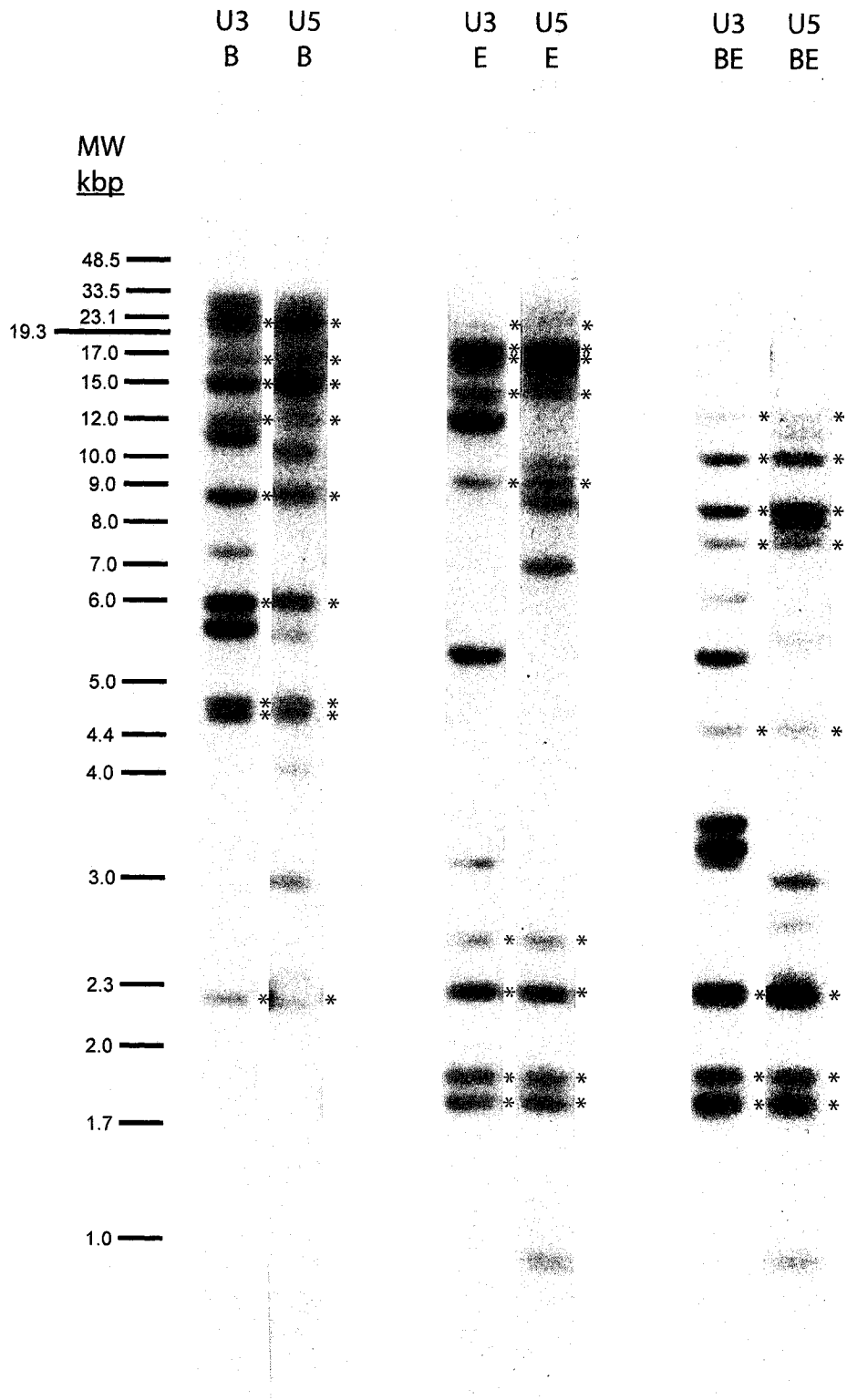

U3 B
tRNA^Arg B
UpStr U3 B
U3 E
tRNA^Arg E
UpStr U3 E
U3 BE
tRNA^Arg BE
UpStr U3 BE
MW
kbp
33.5
23.1
17.0
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7

Figure 3.7 Southern Hybridization Analysis of *Euglena* DNA Reveals Multiple U3 snoRNA-U5 snRNA Gene Linkages. Total *Euglena* DNA, 13 μg per lane, was hydrolyzed with *Bam*H1 (B), *Eco*R1 (E), or with both (*Bam*H1/*Eco*R1; B/E) restriction enzymes, as described in relevant sections of Materials and Methods. Restriction fragments were resolved by electrophoresis in a 0.5% agarose/1X TAE gel prior to Southern transfer and hybridization with *Euglena* U3 snoRNA or U5 snRNA gene probes. (MW; a mixture of Invitrogen™ λ DNA/*Hin*dIII, 1 Kb Plus, and λ DNA/High Molecular Weight Markers). Southern hybridization with a *Euglena* U5 snRNA gene probe reveals 13-15 hybridizing fragments, ranging in size from ~0.9 kbp to ~33.5 kbp, in the *Bam*H1, *Eco*R1, or *Bam*H1/*Eco*R1 lanes. Bands co-hybridizing with different probes are indicated by red asterisks.


U3 B
U5 B
U3 E
U5 E
U3 BE
U5 BE
MW
kbp
48.5
33.5
23.1
19.3
17.0
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7
1.0

3.4: Genomic PCR Confirms Multiple U3 snoRNA-U5 snRNA Gene Linkages in the *Euglena* Genome

The multiple U3 snoRNA-U5 snRNA co-hybridizing bands warranted a more detailed sequence analysis of the putative linkages. Using a genomic PCR strategy, U3 snoRNA-U5 snRNA gene linkages were amplified using the oligonucleotide primer pair o*Eg*U3-F1 and o*Eg*U5-F1, designed to hybridize to the 5′ region of these convergently oriented genes. Multiple PCR amplification products were obtained (see Figure 3.8) and cloned in the TOPO-TA vector system (Invitrogen™). In order to ensure adequate coverage of all U3 snoRNA-U5 snRNA gene linkages, a total of 122 clones, from five different PCR-generated libraries, were sequenced and analyzed.

The cloned U3 snoRNA-U5 snRNA PCR products ranged in size between 422 and 543 bp. Sequence analysis revealed the presence of many different U3 snoRNA-U5 snRNA gene linkages, along with multiple copies of each linkage. Thus, it can be assumed that sequence differences between the different U3 snoRNA-U5 snRNA linkages are legitimate and not attributable to *Taq*-induced errors or to PCR-mediated recombination [215-217].

Detailed analysis of the cloned U3 snoRNA-U5 snRNA PCR products identifies a total of 12 unique U3 snoRNA-U5 snRNA gene linkages. A comparison of the unique sequences reveals that the multiple U3 snoRNA gene sequences are highly similar, as are the multiple U5 snRNA gene sequences, the members of each group displaying only limited sequence heterogeneities (see Figure 3.9). However, significant variability is seen in the size and sequence of the intergenic spacers (IGS) separating the U3 snoRNA and U5 snRNA genes.

Despite the sequence variability seen in the U3 snoRNA-U5 snRNA intergenic spacers, regions of sequence similarity within the IGSs suggest that the gene linkages may be related. Thus, the 12 unique U3 snoRNA-U5 snRNA gene linkages appear to form five families that can be further divided into sub-families (see Figure 3.9).

The U3 snoRNA-U5 snRNA A family, comprising the longest of the linkages, is composed of four members. Based on sequence similarity, the members of this family can be divided into two sub-families, A1 and A2, with the A1 subfamily being further divided

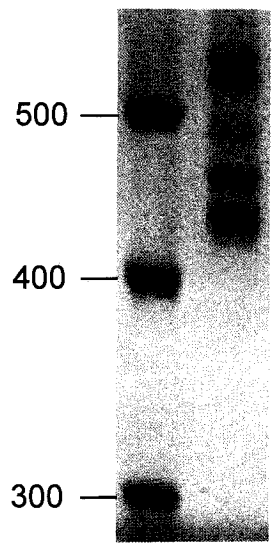
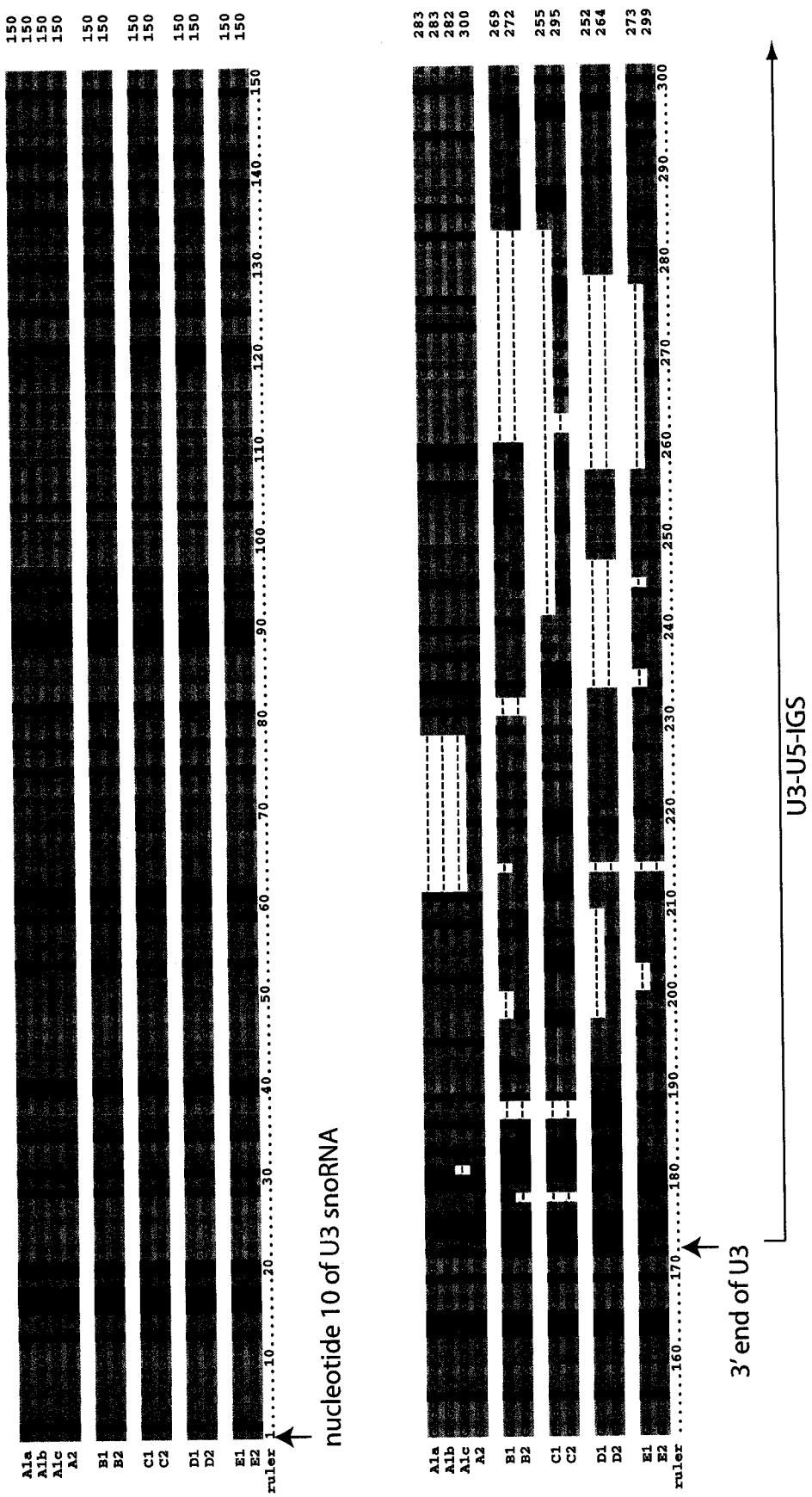
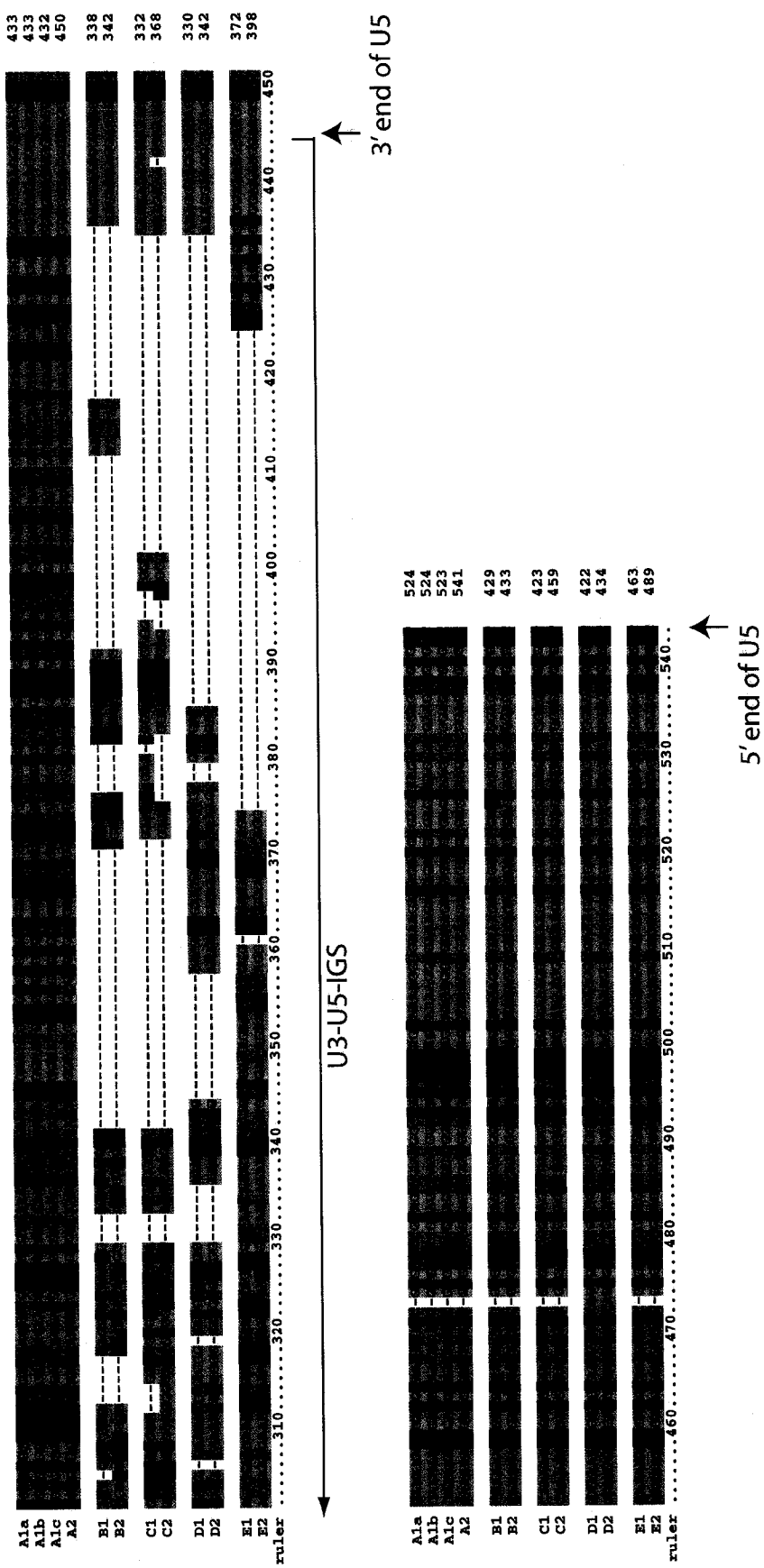


Figure 3.8 Multiple U3 snoRNA-U5 snRNA Amplification Products from *Euglena* Genomic PCR. PCR amplification of total Euglena DNA using oligonucleotide primers designed to the 5'-region of both U3 snoRNA and U5 snRNA (o*Eg*U3-F1 and o*Eg*U5-F1, see Table 2.1) yields the observed multiple banding pattern, with bands ranging in size from ~425- to ~550-bp. The PCR products were resolved by electrophoresis in a 2.5% agarose/1X TBE gel.

Figure 3.9 Sequence Alignment of the Multiple U3 snoRNA-U5 snRNA Gene Linkages Shows Clustering into related Families and Sub-Families. The alignment shows the high sequence conservation of the U3 snoRNA and U5 snRNA genes, both within a linkage family (*i.e.*, compare U3 snoRNA and U5 snRNA genes in A1a and A1b) and between linkage families (*i.e.*, compare U3 snoRNA and U5 snRNA genes A1a and B1). The sequences of the U3 snoRNA-U5 snRNA intergenic spacers (IGSs) show sequence conservation within a linkage family (*i.e.*, compare the U3 snoRNA-U5 snRNA IGS in A1a and A1b) but no sequence similarity when compared between linkage families (*i.e.*, compare the IGS in A1a and B1). The alignment begins at nucleotide 10 of the U3 snoRNA gene as a consequence of PCR amplification using the oligonucleotide primer o*Eg*U3-F1 (see Table 2.1). The 3′-end of the U3 snoRNA gene, 3′-end of the U5 snRNA gene, and the 5′-end of the U5 snRNA gene are indicated.


A1a
A1b
A1c
A2
B1
B2
C1
C2
D1
D2
E1
E2
ruler
150
283
269
255
252
273
nucleotide 10 of U3 snoRNA
3' end of U3
U3-U5-IGS


A1a
A1b
A1c
A2
B1
B2
C1
C2
D1
D2
E1
E2
ruler
433
433
432
450
338
342
332
368
330
342
372
398
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
U3-U5-IGS
3′ end of U5
524
524
523
541
429
433
423
459
422
434
463
489
460
470
480
490
500
510
520
530
540
5′ end of U5

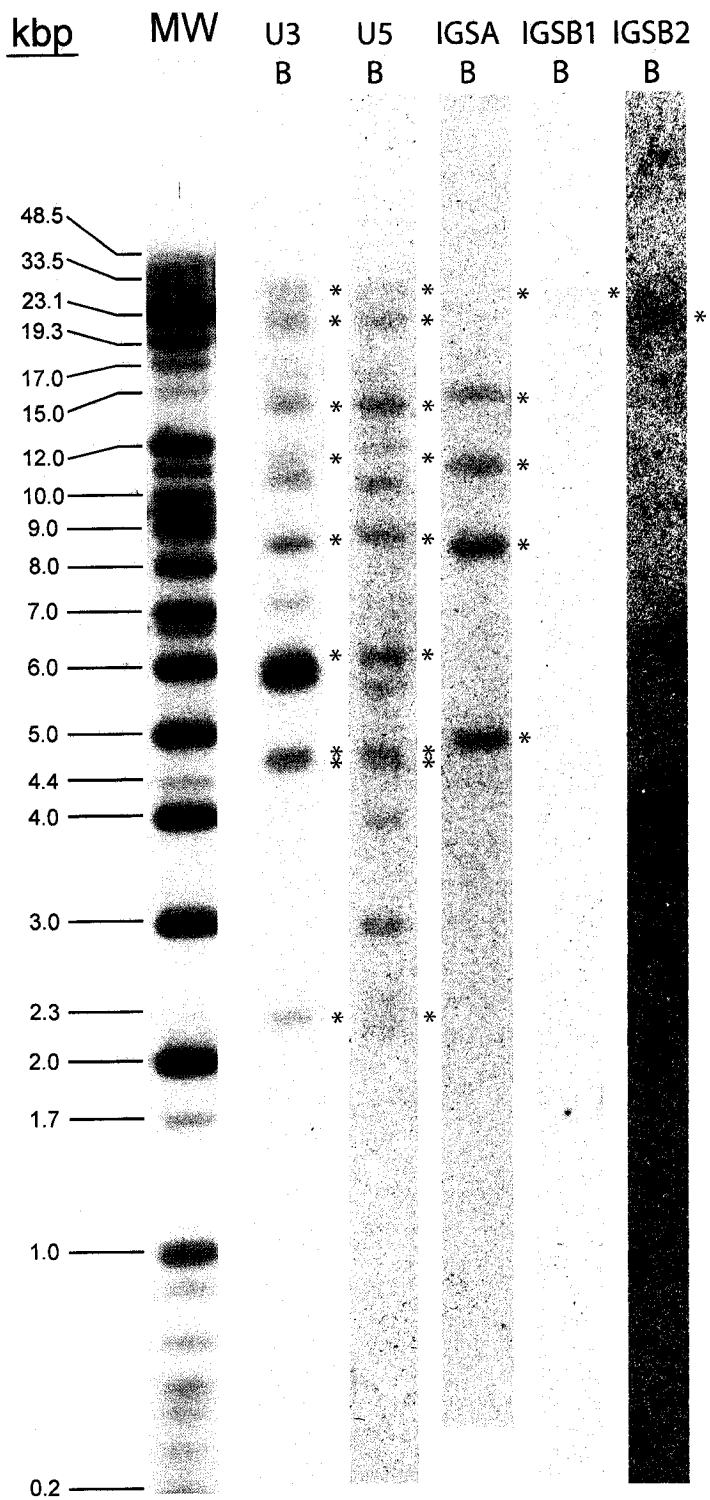
into A1a, A1b and A1c. Members of the A family show the highest level of identity, with only minor nucleotide changes and a 17-nt indel in the U3 snoRNA-U5 snRNA IGS sequence. The U3 snoRNA-U5 snRNA A1a PCR linkage corresponds to the U3 snoRNA-U5 snRNA A linkage identified in the λ genomic clones.

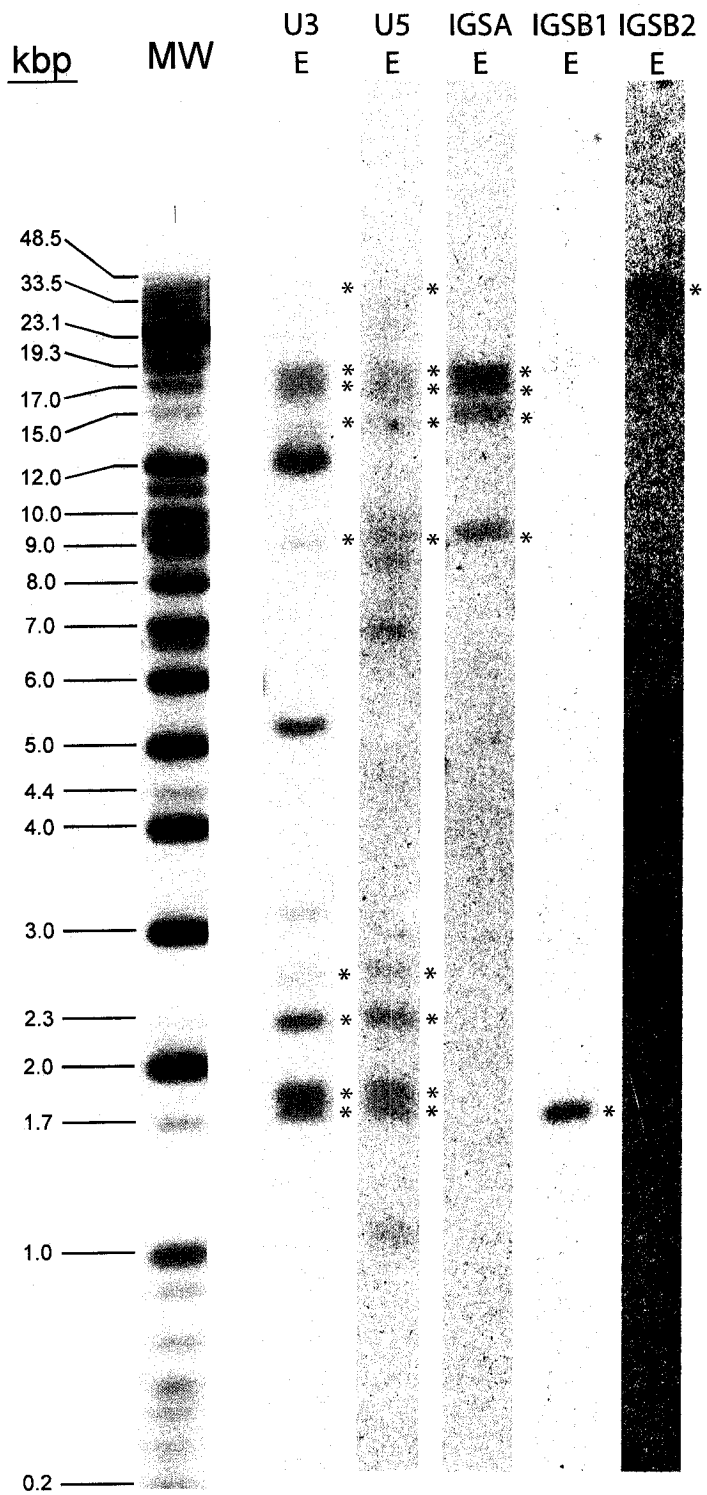
The B-linkage family contains two members, B1 and B2. The IGS sequences of the two linkages show the lowest level of sequence identity of any of the linkage families. The relatedness of the members of the B-linkage family is based on the presence of short regions of sequence identity punctuated by regions of nucleotide differences. Short regions of apparent sequence similarity at both ends of the IGSs, immediately downstream from the 3′-ends of the U3 snoRNA and U5 snRNA genes, may correspond to transcription termination signals; thus, they may not contribute to linkage family relatedness. The B1 PCR linkage corresponds to the U3 snoRNA-U5 snRNA B linkage identified in the λ genomic clones.

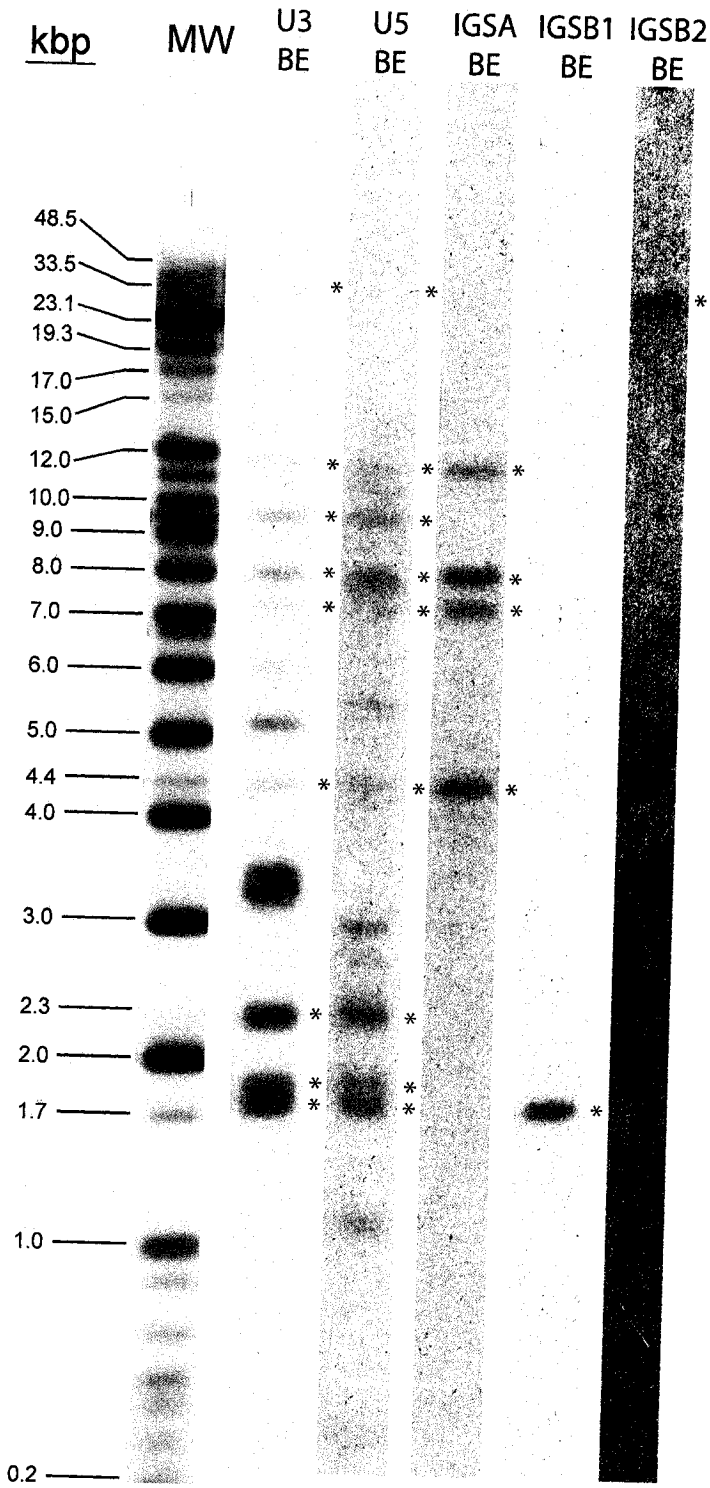
The C-, D-, and E-linkage families each contain two members, C1 and C2, D1 and D2, and E1 and E2. Each linkage family shows regions of sequence similarity punctuated by regions of nucleotide differences and by indels. Overall, the level of sequence identity in the members of the C, D and E linkage families is intermediate between that of the highly similar A family and the very different B family. Many of the U3 snoRNA-U5 snRNA gene linkages were confirmed by Southern hybridization analysis (see Figure 3.10)

A similar genomic PCR approach was used to search for different arrangements of linked U3 snoRNA-U5 snRNA genes, such as divergently (tail-to-tail) and similarly (tail-to-head) oriented genes. The presence of tandem U3 snoRNA genes was also investigated. This strategy was not extensively pursued because all initial attempts yielded no amplification products.

Figure 3.10 Southern Hybridization Analysis of U3 snoRNA, U5 snRNA, and Select U3 snoRNA-U5 snRNA Linkage Intergenic Spacers. Total *Euglena* DNA, 13 μg per lane, was hydrolyzed with *Bam*H1 (B), *Eco*R1 (E), or with both (*Bam*H1/*Eco*R1; B/E) restriction enzymes, as described in relevant sections of Materials and Methods. Restriction fragments were resolved by electrophoresis in a 0.5% agarose/1X TAE gel prior to Southern transfer and hybridization with a relevant probe. The *Euglena* U3 snoRNA and U5 snRNA gene probes used in Figure 3.7 were also used here. In addition, probes corresponding to the intergenic spacer sequence of the U3 snoRNA-U5 snRNA linkage families A and B (both sub-families B1 and B2) were also used. These probes frequently hybridized to large restriction fragments and gave low signal intensities, despite long autoradiographic exposures. The autoradiographs shown here (only for the intergenic spacer probes) were digitally enhanced to assist in the identification of the hybridizing band(s). (MW; a mixture of Invitrogen™ λ DNA/*Hin*dIII, 1 Kb Plus, and λ DNA/High Molecular Weight Markers). Bands co-hybridizing with different probes are indicated by red asterisks.


kbp
MW
U3 B
U5 B
IGSA B
IGSB1 B
IGSB2 B
48.5
33.5
23.1
19.3
17.0
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7
1.0
0.2


kbp
MW
U3
E
U5
E
IGSA
E
IGSB1
E
IGSB2
E
48.5
33.5
23.1
19.3
17.0
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7
1.0
0.2


kbp
MW
U3 BE
U5 BE
IGSA BE
IGSB1 BE
IGSB2 BE
48.5
33.5
23.1
19.3
17.0
15.0
12.0
10.0
9.0
8.0
7.0
6.0
5.0
4.4
4.0
3.0
2.3
2.0
1.7
1.0
0.2

3.5: U3 snoRNA, U5 snRNA, and $tRNA^{Arg}$ Genes are Expressed in *Euglena*

Euglena U3 snoRNA was initially identified at the RNA level [38] and is therefore known to be expressed. $tRNA^{Arg}$ and U5 snRNA are essential components of the translation and splicing machinery, respectively. Thus, there are no *a priori* reasons to assume that these genes not expressed in *Euglena*. Nonetheless, their expression status was determined by northern hybridization analysis.

Total *Euglena* RNA, resolved by electrophoresis and transferred to a nylon membrane, was hybridized with labeled U3 snoRNA, U5 snRNA or $tRNA^{Arg}$ anti-sense oligonucleotide probes (see Figure 3.11). In cells, U3 is the most abundantly transcribed snoRNA. Thus, U3 snoRNA, seen as a ~180-nt band in the ethidium bromide-stained RNA gel, hybridizes strongly to the U3 probe. Hybridization with a U5 snRNA-specific probe revealed a small, ~100-nt hybridizing band of the expected size (98 nt). Similarly, hybridization with a $tRNA^{Arg}$ probe identified a tRNA-sized, ~75-nt hybridizing band of the expected size (76 nt). Thus, U5 snRNA and $tRNA^{Arg}$ genes, as well as U3 snoRNA genes, are abundantly expressed in *Euglena*.

3.6: A Revised Secondary Structure Model for *Euglena* U3 snoRNA

The sequence heterogeneities identified by sequence analysis of cloned U3 snoRNA variants has allowed a re-assessment of the likely secondary structure of *Euglena* U3 snoRNA. Greenwood *et al.* [38] established the identity, basic sequence features and potential secondary structure of the *Euglena* U3 snoRNA. Recent phylogenetic comparison of U3 snoRNA gene sequences from representative taxa has uncovered additional conserved sequence elements and RNA motifs that are also present in the *Euglena* U3 snoRNA (see Figure 3.12). Furthermore, additional information from mutational studies [48,52] and from *in vivo* chemical and enzymatic structure-probing experiments [51,60] has resulted in significant revision of the proposed secondary structure of U3 snoRNA. Thus, a revised conserved sequence element and secondary structure model for *Euglena* U3 snoRNA is presented here (see Figure 3.13).

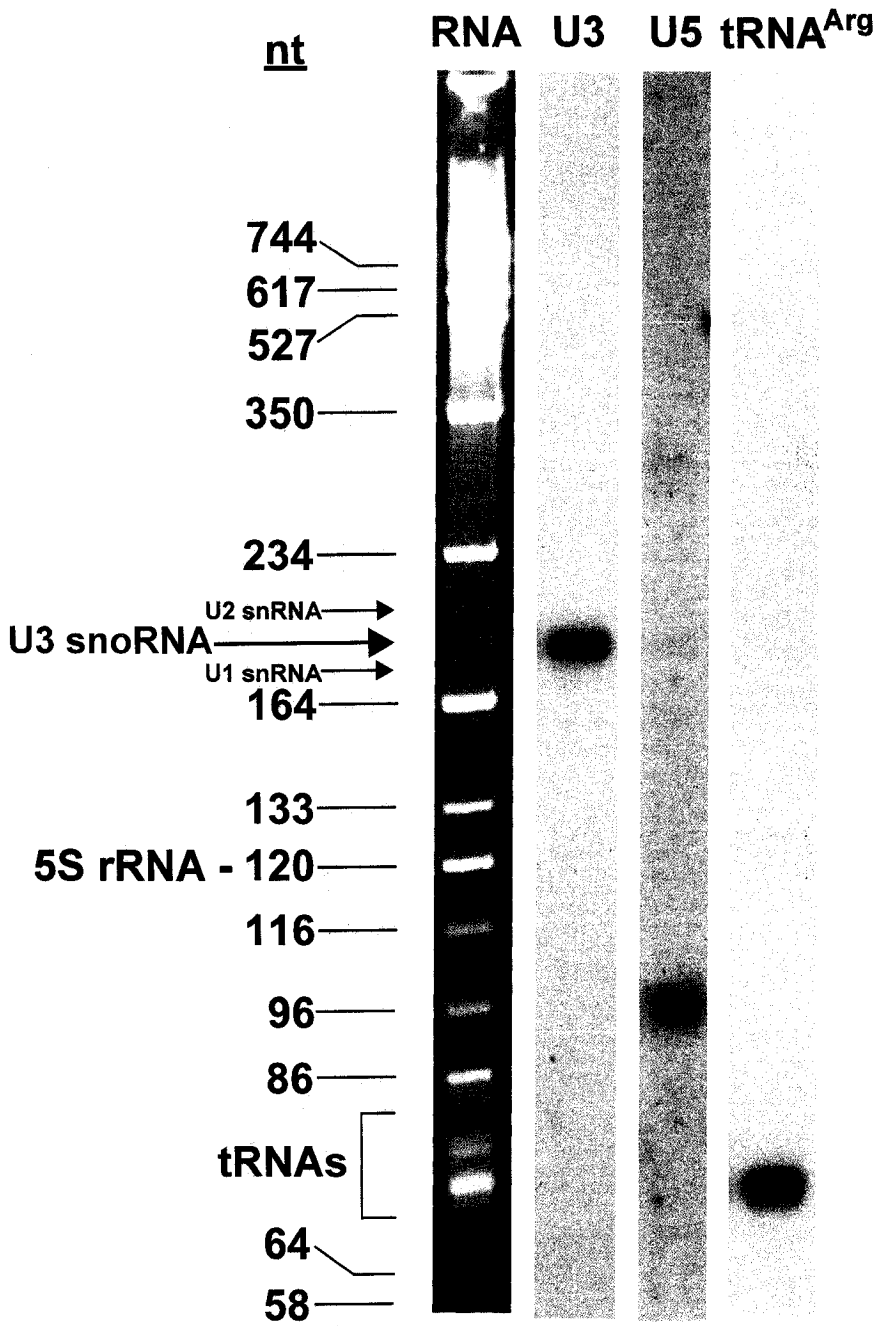
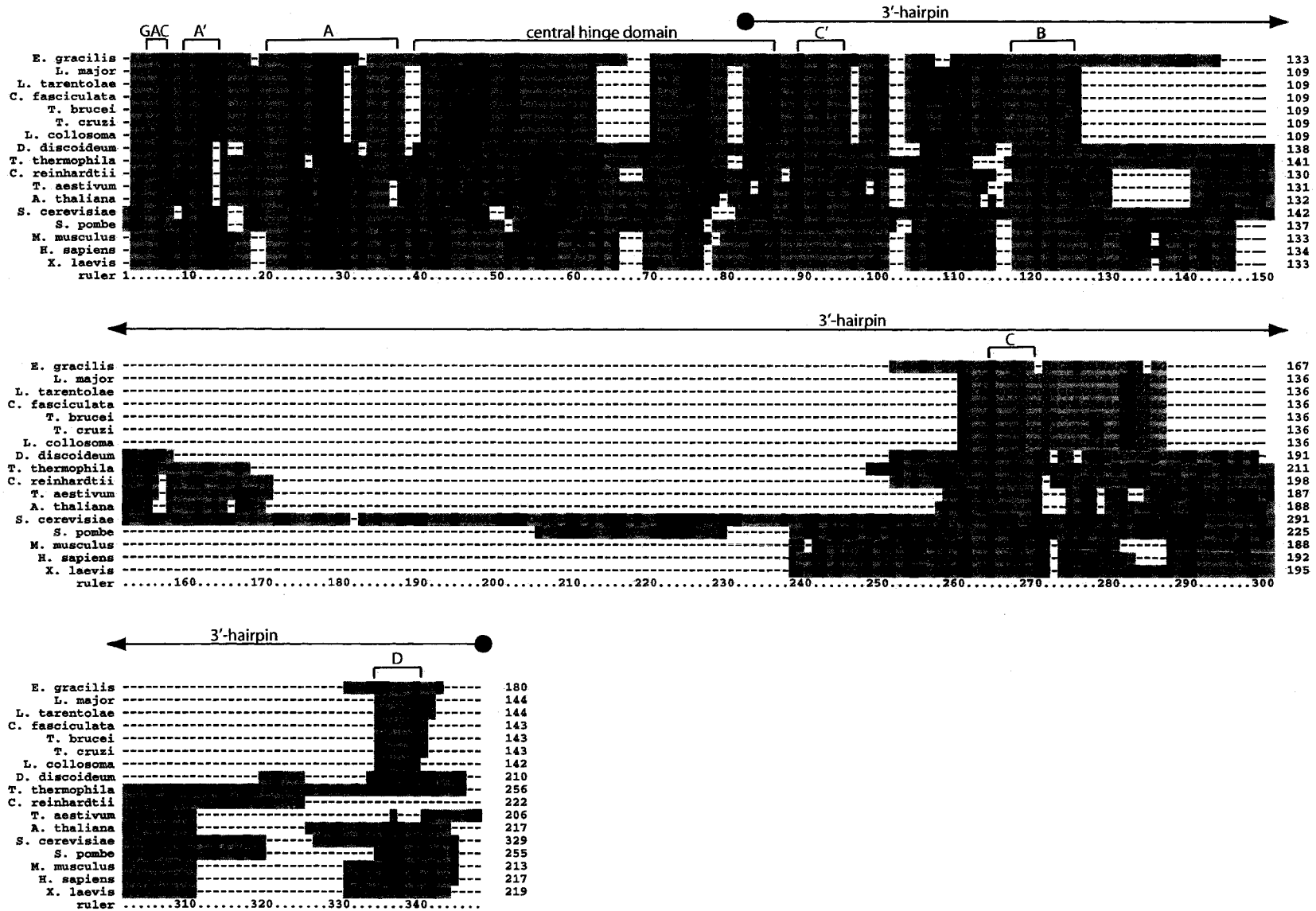


Figure 3.11 Northern Hybridizations Reveal that U3 snoRNA, U5 snRNA, and tRNA^{Arg} Genes are Expressed in *Euglena*. Total *Euglena* RNA was resolved by electrophoresis and visualized by ethidium bromide staining. *Euglena*'s naturally fragmented LSU rRNA produces the observed banding pattern. U1 snRNA (in black), U2 snRNA (in black), and U3 snoRNA (in red), seen as faint bands in total RNA preparations, are identified. Northern hybridization analysis with oligonucleotide probes specific for U3 snoRNA, U5 snRNA or tRNA^{Arg} reveals bands of the expected size for each molecule.

Figure 3.12 Alignment of Known U3 snoRNA Sequences from Representative Organisms. Conserved sequence features in U3 snoRNA, boxes GAC, A′, A, C′, B, C, and D are shown, along with regions of the alignment corresponding to the central hinge and 3′-hairpin domains. All U3 snoRNA sequences were preliminarily aligned using the ClustalX [207] algorithm and manually edited with BioEdit [208], all as described in relevant sections of Materials and Methods. Representative organisms (see Figure 1.6) include: *Chlamydomonas reinhardtii* (AJ001179:171-392), *Crithidia_fasciculata* (AF277396), *Leishmania major* (AQ843909), *Leishmania tarentolae* (L20948:c2128-1984), *Leishmania collosoma* (L32919:391-533), *Tetrahymena thermophila* (X71349), *Trypanosoma brucei* (M25776), *Trypanosoma cruzi*, *Euglena gracilis* (U27297), *Dictyostelium discoideum* (V00190:62-271), *Saccharomyces cerevisiae* (X05498), *Schizosaccharomyces pombe* (X56982:37-291), *Triticum aestivum* (X63065:858-1063), *Arabidopsis thaliana* (X52629:325-541), *Xenopus laevis* (X07318:1-219), *Mus musculus* (X63743:815-1027), and *Homo sapiens* (M14061:277-493).


3′-hairpin
GAC
A′
A
central hinge domain
C′
B
E. gracilis 133
L. major 109
L. tarentolae 109
C. fasciculata 109
T. brucei 109
T. cruzi 109
L. collosoma 109
D. discoideum 138
T. thermophila 141
C. reinhardtii 130
T. aestivum 131
A. thaliana 132
S. cerevisiae 142
S. pombe 137
M. musculus 133
H. sapiens 134
X. laevis 133
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150
3′-hairpin
C
E. gracilis 167
L. major 136
L. tarentolae 136
C. fasciculata 136
T. brucei 136
T. cruzi 136
L. collosoma 136
D. discoideum 191
T. thermophila 211
C. reinhardtii 198
T. aestivum 187
A. thaliana 188
S. cerevisiae 291
S. pombe 225
M. musculus 188
H. sapiens 192
X. laevis 195
ruler .....160.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....290.....300
3′-hairpin
D
E. gracilis 180
L. major 144
L. tarentolae 144
C. fasciculata 143
T. brucei 143
T. cruzi 143
L. collosoma 142
D. discoideum 210
T. thermophila 256
C. reinhardtii 222
T. aestivum 206
A. thaliana 217
S. cerevisiae 329
S. pombe 255
M. musculus 213
H. sapiens 217
X. laevis 219
ruler .....310.....320.....330.....340.....

Sequence elements in the 5′ region of U3 snoRNA are known to interact with the 5′ external transcribed spacer (5′-ETS) of the pre-rRNA and with the 5′ region of the 18S rRNA. The first such element in U3 snoRNA is the GAC box, followed by the A′ box and then the A box [48]. In all U3 snoRNAs, this 5′ region can be folded into a very stable stem-loop structure. However, because this region contains the pre-rRNA/rRNA-interacting GAC, A′, and A elements, an active U3 snoRNA molecule, as shown in Figure 3.13, may adopt an open conformation in which U3 snoRNA-pre-rRNA/rRNA intermolecular interactions displace the intramolecular base-pairing interactions of a closed, inactive U3 snoRNA molecule. Experimental results from *in vivo* chemical and enzymatic structure-probing experiments [60] support this model. In the identified *Euglena* U3 snoRNA genes, one nucleotide heterogeneity has been mapped to this region. The heterogeneity, an A to G transition, is located downstream of the A box sequence.

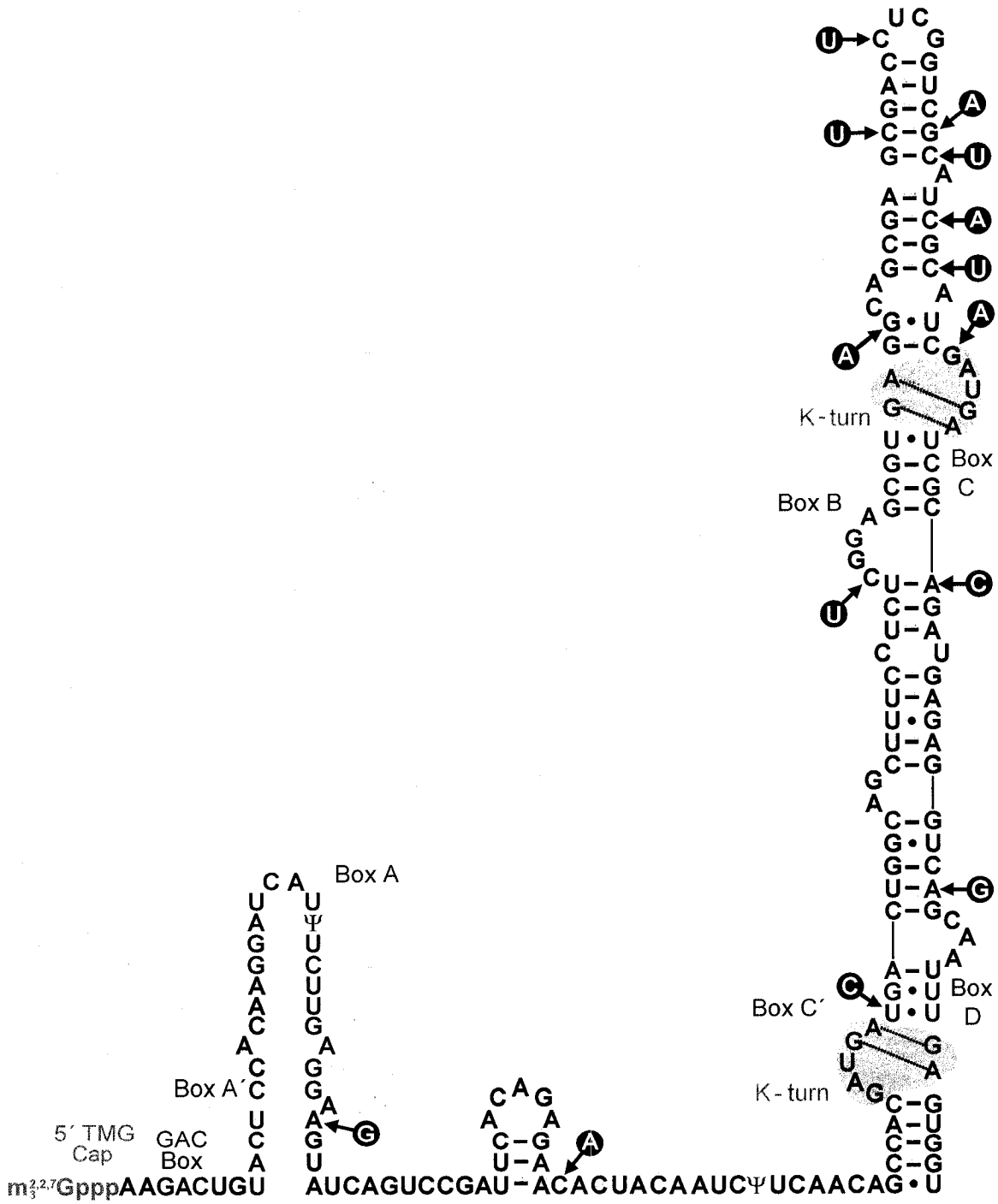
The central hinge region of *Euglena* U3 snoRNA contains a small stem-loop structure consisting of a 3-bp stem with a 5-nt loop. Previous *Euglena* U3 snoRNA secondary structure models also incorporated a similar stem-loop structure in the central hinge region [38]. However, the revised placement of this small stem-loop structure more accurately reflects its phylogenetically conserved position. In addition, as described in other organisms [49], potential base-pairing interactions may occur between the *Euglena* U3 snoRNA hinge domain (both the 5′ and 3′ regions) and the 5′ external transcribed spacer (5′ ETS) of the pre-rRNA [63]. The revised structure presented here is compatible with the proposed U3 snoRNA hinge domain/5′ ETS interactions. In the multiple *Euglena* U3 snoRNA genes, a sequence heterogeneity has been identified in the 3′ hinge region. This A to C transversion is located immediately downstream from the small stem-loop structure.

The 3′-terminal domain of *Euglena* U3 snoRNA contains conserved sequence elements, boxes C′, B, C and D, known to interact with RNA-binding proteins [64,67]. This domain consists of an extended stem-loop structure punctuated by a number of bulge-loop elements. In this revised model, box C′ and box D elements are juxtaposed in the secondary structure context, as are box B and box C elements, even though they are not linearly adjacent in the primary structure. The box elements are mainly single-stranded, corresponding to small bulge-loop structures. However, some regions may form short base-

pairing interactions. More importantly, in this revised model, alternative base-pairing interactions in the 3′-terminal domain allow for the formation of two kink-turn (K-turn) RNA motifs [14,16,17,67] within the juxtaposed box C′/D and B/C elements (compare Figures 1.8 and 3.13).

Nearly all of the sequence heterogeneities identified in the multiple *Euglena* U3 snoRNA variants (12 of the 14) are located in the 3′-extended stem-loop domain. Ten of the 12 sequence heterogeneities are transitions; the remaining two heterogeneities are transversions. The majority of nucleotide heterogeneities occur as shifts from Watson-Crick base-pairings (U-A and G-C) to non-Watson-Crick base-pairings (U•G) (three heterogeneities) or *vice versa* (one heterogeneity). Similarly, three heterogeneities occur in single-stranded regions (none of these is part of the single-stranded conserved box elements). Two nucleotide heterogeneities, present in the same U3 snoRNA variant near the terminal stem-loop structure, form compensatory base changes (C-G to U-A) and thus maintain base-pairing interactions. Other nucleotide heterogeneities (two) form the terminal base pairs of short stem regions, immediately adjacent to the opening or closing of bulge-loop structures. Since neither heterogeneity maintains the terminal base-pairing interactions of the stems, a slight expansion of the adjacent bulge-loop structures results. One of the two heterogeneities is located within the Box C′ element, adjacent to the 5+2 motif of the K-turn [16,17]. Only one heterogeneity, located in a short stem region, disrupts a base-pairing interaction. Thus, the distribution of the 12 sequence heterogeneities identified in the 3′-extended stem-loop domain of *Euglena* U3 snoRNA results in minimal changes to the overall secondary structure of the region.

Figure 3.13 A Revised Secondary Structure Model for *Euglena* U3 snoRNA. In this proposed secondary structure model, the 5′-trimethylguanosine cap (5′-TMG cap) is indicated in orange, the phylogenetically conserved and functionally essential box GAC, A′, A, C′, B, C, and D elements are shown in green. Pseudouridine (Ψ) residues are indicated in blue. Nucleotide heterogeneities identified in the *Euglena* U3 snoRNA gene variants are indicated in red. Nucleotides forming a K-turn motif are highlighted on a gray background. Conventional Watson-Crick base-pairing interactions are depicted as dashes (–) and G/U pairings are identified by dots (•).


Box A
Box A´
5´ TMG Cap
GAC Box
K-turn
Box B
Box C
Box C´
Box D
K-turn
m3 2,2,7GpppAAGACUGU
AUCAGUCCGAU-ACACUACAAUCΨUCAACAG•U

CHAPTER 4: DISCUSSION

4.1: *Euglena* U3 snoRNA is Small in Size and Divergent in Sequence Yet Displays All of the Hallmark Sequence Elements and Secondary Structure Features Typical of U3 snoRNAs

At 180 nt, *Euglena* U3 snoRNA is smaller than most of its counterparts (see Figure 3.12). Some of the largest known U3 snoRNAs are found in the fungi, as in the case of the 329-nt *S. cerevisiae* U3 snoRNA (also known as snR17). Vertebrate U3 snoRNA homologues, as with the 217-nt human U3 snoRNA, are smaller than their yeast counterparts but larger than *Euglena* U3 snoRNA. Since few protist U3 snoRNAs have been identified, it is difficult to determine a phylogenetically meaningful size range for U3 snoRNAs in the broad, diverse, and multi-supergroup assemblage of protists. The smallest known U3 snoRNA homologues have been identified in the trypanosomatids, *Euglena*'s closest relatives, for which there is significant genomic sequence information. The U3 snoRNA of *T. cruzi* is 143 nt, which is slightly smaller than its *Euglena* counterpart. The compact size of *Euglena* U3 snoRNA, and by extension the trypanosomatid U3 snoRNAs, is attributable to the absence of non-conserved hairpins found in the extended stem-loop structure of the 3′-terminal domain of vertebrate and yeast U3 snoRNAs (see Figure 1.3). In *S. cerevisiae*, artificial U3 snoRNA constructs lacking these additional hairpins remain functional [48].

Alignment of U3 snoRNA sequences from representative taxa identifies the conserved sequence box elements GAC, A′, A, C′, B, C and D (see Figure 3.12), diagnostic of U3 snoRNA molecules. All of these conserved box elements are present in *Euglena* U3 snoRNA.

U3 snoRNA consists of three domains. The 5′-domain contains anti-sense guide sequences that interact with the 5′-ETS of the pre-rRNA and with the 5′-region of the 18S rRNA. Using a complex and poorly understood mechanism, this region is thought to guide the multiple pre-rRNA cleavage events that produce the mature 5′-end of the 18S rRNA. Also, through these base pairing interactions, U3 snoRNA is thought to play an RNA chaperone role in the co-transcriptional processing, folding, and assembly of the SSU of the ribosome. The central hinge region of U3 snoRNA also interacts with the 5′-ETS of the pre-rRNA. It may also provide proper spacing between the pre-rRNA-interacting 5′-domain and

the protein-binding 3′-domain [48]. The 3′-domain of U3 snoRNA consists of an extended stem-loop structure. It contains the conserved box C′, B, C, and D elements that act as protein-binding sites. These interactions contribute to U3 snoRNA nucleolar localization, stability, accumulation, and assembly with RNA-binding proteins.

Comparison of U3 snoRNA sequences from phylogenetically disparate organisms (see Figure 3.12) shows that the molecule is poorly conserved over its entire length. However, sequences in the 5′-region are more conserved than those of the central hinge and 3′-regions. The 5′-region contains the conserved sequence elements boxes GAC, A′, and A known to form multiple base-pairing interactions with conserved sequences in the 5′-ETS of the pre-RNA and in the 5′-region of the 18S rRNA. The observed conservation of the box GAC, A′, and A elements over broad phylogenetic distances may be the result of evolutionary constraints imposed by the pre-rRNA. Typically, RNA/RNA-interacting regions follow a co-variation model of evolution. In such a circumstance, a nucleotide change in one member of a base-pairing sequence would result in a disruption of the interaction. A compensatory nucleotide change in the other member of the base-pairing sequence restores the interaction. Similarly, box GAC, A′, and A elements in U3 snoRNA must evolve in concert, or co-evolve, with their target pre-rRNA/rRNA sequences. However, a number of additional constraints may apply to this system. Nucleotide changes in the box GAC, A′, and A elements result in mismatches that would disrupt the interactions between U3 snoRNA and the target pre-rRNA/rRNA sequences. Experimentally, this has been shown to result in the accumulation of unprocessed pre-rRNA transcripts, a halt in ribosome biosynthesis, and ensuing cell death [50,56,57]. Thus, box GAC, A′, and A sequences are not free to change and explore evolutionary space. Furthermore, rRNA sequences are highly conserved, even across broad phylogenetic distances. Thus, their nucleotide sequences change very slowly. This is attributed to an homogenization mechanism that maintains identical copies of the multiple rDNA genes within the nuclear genome [218]. It is assumed that this homogenization mechanism would prevent the appearance of nucleotide changes in the rRNA that would result in compensatory base-pairing interactions with an altered U3 snoRNA sequence, thereby restoring the U3 snoRNA/pre-rRNA interactions and thus pre-rRNA processing. Thus, the sequence of the pre-rRNA/rRNA-interacting regions of U3

snoRNA is highly constrained. Sequence changes in the pre-rRNA/rRNA-interacting regions of U3 snoRNA may only be tolerated as a mechanism to restore the necessary base-pairing interactions when the nucleotide changes have first occurred in the target pre-rRNA/rRNA sequence. Phylogenetic comparison of sequences outside of the box GAC, A′, and A elements reveals extensive sequence variability (see Figure 3.12). Since these sites are not known to interact with the pre-rRNA/rRNA, they may be free to accumulate nucleotide changes. A sequence heterogeneity in *Euglena* U3 snoRNA – an A-to-G transition located immediately downstream from the box A sequence – may represent an example of a site under reduced selection pressures. It is interesting to note that *Euglena* U3 snoRNA contains a pseudouridine (Ψ) in the conserved box A sequence [38]. However, since very few U3 snoRNAs have been studied at the RNA level, it is difficult to assess the degree of phylogenetic conservation of Ψ at this position. Furthermore, while Ψ residues are thought to exert a rigidifying effect on RNA structure [28,42], there is no experimental information on any additional roles played by Ψ in this U3 snoRNA context. However, Ψ residues are known to occur in snRNAs, specifically in functionally important regions that participate in the intermolecular RNA-RNA or RNA-protein interactions involved in the assembly and function of the spliceosome [24,28,37].

The central hinge region of U3 snoRNA reveals surprisingly low sequence conservation (see Figure 3.12). This is unexpected considering this regions's potential interactions with the 5′-ETS of the pre-rRNA. However, sequences in the 5′-ETS of the pre-rRNA are more phylogenetically variable than those of the mature rRNAs. Furthermore, unlike the conserved box GAC, A′, and A interactions with the pre-rRNA/18S rRNA, interactions between the U3 hinge region and the 5′-ETS of the pre-rRNA have been more difficult to identify, both experimentally and phylogenetically. Thus, the U3 snoRNA hinge region may be less phylogenetically conserved owing to its interactions with variable regions of the 5′-ETS of the pre-rRNA. A nucleotide heterogeneity (C-to-A transversion) near the 3′-base of the small stem-loop structure in the central hinge region of *Euglena* U3 snoRNA may represent a site under reduced selection (see Figure 3.13). This heterogeneity is three nucleotides away from sequences in the 3′-central hinge region that potentially interact with the 5′-ETS of the pre-rRNA [63]. *Euglena* U3 snoRNA contains a second Ψ in the 3′-region

of the central hinge domain [38]. As with the previously discussed Ψ in box A, there is little information on the phylogenetic presence of Ψ residues elsewhere in U3 snoRNAs. However, the presence of a Ψ residue in box A and in the 3′-hinge domain, both regions of *Euglena* U3 snoRNA that are postulated to interact with pre-rRNA sequences, is an intriguing commonality. Despite the recent identification of box H/ACA snoRNAs in *Euglena* [219], the box H/ACA snoRNA(s) that may mediate Ψ formation in U3 snoRNA remain(s) unknown.

The 3′-domain of U3 snoRNA shows extensive variability in sequence and in size when compared over wide evolutionary distances (see Figure 3.12). The protein binding box C′, B, C, and D elements are the only conserved sequence elements in this domain. The remainder of the domain shows extensive sequence variability. In some organisms, as with the fungi described above, this domain contains expansion sequences that form additional, non-essential, stem-loop structures. The low sequence conservation of this domain may be due to its functionally different role as a protein-binding domain in comparison to the evolutionary constraints experienced by the RNA-interacting 5′-domain of U3 snoRNA. Furthermore, the protein-binding box C′, B, C, and D elements, while conserved, show some sequence variability when compared over broad evolutionary distances. This may be attributable to a co-variation model, as described above, whereby nucleotide changes in the box elements are compensated by amino acid changes in the binding proteins, or *vice versa*. In the multiple *Euglena* U3 snoRNA gene variants, this domain displays an unexpectedly large number of sequence heterogeneities (see Figure 3.13). Of the 14 nucleotide heterogeneities identified in the multiple *Euglena* U3 snoRNA gene variants, 12 heterogeneities are located in the 3′-extended stem-loop domain. Similar observations have also been made in other organisms, such as humans [220], where multiple U3 snoRNA gene variants have been identified. (These nucleotide heterogeneities will be further explored in the discussion on the secondary structure of *Euglena* U3 snoRNA.) Examination of the sequence and structure context of the nucleotide heterogeneities, in combination with gene expression data from RNA sequencing and 3′ RACE analysis (data not shown), suggests that

none of the *Euglena* U3 snoRNA gene variants represents a pseudogene. Thus, the 3′-domain of U3 snoRNAs in general, and of the *Euglena* U3 snoRNA in particular, is very tolerant of nucleotide changes.

In practical terms, the short and dispersed conserved sequence elements, low sequence conservation over its entire length, and the relatively short length of the molecule result in U3 snoRNA sequences not being identified by BLASTn searches of nucleotide sequence databases such as GenBank. In the case of *Euglena* U3 snoRNA, BLASTn searches of GenBank identify only the *Euglena* U3 snoRNA sequence deposited by Greenwood *et al.* [38]. Thus, nucleotide sequence databases such as GenBank undoubtedly contain numerous unidentified U3 snoRNA sequences from phylogenetically disparate organisms; however, the identification of these sequences by conventional BLASTn searches will be difficult.

The secondary structure of *Euglena* U3 snoRNA is unremarkable (see Figure 3.13). All of the phylogenetically conserved box elements occupy the expected secondary structure locations. By virtue of their small sizes, the *Euglena* and trypanosomatid U3 snoRNAs may represent a minimal U3 snoRNA structure. As previously noted, these snoRNAs do not possess the additional non-conserved, non-essential stem-loop structures present in the 3′-extended stem-loop domain of yeast U3 snoRNAs. The updated *Euglena* U3 snoRNA secondary structure presented here incorporates structural elements identified in U3 snoRNAs from other organisms. These include an unpaired 5′-domain, repositioning of the central hinge region [63], and the incorporation of two K-turn motifs in the 3′-extended stem-loop domain [67]. Thus, over broad evolutionary distances, U3 snoRNAs, though poorly conserved in sequence, show remarkable conservation of their secondary structure.

Nearly all of the sequence heterogeneities identified in *Euglena* U3 snoRNA are located in the 3′-extended stem-loop domain (see Figure 3.13). As previously discussed, this domain appears to be tolerant of nucleotide changes, as seen in its low sequence conservation relative to the 5′-domain of the molecule. The pattern of sequence heterogeneities in *Euglena* U3 snoRNA substantiates this observation.

The distribution of the 12 nucleotide heterogeneities in the 3′-stem-loop domain of *Euglena* U3 snoRNA maintains the overall secondary structure of the region (see Figure 3.13). Ten of the 12 sequence heterogeneities are transitions. The majority of nucleotide heterogeneities occur as shifts from Watson-Crick base-pairings (canonical U-A and G-C) to non-Watson-Crick base-pairings (non-canonical U•G) (three heterogeneities) or *vice versa* (one heterogeneity). Similarly, three heterogeneities occur in single-stranded regions (none of these is part of the single-stranded conserved box elements). Two nucleotide heterogeneities, present in the same U3 snoRNA variant near the terminal stem-loop structure, form compensatory base changes (C-G to U-A) that maintain the base-pairing interactions. Other nucleotide heterogeneities (two) form the terminal base pairs of short stem regions, immediately adjacent to the opening or closing of bulge-loop structures. Since neither heterogeneity maintains the terminal base-pairing interactions of the stems, a slight expansion of the adjacent bulge-loop structures results. One of the two heterogeneities is located within the box C′ element. It is unclear what effect this might have on the binding of proteins that interact with the box C′ motif. Furthermore, the nucleotide heterogeneity is located adjacent to the K-turn motif [16,17]. The consequences of this nucleotide heterogeneity on the structure of the K-turn, and on the interaction of the K-turn-binding protein 15.5-kD/Snu13 [67], are unknown. Only one heterogeneity, located in a stem, disrupts a base-pairing interaction. Thus, the identified sequence heterogeneities result in minimal disruption of the 3′-extended stem-loop domain.

4.2: A Common Gene Organization for U3 snoRNA Genes?

In *Euglena*, I have identified three different U3 snoRNA gene organizations: *i*) solitary, *ii*) linked to $tRNA^{Arg}$ genes, and *iii*) linked to a U5 snRNA gene (see Figure 3.3); only one example of a solitary U3 snoRNA gene was recovered in this analysis of the *Euglena* genome. The sequence of the encoding 12-kbp genomic clone is not known in its entirety; however, analysis of ~1.9 kbp of available upstream and ~0.4 kbp of downstream sequence has not revealed the presence of any additional genes.

Solitary U3 snoRNA genes have been identified in other organisms (see Figure 1.4). In lotus, a solitary U3 snoRNA gene is located in a 77-kbp fragment. Similarly, in *Tetrahymena thermophila*, a solitary U3 snoRNA gene has been identified in a 2.1-kbp fragment [96,97]. In humans, solitary U3 snoRNA genes are encoded in ~45-kbp inverted repeats [83]. Thus, there is precedent in a variety of phylogenetically unrelated organisms for solitary U3 snoRNA genes.

One example of a U3 snoRNA gene linked to two identical, downstream and convergently oriented (relative to the U3 snoRNA gene) $tRNA^{Arg}_{UCG}$ genes has been uncovered in this analysis of the *Euglena* genome (see Figure 3.3). While the sequence of the encoding 14-kbp genomic clone is not known in its entirety, analysis of 0.7 kbp of available sequence upstream of the U3 snoRNA gene and 1.5 kbp downstream from the second tRNA gene identified no additional genes in this clone. A 209-bp intergenic spacer separates the U3 snoRNA gene from the 3′-end of the first tRNA gene; the two tRNA genes are separated by a 70-bp spacer.

Many U3 snoRNA-tRNA gene linkages have been identified in trypanosomatid organisms. The most dramatic example is in *Leishmania tarentolae*, where a 2.6-kbp genomic fragment encodes the U3 snoRNA gene along with 10 tRNA genes, the U1 snRNA gene, and the 5S RNA gene [100] (See Figure 1.4). Sequence data available for a sub-portion of the corresponding region in *Leishmania major* (GenBank Acc: AQ843909) reveals the same gene content organized in the same way. In both *Leishmania* cases, the U3 snoRNA gene is neighbored by one downstream, convergently oriented $tRNA^{Arg}$ gene (possessing the same UCG anticodon as in the *Euglena* case). In *Trypanosoma brucei*, the U3 snoRNA gene is neighbored upstream by a divergently oriented $tRNA^{Arg}$ gene (having a different anticodon, ACG). No genes are encoded downstream of the *T. brucei* U3 snoRNA gene [99,221]. In *Leptomonas collosoma*, a divergently oriented $tRNA^{Lys}$ gene is encoded upstream of the U3 snoRNA gene [222]. No sequence is available downstream of the *Leptomonas* U3 snoRNA gene. Similarly, an upstream, divergently oriented $tRNA^{Lys}$ gene neighbors the U3 snoRNA gene in both *L. tarentolae* and *L. major*, in addition to the previously described downstream, convergently oriented $tRNA^{Arg}$ gene. In all of the trypanosomatid cases described here, the intergenic spacer sequence separating the U3 and

tRNA genes is 71 to 106 bp in size, in comparison to 209 bp in the *Euglena* U3 snoRNA-$tRNA^{Arg}$ linkage. The relevance of the size of the spacer sequence will be discussed further. In the unrelated ciliate *Tetrahymena thermophila*, a U3 snoRNA gene is also neighbored by an upstream, identically oriented $tRNA^{Lys}$ gene. However, unlike the trypanosomatid examples, the intergenic spacer separating the two genes is nearly 500 bp in size [97]. Thus, it appears that close linkage of U3 snoRNA and tRNA genes is a widespread phenomenon in the trypanosomatids and that similar, though less proximal, linkages also occur in *Euglena* and in *Tetrahymena*.

In this thesis, I describe 12 examples of U3 snoRNA-U5 snRNA gene linkages in the *Euglena* genome. In all cases, the linked U5 snRNA gene is downstream and in the opposite transcriptional orientation to the U3 snoRNA gene (see Figures 3.3 and 3.9). The U3 snoRNA-U5 snRNA intergenic spacer sequences vary between 152 and 255 bp in size. To date, no linked U3 snoRNA-U5 snRNA genes have been identified in other organisms. Furthermore, with the exception of tRNA genes, U3 snoRNA genes have not been found directly linked to other small RNA genes (see Figure 1.4). However, other small RNA genes, excluding tRNA genes, have been identified in the vicinity of U3 snoRNA genes. In both *L. tarentolae* and *L. major*, a U1 snRNA and 5S RNA gene has been found near, but not directly linked to, a U3 snoRNA gene (see Figure 1.4). Similarly, in *T. brucei*, a 7SL gene is encoded near the U3 snoRNA gene. Thus, to date, U3 snoRNA-U5 snRNA gene linkages appear to be unique to *Euglena*.

The genes of box C/D and box H/ACA snoRNAs are frequently linked and encoded in clusters. This organization is prevalent in plants [27]. Indeed, an examination of the rice genome revealed the presence of 70 snoRNA clusters encoding a total of 270 snoRNAs [223]. Clustered snoRNA genes have also been identified in trypanosomatid genomes [43,44] and in *Euglena* [219]. However, for unknown reasons U3 snoRNA genes have never been found clustered with modification-guide snoRNA genes, and this also appears to be the case in *Euglena*.

4.3: Influence of U3 snoRNA Gene Organization on U3 snoRNA Gene Expression

The expression of trypanosomatid U3 snoRNA genes has been extensively studied. In all instances examined, trypanosomatid [224], other protist [97], and plant [60,225,226] U3 snoRNA genes are transcribed by RNA polymerase III (RNAP III). Thus, based on the phylogenetic affiliation of trypanosomatids and euglenids [102], U3 snoRNA is assumed to be expressed by RNAP III in *Euglena*. However, in the absence of α-amanitin (an inhibitor of RNAP II) or tagetitoxin (an inhibitor of RNAP III) studies, no definitive conclusions can be drawn in this regard. A 10-bp deletion in an RNAP III promoter can result in polymerase-type switching to an RNA polymerase II (RNAP II) promoter [225]. This appears to have occurred in the common ancestor of fungi and animals. Thus, yeast and human U3 snoRNA genes are transcribed by RNAP II. In all cases, U3 snoRNA genes are transcribed from their own promoter and do not appear to be part of polycistronic transcripts.

In trypanosomatids, the expression of U3 snoRNA is dependent on the linked, upstream, divergently oriented tRNA gene. Two extragenic regulatory elements have been mapped to the A and B boxes of the tRNA [183]. These box elements, corresponding to the D and T-Ψ-C loops of a tRNA, are known to serve as *bona fide* intragenic promoter elements for their host tRNA gene. Furthermore, the spacing of the A and B boxes relative to each other and relative to the U3 snoRNA gene is critical. Indeed, as discussed in prior sections, the spacing between the upstream tRNA gene and the U3 snoRNA gene varies within a narrow range of 93 to 105 bp (see Figure 1.4). The transcription factor TFIIIC is known to bind to the B box element of the tRNA gene, upon which it recruits the transcription factor TFIIIB, which in turn recruits RNAP III for transcription initiation [227,228]. In yeast, transcription of tRNA genes by RNAP III prevents nucleosome assembly in the immediate vicinity of the tRNA gene, and concomitant nucleosome-mediated repression [229]. Thus, the binding of TFIIIC to the B box is postulated to play an indirect role in the expression of trypanosomatid tRNA-linked U3 snoRNA genes through chromatin remodeling, thereby preventing chromatin-mediated repression of transcription [183,224]. This scenario may or may not be relevant to the expression of the tRNA-linked *Euglena* U3 snoRNA gene, owing to a number of differences between the *Euglena* and the trypanosomatid systems. These include the fact that: *i*) the two tRNA genes in the *Euglena* U3 snoRNA-tRNA gene linkage

are downstream instead of upstream of the U3 snoRNA gene and *ii*) the *Euglena* U3 snoRNA-tRNA genes are not as closely linked (209 bp IGS) as they are in trypanosomatids (93-105 bp IGS). However, the presence of two $tRNA^{Arg}$ genes in the *Euglena* U3 snoRNA-tRNA gene linkage, in contrast to one in trypanosomatids, may result in a stronger recruitment of TFIIIC to this site and a concomitant chromatin remodeling effect over a larger region. The inability to genetically manipulate *Euglena* currently hampers further experimental exploration of these issues. At the moment, we know of no obvious advantages that the identified *Euglena* U3 snoRNA-U5 snRNA gene linkage would confer on U3 snoRNA gene expression.

4.4: Could Gene Duplications have Created the Multiple U3 snoRNA-U5 snRNA Gene Linkages in *Euglena*?

The identification of 12 different U3 snoRNA-U5 snRNA gene linkages in the *Euglena* genome was unexpected. The linkages cluster into families, based on regions of sequence conservation in the intergenic spacer sequences. Thus, five different U3 snoRNA-U5 snRNA gene linkage families have been identified (see Figure 3.9). In the "A" family, three linkages form a sub-family (A1a-A1c) and a fourth linkage is the sole member of the second sub-family (A2). The linkage families B, C, D, and E contain two U3 snoRNA-U5 snRNA gene linkages each. No regions of sequence conservation, present in all 12 linkages, have been identified in the intergenic spacers.

The origin and generation of the multiple U3 snoRNA-U5 snRNA gene linkages is an intriguing matter of speculation, but I postulate that one or a few independent ancestral U3 snoRNA-U5 snRNA gene linkages were initially formed. The reasons for the formation and fixation of the ancestral linkage(s) are unclear, but it might have been generated by random gene shuffling and presumably maintained by a favorable effect on gene expression. Multiple rounds of locus, chromosome and/or genome duplication followed by sequence divergence likely then created the observed linkage families and sub-families.

In support of this proposal, there is accumulating evidence from sequence analysis that the *Euglena* genome is highly recombinogenic. A similar arrangement has been described for the spliced leader RNA (SL RNA) gene linked to a 5S rRNA gene [128]. As with the U5 snRNA gene, some SL RNA genes are dispersed throughout the genome.

However, the majority are linked to an identically oriented 5S rRNA gene. The linkage is encoded on a tandemly repeated 0.6-kbp fragment, and it is estimated that up to 300 SL RNA-5S rRNA gene linkages are present in the *Euglena* genome. In the absence of large-scale linkage data, it is not known if the U3 snoRNA-U5 snRNA gene linkages are also encoded as large tandemly repeated arrays in the *Euglena* genome. Limited information on the gene copy number of other *Euglena* snRNAs yields conflicting results. Southern hybridization analysis of *Euglena* DNA with a *Euglena* U1 snRNA [126] gene probe results in a single hybridizing band (unpublished data, Dr. Spencer J. Greenwood, formerly of the Department of Biochemistry and Molecular Biology, Dalhousie University). This result was substantiated by the PCR amplification, cloning, and sequencing of a few *Euglena* U1 snRNA clones, which revealed no nucleotide heterogeneities (my unpublished results). In contrast, 3′ RACE-based PCR amplification was used to obtain the sequence of the *Euglena* U4 snRNA [201]. Sequence analysis of a few *Euglena* U4 snRNA clones revealed at least four differently expressed U4 snRNA sequence variants (my unpublished data). Additional evidence for the recombinogenic and multi-copy nature of the *Euglena* genome has been obtained from *Euglena* modification-guide (box C/D and box H/ACA) snoRNA and cDNA projects. Many instances of multi-copy box C/D and box H/ACA snoRNA genes have been identified in the *Euglena* genome [11,219]. Similarly, in bioinformatic screens for proteins of interest in *Euglena*, multiple allelic variants of typically single-copy protein-coding genes have been identified. Thus, it appears that many, though not all, genes are multi-copy in the genome of *Euglena*.

4.5: Are U3 snoRNA Genes Spatially Organized in the Nucleus?

A recently proposed model for the spatial organization of RNAP III-transcribed genes in the nucleus may be applicable to the organization and expression of U3 snoRNA genes in *Euglena*, trypanosomatids, and other organisms. This model stems from *in situ* hybridization studies that suggest that the 275 tRNA genes of yeast, though dispersed throughout the linear map of the genome, are localized to the nucleolus [230,231]. This situation is analogous to the three-dimensional co-localization of rDNA genes in the nucleolus. Accordingly, it has been proposed that chromosomal loci encoding tRNA genes

also associate in three-dimensional space within the nucleus. Such an arrangement may result in the formation of a tRNA transcription and processing center enriched in RNAP III, transcription, and processing factors [230,232]. Recent experiments suggest that U3 snoRNA genes in human cells may associate in three-dimensional space with coiled bodies in the nucleus [83,233]. Thus, the three-dimensional clustering of genes transcribed by RNAP III may result in the formation of "transcription territories" that could more efficiently recruit transcription complexes to the region and thereby maintain a high level of gene expression. Although it is not known whether this scenario applies to *Euglena* U3 snoRNA genes, the theory is a plausible one, considering that U3 snoRNA genes are *i*) (probably) transcribed by RNAP III, *ii*) linked to tRNA genes, and *iii*) linked to the RNAP III-transcribed U5 snRNA gene.

4.6: Why so Many U3 snoRNA Genes in the *Euglena* Genome?

The *Euglena* genome encodes at least 14 different U3 snoRNA genes. While U3 snoRNA is an essential gene, it is unclear why the *Euglena* genome would encode so many copies. However, several theories can be advanced.

The ribosomal RNA genes of *Euglena* are highly unusual: they are encoded on a plasmid [63,133-135], in contrast to having a chromosomal location, as is the case in nearly all other eukaryotic organisms. Furthermore, a number of additional spacer sequences are removed from the LSU rRNA in *Euglena* thereby creating a naturally fragmented 28 S rRNA equivalent [145,146]. Thus, it has been proposed that the *Euglena* U3 snoRNA may participate in the additional pre-rRNA processing events that excise the additional ITS sequences in the LSU rRNA [38]. At the moment, there is no experimental evidence in support of, or refuting, this proposed additional function for U3 snoRNA in *Euglena* pre-rRNA processing.

Another theory proposes that the presence of the rDNA operon on a plasmid, in conjunction with the additional processing steps required in the formation of the naturally fragmented rRNAs of *Euglena*, may result in the production of a greater proportion of abortive ribosomes. Such ribosomes might be improperly processed, mis-folded, or incorrectly assembled with ribosomal proteins. Thus, in order to maintain an adequate

number of functional ribosomes in the cell, a greater number of pre-rRNA transcripts would need to be processed. This requirement would in turn necessitate a greater number of U3 snoRNA molecules in the cell, which presumably could be achieved by encoding and expressing a greater number of U3 snoRNA genes.

A problem with both of the foregoing theories is that trypanosomatid LSU rRNAs are also fragmented, though not as extensively as those of *Euglena* [147-149]. Furthermore, in all cases examined, U3 snoRNA is a single-copy gene in trypanosomatid genomes [98,100]. Thus, it appears unlikely that the multi-copy nature of *Euglena* U3 snoRNA genes could be a direct consequence of the particularities of rRNA processing in *Euglena.*

Very little is known about the genome of *Euglena.* However, early information points to the possibility that the *Euglena* genome may be very "plastic" and highly recombinogenic. Thus, one might expect frequent gene duplications and rearrangements as part of the normal behavior of the genome, behavior that could explain the multi-copy presence of U3 snoRNA genes in the *Euglena* genome.

The presence of U3 snoRNA-hybridizing bands of varying intensities in Southern analysis of *Euglena* DNA remains enigmatic. Technical considerations in restriction endonuclease digestion, Southern transfer, and hybridization have been eliminated based on the reproducibility of the result under different experimental conditions. The presence of differing numbers of U3 snoRNA genes and/or pseudogenes on the hybridizing fragments could explain the observed results. However, clustered U3 snoRNA genes have so far not been found in the *Euglena* genome.

Aneuploidy appears to be the most likely explanation for the varying hybridizing band intensities. In many organisms, aneuploidy is associated with genome instability. This is the case with human Trisomy-21 (Down's syndrome) and many cancers. Aneuploidy has been postulated to occur in a number of protists [170] and in fungi [168-170], but in these organisms, it does not appear to result in genome instability. Thus, these organisms may require a less rigid control of their chromosome numbers.

In aneuploidy, there is no true ploidy number (*i.e.*, N, 2N, 4N); instead, a cell contains different numbers of one or more chromosomes. Such a phenomena has been reported in a number of protists, including *Trypanosoma cruzi* [170,171,173-175] and *Leishmania* [172],

both distant but specific relatives of *Euglena* [102]. In such cases, results similar to those presented here have been obtained. In pulsed field gel electrophoresis analysis of *T. cruzi* DNA, the ethidium bromide staining intensities of different chromosomal bands varies within single samples. Furthermore, DNA-content variability of up to 70% has been observed in cells derived from a single clone. There is as yet no direct evidence for aneuploidy in *Euglena*; however, the demonstrated presence of aneuploidy in trypanosomatids, in conjunction with the results presented here, make aneuploidy a distinct possibility in *Euglena*.

4.7: Conclusions

A broad phylogenetic comparison of U3 snoRNA sequences reveals that the *Euglena* U3 snoRNA, while among the smallest known examples, possesses all of the hallmark sequence and secondary structure features typical of U3 snoRNAs. Southern hybridization analysis of *Euglena* DNA reveals the presence of at least 13, and up to 18, U3 snoRNA-hybridizing bands. Sequence analysis of *Euglena* U3 snoRNA identified 14 different gene variants, with most nucleotide heterogeneities being located in the 3′-extended stem-loop domain. An examination of the genomic neighborhood of *Euglena* U3 snoRNA genes, based on sequence analysis of obtained genomic λ clones and PCR products, revealed the presence of at least three different genomic organizations: *i*) a stand-alone, *ii*) linked to $tRNA^{Arg}$ genes, and *iii*) linked to a U5 snRNA gene. Only one example of each of the stand-alone and U3 snoRNA-$tRNA^{Arg}$ linkage organizations was recovered from the *Euglena* genome. The sequence features of the stand-alone U3 snoRNA gene are unremarkable. In the U3 snoRNA-$tRNA^{Arg}$ linkage, the U3 snoRNA gene is linked to two identical, downstream and convergently oriented (relative to the U3 snoRNA gene) $tRNA^{Arg}_{UCG}$ genes. This scenario is reminiscent, but different in a number of key ways, to the U3 snoRNA-tRNA gene linkage in trypanosomatids. Lastly, 12 examples of U3 snoRNA-U5 snRNA gene linkages have been identified in which the linked U5 snRNA gene is encoded downstream and in the opposite transcriptional orientation to the U3 snoRNA gene. The multiple U3 snoRNA-U5 snRNA gene linkages cluster into distinct families based on sequence identity within the U3

snoRNA-U5 snRNA intergenic spacer. It is postulated that the multiple U3 snoRNA genes in the *Euglena* genome arose by (multiple) genome, chromosome, and/or locus duplications. The unexpected variability in the signal intensities of the multiple Southern hybridizing bands raises the intriguing possibility that the *Euglena* genome might be aneuploid.

4.8: Future Prospects

Various technical and biological features of the experimental system made this study difficult. The screening of (what was eventually determined to be) an unreliable *Euglena* λ genomic DNA library severely hampered the acquisition of larger-scale genomic sequence neighboring the U3 snoRNA genes. Thus, it would have been preferable to construct a new *Euglena* λ genomic DNA library. However, the *Euglena* genome contains numerous repetitive sequence elements, such as G_{12} homopolymer regions that are recalcitrant to standard sequencing methodologies. This technical difficulty, a biological feature of the *Euglena* genome, would persist in a new *Euglena* λ genomic DNA library.

Information on the large-scale (>10 kbp) genomic organization of U3 snoRNA genes in the *Euglena* genome is lacking. Such questions could be best addressed by the isolation and characterization of a large-insert genomic library, such as a *Euglena* cosmid, fosmid, or yeast artificial chromosome library. However, the repetitive sequence elements in the *Euglena* genome, in addition to causing sequencing difficulties in this context, could mediate aberrent recombination events in the cloned DNA.

This study was also complicated by the possible aneuploid features of the *Euglena* genome. Thus, large-scale studies on U3 snoRNA gene organization in the genome of *Euglena* may be aided by basic studies on the biology of the *Euglena* genome. A characterization of the *Euglena* genome by pulsed-field-gel-electrophoresis could (possibly) give an insight into chromosome number and size, in addition to further substantiating (or refuting) the assumption of aneuploidy in this organism.

REFERENCES

[1] Warner, J.R. 1999. The economics of ribosome biosynthesis in yeast. Trends Biochem. Sci. 24(11): 437-440.

[2] Venema, J. and Tollervey, D. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. Annu. Rev. Genet. 33: 261-311.

[3] Fromont-Racine, M., Senger, B., Saveanu, C. and Fasiolo, F. 2003. Ribosome assembly in eukaryotes. Gene 313: 17-42.

[4] Fatica, A. and Tollervey, D. 2002. Making ribosomes. Curr. Opin. Cell Biol. 14: 313-318.

[5] Dennis, P.P. and Omer, A. 2005. Small non-coding RNAs in Archaea. Curr. Opin. Microbiol. 8: 685-694.

[6] Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J. and Dennis, P.P. 2003. RNA-modifying machines in archaea. Mol. Microbiol. 48: 617-629.

[7] Ziesche, S.M., Omer, A.D. and Dennis, P.P. 2004. RNA-guided nucleotide modification of ribosomal and non-ribosomal RNAs in Archaea. Mol. Microbiol. 54: 980-993.

[8] Aravind, L. and Koonin, E.V. 1999. Novel predicted RNA-binding domains associated with the translation machinery. J. Mol. Evol. 48: 291-302.

[9] Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M. and Kiss, T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell 85(7): 1077-1088.

[10] Nicoloso, M., Qu, L.-H., Michot, B. and Bachellerie, J.-P. 1996. Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-*O*-ribose methylation of rRNAs. J. Mol. Biol. 260: 178-195.

[11] Russell, A.G., Schnare, M.N. and Gray, M.W. 2006. A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: structural, functional and evolutionary insights. J. Mol. Biol. 357(5): 1548-1565.

[12] Liang, X.-H., Uliel, S., Hury, A., Barth, S., Doniger, T., Unger, R. and Michaeli, S. 2005. A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. RNA 11: 619-645.

[13] Aspegren, A., Hinas, A., Larsson, P., Larsson, A. and Söderbom, F. 2004. Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. Nucleic Acids Res. 32(15): 4646-4656.

[14] Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. and Lührmann, R. 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell 103(3): 457-466.

[15] Szewczak, L.B., Gabrielsen, J.S., Degregorio, S.J., Strobel, S.A. and Steitz, J.A. 2005. Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. RNA 11: 1407-1419.

[16] Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. 2001. The kink-turn: a new RNA secondary structure motif. EMBO J. 20(15): 4214-4221.

[17] Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. and Ficner, R. 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol. Cell 6(6): 1331-1342.

[18] Tycowski, K.T., Aab, A. and Steitz, J.A. 2004. Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. Curr. Biol. 14: 1985-1995.

[19] Jády, B.E. and Kiss, T. 2001. A small nucleolar guide RNA functions both in 2′-*O*-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. EMBO J. 20(3): 541-551.

[20] Tollervey, D., Lehtonen, H., Jansen, R., Kern, H. and Hurt, E.C. 1993. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. Cell 72: 443-457.

[21] Mengel-Jorgensen, J., Jensen, S.S., Rasmussen, A., Poehlsgaard, J., Iversen, J.J.L. and Kirpekar, F. 2006. Modifications in *Thermus thermophilus* 23 S ribosomal RNA are centered in regions of RNA-RNA contact. J. Biol. Chem. 281: 22108-22117.

[22] Decatur, W.A. and Fournier, M.J. 2002. rRNA modifications and ribosome function. Trends Biochem. Sci. 27(7): 344-351.

[23] Baxter-Roshek, J.L., Petrov, A.N. and Dinman, J.D. 2007. Optimization of Ribosome Structure and Function by rRNA Base Modification. PLoS ONE 2: e174.

[24] Dönmez, G., Hartmuth, K. and Lührmann, R. 2004. Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. RNA 10: 1925-1933.

[25] Davis, D.R. 1998. Biophysical and conformational properties of modified nucleosides in RNA (nuclear magnetic resonance studies). In: Modification and Editing of RNA; Grosjean, H. and Benne, R., eds., pp. 85-102, American Society for Microbiology, Washington, DC.

[26] Helm, M. 2006. Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res. 34(2): 721-733.

[27] Brown, J.W., Echeverria, M. and Qu, L.H. 2003. Plant snoRNAs: functional evolution and new modes of gene expression. Trends Plant Sci. 8(1): 42-49.

[28] Charette, M. and Gray, M.W. 2000. Pseudouridine in RNA: what, where, how, and why. IUBMB Life 49(5): 341-351.

[29] Balakin, A.G., Smith, L. and Fournier, M.J. 1996. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. Cell 86(5): 823-834.

[30] Ni, J., Tien, A.L. and Fournier, M.J. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell 89(4): 565-573.

[31] Kiss, A.M., Jády, B.E., Darzacq, X., Verheggen, C., Bertrand, E. and Kiss, T. 2002. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res. 30(21): 4643-4649.

[32] Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I. and Yu, Y.-T. 2005. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. EMBO J. 24: 2403-2413.

[33] Atzorn, V., Fragapane, P. and Kiss, T. 2004. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. Mol. Cell. Biol. 24(4): 1769-1778.

[34] Eliceiri, G.L. 2006. The vertebrate E1/U17 small nucleolar ribonucleoprotein particle. J. Cell. Biochem. 98(3): 486-495.

[35] Enright, C.A., Maxwell, E.S., Eliceiri, G.L. and Sollner-Webb, B. 1996. 5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3. RNA 2(11): 1094-1099.

[36] Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. 12: 527-537.

[37] Newby, M.I. and Greenbaum, N.L. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nat. Struct. Biol. 9(12): 958-965.

[38] Greenwood, S.J., Schnare, M.N. and Gray, M.W. 1996. Molecular characterization of U3 small nucleolar RNA from the early diverging protist, *Euglena gracilis*. Curr. Genet. 30(4): 338-346.

[39] Badis, G., Fromont-Racine, M. and Jacquier, A. 2003. A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. RNA 9: 771-779.

[40] King, T.H., Liu, B., McCully, R.R. and Fournier, M.J. 2003. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. Mol. Cell 11: 425-435.

[41] Newby, M.I. and Greenbaum, N.L. 2002. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proc. Natl. Acad. Sci. USA 99(20): 12697-12702.

[42] Kolev, N.G. and Steitz, J.A. 2006. *In vivo* assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site. Nat. Struct. Mol. Biol. 13: 347-353.

[43] Uliel, S., Liang, X.-h., Unger, R. and Michaeli, S. 2004. Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. Int. J. Parasitol. 34: 445-454.

[44] Liang, X.-h., Ochaion, A., Xu, Y.-x., Liu, Q. and Michaeli, S. 2004. Small nucleolar RNA clusters in trypanosomatid *Leptomonas collosoma*. Genome organization, expression studies, and the potential role of sequences present upstream from the first repeated cluster. J. Biol. Chem. 279: 5100-5109.

[45] Dunbar, D.A., Chen, A., A., Wormsley, S. and Baserga, S.J. 2000. The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. Nucleic Acids Res. 28: 2855-2861.

[46] Filippini, D., Renzi, F., Bozzoni, I. and Caffarelli, E. 2001. U86, a novel snoRNA with an unprecedented gene organization in yeast. Biochem. Biophys. Res. Commun. 288: 16-21.

[47] Hodnett, J.L. and Busch, H. 1968. Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. J. Biol. Chem. 243(24): 6334-6342.

[48] Samarsky, D.A. and Fournier, M.J. 1998. Functional mapping of the U3 small nucleolar RNA from the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. 18(6): 3431-3444.

[49] Beltrame, M. and Tollervey, D. 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. EMBO J. 14(17): 4350-4356.

[50] Borovjagin, A.V. and Gerbi, S.A. 2001. *Xenopus* U3 snoRNA GAC-Box A′ and Box A sequences play distinct functional roles in rRNA processing. Mol. Cell. Biol. 21(18): 6210-6221.

[51] Mereau, A., Fournier, R., Grégoire, A., Mougin, A., Fabrizio, P., Lührmann, R. and Branlant, C. 1997. An *in vivo* and *in vitro* structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. J. Mol. Biol. 273(3): 552-571.

[52] Borovjagin, A.V. and Gerbi, S.A. 2000. The spacing between functional *cis*-elements of U3 snoRNA is critical for rRNA processing. J. Mol. Biol. 300(1): 57-74.

[53] Morrissey, J.P. and Tollervey, D. 1997. U14 small nucleolar RNA makes multiple contacts with the pre-ribosomal RNA. Chromosoma 105(7-8): 515-522.

[54] Liang, W.Q., Clark, J.A. and Fournier, M.J. 1997. The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. Mol. Cell. Biol. 17(7): 4124-4132.

[55] Liang, W.Q. and Fournier, M.J. 1995. U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. Genes Dev. 9(19): 2433-2443.

[56] Sharma, K. and Tollervey, D. 1999. Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. Mol. Cell. Biol. 19(9): 6012-6019.

[57] Hughes, J.M. 1996. Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. J. Mol. Biol. 259(4): 645-654.

[58] Kass, S., Tyc, K., Steitz, J.A. and Sollner-Webb, B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell 60(6): 897-908.

[59] Wehner, K.A., Gallagher, J.E. and Baserga, S.J. 2002. Components of an interdependent unit within the SSU processome regulate and mediate its activity. Mol. Cell. Biol. 22: 7258-7267.

[60] Antal, M., Mougin, A., Kis, M., Boros, E., Steger, G., Jakab, G., Solymosy, F. and Branlant, C. 2000. Molecular characterization at the RNA and gene levels of U3 snoRNA from a unicellular green alga, *Chlamydomonas reinhardtii*. Nucleic Acids Res. 28(15): 2959-2968.

[61] Borovjagin, A.V. and Gerbi, S.A. 2004. *Xenopus* U3 snoRNA docks on pre-rRNA through a novel base-pairing interaction. RNA 10(6): 942-953.

[62] Borovjagin, A.V. and Gerbi, S.A. 2005. An evolutionary intra-molecular shift in the preferred U3 snoRNA binding site on pre-ribosomal RNA. Nucleic Acids Res. 33(15): 4995-5005.

[63] Greenwood, S.J., Schnare, M.N., Cook, J.R. and Gray, M.W. 2001. Analysis of intergenic spacer transcripts suggests 'read-around' transcription of the extrachromosomal circular rDNA in *Euglena gracilis*. Nucleic Acids Res. 29(10): 2191-2198.

[64] Cléry, A., Senty-Ségault, V., Leclerc, F., Raué, H.A. and Branlant, C. 2007. Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. Mol. Cell. Biol. 27(4): 1191-1206.

[65] Lee, S.J. and Baserga, S.J. 1999. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. Mol. Cell. Biol. 19(8): 5441-5452.

[66] Baserga, S.J., Yang, X.D. and Steitz, J.A. 1991. An intact Box C sequence in the U3 snRNA is required for binding of fibrillarin, the protein common to the major family of nucleolar snRNPs. EMBO J. 10(9): 2645-2651.

[67] Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R. and Branlant, C. 2003. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. RNA 9(7): 821-838.

[68] Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., Beyer, A.L., Hunt, D.F. and Baserga, S.J. 2002. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature 417(6892): 967-970.

[69] Bernstein, K.A., Gallagher, J.E., Mitchell, B.M., Granneman, S. and Baserga, S.J. 2004. The small-subunit processome is a ribosome assembly intermediate. Eukaryot. Cell 3(6): 1619-1626.

[70] Bleichert, F., Granneman, S., Osheim, Y.N., Beyer, A.L. and Baserga, S.J. 2006. The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. Proc. Natl. Acad. Sci. USA 103(25): 9464-9469.

[71] Grandi, P., Rybin, V., Baßler, J., Petfalski, E., Straub, D., Marzioch, M., Schäfer, T., Kuster, B., Tschochner, H., Tollervey, D., Gavin, A.-C. and Hurt, E. 2002. 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol. Cell 10: 105-115.

[72] Osheim, Y.N., French, S.L., Keck, K.M., Champion, E.A., Spasov, K., Dragon, F., Baserga, S.J. and Beyer, A.L. 2004. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. Mol. Cell. 16(6): 943-954.

[73] Granneman, S. and Baserga, S.J. 2004. Ribosome biogenesis: of knobs and RNA processing. Exp. Cell Res. 296: 43-50.

[74] Granneman, S. and Baserga, S.J. 2005. Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. Curr. Opin. Cell Biol. 17(3): 281-286.

[75] Gallagher, J.E., Dunbar, D.A., Granneman, S., Mitchell, B.M., Osheim, Y., Beyer, A.L. and Baserga, S.J. 2004. RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. Genes Dev. 18: 2506-2517.

[76] Saez-Vasquez, J., Caparros-Ruiz, D., Barneche, F. and Echeverría, M. 2004. A plant snoRNP complex containing snoRNAs, fibrillarin, and nucleolin-like proteins is competent for both rRNA gene binding and pre-rRNA processing *in vitro*. Mol. Cell. Biol. 24: 7284-7297.

[77] Kopp, K., Gasiorowski, J.Z., Chen, D., Gilmore, R., Norton, J.T., Wang, C., Leary, D.J., Chan, E.K., Dean, D.A. and Huang, S. 2006. Pol I Transcription and Pre-rRNA Processing Are Coordinated in a Transcription-dependent Manner in Mammalian Cells. Mol. Biol. Cell In press.

[78] Salim, M. and Maden, B.E. 1973. Early and late methylations in HeLa cell ribosome maturation. Nature 244: 334-336.

[79] Maden, B.E. 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. Prog. Nucleic Acid Res. Mol. Biol. 39: 241-303.

[80] Decatur, W.A. and Fournier, M.J. 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. J. Biol. Chem. 278: 695-698.

[81] Dennis, P.P., Russell, A.G. and Moniz De Sa, M. 1997. Formation of the 5′ end pseudoknot in small subunit ribosomal RNA: involvement of U3-like sequences. RNA 3(4): 337-343.

[82] Schoemaker, R.J. and Gultyaev, A.P. 2006. Computer simulation of chaperone effects of archaeal C/D box sRNA binding on rRNA folding. Nucleic Acids Res. 34(7): 2015-2026.

[83] Gao, L., Frey, M.R. and Matera, A.G. 1997. Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. Nucleic Acids Res. 25(23): 4740-4747.

[84] Mazan, S. and Bachellerie, J.P. 1990. Organization of the gene family encoding mouse U3B RNA: role of gene conversions in its concerted evolution. Gene 94(2): 263-272.

[85] Savino, R., Hitti, Y. and Gerbi, S.A. 1992. Genes for *Xenopus laevis* U3 small nuclear RNA. Nucleic Acids Res. 20(20): 5435-5442.

[86] Denison, R.A., Van Arsdell, S.W., Bernstein, L.B. and Weiner, A.M. 1981. Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. Proc. Natl. Acad. Sci. USA 78(2): 810-814.

[87] Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. 1981. Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell 26(1 Pt 1): 11-17.

[88] Yuan, Y. and Reddy, R. 1989. Genes for human U3 small nucleolar RNA contain highly conserved flanking sequences. Biochim. Biophys. Acta 1008(1): 14-22.

[89] Reddy, R., Henning, D., Chirala, S., Rothblum, L., Wright, D. and Busch, H. 1985. Isolation and characterization of three rat U3 RNA pseudogenes colinear with U3 RNA. J. Biol. Chem. 260(9): 5715-5719.

[90] Stroke, I.L. and Weiner, A.M. 1985. Genes and pseudogenes for rat U3A and U3B small nuclear RNA. J. Mol. Biol. 184(2): 183-193.

[91] Ferrer, P., Qu, L.H., Bouche, G. and Bachellerie, J.P. 1986. A U3 RNA pseudogene in mouse: sequence and organization in genomic DNA. FEBS Lett. 204(2): 307-312.

[92] Mazan, S., Michot, B. and Bachellerie, J.P. 1989. Mouse U3-RNA-processed pseudogenes are nonrandomly integrated into genomic DNA. Implications for the process of retrogene formation. Eur. J. Biochem. 181(3): 599-605.

[93] Kiss, T. and Solymosy, F. 1990. Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes. Nucleic Acids Res. 18(8): 1941-1949.

[94] Vaux, P., Guerineau, F., Waugh, R. and Brown, J.W. 1992. Characterization and expression of U1 snRNA genes from potato. Plant Mol. Biol. 19(6): 959-971.

[95] Wise, J.A. and Weiner, A.M. 1980. *Dictyostelium* small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. Cell 22(1 Pt 1): 109-118.

[96] Orum, H., Nielsen, H. and Engberg, J. 1993. Sequence and proposed secondary structure of the *Tetrahymena thermophila* U3-snRNA. Nucleic Acids Res. 21(10): 2511.

[97] Orum, H., Nielsen, H. and Engberg, J. 1992. Structural organization of the genes encoding the small nuclear RNAs U1 to U6 of *Tetrahymena thermophila* is very similar to that of plant small nuclear RNA genes. J. Mol. Biol. 227(1): 114-121.

[98] Mottram, J., Perry, K.L., Lizardi, P.M., Lührmann, R., Agabian, N. and Nelson, R.G. 1989. Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei*: identification of the U2, U4, and U6 RNA analogs. Mol. Cell. Biol. 9(3): 1212-1223.

[99] Mottram, J.C., Bell, S.D., Nelson, R.G. and Barry, J.D. 1991. tRNAs of *Trypanosoma brucei*. Unusual gene organization and mitochondrial importation. J. Biol. Chem. 266(27): 18313-18317.

[100] Shi, X., Chen, D.-H.T. and Suyama, Y. 1994. A nuclear tRNA gene cluster in the protozoan *Leishmania tarentolae* and differential distribution of nuclear-encoded tRNAs between the cytosol and mitochondria. Mol. Biochem. Parasitol. 65(1): 23-37.

[101] Leedale, G.F. 1967. Euglenida-euglenophyta. Annu. Rev. Microbiol. 21: 31-48.

[102] Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J. and Gray, M.W. 2005. The tree of eukaryotes. Trends Ecol. Evol. 20(12): 670-676.

[103] Archibald, J.M. and Keeling, P.J. 2002. Recycled plastids: a 'green movement' in eukaryotic evolution. Trends Genet. 18: 577-584.

[104] McFadden, G.I. 2001. Chloroplast origin and integration. Plant Physiol. 125: 50-53.

[105] Gibbs, S.P. 1978. The chloroplasts of *Euglena* might have evolved from symbiotic green algae. Can. J. Bot. 56: 2883-2889.

[106] Leander, B.S. 2004. Did trypanosomatid parasites have photosynthetic ancestors? Trends Microbiol. 12: 251-258.

[107] Archibald, J.M. 2005. Jumping genes and shrinking genomes--probing the evolution of eukaryotic photosynthesis with genomics. IUBMB Life 57: 539-547.

[108] Marin, B. 2004. Origin and fate of chloroplasts in the Euglenoida. Protist 155: 13-14.

[109] Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.T., Wu, X., Reith, M., Cavalier-Smith, T. and Maier, U.G. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410: 1091-1096.

[110] Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J. and McFadden, G.I. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc. Natl. Acad. Sci. USA 103: 9566-9571.

[111] Martin, W., Somerville, C.C. and Loiseau-de Goër, S. 1992. Molecular phylogenies of plastid origins and algal evolution. J. Mol. Evol. 35(5): 385-403.

[112] Turmel, M., Otis, C. and Lemieux, C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. Proc. Natl. Acad. Sci. USA 96: 10248-10253.

[113] Palmer, J.D. and Delwiche, C.F. 1998. DNA Sequencing. In: Molecular Systematics of Plants II; Soltis, D.E., Soltis, P.S. and Doyle, J.J., eds., pp. 375-409, Kluwer, Norwell, MA, USA.

[114] Bhattacharya, D. and Medlin, L. 1998. Algal phylogeny and the origin of land plants. Plant Physiol. 116: 9-15.

[115] Hader, D.P., Rosum, A., Schafer, J. and Hemmersbach, R. 1996. Graviperception in the flagellate *Euglena gracilis* during a shuttle space flight. J. Biotechnol. 47: 261-269.

[116] Cramer, M. and Myers, J. 1952. Growth and photosynthetic characteristics of *Euglena gracilis*. Arch. Mikrobiol. 17: 384-402.

[117] Ishii, N. and Uchida, S. 2006. Removal of technetium from solution by algal flagellate *Euglena gracilis*. J. Environ. Qual. 35: 2017-2020.

[118] Mendoza-Cozatl, D.G., Rangel-Gonzalez, E. and Moreno-Sanchez, R. 2006. Simultaneous Cd2+, Zn2+, and Pb2+ uptake and accumulation by photosynthetic *Euglena gracilis*. Arch. Environ. Contam. Toxicol. 51: 521-528.

[119] Millan de Kuhn, R., Streb, C., Breiter, R., Richter, P., Neesse, T. and Hader, D.P. 2006. Screening for unicellular algae as possible bioassay organisms for monitoring marine water samples. Water Res. 40: 2695-2703.

[120] Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. 21(15): 3537-3544.

[121] Thompson, M.D., Copertino, D.W., Thompson, E., Favreau, M.R. and Hallick, R.B. 1995. Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. Nucleic Acids Res. 23(23): 4745-4752.

[122] Frantz, C., Ebel, C., Paulus, F. and Imbault, P. 2000. Characterization of *trans*-splicing in euglenoids. Curr. Genet. 37(6): 349-355.

[123] Tessier, L.H., Keller, M., Chan, R.L., Fournier, R., Weil, J.H. and Imbault, P. 1991. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. EMBO J. 10(9): 2621-2625.

[124] Bonen, L. 1993. *Trans*-splicing of pre-mRNA in plants, animals, and protists. FASEB J. 7(1): 40-46.

[125] Liang, X.-h., Haritan, A., Uliel, S. and Michaeli, S. 2003. *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot. Cell 2(5): 830-840.

[126] Breckenridge, D.G., Watanabe, Y., Greenwood, S.J., Gray, M.W. and Schnare, M.N. 1999. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. Proc. Natl. Acad. Sci. USA 96(3): 852-856.

[127] Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I., Stern, L.S., Wang, Z., Ullu, E. and Tschudi, C. 2000. A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. RNA 6(2): 163-169.

[128] Keller, M., Tessier, L.H., Chan, R.L., Weil, J.H. and Imbault, P. 1992. In *Euglena*, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated. Nucleic Acids Res. 20(7): 1711-1715.

[129] Tessier, L.H., Chan, R.L., Keller, M., Weil, J.H. and Imbault, P. 1992. The *Euglena gracilis rbcS* gene contains introns with unusual borders. FEBS Lett. 304(2-3): 252-255.

[130] Stutz, E. and Vandrey, J.P. 1971. Ribosomal DNA satellite of *Euglena gracilis* chloroplast DNA. FEBS Lett. 17(2): 277-280.

[131] Nass, M.M. and Ben-Shaul, Y. 1972. A novel closed circular duplex DNA in bleached mutant and green strains of *Euglena gracilis*. Biochim. Biophys. Acta 272(1): 130-136.

[132] Curtis, S.E. and Rawson, J.R. 1981. Characterization of the nuclear ribosomal DNA of *Euglena gracilis*. Gene 15(2-3): 237-247.

[133] Cook, J.R. and Roxby, R. 1985. Physical properties of a plasmid-like DNA from *Euglena gracilis*. Biochim. Biophys. Acta 824(1): 80-83.

[134] Ravel-Chapuis, P., Nicolas, P., Nigon, V., Neyret, O. and Freyssinet, G. 1985. Extrachromosomal circular nuclear rDNA in *Euglena gracilis*. Nucleic Acids Res. 13(20): 7529-7537.

[135] Ravel-Chapuis, P. 1988. Nuclear rDNA in *Euglena gracilis*: paucity of chromosomal units and replication of extrachromosomal units. Nucleic Acids Res. 16(11): 4801-4810.

[136] Findly, R.C. and Gall, J.G. 1978. Free ribosomal RNA genes in *Paramecium* are tandemly repeated. Proc. Natl. Acad. Sci. USA 75(7): 3312-3316.

[137] Cockburn, A.F., Taylor, W.C. and Firtel, R.A. 1978. *Dictyostelium* rDNA consists of non-chromosomal palindromic dimers containing 5S and 36S coding regions. Chromosoma 70(1): 19-29.

[138] Campbell, G.R., Littau, V.C., Melera, P.W., Allfrey, V.G. and Johnson, E.M. 1979. Unique sequence arrangement of ribosomal genes in the palindromic rDNA molecule of *Physarum polycephalum*. Nucleic Acids Res. 6(4): 1433-1447.

[139] Engberg, J., Andersson, P., Leick, V. and Collins, J. 1976. Free ribosomal DNA molecules from *Tetrahymena pyriformis* GL are giant palindromes. J. Mol. Biol. 104(2): 455-470.

[140] Bhattacharya, S., Bhattacharya, A., Diamond, L.S. and Soldo, A.T. 1989. Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. J. Protozool. 36(5): 455-458.

[141] Huber, M., Koller, B., Gitler, C., Mirelman, D., Revel, M., Rozenblatt, S. and Garfinkel, L. 1989. *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules. Mol. Biochem. Parasitol. 32(2-3): 285-296.

[142] Bagchi, A., Bhattacharya, A. and Bhattacharya, S. 1999. Lack of a chromosomal copy of the circular rDNA plasmid of *Entamoeba histolytica*. Int. J. Parasitol. 29(11): 1775-1783.

[143] Clark, C.G. and Cross, G.A. 1987. rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. Mol. Cell. Biol. 7(9): 3027-3031.

[144] Clark, C.G. and Cross, G.A. 1988. Circular ribosomal RNA genes are a general feature of schizopyrenid amoebae. J. Protozool. 35(2): 326-329.

[145] Schnare, M.N., Cook, J.R. and Gray, M.W. 1990. Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis*. J. Mol. Biol. 215(1): 85-91.

[146] Schnare, M.N. and Gray, M.W. 1990. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. J. Mol. Biol. 215(1): 73-83.

[147] White, T.C., Rudenko, G. and Borst, P. 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res. 14(23): 9471-9489.

[148] Campbell, D.A., Kubo, K., Clark, C.G. and Boothroyd, J.C. 1987. Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J. Mol. Biol. 196(1): 113-124.

[149] Spencer, D.F., Collings, J.C., Schnare, M.N. and Gray, M.W. 1987. Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*. EMBO J. 6: 1063-1071.

[150] D'Alessio, J.M., Harris, G.H., Perna, P.J. and Paule, M.R. 1981. Ribosomal ribonucleic acid repeat unit of *Acanthamoeba castellanii*: cloning and restriction endonuclease map. Biochemistry 20: 3822-3827.

[151] Engberg, J., Nielsen, H., Lenaers, G., Murayama, O., Fujitani, H. and Higashinakagawa, T. 1990. Comparison of primary and secondary 26S rRNA structures in two *Tetrahymena* species: evidence for a strong evolutionary and structural constraint in expansion segments. J. Mol. Evol. 30(6): 514-521.

[152] Lenaers, G., Maroteaux, L., Michot, B. and Herzog, M. 1989. Dinoflagellates in evolution. A molecular phylogenetic analysis of large subunit ribosomal RNA. J. Mol. Evol. 29(1): 40-51.

[153] Lenaers, G., Scholin, C., Bhaud, Y., Saint-Hilaire, D. and Herzog, M. 1991. A molecular phylogeny of dinoflagellate protists (pyrrhophyta) inferred from the sequence of 24S rRNA divergent domains D1 and D8. J. Mol. Evol. 32(1): 53-63.

[154] Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. 1979. Sequence and secondary structure of *Drosophila melanogaster* 5.8S and 2S rRNAs and of the processing site between them. Nucleic Acids Res. 7(8): 2213-2238.

[155] de Lanversin, G. and Jacq, B. 1989. Sequence and secondary structure of the central domain of *Drosophila* 26S rRNA: a universal model for the central domain of the large rRNA containing the region in which the central break may happen. J. Mol. Evol. 28(5): 403-417.

[156] Avadhani, N.G. and Buetow, D.E. 1972. Isolation of active polyribosomes from the cytoplasm, mitochondria and chloroplasts of *Euglena gracilis*. Biochem. J. 128(2): 353-365.

[157] Heizmann, P. 1970. Properties of *Euglena gracilis* ribosomes and ribosomal RNA. Biochim. Biophys. Acta 224(1): 144-154.

[158] Smallman, D.S., Schnare, M.N. and Gray, M.W. 1996. RNA: RNA interactions in the large subunit ribosomal RNA of *Euglena gracilis*. Biochim. Biophys. Acta 1305(1-2): 1-6.

[159] Eisenstadt, J.M. and Brawerman, G. 1964. The protein-synthesizing systems from the cytoplasm and the chloroplasts of *Euglena gracilis*. J. Mol. Biol. 10: 392-402.

[160] Eisenstadt, J. and Brawerman, G. 1964. Characteristics of a cell-free system from *Euglena gracilis* for the incorporation of amino acids into protein. Biochim. Biophys. Acta 80: 463-472.

[161] Avadhani, N.G. and Buetow, D.E. 1974. Mitochondrial and cytoplasmic ribosomes. Distinguishing characteristics and a requirement for the homologous ribosomal salt-extractable fraction for protein synthesis. Biochem. J. 140(1): 73-78.

[162] Simpson, L., Sbicego, S. and Aphasizhev, R. 2003. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. RNA 9: 265-276.

[163] Madison-Antenucci, S., Grams, J. and Hajduk, S.L. 2002. Editing machines: the complexities of trypanosome RNA editing. Cell 108: 435-438.

[164] Tessier, L.H., van der Speck, H., Gualberto, J.M. and Grienenberger, J.M. 1997. The cox1 gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. Curr. Genet. 31: 208-213.

[165] Yasuhira, S. and Simpson, L. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. J. Mol. Evol. 44: 341-347.

[166] Rawson, J.R., Eckenrode, V.K., Boerma, C.L. and Curtis, S. 1979. DNA sequence organization in the alga *Euglena gracilis*. Biochim. Biophys. Acta 563(1): 1-16.

[167] Rawson, J.R. 1975. The characterization of *Euglena gracilis* DNA by its reassociation kinetics. Biochim. Biophys. Acta 402(2): 171-178.

[168] Pitt, C.W., Moreau, E., Lunness, P.A. and Doonan, J.H. 2004. The *pot1*$^{+}$ homologue in *Aspergillus nidulans* is required for ordering mitotic events. J. Cell. Sci. 117: 199-209.

[169] Carr, J. and Shearer, G.J. 1998. Genome size, complexity, and ploidy of the pathogenic fungus *Histoplasma capsulatum*. J. Bacteriol. 180(24): 6697-6703.

[170] Kondrashov, A.S. 1997. Evolutionary genetics of life cycles. Annu. Rev. Ecol. Syst. 28: 391-435.

[171] Perez-Castineira, J.R., Alvar, J., Ruiz-Pérez, L.M. and Serrano, A. 2002. Evidence for a wide occurrence of proton-translocating pyrophosphatase genes in parasitic and free-living protozoa. Biochem. Biophys. Res. Commun. 294: 567-573.

[172] Lighthall, G.K. and Giannini, S.H. 1992. The chromosomes of *Leishmania*. Parasitol. Today 8(6): 192-199.

[173] El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D.C., Haas, B.J., Tran, A.-N., Wortman, J.R., Alsmark, U.C.M., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J.M., Cerqueira, G.C., Creasy, T., Delcher, A.L., Djikeng, A., Embley, T.M., Hauser, C., Ivens, A.C., Kummerfeld, S.K., Pereira-Leal, J.B., Nilsson, D., Peterson, J., Salzberg, S.L., Shallom, J., Silva, J.C., Sundaram, J., Westenberger, S., White, O., Melville, S.E., Donelson, J.E., Andersson, B., Stuart, K.D. and Hall, N. 2005. Comparative genomics of trypanosomatid parasitic protozoa. Science 309: 404-409.

[174] McDaniel, J.P. and Dvorak, J.A. 1993. Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. Mol. Biochem. Parasitol. 57: 213-222.

[175] Henriksson, J., Åslund, L. and Pettersson, U. 1996. Karyotype variability in *Trypanosoma cruzi*. Parasitol. Today 12(3): 108-114.

[176] Parenti, F., Brawerman, G., Preston, J.F. and Eisenstadt, J.M. 1969. Isolation of nuclei from *Euglena gracilis*. Biochim. Biophys. Acta 195(1): 234-243.

[177] O'Donnell, E.H.J. 1965. Nucleolus and chromosomes in *Euglena gracilis*. Cytologia 30: 118-154.

[178] Dooijes, D., Chaves, I., Kieft, R., Dirks-Mulder, A., Martin, W. and Borst, P. 2000. Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. Nucleic Acids Res. 28: 3017-3021.

[179] Leedale, G.F. 1958. Mitosis and chromosome numbers in the Euglenineae (Flagellata). Nature 181(4607): 502-503.

[180] Rizzo, P.J. 2003. Those amazing dinoflagellate chromosomes. Cell Res. 13: 215-217.

[181] Rizzo, P.J. 1991. The enigma of the dinoflagellate chromosome. J. Protozool. 38(3): 246-252.

[182] Doetsch, N.A., Favreau, M.R., Kuscuoglu, N., Thompson, M.D. and Hallick, R.B. 2001. Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic *psbK* operon. Curr. Genet. 39: 49-60.

[183] Nakaar, V., Dare, A.O., Hong, D., Ullu, E. and Tschudi, C. 1994. Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol. Cell. Biol. 14(10): 6736-6742.

[184] Nakaar, V., Tschudi, C. and Ullu, E. 1995. An unusual liaison: Small nuclear and cytoplasmis RNA genes team up with tRNA genes in trypanosomatid protozoa. Parasitology Today 11(6): 225-228.

[185] Provasoli, L., Hutner, S.H. and Schatz, A. 1948. Streptomycin-induced chlorophyll-less races of *Euglena*. Proc. Soc. Exptl. Biol. Med. 69: 279-282.

[186] Heizmann, P., Doly, J., Hussein, Y., Nicolas, P., Nigon, V. and Bernardi, G. 1981. The chloroplast genome of bleached mutants of *Euglena gracilis*. Biochim. Biophys. Acta 653(3): 412-415.

[187] Cook, J.R. 1974. Irreversible Plastid Loss in *Euglena gracilis* under Physiological Conditions. Plant Physiol. 53(2): 284-290.

[188] Rodriguez-Zavala, J.S., Ortiz-Cruz, M.A. and Moreno-Sanchez, R. 2006. Characterization of an aldehyde dehydrogenase from *Euglena gracilis*. J. Eukaryot. Microbiol. 53(1): 36-42.

[189] Danilov, R.A. and Ekelund, N.G. 2001. Effects of pH on the growth rate, motility and photosynthesis in *Euglena gracilis*. Folia. Microbiol. (Praha) 46(6): 549-554.

[190] Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. 3: 208-218.

[191] Kirby, K.S. 1965. Isolation and characterization of ribosomal ribonucleic acid. Biochem. J. 96: 266-269.

[192] Lis, J.T. and Schleif, R. 1975. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. Nucleic Acids Res. 2(3): 383-389.

[193] Spencer, D.F., Gray, M.W. and Schnare, M.N. 1992. The isolation of wheat mitochondria DNA and RNA. In: Seed analysis; Modern Methods of Plant Analysis Linskens, H.F. and Jackson, J.F., eds. vol.14, pp. 347-360, Springer, Berlin, Heidelberg.

[194] Sherwood, A.L. 2003. Virtual elimination of false positives in blue-white colony screening. Biotechniques 34(3): 644-647.

[195] Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[196] Gaillard, C. and Strauss, F. 1990. Ethanol precipitation of DNA with linear polyacrylamide as carrier. Nucleic Acids Res. 18(2): 378.

[197] Feinberg, A.P. and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132(1): 6-13.

[198] Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98(3): 503-517.

[199] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. eds. 1989. Current Protocols in Molecular Biology. John Wiley & Sons, Toronto.

[200] Rasband, W.S. 2006. ImageJ. U.S. National Institutes of Health, Bethesda, Maryland, USA. http://rsb.info.nih.gov/ij/index.html

[201] Russell, A.G., Watanabe, Y., Charette, J.M. and Gray, M.W. 2005. Unusual features of fibrillarin cDNA and gene structure in *Euglena gracilis*: evolutionary conservation of core proteins and structural predictions for methylation-guide box C/D snoRNPs throughout the domain Eucarya. Nucleic Acids Res. 33(9): 2781-2791.

[202] Clarke, L. and Carbon, J. 1976. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell 9(1): 91-99.

[203] Staden, R., Judge, D.P. and Bonfield, J.K. 2001. Sequence assembly and finishing methods. Methods Biochem. Anal. 43: 303-322.

[204] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215(3): 403-410.

[205] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

[206] Dsouza, M., Larsen, N. and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13(12): 497-498.

[207] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25(24): 4876-4882.

[208] Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95-98.

[209] Weiser, B. and Noller, H.F. 1995. XRNA: Auto-interactive program for modeling RNA. The Center for Molecular Biology of RNA, University of California, Santa Cruz. Santa Cruz, California. http://rna.ucsc.edu/rnacenter/xrna/xrna.html

[210] Zhang, W.J., Yang, J., Yu, Y.H. and Shen, Y.F. 2005. Identification, cloning, and characterization of microsatellite DNA in *Euglena gracilis*. J. Eukaryot. Microbiol. 52(4): 356-359.

[211] Canaday, J., Tessier, L.H., Imbault, P. and Paulus, F. 2001. Analysis of *Euglena gracilis* alpha-, beta- and gamma-tubulin genes: introns and pre-mRNA maturation. Mol. Genet. Genomics 265(1): 153-160.

[212] Hopper, A.K. and Phizicky, E.M. 2003. tRNA transfers to the limelight. Genes Dev. 17(2): 162-180.

[213] Newman, A.J. 1997. The role of U5 snRNP in pre-mRNA splicing. EMBO J. 16(19): 5797-5800.

[214] Reddy, R. and Busch, H. 1988. Small nuclear RNAs: RNA sequences, structure, and modification. In: Structure and Functions of Major and Minor Small Nuclear Ribonucleoprotein Particles.; (Birnstiel, M.L., ed., pp. 1-37, Springer-Verlag, Berlin, Germany.

[215] Judo, M.S., Wedel, A.B. and Wilson, C. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26(7): 1819-1825.

[216] Bradley, R.D. and Hillis, D.M. 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14(5): 592-593.

[217] Meyerhans, A., Vartanian, J.-P. and Wain-Hobson, S. 1990. DNA recombination during PCR. Nucleic Acids Res. 18(7): 1687-1691.

[218] Hillis, D.M., Moritz, C., Porter, C.A. and Baker, R.J. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251(4991): 308-310.

[219] Russell, A.G., Schnare, M.N. and Gray, M.W. 2004. Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. RNA 10(7): 1034-1046.

[220] Matera, A.G., Tycowski, K.T., Steitz, J.A. and Ward, D.C. 1994. Organization of small nucleolar ribonucleoproteins (snoRNPs) by fluorescence *in situ* hybridization and immunocytochemistry. Mol. Biol. Cell. 5(12): 1289-1299.

[221] Marchetti, M.A., Tschudi, C., Silva, E. and Ullu, E. 1998. Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersionof RNA polymerase II- and III-transcribed genes. Nucleic Acids Res. 26(15): 3591-3598.

[222] Hartshorne, T. and Agabian, N. 1994. A common core structure for U3 small nucleolar RNAs. Nucleic Acids Res. 22(16): 3354-3364.

[223] Chen, C.L., Liang, D., Zhou, H., Zhuo, M., Chen, Y.Q. and Qu, L.H. 2003. The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. Nucleic Acids Res. 31(10): 2601-2613.

[224] Fantoni, A., Dare, A.O. and Tschudi, C. 1994. RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. Mol. Cell. Biol. 14(3): 2021-2028.

[225] Kiss, T., Marshallsay, C. and Filipowicz, W. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and *in vitro*. Cell 65(3): 517-526.

[226] Marshallsay, C., Connelly, S. and Filipowicz, W. 1992. Characterization of the U3 and U6 snRNA genes from wheat: U3 snRNA genes in monocot plants are transcribed by RNA polymerase III. Plant Mol. Biol. 19(6): 973-983.

[227] Geiduschek, E.P. and Kassavetis, G.A. 2001. The RNA polymerase III transcription apparatus. J. Mol. Biol. 310(1): 1-26.

[228] Schramm, L. and Hernandez, N. 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 16(20): 2593-2620.

[229] Morse, R.H., Roth, S.Y. and Simpson, R.T. 1992. A transcriptionally active tRNA gene interferes with nucleosome positioning *in vivo*. Mol. Cell. Biol. 12(9): 4015-4025.

[230] Thompson, M., Haeusler, R.A., Good, P.D. and Engelke, D.R. 2003. Nucleolar clustering of dispersed tRNA genes. Science 302(5649): 1399-1401.

[231] Haeusler, R.A. and Engelke, D.R. 2006. Spatial organization of transcription by RNA polymerase III. Nucleic Acids Res. 34(17): 4826-4836.

[232] Kosak, S.T. and Groudine, M. 2004. Gene order and dynamic domains. Science 306(5696): 644-647.

[233] Schul, W., Adelaar, B., van Driel, R. and de Jong, L. 1999. Coiled bodies are predisposed to a spatial association with genes that contain snoRNA sequences in their introns. J. Cell. Biochem. 75(3): 393-403.