

1988

The mRNA Cap and Discontinuous Transcription in *Trypanosoma brucei*

Marion Struik Freistadt

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations



Part of the [Life Sciences Commons](#)

Recommended Citation

Freistadt, Marion Struik, "The mRNA Cap and Discontinuous Transcription in *Trypanosoma brucei*" (1988). *Student Theses and Dissertations*. 384.

http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/384

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.



**The mRNA Cap and Discontinuous Transcription
in *Trypanosoma brucei***

**A thesis submitted to the Faculty of The Rockefeller University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy**

by Marion Struik Freistadt

March 17, 1988

The Rockefeller University

New York, N.Y.

Table of Contents

Abbreviations Used	v
Acknowledgements.....	vi
Abstract	1
Chapter I.Introduction	2
Background: Gene Structure and Function	2
Discontinuous Transcription.....	6
Trypanosomes	8
Polycistronic Transcripts.....	10
The Mini-exon Problem.....	11
Caps	14
Objectives of this thesis	15
Figures 1-7; Table I	17-32
Chapter II.Materials and Methods	33
Specificity and Polarity of Nuclease Cleavages.....	36
Chapter III.Purification and Analysis of medRNA.....	37
Results	37
Labeling.....	37
Hybrid Selection.....	38
Fingerprint Analysis of medRNA	39
Summary.....	42
Figures 8-13; Tables II & III	43-58
Chapter IV.Detection of a Free Mini-exon	59
Results	59
Other Small Selected RNAs.....	59
37mer.....	60
39mer.....	60
Secondary Structure Models.....	61
Summary.....	62
Figures 14-17; Table IV.....	63-72
Chapter V.Initial Analyses of the Cap.....	73
Results	73
Use of RNase T1 Fingerprints to Find the mRNA Cap	73

Cyclizing RNases	74
Initial Analysis with Nuclease P1	75
Artifact	75
Summary.....	76
Figures 18-20; Table V.....	77-84
Chapter VI.Unfingerprinted Cap.....	85
Results	85
Accelerated Purification of the Cap	85
Complete Nuclease P1 Digest of the Cap	86
Partial Nuclease P1 Digests of the Cap	88
i = iv + pA*	89
ii = iii + pC*	89
i + ii = Cap?	90
Venom Phosphodiesterase Analysis of the Cap	90
Partial Venom Phosphodiesterase Digests.....	91
Summary.....	92
Figures 21-34; Table VI.....	93-122
Chapter VII.Discussion.....	123
Figure 35	131
Chapter VIII.References.....	133

Abbreviations Used¹

~	approximately	N	any nucleotide
*	any modification	ND	no data
		NP	nucleotide
BC	Basic Copy		pyrophosphatase
bp	base pairs	PFG	Pulsed Field Gel
CAT	chloramphenicol	R	any purine
	acetyl transferase		
CIAP	calf intestinal	SAM	S-adenosyl
	alkaline phosphatase		methionine
cpm	counts per minute	SDM	semi-defined media
EC	Expressed Copy	SDS	sodium dodecyl sulfate
ESAG	Expression Site	SL	spliced leader (mini-exon)
	Associated Gene	snRNA	small nuclear RNA
hr	hour		
		snRNP	small nuclear
hnRNPs	heterogeneous nuclear		ribonucleoprotein
	ribonucleoprotein		
		TAPase	tobacco acid
HVE	high voltage		pyrophosphatase
	electrophoresis	TCA	trichloro-acetic acid
kbp	kilobase pairs	TES	N-tris(hydroxymethyl)
			methyl-2-aminoethane
			sulfonic acid
		TLC	thin layer
kDa	kilodaltons		chromatography
l	liter	Tm	melting temperature
		VP	venom
m ⁷ G	7-methylguanosine		phosphodiesterase
		VSG	Variant-specific
m ^{2,2,7} G	2,2,7-trimethylguanosine		Surface Glycoprotein
medRNA	mini-exon donor RNA	Y	any pyrimidine

¹Unless otherwise specified, Journal of Biological Chemistry standard abbreviations are used.

Acknowledgements

Many people contributed towards the completion of this work. I thank Drs. Hugh Robertson and George Cross for their continual insight and encouragement. I wish to thank The Rockefeller University graduate program, the Dean's Office, members of the faculty and all the members of the Cross and Robertson laboratories for providing me with a pleasant and productive working environment.

Abstract

The protozoan parasite, *Trypanosoma brucei*, and other members of the phylogenetic order, Kinetoplastida, synthesize mRNA by a novel pathway that joins RNAs from physically unlinked gene segments. This results in the presence of a short non-coding RNA leader, called the mini-exon, at the 5' end of most, probably all, mRNA. Indirect evidence suggests that trans-splicing, rather than RNA priming of transcription, is the mechanism by which the mini-exon is joined to RNA containing protein-coding sequences. *In vivo* ³²P-labeled RNA was purified by hybrid selection with mini-exon genomic DNA, in order to analyze the structure of medRNA, the 140-nucleotide precursor of the mini-exon in mRNA. Some possible intermediates in the joining reaction and the 5' mini-exon cap structure were detected and characterized.

Using this direct RNA analysis, the 5' and 3' ends of medRNA were precisely mapped and an internal modification was detected. Two small mini-exon-selected RNAs, a 39mer and a 37mer, were purified, sequenced and found to consist of the free mini-exon and a 3' - truncated version, respectively. Both ended with a 3' OH. The 39mer is a candidate biosynthetic intermediate because its structure is consistent with the proposed trans-splicing mechanism of joining.

The cap structure from *T. brucei* mRNA was found to consist of a novel "cap 4" structure: m⁷GpppA^{*}A^{*}C(2'-O-m)U^{*}A, where asterisks denote modification. The modified nucleotides in the cap were sequenced by a combination of partial and complete digests with non-specific nucleases. The four encoded, modified nucleotides are likely to have 2'-O-methylations, and possibly two other unidentified base modifications. The mRNA cap was indistinguishable from the medRNA cap. The cap structure is likely to be similar in the other Kinetoplastidae. This hypermethylated cap structure may have a functional role in discontinuous mRNA synthesis. These findings will be useful in future *in vitro* and *in vivo* studies. In particular, an authentic trypanosome cap may be necessary to obtain an *in vitro* mini-exon joining reaction.

CHAPTER I INTRODUCTION

BACKGROUND: GENE STRUCTURE AND FUNCTION

The concept of a gene, an entity that stores genetic information but is separate from the soma, can be traced back to Mendel (1865). The description of genes as linear arrays on chromosomes, such as beads on a string, was explicitly articulated and experimentally tested by Morgan (1910). This fundamental concept has withstood experimentation since that time, and, with modification, is considered the appropriate way to view the repository of genetic information in both prokaryotes and eukaryotes. Other patterns of genetic storage, such as two-dimensional networks, have been largely discounted. This thesis is concerned with one of several significant exceptions to the concept of genes as continuous linear units.

The first experimental evidence for the idea of "one gene, one enzyme" came from Beadle and Tatum (1941). Nutritional requirements in *Neurospora crassa* were shown to be inherited and recombined in a discrete fashion, implying the existence of discrete units of genetic information. Jacob and Monod (1961) refined the concept of the gene in two ways: they inferred the existence of genes whose products modulate the basal level of the expression of nutritional genes and they expanded it to include specific regulatory regions, which respond to the regulatory proteins. Although the spatial relationship of the regulator to the regulated was unknown, genetic information was still considered linear. Benzer (1959) elevated the concept of linearity to a fine art with the detailed deletion mapping of mutations within a bacteriophage gene. The subunits of genes, nucleotides, became the "beads" on the string, but a gene was still considered a continuous linear unit of information. Current views about the biochemical basis for the linear nature of genetic information rely on studies of genomic organization and of the enzymes involved in nucleic acid metabolism.

The C (for DNA content) value paradox (Gall, 1981; Lewin, 1980) was recognized when it was found that there is often more DNA in a eukaryotic genome than is necessary to encode the polypeptides that make up the organism. Some structural genes may be repetitive, or amplified, because large amounts of a product are necessary at particular times (Brown and Dawid, 1968; Laird et al, 1974). However, there is no straightforward relationship between DNA content and other physical parameters, such as number of cells, number of proteins or evolutionary position. The excess DNA consists of several classes of non-coding, repetitive or satellite DNA (Britton and

Kohne, 1968). Some of this DNA has a role in chromosome structure, such as telomeres (Shampay et al, 1984; Bernards et al, 1984) or centromeres (Radic et al, 1987). Regulatory functions have been postulated for RNA transcribed from satellite DNA (Britton and Davidson, 1969; Robertson and Dickson, 1974; Jelinek et al, 1980) and, indeed, novel roles, some perhaps regulatory, have been found (Walter and Blobel, 1982; Steitz, 1986). However, the function of one type of repetitive DNA sequences, the mobile "retrotransposons" is not clear (Jelinek and Schmidt, 1982).

The definition of a gene as a "transcription unit," coding for a product, either a protein or an RNA, emerged as the biochemical requirements--polymerases, promoters, accessory factors, terminators, and chromatin structure--for constitutive and regulated gene expression were elucidated (Darnell et al, 1986). Prokaryotes have one polymerase to transcribe all genes; sigma factors can regulate transcription initiation (McClure, 1985; Darnell et al, 1986). Many prokaryotic transcription units are rapidly induced or repressed by regulatory proteins in response to environmental changes (Jacob and Monod, 1961; Ptashne, 1986; Darnell et al, 1986). Upstream regions that are responsible for basal and regulated gene expression have been studied by genetic mutations, sequence comparisons and DNA-binding assays (Jacob and Monod, 1961; Pribnow, 1975; McClure, 1985; Ptashne, 1986). Other mechanisms, such as termination/anti-termination, also control the utilization of prokaryotic transcription units (Darnell et al, 1986, Ptashne, 1986).

Eukaryotes have three polymerases, polymerases I, II and III, which transcribe ribosomal RNA (rRNA), mRNA and small RNAs (such as tRNA and 5S RNA), respectively (Roeder, 1977; Breathnach and Chambon, 1981). Eukaryotic transcription units are mapped by a variety of methods including DNA and RNA sequencing (Barrell, 1971; Brownlee, 1972; Maxam and Gilbert, 1977; Sanger 1981; Donis-Keller et al, 1977), Northern blots (Thomas, 1980), *in vitro* transcription in isolated nuclei (Manley et al, 1979), primer extension (Taylor, 1976), S-1 mapping (Berk and Sharp, 1977), and RNase protection (Zinn et al, 1983) experiments. Regulatory regions (for all three classes of genes), consisting of proximal and distal upstream regulatory elements (McKnight and Kingsbury, 1982; Reeder, 1984; Zinn et al, 1983; Darnell et al, 1986; Sollner-Webb, 1988), internal control regions (Sakonju et al, 1980; Hofstetter et al, 1981) and enhancers (Khoury and Gruss, 1983; Reeder, 1984) have been delineated by many experiments. Accurate *in vitro* and *in vivo* functional assays, such as *in vitro* transcription (Roeder et al, 1977; Sakonju et al, 1980; Hofstetter et al, 1981), transfection (Wigler et al, 1977; Zinn et al, 1983), micro-injection of oocytes (Hofstetter et al, 1981; Birchmeier et al, 1982), CAT assays (Gorman et al, 1987) and transgenic mice (Palmiter and Brinster, 1985) combined with site specific mutagenesis (McKnight

and Kingsbury, 1982), have been essential in the elucidation of these sequences. Experiments concerning DNase I sensitive and hypersensitive sites have shown that chromatin structure is an important aspect of accessibility of genes to polymerases (Wu, 1984; Darnell et al, 1986). Constitutive and regulatory factors that provide the specificity for the interaction between the polymerases and their target sequences are currently being purified and characterized in many laboratories (Engelke et al, 1980; Nabel and Baltimore, 1987). A unifying theme is developing: the positioning of short regulatory sequence motifs of 10-15 nucleotides combined with the availability of cognate recognition proteins likely provides the differential regulation of transcription initiation (Baldwin and Sharp, 1988). Eukaryotic transcription termination is not well delineated (Birnstiel et al, 1985); however, there are some known cases of differential transcription termination (Hay et al, 1982; Bentley and Groudine, 1986).

Further advances in the biochemistry of gene expression revealed novel processes, which necessitated modification of the concept of genetic linearity: "transcription unit" was redefined to mean an area, whose primary transcript is post-transcriptionally modified, or processed, to form mature mRNA. The processing can include capping (Shatkin, 1976; Banerjee, 1980), nucleotide modification (Perry et al, 1975), cleavage (Abelson, 1979), splicing (Abelson, 1979; Breathnach and Chambon, 1981; Padgett et al, 1986), 3' processing (Birchmeier et al, 1982) and polyadenylation (Darnell et al, 1986). The most unusual of these discoveries was that of intervening sequences, or introns (Berget et al, 1977; Chu et al, 1977), interruptions of variable size in the coding regions, or exons. Introns, which are removed at the RNA level by splicing, have been found in virtually every type of transcript: rRNA, mRNA and tRNA transcripts from organelles and nuclei can be spliced (Abelson, 1979; Michel and Dujon, 1983; Shinozaki et al, 1986; Padgett et al, 1986). Splicing and polyadenylation probably occur in prokaryotes as well as eukaryotes (Chu et al, 1984; Karnik et al, 1986).

There are three categories of introns (and thus three splicing mechanisms): Group I, Group II and tRNA (Michel and Dujon, 1983; Padgett et al, 1986; Cech and Bass, 1986). The distinction between Group I and II, first applied to yeast mitochondrial introns, now appears to be a more general grouping, reflecting two quite different intron folding and catalysis mechanisms (Michel and Dujon, 1983; Padgett et al, 1986). Autocatalytic examples of both groups are now known (Kruger et al, 1982; Waring and Davies, 1984; Peebles et al, 1986; Cech and Bass, 1986). Due to the development of accurate and efficient *in vitro* splicing reactions (Padgett et al, 1986), many of the mechanistic aspects of splicing are now understood and, despite the difference between these groups of reactions, certain underlying mechanisms, in particular, the trans-esterification

reactions (see Fig. 1 and below), are fundamentally similar and may be derived from a single ancestral mechanism (Sharp, 1985; Cech and Bass, 1986).

For Group I splicing, a guanosine nucleotide co-factor attacks the phosphodiester bond at the 5' exon/intron boundary, releasing the 5' exon and the intron/3' exon with the guanosine residue at its new 5' end. In the second step, the 3' OH of the 5' exon attacks the 3' intron/exon boundary, causing the ligation of the exons and the release of the intron. Some interesting side reactions, such as circularization of the intron, can also occur. There is a spectrum of secondary and tertiary structures that an intron RNA can form: the closer the secondary structure approximates an optimal structure for cleavage, the higher the likelihood is that it will be autocatalytic (Warren and Davies, 1983). This is why not all group I introns are autocatalytic, although, by definition, they all have a canonical secondary structure.

The Group II introns, which include the introns from nuclear mRNA, are spliced in two steps (Fig. 1; Padgett et al, 1986), involving the formation of a novel branch-point, a 2', 3', 5' triphosphate nucleotide (Wallace and Edmonds, 1983). The 2' OH of the nucleotide destined to become the branch point nucleotide first attacks the phosphodiester bond at the 5' exon/intron boundary, releasing the 5' exon with a free 3' OH, while a lariat consisting of the intron and the 3' exon is formed. In the second step, the free 3' OH of the 5' exon attacks the 3' exon/intron boundary, releasing the intron-lariat as the exons are ligated.

Small nuclear ribonucleoproteins (snRNPs) are likely to be the catalytic molecules for mRNA splicing (Padgett et al, 1986; Steitz, 1986) and possibly polyadenylation (Moore et al, 1988; Hart et al, 1985). Although the actual mechanisms of catalysis are not understood, it is probable that snRNPs facilitate the folding of the substrate RNA into a specific shape for catalysis. Similar to the exon-binding sites in the intron RNA from *Tetrahymena* (Cech and Bass, 1986), the RNA moieties probably hybridize to complementary regions, bridging the exons, while the protein components of the snRNPs and hnRNPs position or stabilize the RNA structures for cleavage and ligation.

tRNA splicing uses completely different mechanisms, sometimes involving a 2', 3' cyclic phosphorylated intermediate (Abelson, 1979; Greer et al, 1983; Filipowicz and Shatkin, 1983; Laski et al, 1983). Some of the enzymes responsible for yeast tRNA splicing have been purified (Abelson, 1979; Greer et al, 1983; Peebles et al, 1983) and do not appear to utilize snRNPs. The structure of the exons, rather than the introns seems important for specificity. Splicing, as

opposed to ligation, refers to the joining of pieces that were previously in a single molecule. However, the ligation during splicing requires the joining of physically distinct entities.

Prokaryotic and eukaryotic transcription units can be simple or complex; that is, they can code for single or multiple products (Leff et al, 1986). Organelles often have polycistronic transcripts that are cleaved (Montoya et al, 1983; Shinozaki, 1986). Prokaryotic and eukaryotic rRNA and tRNA transcripts are processed into functional RNAs (Dunn and Studier, 1973; Abelson, 1979; Lewin, 1980; Darnell et al, 1986). Coordinate expression of functionally related polypeptides in prokaryotes is achieved by operons, which are clusters of open reading frames co-transcribed under the control of a single, regulated promoter (Jacob and Monod, 1961; Darnell et al, 1986; Ptashne, 1986). Each polypeptide from a polycistronic message is translated by an independent initiation event. In eukaryotes, complex transcription units are formed by alternative splicing or polyadenylation or both. These processes are highly regulated, both in tissue specific and in developmental manners (Leff et al, 1986). Examples of complex transcription units are the major late transcript of adenovirus (Ziff, 1980) and the troponins (Breitbart et al, 1985). Although the genetic locus may be quite complex (for example, the rat troponin T gene has the potential to code for 64 distinct mRNAs), each cellular transcript consists of a single open reading frame: examples of eukaryotic (non-organelle) polycistronic transcription units are not known, except in certain viruses and the Kinetoplastidae (see below).

Other factors, such as RNA turnover, differential translation and proteolytic processing, can affect gene expression. In prokaryotes, differential message stability does not occur because mRNA has a very short half-life (Darnell et al, 1986); the turnover is used to permit rapid response and recovery to environmental changes. In contrast, eukaryotic mRNA is likely turned over by very controlled mechanisms. Several cases of regulatory differential message stability are known (Jefferson et al, 1984). The stem-loop structure, instead of polyadenylation (Birchmeier et al, 1982), at the 3' end of histone mRNA is likely to be related to its tight cell-cycle regulation. However, the full impact of mRNA stability as a regulatory component is still unknown. The best characterized example of regulated control of translation is attenuation (Yanofsky, 1981; Hay et al, 1982). Many peptide hormones are synthesized as poly-protein precursors that are cleaved in tissue specific manners (Leff et al, 1986).

DISCONTINUOUS TRANSCRIPTION

A level of complexity superimposed upon those described is physical discontinuity within genes. Recent findings have revealed situations where not only can a gene be interrupted, but the

interrupted pieces were not previously linked. Although such processes might appear to introduce chaos into regulatory mechanisms, and hence expected to be quite rare, Figure 2 presents ten examples. In these cases, a gene is difficult to define because it is not possible to specifically determine which previously unlinked coding regions, or transcripts, contributed to a particular mRNA. The phrase "discontinuous transcription" is used in this thesis to refer to any process by which unlinked gene segments result in a single RNA.

In the first three cases in Figure 2, several phylogenetically unrelated viruses use primers to facilitate initiation of transcription. The mRNAs of each of these viruses contain discontinuously encoded leader sequences. The nonsegmented coronaviruses, a group of plus strand RNA viruses, utilize an ~70 nucleotide viral-encoded primer to prime mRNA transcription (Lai, 1986). UV inactivation studies were required to distinguish between a priming and splicing mechanism (Jacobs et al, 1981). Influenza virus and bunyavirus, segmented, single stranded, negative RNA viruses, parasitize host messages as primers for mRNA transcription (Plotch et al, 1979; Robertson et al, 1980; Patterson et al, 1984). These two types of cytoplasmic viruses carry a viral encoded RNA-dependent RNA polymerase when they enter the host cell (Plotch et al, 1979). An endonucleolytic activity cleaves the caps plus 10-15 nucleotides from any host messages (Robertson et al, 1980). Coronaviruses differ from the latter two viruses in that the coronavirus leader is homogeneous and viral-encoded, while the influenza virus and bunyavirus leaders are heterogeneous and host-encoded. The coronavirus primer may have differential affinity, based on sequence complementarity, for the regions upstream of the open reading frames (Budzilowicz et al, 1985; Makino et al, 1986). Thus, in these three cases, the viral messages are encoded by previously unlinked pieces of genetic information. It is possible that the use of RNA primers by viruses may serve to simplify the number of factors necessary to initiate and regulate transcription.

In the next four cases (Fig. 2), it is likely that post-transcriptional mechanisms join previously unlinked gene segments. These situations have been called "trans-splicing," in contrast to intramolecular splicing, or "cis-splicing". In the 155 kbp circular genome of tobacco (and other) chloroplasts, the gene for ribosomal protein S12 appears to be split into three exons (Shinozaki et al 1986; Zaita et al, 1987). A continuous copy of the sequences encoding the two 3' exons is found in each of two large (25 kbp) inverted repeats. The two exons, in each repeat, are separated from one another by a conventional intron. However, the single copy of the 5' exon is located 86 or 28 kbp (depending on the direction) from either pair of 3' exons (Zaita et al, 1987). Molecules with structures that could be primary transcripts from the two different regions, as well as a fused mRNA, have been detected by Northern blots, primer extension (Zaita et al, 1987), heteroduplex analysis (Koller et al, 1987) and S-1 mapping (Hildebrand et al, 1988). These

combined indirect data argue strongly that the mature mRNA is formed by a post-transcriptional joining of two separate RNAs.

Messenger RNA formation in parasitic Kinetoplastidae consists of the joining of RNAs from two chromosomally unlinked gene segments (reviewed in Borst, 1986; Boothroyd, 1985; Donelson and Rice-Ficht, 1985). This thesis is concerned with this process in *Trypanosoma brucei*. Similar mechanisms have also been detected in unrelated organisms. In the nematode, *Caenorhabditis elegans*, some, but not all, messages have an identical 22-nucleotide leader sequence, which is encoded separately from protein-coding regions (Krause and Hirsch, 1987). An identical leader has been detected at the 5' end of an mRNA from *Brugia malayi* (T. Nilsen, unpublished observations). In another case, RNA transcribed from distant regions and opposite strands of Vaccinia virus are found in single messages (Bertholet et al, 1987; Schwer et al, 1987).

In three different systems, artificial trans-splicing of naturally cis-spliced substrates was achieved. Separate *in vitro* synthesized transcripts, each containing single exons with partial introns, of globin (Solnick, 1985) and adenovirus (Konarska et al, 1985) pre-mRNAs were mixed, incubated in Hela cell extracts, and assayed for splicing. The efficiency of trans-splicing was increased when the half-introns contained regions of complementarity (Solnick, 1985). The exons flanking a group I intron from the yeast mitochondrial gene, *oxi 3*, can undergo autocatalytic trans-splicing, even with remarkably large deletions from the intron portions (Jacquier and Rosbash, 1986). Three kinds of experiments using artificial substrates from the autocatalytic *Tetrahymena thermophilus* rRNA intron have shown that trans-splicing activity is intrinsic in some very small sequences. Small pyrimidine-rich oligoribonucleotides functioned as autocatalytic 5' trans-exons (Inoue et al, 1985) and, in separate experiments, portions of the intron displayed an autocatalytic oligomerization (Zaug and Cech, 1986a) or nucleotidyl transferase activity (Zaug and Cech, 1986b). Although the physiological significance of these *in vitro* experiments is not clear, the results suggest that cis- and trans-splicing may be variations of a similar underlying mechanism.

TRYPANOSOMES

Discontinuous synthesis of a cellular mRNA was unexpectedly discovered in the protozoan parasite, *Trypanosoma brucei*. This African trypanosome is from the phylogenetic order, Kinetoplastida (Fig. 3), which branched from the other protozoa quite early in evolution (Sogin et al, 1986) and is divergent from other eukaryotes in several regards. *T. brucei* and related species cause trypanosomiasis (sleeping sickness) in humans, nagana in cattle and corresponding diseases in many types of mammals (Kuzoe, 1986). Other Kinetoplastidae are responsible for

several major human diseases, Chagas' disease and leishmaniasis. Eukaryotic pathogens have evolved complex strategies (when compared to prokaryotic pathogens) for survival in their respective hosts (Trager, 1986) and present serious medical, scientific and socioeconomic problems (Kuzoe, 1986). Discontinuous mRNA synthesis has been detected in all the Kinetoplastidae that have been examined (Nelson et al, 1983b; DeLange et al, 1984a; Landfear and Wirth, 1985; Gonzales et al, 1985; Muhich et al, 1987; Cook and Donelson, 1987; Aksoy et al, 1987). Other fascinating biochemical peculiarities of trypanosomes, such as the glycosome and its developmental regulation (Oppenheimer et al, 1984; Clayton, 1985), the kinetoplast and its DNA (Englund et al, 1982), a novel post-transcriptional in-frame uridylation (Hensgens et al, 1984; Feagin and Stuart, 1988), the phosphoglycolipid membrane anchor (Ferguson and Williams, 1988) and trypanothione (Fairlamb et al, 1985) are outside the scope of this thesis.

The life cycle (Fig. 4) of the African trypanosome alternates between the mammalian host, where it can cause disease, and the insect vector, *Glossina* species (the tse-tse fly). The form of the parasite in mammals is called the bloodstream form and the form from the midgut of insects is called the procyclic form, or procyclics (Schmidt and Roberts, 1985).

Many trypanosome species utilize a unique mechanism to evade the mammalian host's immune response, which is largely humoral (Vickerman and Barry, 1984). An infection is characterized by successive waves of antigenically distinct parasites in the bloodstream and interstitial fluids. The surface of the parasite is covered with a dense coat, which is composed of a homogeneous layer of a single protein, the Variant-specific Surface Glycoprotein, or VSG (Cross, 1975). A sequential switching in the particular species of VSG protein that is expressed at the cell surface explains the immunological phenomenon of antigenic variation (Cross, 1975). The VSGs, which are about 60 kDa in size, have several well conserved structural features (Cross, 1978; Cross, 1984). Parallel changes in the expression of individual VSG genes accompany the antigen switches (Borst and Cross, 1982). However, despite much research, the actual trigger for changes in VSG gene expression is not understood (Hoeijmakers et al, 1980a; Hoeijmakers et al, 1980b; Borst, 1986; Boothroyd, 1985; Donelson and Rice-Ficht, 1985).

The diploid genome size of *T. brucei* is approximately 4×10^7 base pairs (Borst, 1986). The relationship between ploidy and life cycle is not clear, and there is no classical genetic system available. Like yeast, trypanosomes have a "closed" mitosis and do not form cytologically visible chromosomes (deSouza, 1984). Four size classes of linear chromosomes have been characterized by Pulsed Field Gel (PFG) electrophoresis (Van der Ploeg et al, 1984c; Van der Ploeg et al, 1984b). There are many, perhaps several hundred, mini-chromosomes (50-100 Kb),

about 6 chromosomes in the 200-500 kbp range, several megabase chromosomes and "slot" DNA (too large to enter the gel). The mini-chromosomes occur only in Kinetoplastidae that undergo antigenic variation and do not appear to contain any expressed DNA (Van der Ploeg et al, 1984a; Gibson and Miles, 1986) so they may function in generating or maintaining VSG gene diversity; for instance, they may be intermediates in gene re-arrangement. Portions of the *T. brucei* genome may be quite unstable; the advantages and compensatory stabilizing factors are not clear.

Unlike antibody gene diversity, the entire repertoire of VSG genes, perhaps consisting of 1000 genes each about 2 kbp in size, is carried in the genome (Borst, 1986). The VSG gene repertoire is highly mobile and may be constantly evolving (Longacre and Eisen, 1986; Roth and Eisen, 1986). The unexpressed VSG genes, known as basic copies (BC), reside in clusters. DNA rearrangement frequently, but not absolutely, accompanies the activation of a VSG gene at a telomeric "expression site". The expressed copy of a VSG gene is called the expression copy (EC). Although only one VSG is expressed at a time, it is not known how many expression sites there are. The ECs are flanked by "barren" regions, repetitive DNA lacking restriction sites. 3' to the VSG expression site coding regions are telomeric repeats (Bernards et al, 1983). 5' to these coding regions are long regions which probably mediate homologous recombination or gene conversion into the expression site (Campbell et al, 1984b). Other genes and gene families, such as the "Expression Site Associated Genes" (ESAGs) have been detected upstream of the barren regions (Cully et al, 1983; Kooter et al, 1987).

Several classes of repetitive DNA have been detected in trypanosomes. In addition to telomeric repeats and the barren regions, several repetitive elements have been cloned (Hasan et al, 1984; Murphy et al, 1987; Aksoy et al, 1987). These appear to be retroposon-like elements because they have open reading frames, direct repeats at their termini and poly-A tracts at one end (Jelinek and Schmidt, 1982). Some of these elements appear to be transcribed (Murphy et al, 1987).

POLYCISTRONIC TRANSCRIPTS

Two peculiarities of trypanosome mRNA biogenesis have hampered progress in studying gene expression. These are the apparent utilization of polycistronic transcripts and discontinuous mRNA synthesis. No conventional eukaryotic promoters, nor a novel trypanosome consensus promoter sequence, have been detected upstream of expressed VSG genes or other structural genes. Recent experimental evidence suggests that the primary transcripts for VSG genes and ESAGs may be very long, perhaps up to 50 kbp (Kooter et al, 1987; Johnson et al, 1987). These

long primary transcripts must be processed at their 5' and 3' ends. Some VSG genes (but not other trypanosome structural genes) may be transcribed by polymerases similar to polymerase I from other eukaryotes (Shea et al, 1987). Two types of evidence, insensitivity to α -amanitin during transcription in nuclei *in vitro* (Kooter et al, 1984; Kooter and Borst, 1984) and putative homology to rRNA gene promoters (Shea et al, 1987) support this hypothesis. Although very little biochemical purification of trypanosome polymerases has been carried out, recent work suggests that trypanosomes have several different polymerases, with both conventional and nonconventional properties (Kitchen et al, 1984; Earnshaw et al, 1987; Tittawella, 1988).

Many trypanosome protein-coding regions, in addition to those for VSG, occur in tandem repeats (Sather and Agabian, 1985; Tschudi et al, 1985; Clayton, 1985; Glass et al, 1986; Osinga et al, 1985). Although tubulin genes appear to be transcribed by a polymerase II-like activity, by sensitivity to α -amanitin in nuclei *in vitro* (Kooter et al, 1984), these and other genes may also be synthesized as polycistronic transcripts (Gonzales et al, 1985; Imboden et al, 1988; Tschudi and Ullu, 1988), which would have to be processed into mRNA.

THE MINI-EXON PROBLEM

When the 5' end of a VSG gene was studied by S1 mapping and primer extension of the mRNA, a slight discrepancy in length between the genomic clone and the cDNA was detected in S1 mapping experiments (Bernards et al, 1981). Primer extension revealed a short non-coding upstream sequence (Van der Ploeg et al, 1982; Boothroyd and Cross, 1982). Since this segment appeared to be a conventional exon, it was termed the "mini-exon". The sequence is also referred to as the "spliced leader" (Nelson et al, 1983a; Dorfman and Donelson, 1984). It appeared to be 35 nucleotides long. It was unusual in that several different VSG mRNAs had the same sequence. However, the most unusual aspect of this leader was that it was not detectable 50 kbp upstream of VSG genes (DeLange et al, 1983). When DNA encoding the leader was cloned, it was found to be unlinked to VSG genes. In several cases, using PFG electrophoresis, it has been demonstrated that expressed VSG genes are on chromosomes containing no mini-exon sequences (Van der Ploeg et al, 1984b; Guyaux et al, 1985). Although DNA rearrangement occurs frequently in trypanosomes, it cannot explain the presence of the mini-exon at 5' ends of mRNAs.

The *T. brucei* mini-exon is encoded in a 1.35 kbp DNA segment which is present in about 200 copies, many of which are tandem repeats (DeLange et al, 1983; Nelson et al, 1983a). The primary transcript from this region is a 140-nucleotide RNA, termed mini-exon donor RNA, or

medRNA (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984). medRNA is abundant and has a half-life of about 6 minutes (Ehlers et al, 1987; Laird et al, 1987; Laird et al, 1988). Consistent with discontinuous transcription, the sensitivity of medRNA transcription to α -amanitin in nuclei *in vitro* is different than that of VSG, as well as other genes (Kooter et al 1984, Laird et al, 1985). medRNA is encoded in the megabase chromosomes and in the "slot" DNA (Van der Ploeg et al, 1984b). Data from *in vitro* nuclear transcription experiments suggest that the regions flanking medRNA in the genomic repeats are not transcribed (Kooter and Borst, 1984).

Since this early work, the mini-exon has been detected at the 5' end of every kinetoplastid mRNA that has been examined (Sather and Agabian, 1985; Tschudi et al, 1985; Clayton, 1985; Gonzales et al, 1985; Landfear and Wirth, 1985; Osinga et al, 1985; Swinkels et al, 1986). A Northern blot with a mini-exon probe yields a smear (DeLange et al, 1983, Nelson et al, 1983a). Estimates based on its frequency in random cDNA clones predict its existence on every trypanosome mRNA (DeLange et al, 1984b, Parsons et al, 1984). Experiments in which oligodeoxynucleotides complimentary to the mini-exon were used to arrest cell-free translation also suggest that every trypanosome mRNA contains the mini-exon (Walder et al, 1986; Cornelissen et al, 1986). Although it is difficult to definitively establish whether the mini-exon is on every mRNA, it is likely. One explanation for the repetitive nature of the medRNA genomic sequences is that the mini-exon RNA is highly utilized. Several mRNAs appear to have the mini-exon joined to different sites (Tschudi et al, 1985; Layden and Eisen, 1988). mRNA from procyclic trypanosomes has the mini-exon, so the problems of antigenic variation and discontinuous mRNA biosynthesis are separate problems. Thus, some of the peculiarities of VSG gene expression, such as telomeric expression sites, are specific for VSG genes, while other aspects, such as polycistronic primary transcripts and the presence, in the mature mRNA, of the mini-exon, are particular to trypanosome, and perhaps kinetoplastid gene expression.

Table I presents a compilation of information about the mini-exon genomic sequences from the different kinetoplastid species that have been examined. The genomic organization of mini-exon sequence and the medRNA length differs among the Kinetoplastidae; however, certain features, such as the length of the mini-exon and the presence of tandem repeats, are conserved (Milhausen et al, 1984; DeLange et al, 1984a; V. Bellofatto, unpublished observations; Miller et al, 1986; Muhich et al, 1987; Cook and Donelson, 1987).

Two models, priming and trans-splicing, for discontinuous mRNA biosynthesis have emerged. As in the viral cases, if medRNA or a cleavage product primed transcription of structural genes, the joining of the gene segments would precede transcription of the 3' segment. However, if trans-

splicing were operative, the joining would be post-transcriptional. The data currently favor the trans-splicing model. Considering that the mini-exon portion of medRNA is analogous to a conventional 5' exon, while the 3' portion of medRNA is like a 5' half-intron, I propose the term "extran" (5' and 3') to refer to each of the trans-exons and "intran" (5' and 3') to refer to the half-intron-like portions of trans-spliced RNAs. Some aspects of conventional splice consensus sequences have been found in *T. brucei*, such as GU at the 5' extran/intran boundary as well as pyrimidine-rich stretches and AG at the 3' boundaries, but TACTAAC-type consensus sequences (Langford and Gallwitz, 1983) have not been detected. A U2-type RNA has been detected and cloned (Tschudi et al, 1986). No conventional introns interrupting the coding sequences have been detected in any genes of the Kinetoplastida. The conventional eukaryotic polyadenylation sequence AAUAAA (Proudfoot and Brownlee, 1976) has not been detected in any trypanosome cDNA.

The trans-splicing model is supported by indirect evidence. If mechanisms similar to cis-splicing were operative in trans-splicing, the reaction would occur in two steps (Fig. 5). In the first step, the 5' extran, terminating in a 3' OH, would be released. At the same time, a Y-shaped structure, similar to a nicked lariat, containing the 5' and 3' intrans and the 3' extran, would be formed. In the second step, the extrans would be ligated as the intrans structure is released. In trypanosomes, the long polycistronic transcripts for structural genes must be processed at their 5' and 3' ends. Polycistronic transcription is not inherent in the discontinuity problem, but makes it more difficult to analyze. Experiments from several labs (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987) showed that the 3' portion of medRNA, the 5' intran, can be released from trypanosome poly(A)⁺ RNA by a Hela cell debranching enzyme (Arenas and Hurwitz, 1987). Figure 6 depicts a model for trypanosome gene expression based on these experiments and what is known about cis-splicing. More recently, a putative branch point nucleotide and a debranching activity have been purified from trypanosomes (Sutton and Boothroyd, 1988b). These data support the trans-splicing model for the mechanism of mini-exon addition to mRNA. However, since these experiments did not determine from what RNA the 5' intran was released, it is formally possible that another RNA, not a message precursor, cleaved medRNA, releasing a free mini-exon to prime transcription or trans-splice to a message precursor. The relative order of polyadenylation and mini-exon joining is not inherent in the model, but is consistent with experimental finding that the debranchable intran was in poly(A)⁺ RNA (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987). However, the fact that the branch point nucleotide was isolated from poly(A)⁻ RNA (Sutton and Boothroyd, 1988b) suggests that there is no simple kinetic relationship between mini-exon joining and polyadenylation.

CAPS

All studied eukaryotic mRNAs, except poliovirus (Hewlett et al, 1976; Nomoto et al, 1976), are "capped" in a co-transcriptional reaction, which results in a 5'-5' triphosphate linkage of an unencoded m⁷G to the first encoded nucleotide (Shatkin, 1976; Banerjee, 1980). A diagram of a typical cap structure and the specificities of enzymes used in this work is presented in Figure 7. The first few encoded nucleotides can have zero (cap 0), one (cap 1) or two (cap 2) 2'-O-methylations. The term "cap structure" will be used in this thesis to refer to the entire modified structure. Capping occurs early during transcription (Salditt-Georgieff et al, 1980). mRNA caps containing more than two 2' OH methylations have not been reported previously. Most of the small nuclear RNAs have m^{2,2,7}G caps (Reddy et al, 1972; Reddy et al, 1974; Steitz, 1986).

No single function can be exclusively assigned to caps. They appear to provide enhancement of RNA function in a number of metabolic steps. For instance, caps may have a role in overcoming premature termination of transcription (Jove and Manley, 1984). Caps clearly provide stability to mRNA, probably by protection against 5' exonucleolytic degradation (Furuichi et al, 1977; Shimotohno et al, 1977; Green et al, 1983). Independent of the effect on stability, *in vitro* splicing occurs more efficiently with capped than with uncapped pre-mRNA substrates (Krainer et al, 1984; Konarska et al, 1984; Ederly and Sonenberg, 1985). Although this effect is greater in whole cell extracts than in nuclear extracts (Krainer et al, 1984; Konarska et al, 1984), it is specific for cap structures and may be due to a requirement for a cap during splicing complex formation (Ederly and Sonenberg, 1985; Patzelt et al, 1987) or recognition of the proximal intron (Ohno et al, 1987). 3' processing and polyadenylation of an adenoviral mRNA *in vitro* was reported to require a cap structure in one system (Hart et al, 1985); in another system, the cap became associated with snRNPs known to be necessary for histone mRNA 3' processing (Mowry and Steitz, 1987). A cap structure enhanced *in vitro* priming of influenza virus transcription about 50-fold relative to an oligonucleotide primer (Bouloy et al, 1979). A cap 1 structure was about 10-fold more efficient than a cap 0 structure (Bouloy et al, 1980). An inhibitor of RNA methylation, Neplanocin A, was shown to inhibit influenza virus transcription (Ransohoff et al, 1987), suggesting that the 2'-O-methylation was necessary for priming.

A cap-dependent, ~4- to 12-fold enhancement of translation is observed in several cell-free or micro-injection systems (Muthukrishnan et al, 1975; Shatkin, 1976; Shatkin, 1985). 7-Methylation of the guanosine residue is necessary for this effect (Furuichi et al, 1977), but the 2'-O-methylations do not contribute (Hickey et al, 1977). The binding of the cap to a eukaryotic

initiation factor, the Cap Binding Protein, may be necessary for an unwinding of mRNA 5' secondary structure in the early steps of translation (Sonenberg et al,1978; Shatkin,1985).

A knowledge of the structure of the 5' end group of an mRNA is important for several reasons. It is necessary to precisely determine the initiating nucleotide. In addition, in order to specifically label the 5' end of an RNA, a knowledge of its structure is required. Indirect experiments in which medRNA and trypanosome mRNA were chemically decapped and recapped suggested that they contained a cap structure (Laird et al,1985; Lenardo et al,1985a). However, the type of cap and the relationship of the two caps was not determined.

OBJECTIVES OF THIS THESIS

A direct analysis of medRNA and other RNAs in this unusual biosynthetic pathway was sought for a number of reasons. The previous analyses of medRNA and trypanosome mRNA used indirect methods such as primer extension. Methods to directly analyze phosphate-labeled RNA have been historically used to analyze putative precursors and products in RNA processing reactions (Curtis et al, 1978; Lai et al,1982). It was of particular interest to determine the cap structure for a number of reasons. One interpretation of the trans-splicing pathway is that the mini-exon is similar to snRNAs from other systems (Steitz,1986), but that it is used stoichiometrically rather than catalytically (Sharp,1987). If true, the medRNA might have a m^{2,2,7}G cap. If medRNA is the precursor of the mini-exon in mRNA, the medRNA cap should be identical to, or a precursor of, the mRNA cap. If every trypanosome mRNA has the mini-exon, then the mini-exon should be the only cap in mRNA cap.

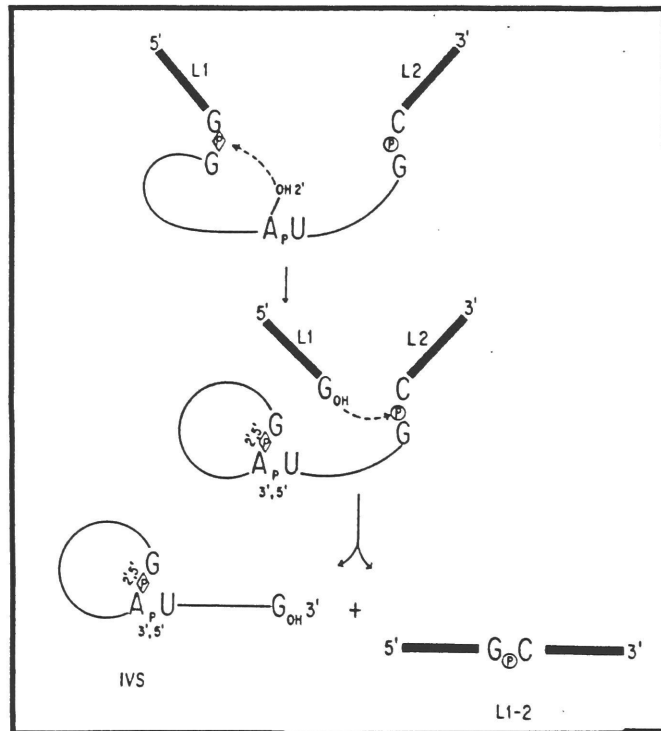
Questions regarding the actual mechanism of joining, its regulation and function, will likely be answered only after an efficient homologous or heterologous *in vitro* mini-exon joining reaction is developed. No *in vivo* assay system, such as a homologous or heterologous transfection system, is yet available. A detailed knowledge of the structure of the molecules in this process; that is, precursors, products and putative intermediates, may be required before such assay systems can be established. For example, an *in vitro* joining reaction may require a coupled transcription and processing reaction, or it may require an authentic trypanosome cap structure.

The work described in this thesis consists of the development of labeling, purification and direct analysis procedures for RNAs involved in this novel metabolic pathway; a fine structural analysis of medRNA; the detection, purification and structural analysis of two small RNAs, which may be

intermediates in this pathway; and the characterization of a novel highly methylated cap structure from medRNA and trypanosome mRNA (Freistadt et al, 1987; Freistadt et al, 1988).

Figure 1
Mechanism of mRNA Splicing

A schematic diagram of the mechanism of eukaryotic mRNA splicing (Sharp, 1985) is shown here. The thick lines, designated "L1" and "L2", represent exons, while the thin line ("IVS") represents an intron.



from Sharp, P.A. (1985) Cell 42, 398

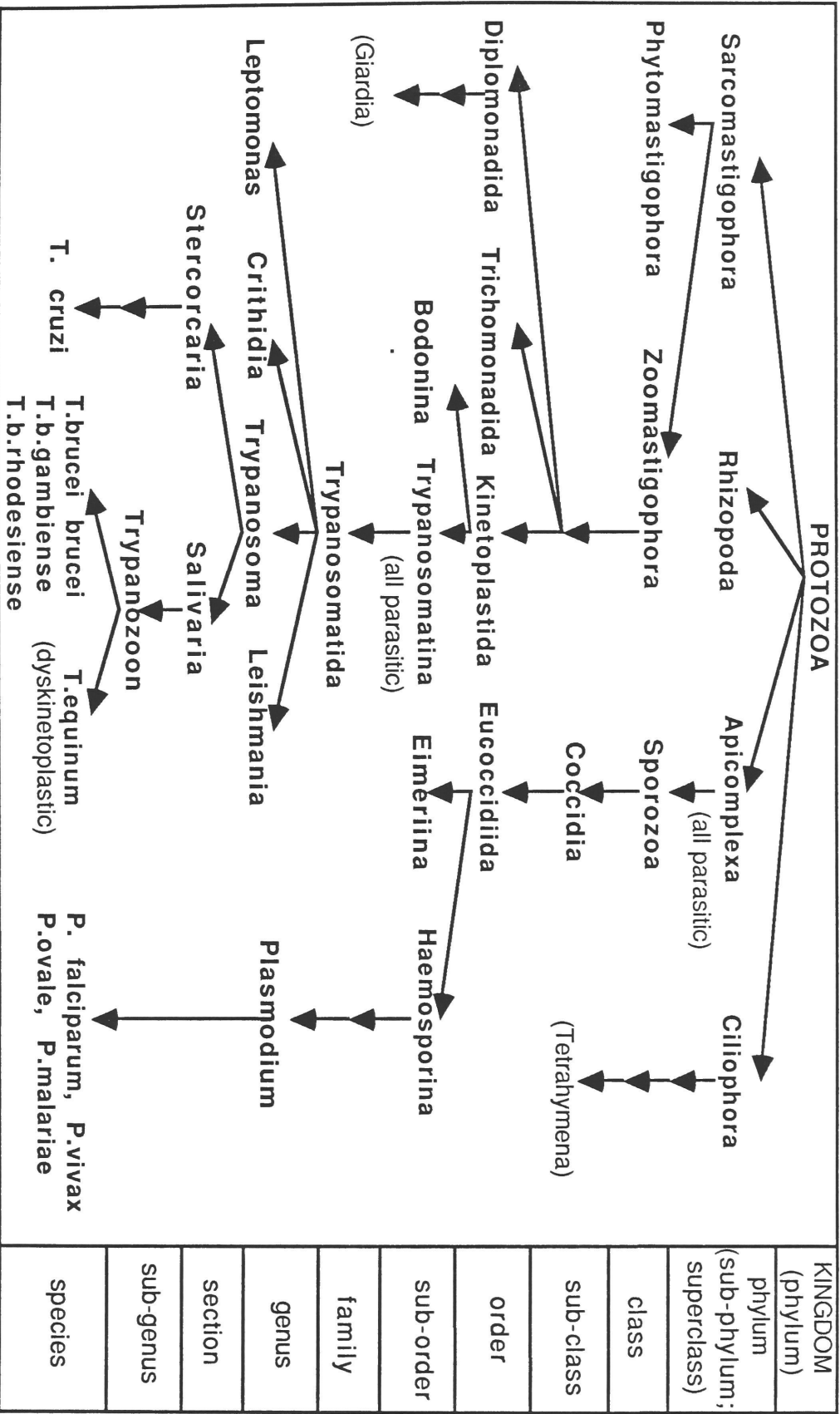
Figure 2
Cases of Discontinuous RNA synthesis

The examples listed here are classified according to putative mechanisms. See text for specific references. Whether each gene segment encodes protein is indicated.

<u>Discontinuous RNA synthesis</u>		
	<u>5' segment</u> Coding?	<u>3' segment</u> Coding?
<u>Priming</u> Influenza Virus Bunyavirus Coronavirus	-	+
	-	+
	-	+
<u>Trans-splicing</u> Kinetoplastida Nematodes C. elegans; B. malayi Chloroplasts Rib. prot. S12; Tobacco; et al <u>In Vitro only</u> Hela Cell extracts Globin, Adeno Yeast Mitochondria Oxi3 (Group II; Autocatalytic) Tetrahymena rRNA (Group I; Autocatalytic)	-	+
	-	+
	+	+
	+	+
	+	+
	rRNA	rRNA
<u>?</u> Vaccinia Virus	?	+

Figure 3
Phylogenetic Classification of the Order, Kinetoplastida

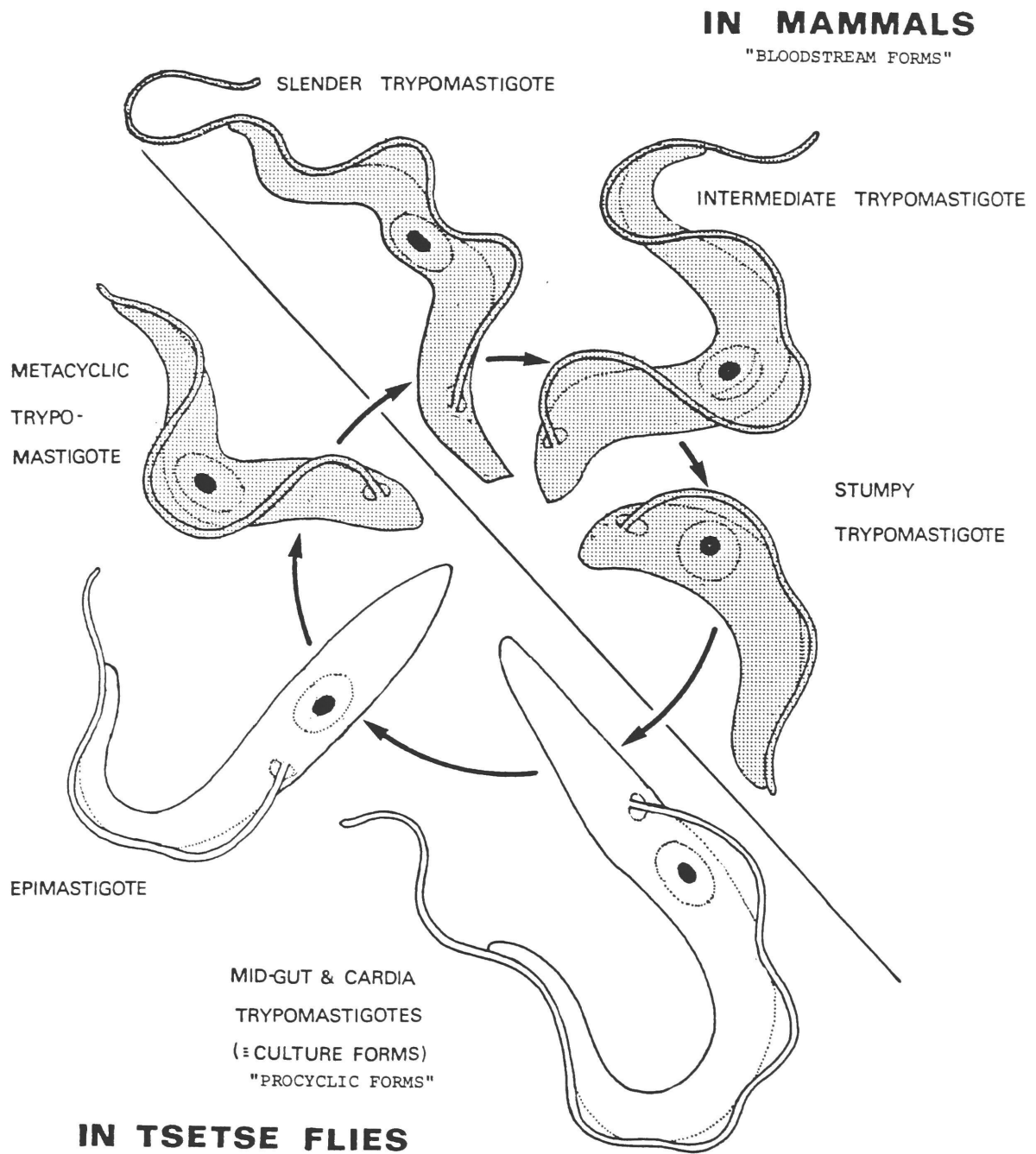
The phylogenetic relationship of trypanosomes to selected other protozoa is shown. Some major human pathogens are included. *T. congolense* and *T. vivax* are classified in the Duttonella and Nannomonas, respectively, sub-genera (not shown in this diagram) of the salivarian trypanosomes. This diagram is adapted from Schmidt and Roberts (1985).



KINGDOM (phylum)
phylum (sub-phylum; superclass)
class
sub-class
order
sub-order
family
genus
section
sub-genus
species

Figure 4
Life Cycle of the African Trypanosome

The morphological changes of *Trypanosoma brucei* are depicted as it cycles between the mammalian host and the insect vector. The forms that exhibit antigenic variation are shaded. Discontinuous RNA synthesis occurs in all stages.



Vickerman, in *Ecology and Physiology of Parasites*, 1971.

Table I
Comparison of Kinetoplastid Mini-exon Genomic Organization

Aspects of the genomic organization of mini-exon sequences and the putative medRNA-equivalent from the indicated kinetoplastid species are presented. ND stands for no data. Repetitive nucleotides following 3' termini are indicated by n. The species are grouped (A,B, and C) according to the results of the secondary structure studies (see Chapter IV, especially Fig. 17). The phylogenetic classification of the species is presented in Figure 3. The repeats, in most cases, occur in one or two large clusters of tandem arrays. Orphans are seen in *T. brucei*^{4,5}, *T. vivax*^{4,9}, *T. cruzi*^{4,9} and *T. congolense*^{9,13}. Retroposon-like repetitive sequence elements interrupting the mini-exon sequence have been detected in the genomic sequence of *T. gambiense*⁷ and *L. seymori*¹⁰.

¹The length of the genomic repeat unit is given in nucleotides.

²Sequences flanking the actual or putative (determined by homology) 3' terminus (designated ↓) of medRNA.

³Campbell et al,1984a.

⁴DeLange et al,1984a.

⁵Milhausen et al,1984.

⁶Dorfman and Donelson,1984.

⁷Aksoy et al,1988.

⁸Putative length determined by homology.

⁹Nelson et al,1983b.

¹⁰V.Bellofatto, unpublished observations. *L.seymori* has two sets of mini-exon genomic repeats.

¹¹Miller et al,1986.

¹²Muhich et al,1987.

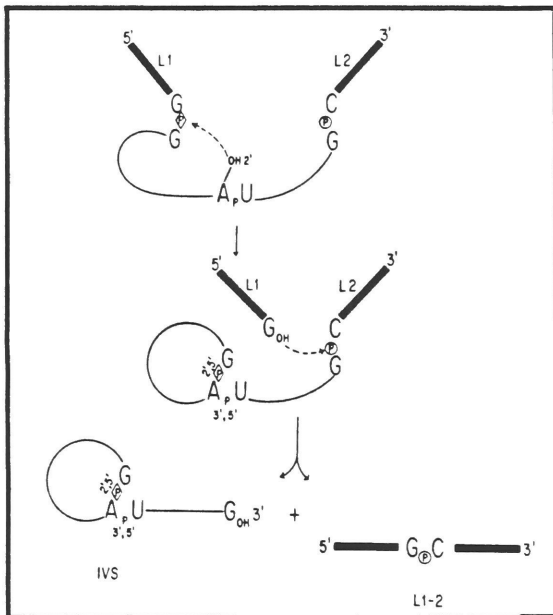
¹³Cook and Donelson,1987.

Table I. Comparison of Mini-exon Genomic Organization					
Species	Antigenic Variation?	Repeat Unit Length ¹	Number of Repeats	Size of medRNA	Termination Sequence ²
A.					
<i>T. brucei</i> ^{3,4,5,6}	+	1361	~200	140	GACCTCC↓ACTCT _n ³ GACCTTT↓ATCTCT _n ⁴
<i>T. gambiense</i> ⁷	+	~1400	ND	140 ⁸	GACCTCC↓ACTCTT _n
<i>T. equinum</i> ⁹	+	~1500	ND	ND	ND
<i>T. cruzi</i> ⁹	-	609	ND	112	GGAC↓CCT _n
<i>Leptomonas seymori</i> ¹⁰	-	757,1090	~425	86	TGGGTA↓GGGCGGAG
<i>L. collosoma</i> ^{5,9}	-	640	ND	92	GACCTTTCGGGC↓T _n
B.					
<i>Leishmania enrietti</i> ¹¹	-	438	~150	~85	AGCGCGG↓AACGGT _n
<i>Crithidia fasciculata</i> ¹²	-	417	250	88	GCGGAAG ↓CTGCT _n
C.					
<i>T. congolense</i> ^{9,13}	+	760	ND	140 ⁸	CCCTC↓ATGGTGC(GT) _n
<i>T. vivax</i> ^{4,9}	+	683	ND	~120 ⁸	ACCAAAAC↓AAAT _n

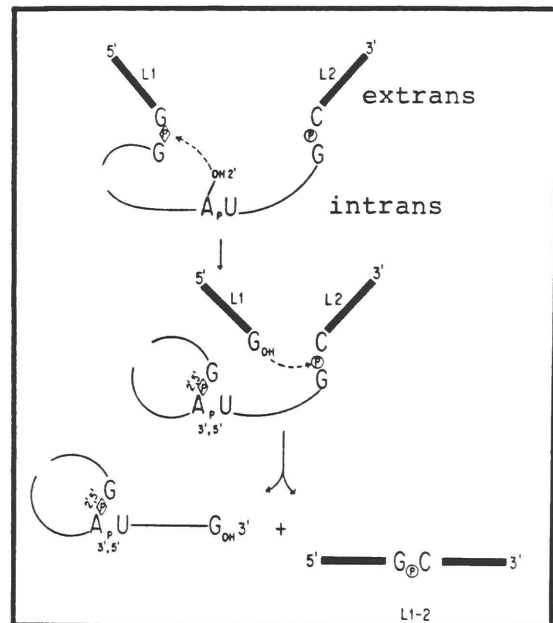
Figure 5
Cis- versus Trans-Splicing

A comparison of intramolecular ("cis-") splicing (see Fig. 1; Sharp, 1985) to the putative trans-splicing mechanism is shown here. If the mechanisms were similar, a Y- shaped structure, similar to a nicked lariat, would be an intermediate. The terms "extran" and "intran" are being proposed in this thesis to refer to the trans-splicing counterparts of exon and intron.

Cis-splicing



Trans-splicing



from Sharp, P.A. (1985) Cell 42, 398

Figure 6
A Trans-splicing Model for Trypanosome mRNA Biosynthesis

This model is based on evidence from several kinds of experiments (see text). DNA is depicted by solid bars, RNA by open or hatched bars. The open bar represents the mini-exon sequence only. The diagram reads downward from the top. medRNA and protein-coding regions are likely transcribed separately. The polycistronic primary transcripts for structural genes must be processed at their 3' (polyadenylation) and at their 5' ends (mini-exon joining). By analogy to cis-splicing, snRNP complexes (open circles) probably regulate and catalyze the intermolecular reaction. The actual joining may take place in two steps: first, branch formation coupled with release of the free mini-exon, and secondly, ligation of the "extrons", as the Y-shaped "intrans" structure is released.

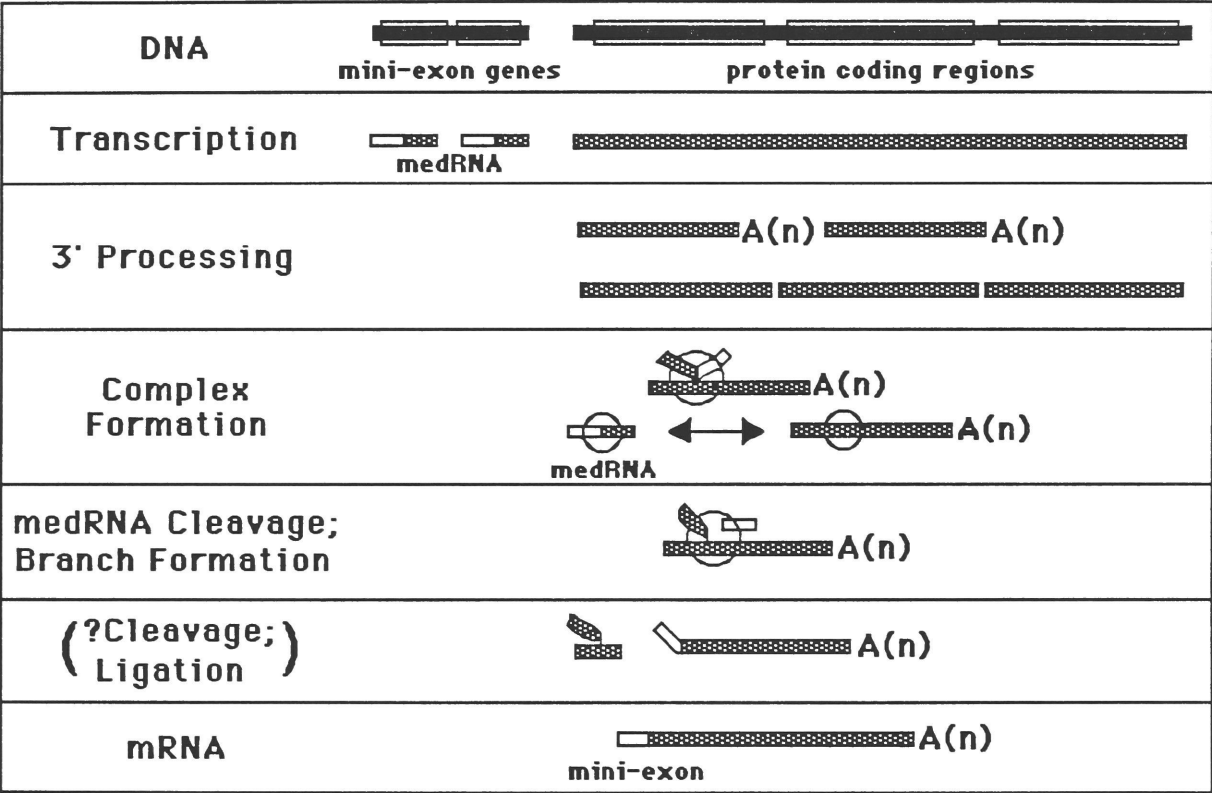
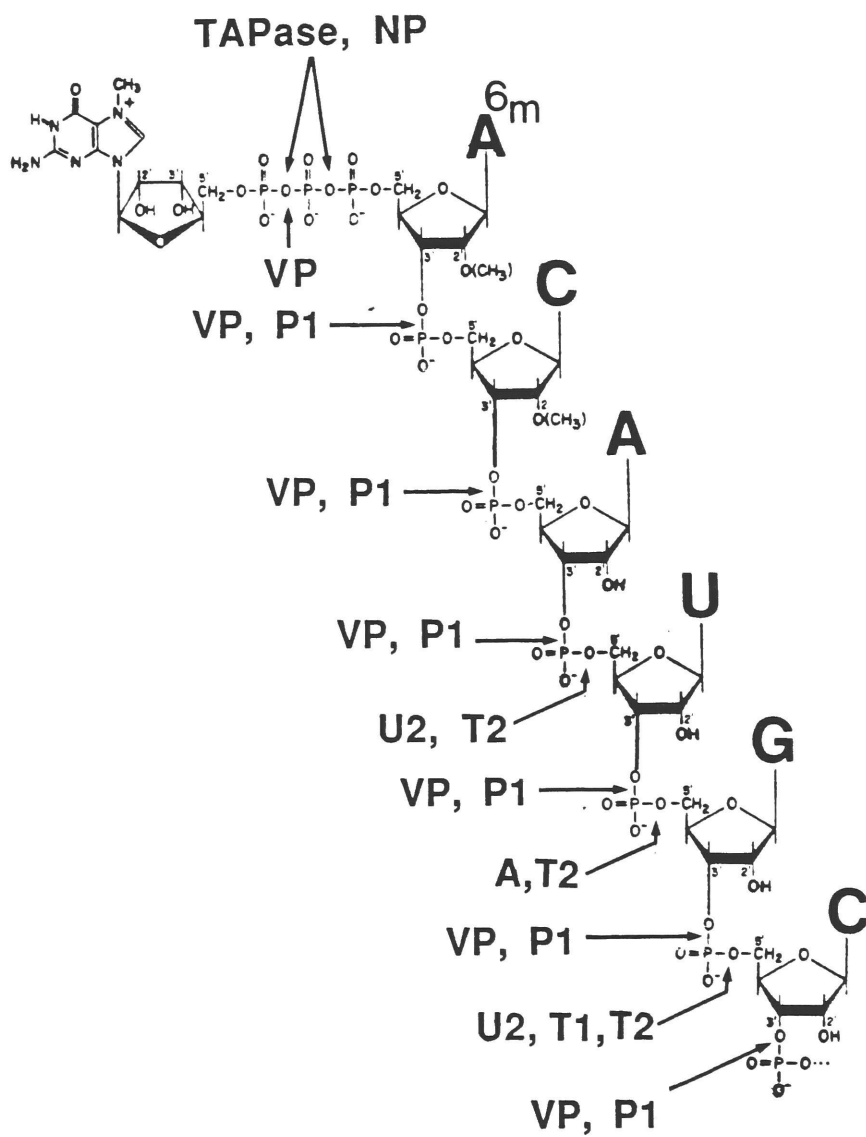


Figure 7
Model of a Typical Cap Structure

A conventional mRNA cap is shown (adapted from Shatkin, 1976). This sequence is hypothetical. The specificities of enzymes used throughout the work are depicted here and described in Materials and Methods. The cyclizing RNases, U2 (specific for purines), T2 (lacks specificity), A (cleaves after pyrimidines) and T1 (cleaves after guanosine residues) yield 3' terminal phosphates. An exonuclease, venom phosphodiesterase (VP), and the endonuclease, P1, can cleave the phosphodiester bonds of modified nucleotides, yielding 5' phosphorylated products. Tobacco acid pyrophosphatase (TAPase), nucleotide pyrophosphatase (NP) as well as VP can cleave pyrophosphate linkages. A 5' terminal adenosine is often methylated at the N⁶ position. Note the positive charge on guanosine caused by the methylation at the 7 position.



CHAPTER II MATERIALS AND METHODS

TRYPANOSOMES

A cloned line of *T. brucei* strain 427, 117B was used. Standard methods were employed for the growth and isolation of bloodstream forms (Cross, 1975). The procyclic forms (procyclics) had been transformed from bloodstream forms *in vitro* (Brun and Schönenberger, 1979) by G. Lamont. The procyclics were maintained in a CO₂ incubator at 27° C in SDM-79 (Brun and Schönenberger, 1981) containing 10% fetal calf serum, 7.5 mg/l heme, 100 U/ml penicillin and 100 µg/ml streptomycin.

LABELING CONDITIONS

Unless otherwise stated, the medium of stationary phase procyclics (at approximately 2×10^7 trypanosomes per ml) was changed to low phosphate (0.2 mM) SDM-79 made with dialyzed serum (Gibco). After 2 days, [³²P]orthophosphate (NEN) was added to 0.5-1 mCi/ml. Nucleic acids were isolated after 20-24 hr additional incubation unless otherwise specified.

RNA ISOLATION

Glassware for RNA purification was siliconized and baked. Plasticware was treated with diethylpyrocarbonate or autoclaved. After labeling, cells were pelleted for 10 minutes at room temperature and then lysed in 0.8 ml (per 20 ml cell culture) of 6 M guanidinium isothiocyanate (Fluka AG), 0.05 M Tris (pH 7.4), 0.01 M EDTA (pH 8.0), 2% SDS, 2% Triton X-100, 2% Sarkosyl and 50 µg/ml *E. coli* tRNA (Sigma). The lysate was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) containing 0.1% 8-hydroxyquinoline and re-extracted until the aqueous phase was clear. After ethanol precipitation, the nucleic acids were loaded onto a CF11 cellulose (Whatman) column (Franklin, 1966) in 0.2 M NaCl, 0.1 M Tris (pH 7.4), 0.2 mM EDTA and 50% ethanol, rinsed 10 times in the loading solution, rinsed with 100% ethanol, eluted in water and lyophilized. Typically, about 3×10^9 cpm of nucleic acids, at about 10^7 cpm per µg, were obtained from about 5×10^8 trypanosomes. Estimates of specific activity were based on the assumption that recovery of RNA from labeled and unlabeled trypanosomes was similar.

HYBRID SELECTION

The plasmid pMX117.13 (constructed by P. Hevezi), a genomic clone of the *T. brucei* mini-exon repeat unit, was used for hybrid selection. It contains two tandem copies of the unit length (1.35 kbp) (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984; Dorfman and Donelson, 1984) Sau3AI fragment inserted into the Bam HI site of pAT153, a derivative of pBR322. Linearized plasmid DNA was boiled in 0.1 N NaOH for 15 min, neutralized and diluted into cold 2 M NaCl. Using a "dot blot" apparatus (Schleicher and Schuell, Minifold), the DNA was bound to nitrocellulose [that had been wetted in water, then in 5 X SSC (SSC is 0.15 M NaCl, 15 mM trisodium citrate)] at not greater than 50 µg/cm² (usually 20 µg/dot). The filters were then rinsed once in 5XSSC, then water, baked for 2 hrs and prehybridized in the hybridization solution: 50% formamide, 0.6 M NaCl, 0.01 M EDTA (pH 8.0), 0.02 M TES (pH 7.4, Sigma), 0.2% SDS, 1 mg/ml poly(A)(Sigma), and 0.2 mg/ml tRNA for 2-16 hrs at 15° C. A 5 to 10 fold molar excess of DNA to expected RNA was used. The RNA was selected for two days at 15° C in a minimal volume (approximately 0.125 ml per cm² of nitrocellulose) of fresh hybridization solution. The filters were rinsed in 50% formamide, 0.6 M NaCl, 0.01 M EDTA (pH 8.0), 0.02 M TES (pH 7.4), 0.1% SDS, 0.1 mg/ml poly (A) and 10 µg/ml tRNA at 15° C. RNA was eluted 3 times for 10 min each at 90° C, in a minimal volume (usually 0.1 ml per DNA dot) of 99% formamide, 0.2% SDS and 50 µg/ml tRNA. The pooled eluates were ethanol-precipitated and chromatographed on CF11 cellulose.

GEL ELECTROPHORESIS AND RECOVERY OF RNA

RNA samples were lyophilized, resuspended in 90% formamide, 1/2 X TBE (1 X TBE is 8.9 mM Tris, 8.9 mM boric acid, 0.25 mM Na₂EDTA, pH 8.3 with HCl), 0.02% xylene cyanol, 0.02% bromphenol blue and electrophoresed in 0.5 or 1.0 mm, 10% polyacrylamide, 7 M urea gels. RNA was extracted from gels by soaking finely chopped, excised gel bands in a minimal volume (usually 0.1 ml) of 0.2 M NaCl, 0.1 M Tris (pH 7.4), 2 mM EDTA (pH 8.0), 0.2% SDS and 50 µg/ml tRNA at 65° C four times for 30 min each, or overnight. The pooled RNA was then chromatographed on CF11 cellulose. Typically, about 3 x 10⁴ cpm of labeled medRNA were obtained from about 5 x 10⁸ trypanosomes.

OLIGO (dT) SELECTION

Oligo(dT) cellulose (Collaborative Research) was washed with 10 volumes of 0.01 M Tris (pH 7.4), 1 mM EDTA and 0.1% SDS. It was then equilibrated with 10 column volumes of 0.4 M NaCl, 0.01

M Tris (pH 7.4), 1 mM EDTA and 0.1% SDS containing 500 µg of yeast tRNA. The RNA was heated to 65° C for 5 min, brought to 0.4 M NaCl and loaded onto the column. The flow-through fraction was re-applied. After 10 rinses with the binding buffer, the poly (A)⁺ RNA was eluted with 37° C wash buffer. The RNA was then chromatographed on CF11 cellulose and lyophilized.

NORTHERN BLOTS

After electrophoresis, the gel was assembled into a "Transblot" apparatus (Biorad) with Gene Screen Plus (NEN) and the RNA was electrophoresed at 5V, 40 mA, overnight at 4° C. The filter was prehybridized in 50% formamide, 1% SDS, 10% dextran sulfate, 1 M NaCl (for 1 hr) and then hybridized in fresh hybridization solution containing nick translated probe for 2 days at room temperature. The filter was rinsed 3 times in 2 X SSC at room temperature.

RNA FINGERPRINTING

RNA samples, with 10 µg of tRNA, were incubated at 37° C in 2 µl of 0.01 M Tris (pH 7.6), 1 mM EDTA, containing 1 mg/ml RNase T1 (Sankyo, Calbiochem) or RNase A (Worthington) for 40 min. Oligonucleotides were fractionated in the first dimension in 10% polyacrylamide gels containing 6M urea and 0.025 M citric acid (pH 3.5) (DeWachter and Fiers, 1971) and then blotted onto DEAE cellulose plates (Brinkmann) and fractionated in the second dimension by ascending homochromatography (Barrell, 1971; Brownlee, 1982; Robertson et al, 1980). Oligonucleotides were eluted from fingerprints in 30% triethylammonium carbonate and analyzed further.

ANALYTICAL DIGESTIONS AND CHROMATOGRAPHY

The following enzymatic digestions were carried out in 3 µl volumes in the presence of 10 µg of tRNA at 37° C for one hr (except RNase U2): RNase T1, 1 mg/ml in 0.01 M Tris (pH 7.6), 1 mM EDTA; RNase A, 1 mg/ml in 0.01 M Tris (pH 7.6), 1mM EDTA; RNase U2 (Sankyo, Calbiochem), 10 U/ml in 0.05 M sodium acetate (pH 4.5), 2mM EDTA, for 2 hrs; RNase T2 (Sankyo, Calbiochem), 2 U/ml in a solution containing 0.05 mg/ml RNase T2, 0.05 mg/ml RNase A, 0.05 mg/ml RNase T1 and 0.05 M ammonium acetate (pH 4.5); nuclease P1 (Calbiochem) (standard conditions), 1 mg/ml in 0.01 M sodium acetate (pH 6.0); calf intestinal alkaline phosphatase (Boehringer Mannheim), in 0-17% glycerol, 0.05 M Tris (pH 8.0), 0.25 M NaCl, and nucleotide pyrophosphatase (Sigma) in 2 mM ATP, 0.02 M Tris (pH 7.5), 1mM MgCl₂ (2 hrs).

Complete venom phosphodiesterase (Calbiochem) digestion (1 µg/sample, no tRNA present) was in 50 mM Tris (pH 9.0), 1 mM KPO₄ and 5 mM MgCl₂ for one hr at 37° C. Partial venom phosphodiesterase digestions were in the same buffer with 10 µg of tRNA per µg of enzyme. Samples at each time-point were brought to 2mM EDTA and boiled for two minutes. Tobacco acid pyrophosphatase (BRL) digestion (10 units/sample) was in 50 mM sodium acetate (pH 6.0) and 10 mM 2-mercaptoethanol for one hr at 37° C. Alkaline treatment was in 10 M NH₄OH for one hr at 37° C.

Digestion products were analyzed by high voltage electrophoresis (HVE) on 3MM or 540 paper (Whatman) or DE81 DEAE paper (Whatman) in pyridine acetate buffer, pH 3.5 (Barrell, 1971), or in 2.5% formic acid, 8.7% acetic acid (pH 1.9). DEAE paper HVE separates on the basis of size and sequence; it does not separate the mononucleotides well and was used primarily to identify oligonucleotides, while Whatman paper HVE (540 and 3MM) separate only on the basis of charge and are used primarily to identify individual mononucleotides. A slight acceleration in mobility due to methylations that do not alter charge was found on DEAE paper HVE, pH 1.9, while almost no change in mobility was found on Whatman 540 or 3MM paper HVE (Cory and Adams, 1975). However, in both systems, methylations tended to cause diffusion of spots. Thin-layer chromatography (TLC) was on 20 x 20 cm² cellulose sheets (Kodak) developed in isobutyric acid (Sigma) and 0.5 M NH₄OH, at a ratio of 10:6 (for one-dimensional TLC or for the first dimension of two-dimensional TLC) and saturated (NH₄)₂SO₄/1 M Sodium Acetate/Isopropanol, 80:16:4 (v/v) for the second dimension (Marcu et al, 1978). Cap markers (m⁷GpppA_{2'},3'-OH, m⁷GpppA_{2'm},3'-OH, and m⁷GpppG_{2'},3'-OH) (Pharmacia) and modified nucleotide markers (Sigma) were dissolved in water and stored at -20° C.

SPECIFICITY AND POLARITY OF NUCLEASE CLEAVAGES

The specificities of RNases that cleave via 2', 3' cyclic intermediates to yield 3' phosphorylated termini are as follows (Barrell, 1971; Uchida and Egami, 1971; see Fig. 7): RNase T1 cleaves after G residues, RNase A cleaves after pyrimidine residues, RNase U2 cleaves after purine residues and RNase T2 cleaves after all conventional nucleotides. RNA products discussed in this thesis have 5' hydroxyl and 3' phosphate termini unless otherwise noted. Nuclease P1 and venom phosphodiesterase lack specificity but yield 5' monophosphate 3' hydroxyl, nucleotides (Fujimoto et al, 1974), abbreviated pN. However, nuclease P1 will also yield the modified 5' end groups from RNA molecules that lack a free 5' phosphate.

CHAPTER III

PURIFICATION AND ANALYSIS OF MEDRNA

The mini-exon donor RNA, medRNA, was originally detected by the indirect analyses of Northern blots, S-1 protection and primer extension (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984). A direct analysis was carried out in order to more fully characterize this abundant RNA which is a precursor in *T. brucei* mRNA biosynthesis. *In vivo* phosphate labeling of RNA was chosen to study medRNA in detail for several reasons. Enrichment for rare or unstable intermediates in the joining reaction may be obtained by phosphate labeling. A complete analysis of medRNA could then be used as a guide to analyze related RNAs in the same pathway. A fingerprint analysis would permit the identification of the medRNA capped oligonucleotide, a determination of its structure and comparison to the mRNA cap structure.

The general strategy for direct analysis was to develop methods for *in vivo* phosphate labeling of *T. brucei* RNA and then purify mini-exon-containing sequences by hybrid selection. Specific selected RNAs were then preparatively electrophoresed. These purified species were next analyzed by classical fingerprint methods.

RESULTS

LABELING

Pilot labeling experiments on bloodstream trypanosomes, freshly obtained from animals, yielded very inefficient labeling of RNA with $^{32}\text{PO}_4$. This was likely caused by the fact that the trypanosomes, after removal from the animal host, were out of their normal environment. Transcription of VSG genes in isolated nuclei is abolished when the trypanosomes are chilled (unpublished observations). Since procyclic forms are easier to grow in culture than bloodstream forms, after these initial results, procyclic forms in culture were studied. Although some stage-specific differences in restriction endonuclease fragmentation of mini-exon DNA have been reported (Parsons et al, 1986), these apparently are not reflected in the RNA metabolism (DeLange et al, 1983, Nelson et al, 1983), and probably involve differential methylation of a few orphan medRNA genes.

Trypanosomes were incubated in low phosphate (0.2 mM) SDM-79 (normal SDM-79 contains ~1mM phosphate) for increasing lengths of time before the labeling period (Fig. 8A). A maximum

incorporation of phosphate into total RNA was obtained with 48 hours of pre-incubation in low phosphate medium. After twenty to twenty-four hours of incubation with phosphate, incorporation into acid precipitable material reached a plateau (Fig. 8B). The highest incorporation of $^{32}\text{PO}_4$ into RNA as a function of trypanosome density was obtained at $1-2 \times 10^7$ trypanosomes/ml (Fig. 8C), the saturation density of this culture system.

HYBRID SELECTION

Linearized, denatured, double-stranded plasmid DNA containing two copies of the mini-exon genomic repeat was immobilized on nitrocellulose and used to hybrid-select labeled RNAs containing mini-exon sequences. Since the mini-exon is very short, it was important to use hybridization conditions of low stringency. Melting temperature (T_m) was calculated from the following formula: T_m (in $^{\circ}\text{C}$) = $81.5 + 16.6(\log M) + 0.41(\%G/C) - 0.72 (\%F) - 500/n$, where M = ionic strength, F = formamide concentration and n = length of hybrid in nucleotides (adapted from Meinkoth and Wahl, 1984). The G/C content of the 35mer is 28.5%, so the temperature of 15 degrees (25 degrees below T_m) was chosen for hybridization. A time-course experiment showed that a minimum of 24 hours of selection was necessary to optimize recovery (Table II). An average of ~2% of the total radioactive RNA was recovered. This amount is a reasonable approximation of trypanosome poly(A)⁺ RNA (1-5% of total [^{32}P]RNA; Table II).

The selected RNAs were eluted from the DNA and analyzed on polyacrylamide/urea gels. Figure 9 shows a comparison of the *in vivo* labeled RNAs to a Northern blot from the same gel. Five of the six small ribosomal RNAs (Cordingley and Turner, 1980; Schnare et al, 1983; Lenardo et al, 1985b; Dorfman et al, 1985; White et al, 1986; Campbell et al, 1987) are indicated by their size, in nucleotides (lane a). The integrity of these small RNAs was used to determine that the RNA was not degraded. The mini-exon-selected material is shown in lane b. A major single component was an RNA of approximately 140 nucleotides which co-migrated with medRNA detected by Northern blot (lane d). This RNA was tentatively identified as medRNA. RNAs trapped near the top of the gel were considered to be mRNA, having been selected by the mini-exon at their 5' ends. Two RNAs the size of two of the small rRNAs were also purified by this procedure; these RNAs were not analyzed further. An RNA slightly smaller than medRNA (medRNA') and several smaller RNAs (see below) were also co-selected. Each of the RNAs was excised from gels and subjected to further analyses. Lane c shows an aliquot of gel-purified medRNA.

FINGERPRINT ANALYSIS OF MEDRNA

A complete fingerprint analysis of medRNA was carried out. Separate aliquots of the purified RNA were digested to completion with RNases of differing specificities. Each digest was fractionated on a two-dimensional chromatography system designed to maximize separation by size and sequence: the resulting pattern is called a "fingerprint". The oligonucleotides were then recovered, redigested with different RNases, analyzed on chromatographic systems and sequenced.

Sequence assignments for each oligonucleotide were based on: (i) position in the fingerprint, (ii) secondary digestion products, (iii) comparison to the DNA sequence of Campbell et al (1984a) and (iv) other indirect data (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984; Dorfman and Donelson, 1984; Laird et al, 1985; Lenardo et al, 1985a). Figure 10 shows RNase T1 (Panel A) and RNase A (Panel C) fingerprints of medRNA. Each oligonucleotide was numbered (Panels B and D), eluted and further digested with RNase A or T1, U2 and T2. Products obtained from this analysis were identified in suitable chromatography systems. The data obtained from this analysis are presented in Table III and compiled in Figure 11. The sequencing strategy often relied on the specificities of the enzymes used and the well-characterized mobilities of oligonucleotides in these chromatographic systems (Barrell, 1971). An example of how sequence data was obtained is the following (Table III): spot 18 likely consists of UUG because: (i) RNase T1-generated spots end in G, (ii) it migrated with the U-rich trinucleotides in the fingerprint, (iii) its base composition is U and G, with more U than G, (iv) RNase A digestion produced U and G and v) it was resistant to RNase U2 digestion. The mobility of spot 18 then provided a marker for one of the RNase U2 products of spot 11, which, considering: (i) its base composition, (ii) its RNase U2 products and (iii) its RNase A products, must be CAUUG. In this work, oligonucleotides of 5 or less could be sequenced without the aid of the DNA sequence. For the larger oligonucleotides, the DNA sequence was necessary to eliminate some possible sequences. However, all the products obtained were consistent with the proposed sequence.

Every oligonucleotide detected could be fit into the sequence of Campbell et al (1984a) and every oligonucleotide predicted by that DNA sequence was detected by this analysis, with the exceptions listed below. The RNA sequence differed from the published sequence of Milhausen et al (1984) at position 71 (their numbering), where the presence of a G would have caused the disappearance of spot 1. This overlapping RNase analysis provided positive identification of medRNA.

The DNA sequence around the predicted 3' end of medRNA is GACCTCCACTTT in Campbell et al (1984a). Spot 9 from the RNase T1 fingerprint of medRNA was determined to be the 3' terminal oligonucleotide. It was less compact than the other oligonucleotides, suggesting a 3' OH. Every oligonucleotide from an RNase T1 fingerprint would be expected to contain one G, except a 3' terminal oligonucleotide. Spot 9, but no other RNase T1-generated oligonucleotide, yielded no G when digested with RNase T2 (Table III and Fig. 12). The base composition indicated C, U and A (Fig. 12, lane d). RNase A treatment of 9 yielded AC, C and U (Table III). The intensity of AC relative to C and U suggested only one copy of AC, limiting the sequence to ACCUCCA_{OH} or something shorter. RNase U2 treatment produced an A and a spot whose sequence was not determined, but, from its mobility on DEAE paper, pH 1.9, had the likely base composition (C₄,U) (data not shown; Barrell,1971). The combined data suggest that the most likely sequence for spot 9 is ACCUCC_{OH}. Because the 3' end is pyrimidine rich, an overlapping analysis using RNase A was not possible.

The oligonucleotides from the mini-exon were identified. These were RNase T1 spots 2, 3, 7 and 10. Both the RNase U2 and RNase A products of spot 2 yielded two Gs, suggesting that it consisted of two co-migrating decanucleotides (Table III and Fig. 13A): they were designated 2a and 2b. The two spots were not separately sequenced but likely correspond to two predicted oligonucleotides of identical base composition, but different sequence, both from the mini-exon: the 3' terminal G of 2b is the 3' terminus of the mini-exon (Fig. 11).

AACG was predicted to be the 5' end of the medRNA by the earlier studies using S-1 mapping and primer extension (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984). A similar 5' end had been found by primer extension of the mini-exon on VSG mRNA (Van der Ploeg, et al, 1982; Boothroyd and Cross, 1982). No oligonucleotide consisting solely of AACG was detected. Spots 3 and D were found to be the 5' terminal oligonucleotides. As will be described below, the 3' end of each of these spots, (AACG for spot 3 and AAC for spot D) was consistent with the previously mapped 5' end but, in addition, each of these oligonucleotides uniquely contained a cap.

The presence of a cap was first suggested by the resistance of the oligonucleotides 3 and D to RNases specific for both purines and pyrimidines. RNase A treatment of spot 3 released only G and yielded a spot with behavior similar to spot D. These and other data indicated that spot 3 differed from spot D only by the presence of a G at its 3' end (Table III, Figs. 13B and C, lanes a and b). An identical pair of slow migrating doublets, as well as other products (Table III) were produced when spots D and 3 were treated with RNase U2 and analyzed on DEAE paper HVE,

pH 1.9 (Figs. 13B and C, lanes c). If medRNA initiated with pppAACG, RNase U2 treatment of the 5' oligonucleotide would have yielded pppAp. The fastest member of the doublet migrated much slower than this tetraphosphate would migrate in this system. The oligonucleotides near the origin, because they were obtained by RNase U2 treatment of an RNase A-generated oligonucleotide, likely contain the modified bonds typical of a cap structure.

The sequence at the 3' end of the capped oligonucleotides was obtained as follows (Table III). The release of G by RNase A from spot 3 indicated that the 3' sequence of spot 3 was pyrimidine-G. RNase U2 released C and AC from spot D and CG and ACG from spot 3. RNase T2 released A and C from spot D and A, C and G from spot 3, in addition to the cap portion. These combined data indicated that the 3' end of spot 3 was AACG, consistent with the predicted 5' end of medRNA.

The release of ACG and AC by RNase U2 treatment of spots 3 and D, respectively, in addition to CG and C (Table III), was unexpected. ACG and AC were likely to be incomplete digestion products. Since the bond between A and C was susceptible to cleavage by RNase T2 (Table III), the existence of a base modification other than 2' OH methylation was suggested. Methylation at the 6 position of adenosine is likely to be the cause of the inhibition (Uchida and Egami, 1971). This adenosine is in a consensus sequence (G/A A(m⁶)C) for N⁶-methylation of adenosine (Padgett et al, 1986; Darnell et al, 1986).

RNase T2, because it lacks nucleotide specificity, is used to identify 5' cap structures with a minimum of unmodified nucleotides (Shatkin, 1976). As can be seen in Figure 7, the first unmodified residue will be the one recognized by RNase T2. Treatment of medRNA oligonucleotides 3 and D with RNase T2 generated, in each case, a pair of resistant oligonucleotides with mobility identical to those generated by RNase U2 (Figs. 13B and C). Since RNase U2 is specific for adenosines in RNase T1-generated oligonucleotides (Fig. 7), the co-migration suggested that the 3' end of the cap was an A.

If the previously mapped 5' end was correct, these results (the release of A, C and G by RNase T2 and of ACG and CG by RNase U2 from spot 3) suggested that medRNA would have a cap 0 structure or a mixture of cap 0 and cap 1. The presence of the doublet was consistent with this hypothesis. However, subsequent experiments showed that neither alternative was correct.

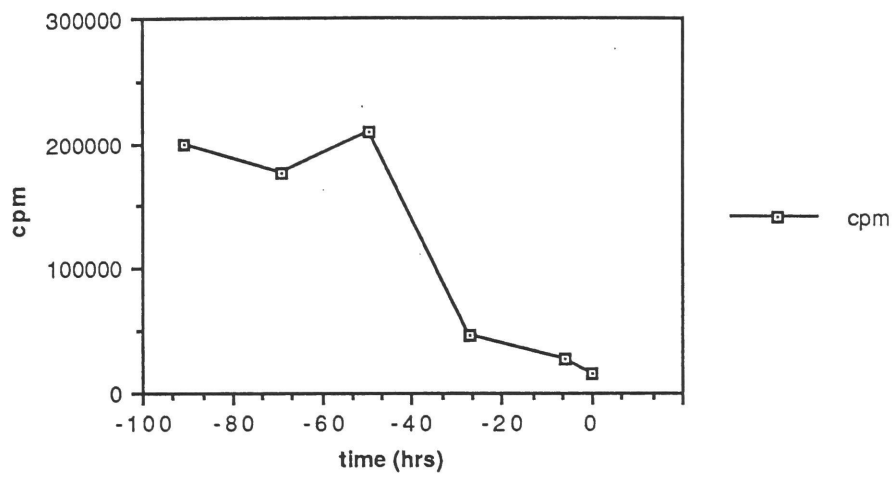
SUMMARY

These initial experiments yielded conditions to $^{32}\text{PO}_4$ label, isolate and analyze trypanosome RNA. A relatively long pre-incubation in low PO_4 medium was necessary to optimally label RNA. This may be due to the reported large PO_4 pools in trypanosomes (Laird et al, 1987). A selection procedure of very low stringency was developed to purify RNAs involved in discontinuous mRNA synthesis. medRNA, the precursor of the mini-exon in mRNA, was studied in detail. Using standard RNA fingerprinting methods, this *in vivo* labeled RNA was sequenced. The 3' end of medRNA was delineated. A nucleotide modification, likely to be N^6 -methylation, was detected. The capped oligonucleotide was identified.

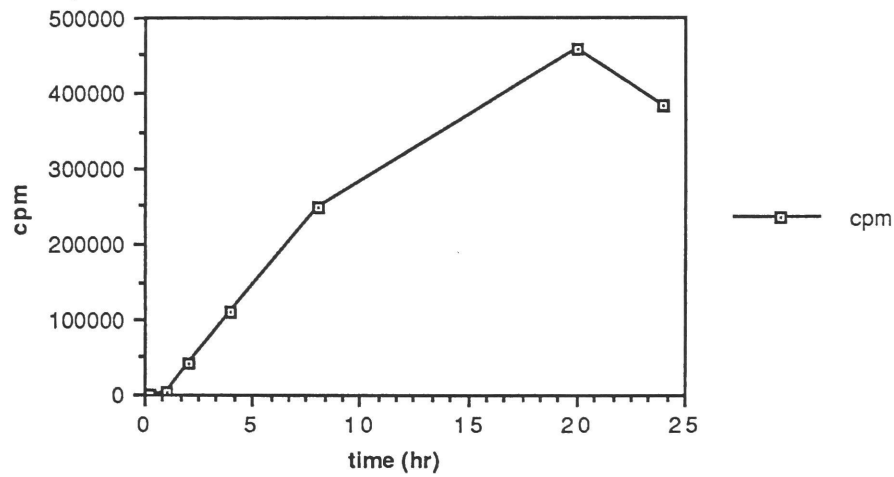
Figure 8
Parameters of Phosphate Labeling of Trypanosome RNA

Panel A: Logarithmically growing procyclic forms were washed and pre-incubated in low phosphate medium for varying amounts of time before beginning the labeling at time 0. The time (in hours) of pre-incubation is indicated by the negative numbers. After labeling for 2 hours, incorporation into TCA-insoluble material was determined. Panel B: RNA was isolated after the indicated labeling periods. Panel C: Incorporation of $^{32}\text{PO}_4$ into RNA was determined as a function of trypanosome concentration. Logarithmically growing cells were diluted to achieve the half the indicated concentration, incubated overnight and then labeled for 2 hours with $^{32}\text{PO}_4$. RNA was then isolated and quantitated. Cpm, per 0.1 ml of trypanosomes, is shown as a function of trypanosome concentration. "Tryps" stands for trypanosomes.

A. Effect of Pre-incubation



B. Time-course of Labeling



C. Effect of Cell Concentration

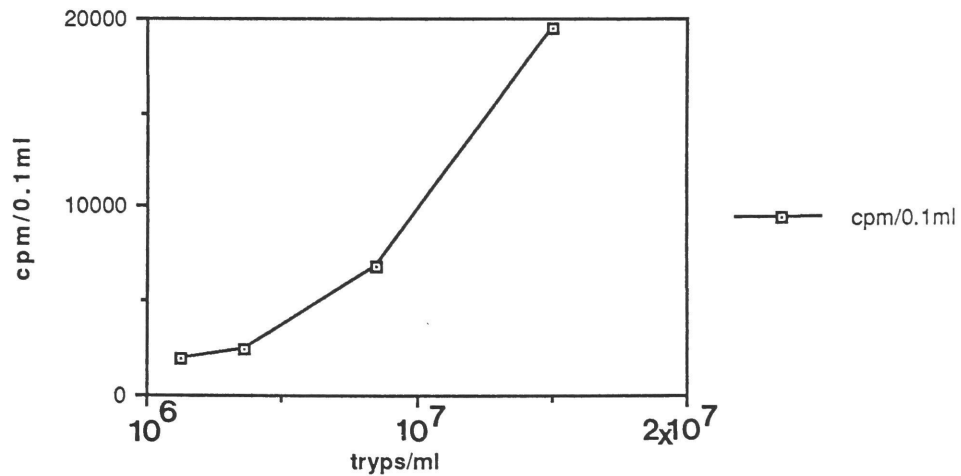


Table II
Recovery of RNA by Different Selection Procedures

5' selected (Part A) and 3' selected (Part B) RNA was quantitated by Cerenkov counting after the indicated procedures.

¹RNA from bloodstream trypanosomes.

²RNA from procyclic trypanosomes.

Table II. Recovery of RNA by Different Selection Procedures

A. Mini-exon Selection

<u>Time (hrs)</u>	<u>Time-course Experiment</u> ¹	<u>Recovery (% of total)</u>
12		4.2
24		6.1
48		6.6

Other experiments (% Recovery)

2.9¹
6.3¹
0.12¹

4.75²
0.15²
0.25²

B.Oligo d(T) Selection (% of total)

1.25²
6.3²
7.6²

Figure 9
Polyacrylamide Gel Analysis of *In Vivo* Labeled RNA

In vivo labeled trypanosome RNA was compared to unlabeled RNA on the same 10% polyacrylamide, 7 M urea gel. Total labeled RNA is shown in lane a. Five of the six small ribosomal RNAs are indicated by their size in nucleotides (Cordingley and Turner, 1980; Schnare et al,1983; Dorfman et al,1985; Lenardo et al,1985b; White et al,1986; Campbell et al,1987). Lane b shows an aliquot of mini-exon selected RNA. An aliquot of gel-purified medRNA, which was used in subsequent analyses, is shown in lane c. The major selected component, medRNA, as well as medRNA' are indicated. A Northern blot probed with the mini-exon genomic sequences is depicted in lane d. mRNA does not appear in the Northern blot because the transfer was optimized for small RNA.

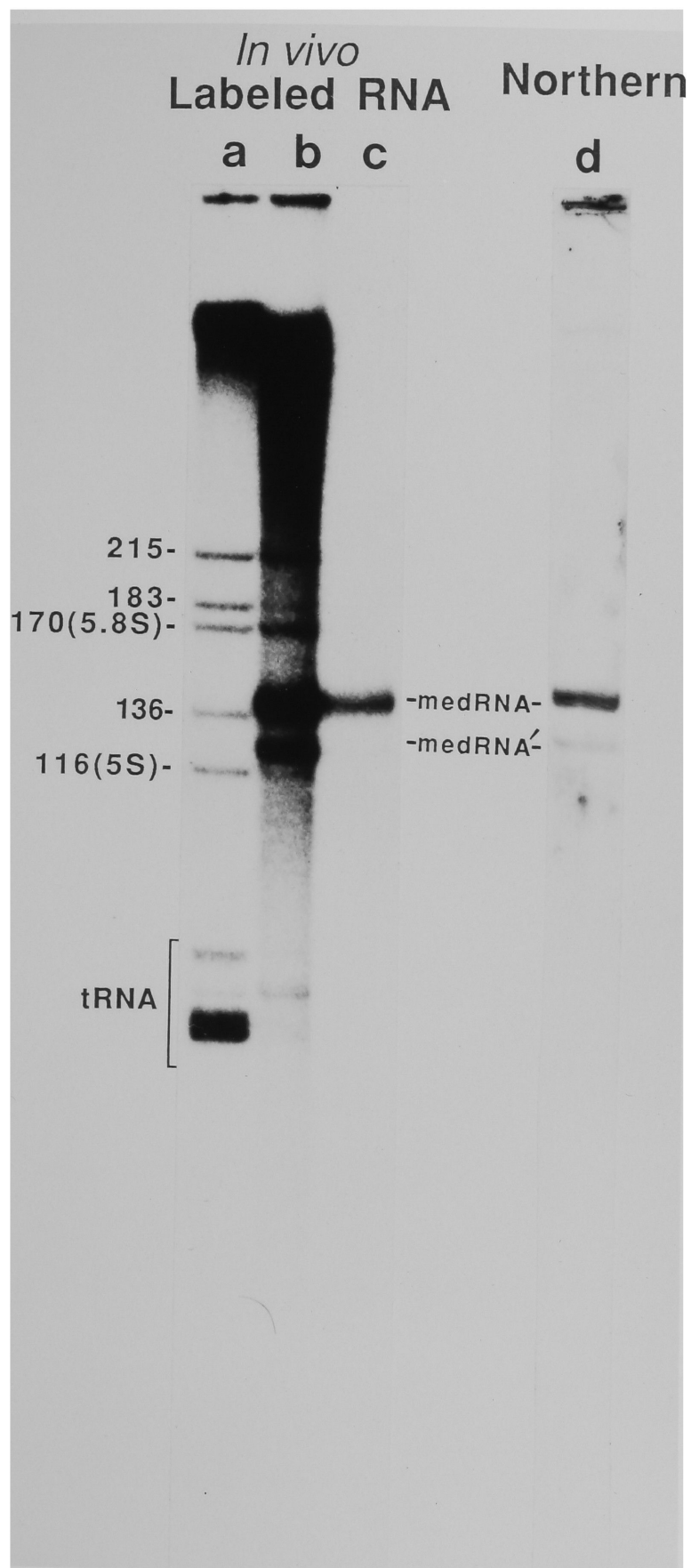


Figure 10
Fingerprint Analysis of medRNA

RNase T1 (panel A) and RNase A (panel C) fingerprints of gel-purified medRNA are shown. The origin of the fingerprints is at the lower right. Each oligonucleotide was given a designation, eluted and analyzed further. Schematic diagrams of the numbered oligonucleotides, from the RNase T1 and A fingerprints are shown in panels B and D, respectively.

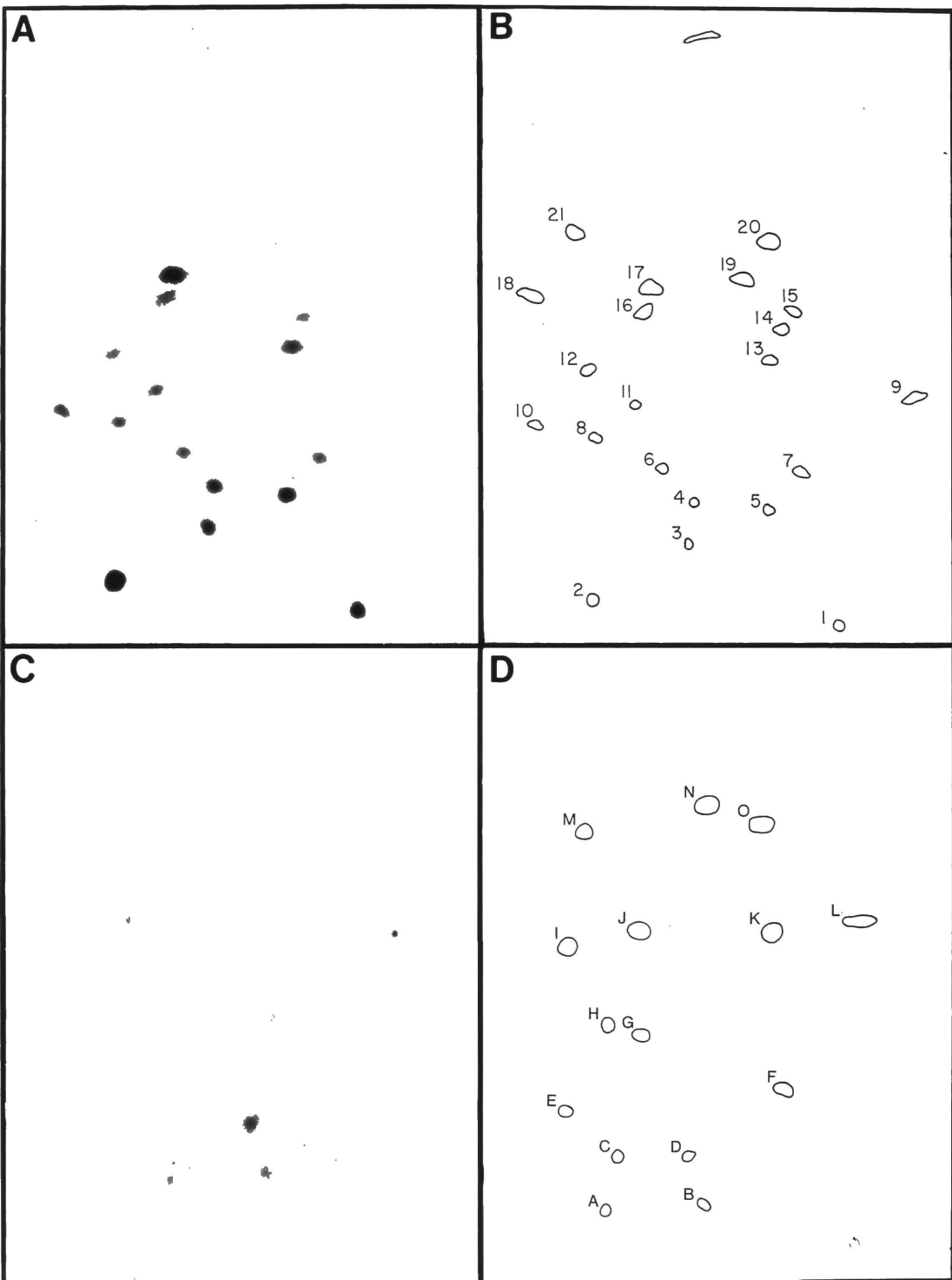


Table III
MedRNA Analytical Digestion Products

Oligonucleotide (spot) designations refer to the fingerprints in Figure 10. Digestion conditions and specificities are given in Materials and Methods. Products were analyzed on Whatman 3MM or 540 paper HVE, pH 3.5 (RNase T2), DEAE paper HVE, pH 3.5 (RNase T1 and A), or DEAE paper, HVE pH 1.9 (RNase U2). The cap-containing oligonucleotides from the RNase T2, T1 and A digests were also analyzed on DEAE paper HVE, pH 1.9. Products are listed in order of mobility, beginning at the origin. K refers to products which were subsequently shown to contain a cap. Underlining indicates that the species were obtained in relatively high yield; parenthesis, low yield; a slash, co-migration on the analytical system used; brackets, the likely base composition, but not sequence, of a species; "uncut," no change in mobility after RNase U2 treatment; ND, no data; n, an unknown number. The process by which the proposed sequence was derived is described in the text. Numbers in parentheses indicate the number of times the sequence occurs in medRNA. Spot 2 was found to consist of two co-migrating decanucleotides, 2a and 2b.

¹These products were not positively identified, due to the absence of internal markers. However, the likely sequence, derived from the base composition and published mobilities (Barrell, 1971), is indicated.

²See Figure 13.

³The loss of the 3' phosphate caused a dramatic mobility difference when comparing [C₄U] from different sources.

⁴See Figure 12.

⁵Partial digestion product.

⁶RNase T1 was found to cleave after AA under the conditions used.

⁷The small amount of GAU present in spot J is from an unknown source, possibly a variant form of medRNA.

Table III. MedRNA Analytical Digestion Products

Spot	RNase T2	RNase A	RNase T1	RNase U2	Proposed Sequence
RNase T1 fingerprint					
1	<u>C</u> ,A,G	AAC ¹ ,AC,G,C	-	CG,CCA ¹ ,CA,C	CCAACACACG
2 ²	C,A,G, <u>U</u>	AG,AU,AC,G,C,U	-	UUG,CUA,UUA,UA,G	CUAUUAUUAG(2a) UACUAUAUUG(2b)
3 ²	C,A,K/G	K(origin),G	-	K(origin doublet), ACG,CG	CapAACG
4	C,A,G,U	AU,AC,G,C,U	-	[CU]G,UA,CA	CAUACUG
5	<u>C</u> ,A,G,U	AG, <u>C</u> ,U	-	[C ₄ U], ^{1,3} AG	CUCCCAG
6	C, <u>A</u> ,G,U	AAU,G,C,U	-	UCUG, ¹ AA ¹	AAUCUG
7	C, <u>A</u> ,G ⁴	AAC ¹ ,AG	-	CA,G,A	AACAG
8	C,G, <u>U</u> ⁴	G,C, <u>U</u>	-	uncut	UUCCUG
9	<u>C</u> ,A,U ⁴	AC,C,U	-	[C ₄ U] _{OH} ^{1,3} A	ACCUCC _{OH}
10	C,G, <u>U</u> ⁴	G,C,U	-	uncut	UUUCUG
11	C,A,G, <u>U</u> ⁴	AU,G,C,U	-	CA,UUG	CAUUG
12	A,G, <u>U</u> ⁴	AU,G,U	-	UA,UG	UAUG
13	<u>A</u> ,G	AAG	-	A,G	AAG(2)
14	C,A,G	AG,C	-	CA,G	CAG
15	C, <u>G</u>	G, <u>C</u>	-	CCG	CCG
16	A,G,U	AG,AU,G,U	-	UG,UA,G,A	UAG,AUG
17	C,G,U	G,C,U	-	[UC]G	UCG,CUG(2)
18	G, <u>U</u>	G, <u>U</u>	-	UUG	UUG
19	A,G	AG	-	G,A	AG
20	C,G	G,C	-	CG	CG
21	G,U	G,U	-	UG	UG
RNase A fingerprint					
A	A,G,U	-	AAG,G	G,A	GGAAGGU
B	A,G,C	-	AAG,AG,G,(C)	G,A,C	GAGAAGC
C	A,G,U	-	AAU, <u>G</u>	<u>G</u> ,A,U	GGGAUU
D	C,A,K	-	K(origin)	K(origin doublet), AC,A/C	CapAAC
E	ND	-	G _n ^{1,5} ,G,U	G _n ^{1,5} ,G,U	GGGGU
F	C, <u>A</u> ,G	-	AAC,AG,AA ⁶ ,C ⁶	G,A,C	AGAAC
G	C, <u>G</u>	-	<u>G</u> ,C	(GG), ^{1,5} <u>G</u> ,C	GGGC
H	A,G,U	-	AU,G	(GG), ^{1,5} A,U	GGAU
I	G,U	-	<u>G</u> ,U	(GG), ^{1,5} G,U	GGU(2)
J	A,G,U	-	AG (AU),(G),U	G,A,U	AGU,GAU ⁷
K	C,A,G	-	AG,AC,G,C	G,A,C	AGC,GAC
L	ND	-	ND	ND	AAC
M	G,U	-	G,U	G,U	GU(3)
N	A,U	-	AU	A,U	AU(7)
O	G,C	-	G,C	G,C	GC(6)

Figure 11
Alignment of Mini-exon RNA Sequences

medRNA (top sequence of both tiers) was sequenced by an alignment of the oligonucleotides (Table III) with the published DNA sequence (Campbell et al,1984a). RNase T1 oligonucleotides are represented above the medRNA sequence, RNase A, below. The determination of oligonucleotide sequences and the detection of a modification in spot 3 (likely to be N⁶-methylation) was carried out as described in the text and in Table III. medRNA was used as a guide to analyze the other mini-exon-selected RNAs isolated from gels such as the ones depicted in Figures 9 and 14: medRNA', ~100mer, the 39mer, the 37mer and the mini-exon from mRNA. The fingerprints and sequence analysis of other selected RNAs are presented in Chapters IV and V, but are shown here for comparison to the medRNA sequence. The description of the cap structure is presented in Chapter VI. Only oligonucleotides that were analyzed are presented. The continuous lines representing medRNA' and the ~100mer signify identity to the medRNA oligonucleotides. The dotted lines at their 3' termini show the region in which the molecules could terminate. The uncertainty is due to the redundancy of short oligonucleotides or mononucleotides in the specified areas.

Figure 12
RNase T2 Digests of medRNA

RNase T2 digests of RNase T1 oligonucleotides, spots 12 (lane a), 11 (lane b), 10 (lane c), 9 (lane d), 8 (lane e) and 7 (lane f) were analyzed on Whatman 3MM paper HVE, pH 3.5. The origin (O) is at the bottom of the diagram; XC signifies the position of the xylene cyanol dye marker.

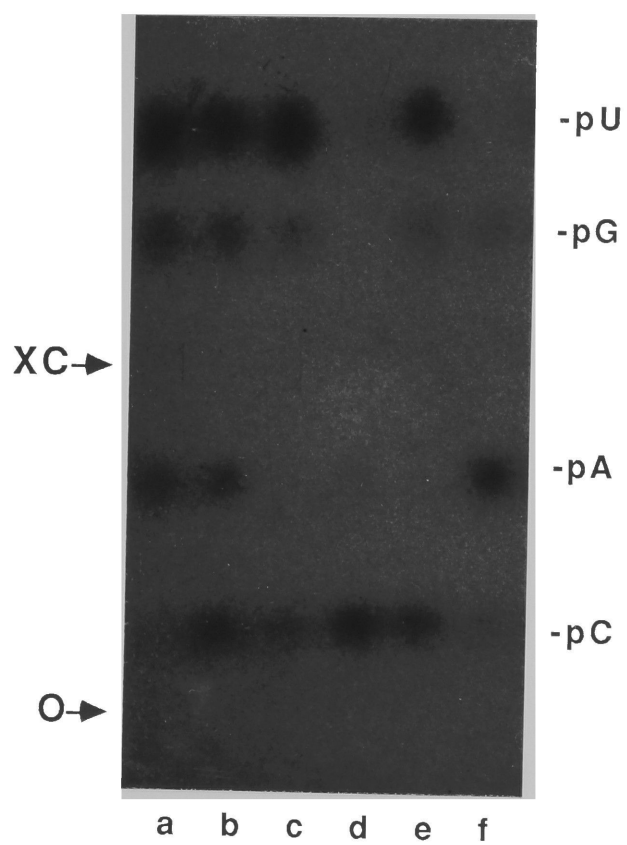
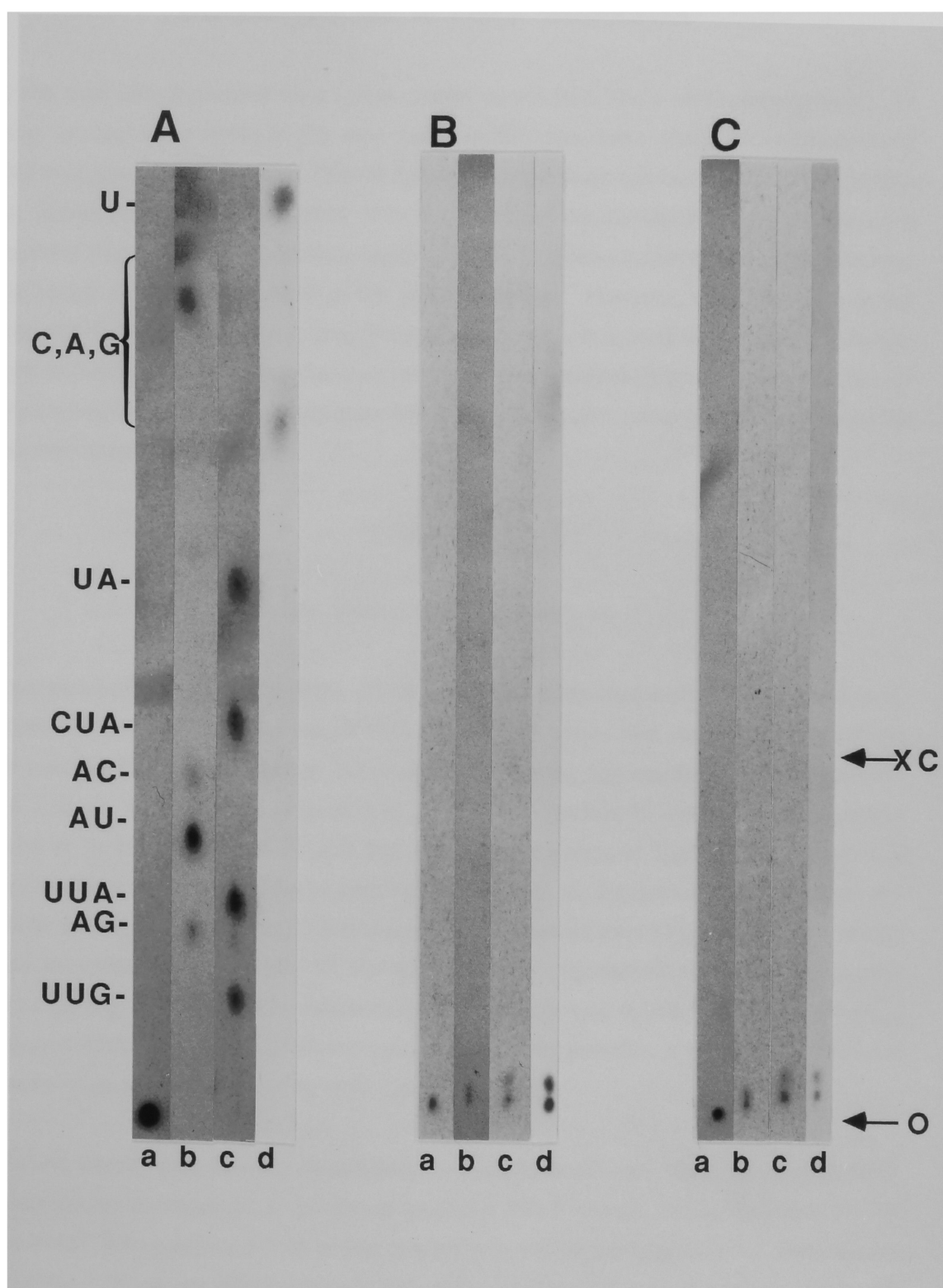


Figure 13
Identification of the Capped Oligonucleotides

Purified medRNA oligonucleotides 2 (panel A), 3 (panel B) and D (panel C) were analyzed after no treatment (Panels A and B, lanes a and panel C, lane b), or after further digestion with RNase T1 (C, lane a), RNase A (A and B, lanes b), RNase U2 (lanes c) or RNase T2 (lanes d) on DEAE paper HVE, pH 3.5 (Panel A, lane b) or pH 1.9 (other lanes). Migration of the markers is indicated. The origin (O) is at the bottom of the diagram; XC signifies the position of the xylene cyanol dye marker.



CHAPTER IV

DETECTION OF A FREE MINI-EXON

Using the medRNA fingerprint as a guide, other co-selected RNAs were fingerprinted. Of particular interest were RNAs in the size range of 35 nucleotides, since these represented potential processing intermediates. Certain splicing intermediates can be detected with difficulty *in vivo* (Zeitlin and Efstradiadis, 1984). The abundance of the mini-exon suggested that such intermediates may be easily detectable in trypanosomes. Characterization of such molecules may provide insight into the mechanism of the joining reaction. However, their existence is not evidence that they are used in the joining process; such proof will depend on the results of *in vitro* reactions or possibly *in vivo* pulse-chase experiments. The experiments presented in this chapter are structural determinations of steady state RNAs detected *in vivo* under conditions in which the RNA was not detectably degraded.

RESULTS

OTHER SMALL SELECTED RNAS

The qualitative pattern of selected RNAs did not vary significantly from experiment to experiment. In addition to the higher molecular weight RNA (mRNA--see below) and medRNA, several RNAs smaller than medRNA were detected. An RNA slightly shorter than medRNA, called medRNA' (Figs. 9 and 14, band 3), was fingerprinted (Fig. 15) and found to consist of the medRNA truncated at its 3' end (Table IV and Fig. 11). The absence of spot 9, the 3' terminal oligonucleotide of medRNA, and the absence of AUG in spot 16, suggested that a minimum of 7 nucleotides were missing from the 3' end of medRNA'. No more than 10 nucleotides are lacking because medRNA' contains spot 17, the next upstream oligonucleotide, at the appropriate intensity (see Fig. 10). Due to the redundancy of the nucleotides in this region of medRNA, a more precise designation of the 3' end of medRNA' was not possible. A similar RNA has been detected by other workers (Laird et al, 1985; Laird et al, 1988).

Debranching experiments (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987) predicted that the debranched, 3' portion of medRNA (the 5' intran) should be present in the selected RNA. When selected RNA of this size class (~100 nucleotides, Fig. 14, band 4) was analyzed (Fig. 15C), it was found to consist only of 5' medRNA sequences, as indicated by the presence of spot 4 but absence of spot 6 (Figs. 15C and 11).

37MER

RNAs in the size class of ~ 35 nucleotides were also selected (Fig. 14). One RNA, designated "band 6", was abundant in the RNA that had been labeled for long periods, such as 20 hours. This RNA was fingerprinted using RNases T1 (Fig.15E) and A (data not shown). The fingerprints of this RNA consisted almost exclusively of mini-exon-specific oligonucleotides (Fig.15E). Each oligonucleotide was eluted and subjected to further analytical digests (Table IV). This analysis was dependent upon the initial studies of medRNA.

Spots 3, 7 and 10 were identical to the comparable medRNA spots (Table IV). RNase T1 spot 2 consisted solely of oligonucleotide 2a. A novel spot, 23, appeared by its mobility on the fingerprint and by its digestion products to be a truncated version of spot 2b. It released the RNase U2 products CUA and UA; and the RNase A products AU and AC, suggesting that it extended minimally to UACUAU. The lack of G from RNase A and UUG from RNase U2 suggested that its maximum extent was UACUAUAUU. The intensity of UA to CUA in the RNase U2 digest suggested that two copies of UA were present; while the intensity of AC to AU in the RNase A digest suggested that only one copy of AU was present, thus limiting the sequence to UACUAUAU_{OH} or something shorter. The mobility on the fingerprint suggested the specified sequence. Although spot 23 migrates to the same position as spot 11, it is clearly a different oligonucleotide.

These analyses of spot 23 indicated that the RNA from which it was produced, the "37mer," consisted of a version of the mini-exon that was truncated by 2 nucleotides. The reason that it was considered to be 37 rather than 33 nucleotides long is presented in Chapter VI.

39MER

In order to study the kinetics of formation of the small RNAs, a time course experiment was performed. Procyclic trypanosomes were grown in low phosphate medium as before and RNA was isolated at different times after adding ³²PO₄. RNA from each time-point was then selected as before (Fig. 16). Two RNAs in the size class of ~35 nucleotides are indicated. The faster-migrating RNA was analyzed in the previous section ("37mer" or band 6 from Fig. 14). The RNA species directly above it was enriched in the RNA labeled for the shorter time. This slightly larger RNA ("39mer") was eluted, fingerprinted and analyzed by standard techniques.

Its fingerprint was similar to that of the 37mer: it contained primarily oligonucleotides from the mini-exon (Fig. 15D). However, a spot that was designated spot 2b' replaced spot 23 (Table IV). It appeared to consist of spot 2b except that it lacked the 3' terminal phosphate, which, in the medRNA fingerprint, is produced by the RNase T1 cleavage. Spot 2a from this fingerprint was sensitive to CIAP, while spot 2b' was not (data not shown). RNase U2 treatment of spot 2a produced CUA, UUA and G, while digestion by RNase U2 of spot 2b' produced CUA, UA and a spot likely to be UUGOH (Table IV). The combined data support the interpretation that the sequence of spot 2b' is UACUAUAUUGOH. These findings indicate that the RNA which migrates slightly slower than the 37mer consists precisely of the 39-nucleotide free mini-exon terminating with a 3' OH.

SECONDARY STRUCTURE MODELS

Kinetoplastid medRNA sequences were analyzed by an RNA-folding computer program (Zuker and Stiegler, 1981). Three types of structures, roughly falling into phylogenetic groups (Fig. 3), were generated (Figs. 17A, B and C). Although the structures differ from one another, some conserved features of medRNA structure are apparent. The mini-exon can form either an internal stem loop structure (Figs. 17B and C), or a larger fold with the intran (Fig. 17A). The mini-exon cleavage point is either at the apex of the stem-loop (Fig. 17A), at the base (Fig. 17B) or at a bulge (Fig. 17C), but never in a base-paired region.

A recent report presenting the sequence of the *T. brucei* U2 RNA speculates that U2 may bind to medRNA via regions of complementarity immediately 5' to the mini-exon cleavage site (Tschudi et al, 1986; Fig. 17G). Yeast U2 RNA has been shown to bind to the pre-mRNA branch site by binding studies and mutational analysis (Parker et al, 1987; Fig. 17F). If trans-splicing is mechanistically similar to cis-splicing (Fig. 5), branch formation would occur on the 3' intran; that is, on the RNA containing the coding sequence, slightly upstream of the joining site. Hence, one would expect U2 to bind the 3' intran, not medRNA (the 5' extran). Thus, the proposal that the *T. brucei* U2 RNA binds medRNA would ostensibly seem unlikely. However, the *T. brucei* U2 RNA differs from the yeast U2 RNA at the putative branch-binding site (Figs. 17F and G). This is not particularly surprising, since no branch site consensus has been detected in trypanosomes. The sequence YYRAY has been proposed, by sequence studies (Borst, 1986), to function analogously. The portion of the *T. brucei* U2 RNA which corresponds to the putative branch-binding site occurs in a large bulge in the model (Tschudi et al, 1986; Fig. 17G). There is some complementarity between YYRAY and this site. In addition, regions of complementarity between

a purine-rich part of the medRNA intran and the pyrimidine-rich tract near the joining site in pre-mRNA, called "med-comp," have been noted (Layden and Eisen, 1988).

I propose that a trimolecular complex, in which each molecule is bound to the other two, is the active complex during branch formation (Fig. 17E). U2 RNA binding to both the branch point and medRNA could bridge the two molecules, bring the 5' intran in close proximity to the 3' intran. Because the complementarity between the U2 RNA and medRNA is weakest at the putative branch-binding site in U2 RNA, that region could be weakly bound; it could have a higher on-off rate than the rest of the hybrid, permitting the substitution of the pre-mRNA branch site for medRNA. If snRNPs are active here, the 39mer would be held in place and, therefore, in close proximity to the 3' extran for the second step, ligation.

Another computer study (Fig. 17D), revealed that the mini-exon (the 39mer) can form a stable stem-loop, but that the two 3' nucleotides do not fall in the loop. This could explain how the two 3' most nucleotides could be easily removed to form the 37mer.

SUMMARY

Fingerprint analyses were performed on several small RNAs purified by the mini-exon selection procedure. medRNA' is a form of medRNA lacking 6-10 nucleotides from its 3' end. An RNA of about 100 nucleotides was also found to be a 3' truncated form of medRNA. Two RNAs consisting of the free mini-exon with a 3'-OH (the 39mer) and a slightly truncated version (the 37mer) were characterized. The analysis of these RNAs was dependent upon the use of the medRNA fingerprint as a guide. The enrichment of the 37mer in the longer labeling times may be due to an increase in intracellular nucleases during the labeling period.

The proposed trans-splicing mechanism (Figs. 5 and 6), if analogous to conventional intramolecular cis-splicing, would yield a 5' extran with a 3' OH. Since the 39mer corresponds to this structure, it is possible that it is an intermediate in the joining reaction. This argument cannot be used to prove a mechanism: the detection of this molecule is simply consistent with the proposed mechanism. Such a molecule is also consistent with the priming model.

Some computer-generated models of the medRNA secondary structure were examined, revealing structural conservation as well as a possible explanation for the presence of the 37mer. A model for a trimolecular trans-splicing complex, possibly involved in branch formation and production of the 39mer, was presented.

Figure 14
Detection of Small Mini-exon Selected RNAs

In vivo labeled RNA (lane a) was hybrid-selected as described. Five of the six small ribosomal RNAs are indicated by their size in nucleotides. Lane b shows an aliquot of the selected material and lane c shows a longer exposure of the same 10% polyacrylamide, 7 M urea gel. RNAs that were fingerprinted are: Area 1 (mRNA, Fig. 18), band 2 (medRNA, Figs. 10 and 15); band 3, medRNA' (Fig. 15), band 4, the ~100mer (Fig. 15), band 5 (not shown) and band 6, the 37mer (Fig. 15).

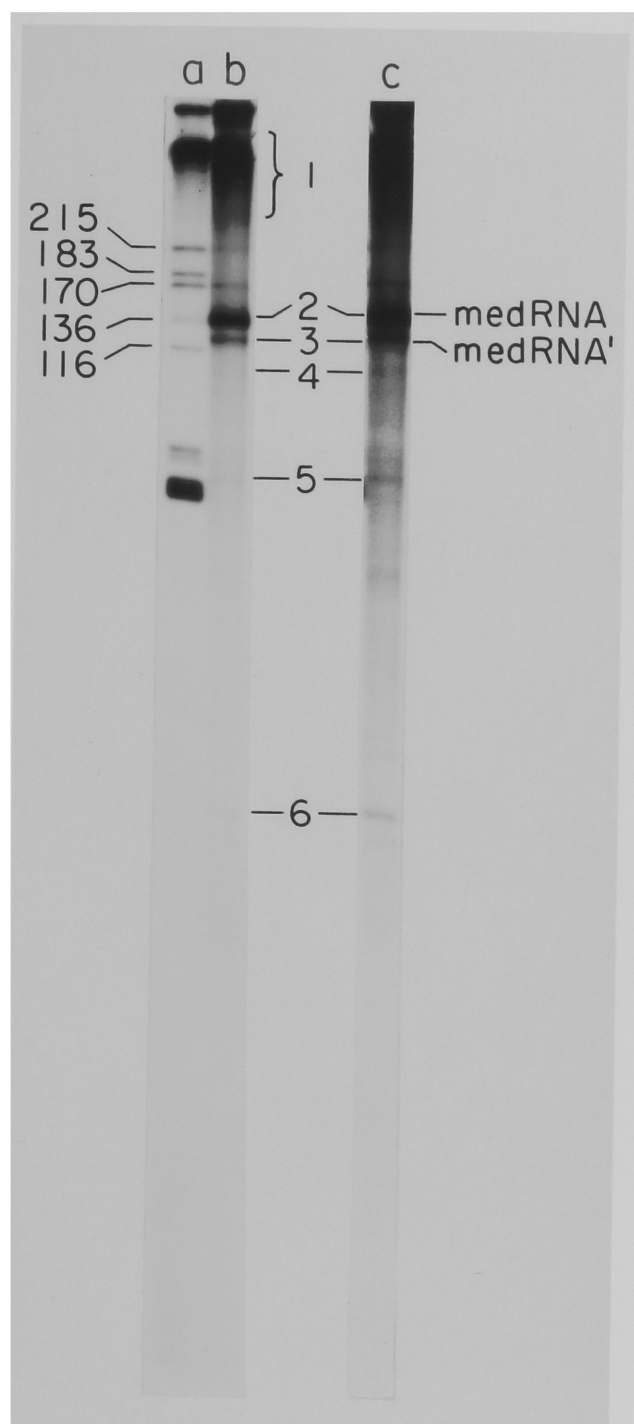


Figure 15
RNase T1 Fingerprints of Small Mini-exon Selected RNAs

MedRNA (Panel A), medRNA' (Panel B), the ~100mer, band 4 from Figure 14 (Panel C), the 39mer from Figure 16 (Panel D) and the 37mer, band 6 from Figure 14 and from Figure 16 (Panel E) were fingerprinted as described. The gel mobility of the RNAs is shown in Figures 9, 14 and 16. An alignment of the oligonucleotides with the linear sequence is presented in Figure 11. The mini-exon specific oligonucleotides (2, 3, 7 and 10) as well as spots 9, 16 and 6 (the 3' terminal oligonucleotides) are indicated. The open arrow in Panel C indicates the expected position of spot 6. Spots 2 in the fingerprints of the 39mer and the 37mer were found to consist only of 2a (Table IV), while the novel spots, 2b' and 23, respectively, replaced spot 2b. The origin of the fingerprints is at the lower right.

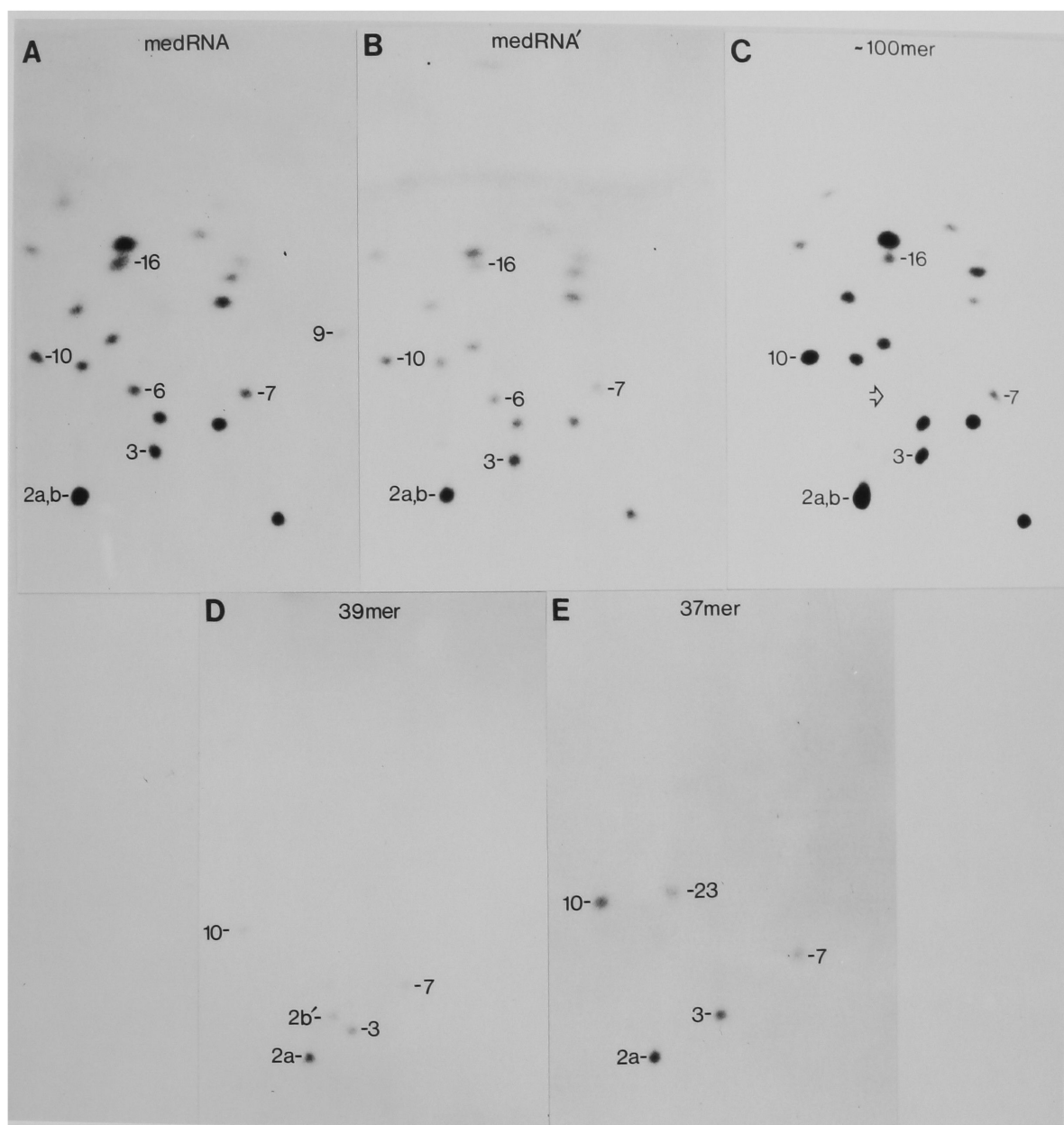


Table IV
Small RNAs: Analytical Digestion Products

Except for the data from the RNase A fingerprint of the 37mer, which is not shown, the oligonucleotide (spot) designations are from the fingerprints in Figure 15. See also legend and footnotes, Table III. Figure 11 shows the alignment of the oligonucleotides with medRNA sequences.

Table IV. Small RNAs: Analytical Digestion Products					
medRNA'					
Spot	RNase T2	RNase A	RNase T1	RNase U2	Proposed Sequence
1	<u>C</u> ,A,G	AAC ¹ ,AC,G,C	-	CG,CCA ¹ ,CA,C	CCAACACACG
2	C,A,G, <u>U</u>	AG,AU,AC,G,C,U	-	UUG,CUA,UUA,UA,G	CUAUUUAUUAG (2a) UACUAUAUUUG(2b)
3	C,A,K	K(origin),G	-	K(origin doublet)	CapAACG
4	C,A,G,U	AU,AC,G,C,U	-	ACG,CG	CAUACUG
5	<u>C</u> ,A,G,U	AG, <u>C</u> ,U	-	[CU]G,UA,CA	CUCCCAG
6	C, <u>A</u> ,G,U	AAU,G,C,U	-	[C ₄ U], ¹ AG	AAUCUG
7	C, <u>A</u> ,G	AAC,AG	-	UCUG, ¹ AA ¹	AACAG
8	C,G, <u>U</u>	G,C, <u>U</u>	-	CA,G,A	UUCCUG
10	C,G, <u>U</u>	G,C,U	-	uncut	UUUCUG
11	C,A,G, <u>U</u>	AU,G,C,U	-	uncut	CAUUG
12	A,G, <u>U</u>	AU,G,U	-	CA,UUG	UAUG
13	<u>A</u> ,G	AAG	-	UA,UG	AAG(2)
14	C,A,G	AG,C	-	A,G	CAG
15	C,G	G, <u>C</u>	-	CA,G	CCG
16	A,G,U	AG,U	-	CCG	UAG
17	C,G,U	G,C,U	-	UA,G	UCG,CUG(2)
18	G, <u>U</u>	G, <u>U</u>	-	[UC]G	UUG
19	A,G	AG	-	UUG	AG
20	C,G	G,C	-	G,A	CG
21	G,U	G,U	-	CG	UG
*37mer					
RNase T1 fingerprint					
2a	C,A,G, <u>U</u>	AG,AU	-	UUA,CUA,G	CUAUUUAUUAG
3	C,A,K/G	K(origin),G	-	K(origin doublet),CapAACG	ACG,CG
7	C,A,G	(AAC),AG	-	CA	AACAG
10	C,U	U	-	uncut	UUUCUG
23	C,A,U	AU,AC,U	-	CUA,UA	UACUAUAU _{OH}
RNase A fingerprint					
D	C,A,K	-	K(origin)	K(origin doublet)	CapAAC
E	C,A,G	-	AAC,AG,AA ⁶ C	G,C,A	AGAAC
J	A,G,U	-	AG,U	A/G,U	AGU
M	G,U	-	G,U	G,U	GU
N	A,U	-	AU	A,U	AU
O	C,G	-	G,C	G,C	GC
39mer					
2a	C,A,G,U	AG,AU	-	CUA,UUA,G	CUAUUUAUUAG
2b'	C,A,U	AU,AC,U	-	CUA,UA,	UACUAUAUUUG _{OH}
				UUG _{OH} ¹	
3	C,A,K/G	K(origin),G	-	K(origin doublet),	CapAACG
7	C,A,G	AAC,AG	-	CA,A,G	AACAG
10	C,G,U	G,C,U	-	uncut	UUUCUG

Figure 16
Time-course of Appearance of Small RNAs

Procyclic trypanosomes were labeled for varying amounts of time with $^{32}\text{PO}_4$. Lane a is unselected RNA from the 24 hour time-point. The migration of tRNA is indicated. Mini-exon sequences were selected from RNA isolated at each time-point, [3 (lane b), 6 (lane c), 11 (lane d) and 24 (lane e) hours] and aliquots were electrophoresed in a 10% polyacrylamide, 7M urea gel. The mobilities of medRNA, the 39mer and the 37mer are indicated. The fingerprints of these RNAs and the analytical digests are presented in Figure 15 and Table IV, respectively.

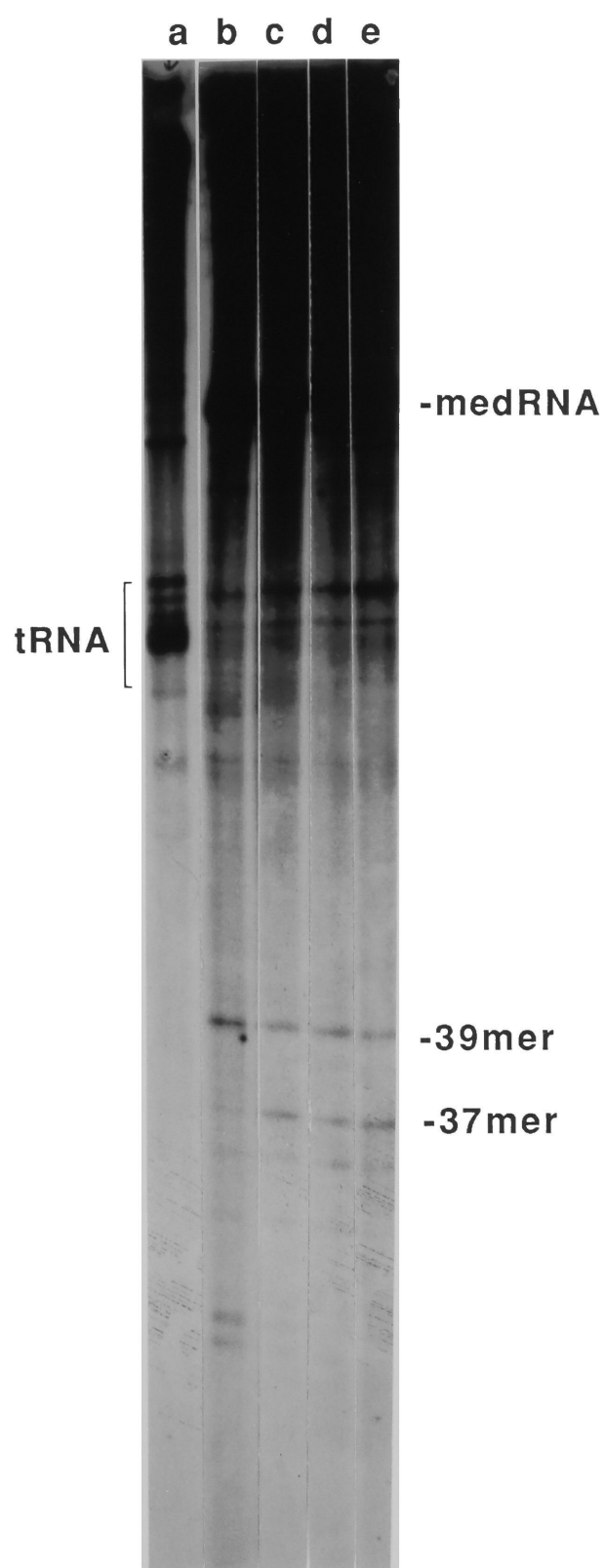
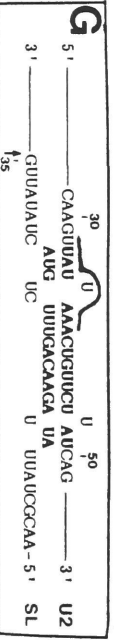
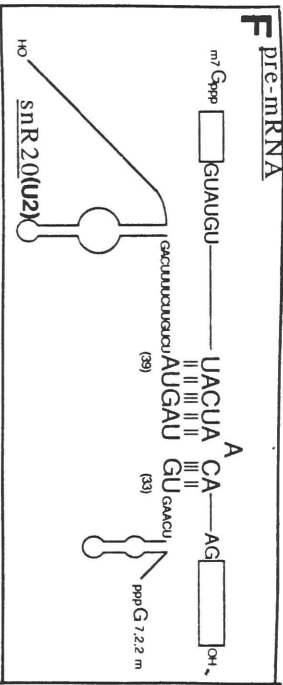
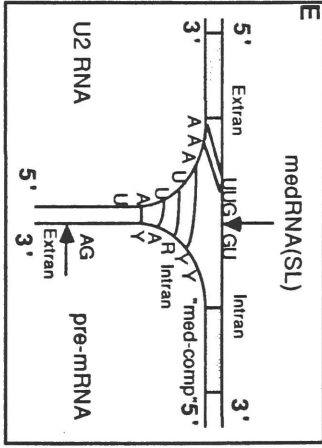
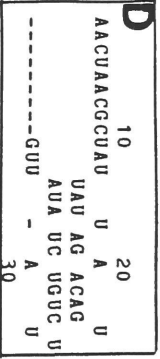
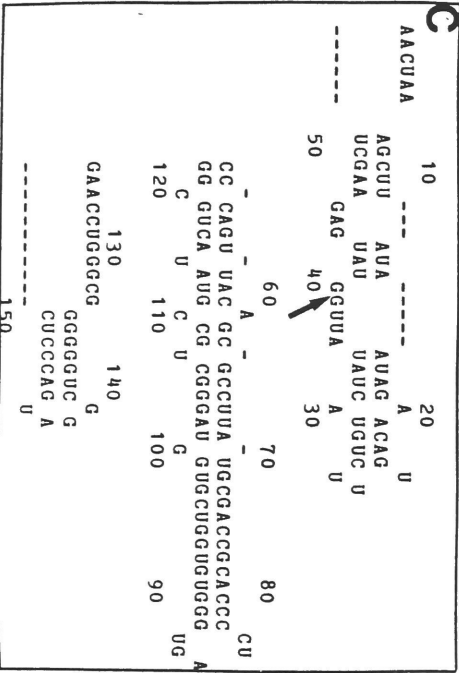
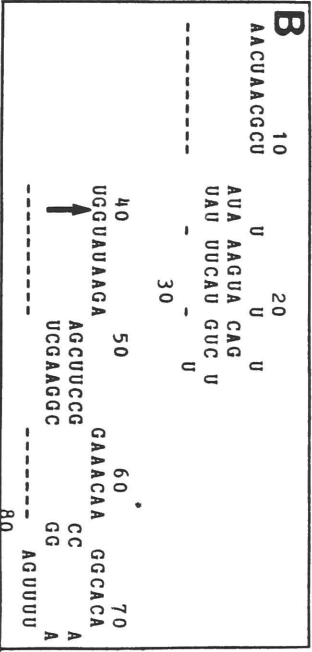
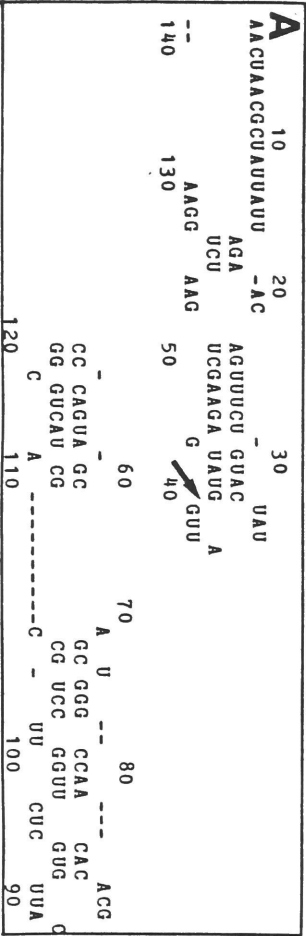


Figure 17
Models for medRNA Cleavage During Trans-splicing

Computer-generated secondary structure models (Zuker and Stiegler, 1981) of medRNA from different Kinetoplastidae are presented in Panels A, B, and C. Panel A is the *T. brucei* (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984; Dorfman and Donelson, 1984) model and is representative of the results obtained from the *T. gambiense* (Aksoy et al, 1988), *T. cruzi* (DeLange et al, 1984a), *L. seymori* (V. Bellofatto, unpublished observations) and *L. collosoma* (Nelson et al, 1983b) medRNAs. The model of *C. fasciculata* (Muhich et al, 1987) medRNA is presented in Panel B; *L. enrietti* (Miller et al, 1986) medRNA yielded a similar structure. Panel C presents that for *T. congolense* (Cook and Donelson, 1987); *T. vivax* (DeLange et al, 1984a) was similar. See Table I and Figure 3 for phylogenetic information about these sequences. Panel D is a similar analysis of the free mini-exon from *T. brucei*, showing that the two 3' nucleotides may be susceptible to nucleases.

Panel E is a proposed model for a trimolecular complex between medRNA (also referred to as "SL" in Panel G), U2 RNA and a hypothetical primary transcript for protein-coding sequences ("pre-mRNA"). Regions of possible base pairing are indicated by lines or hatching. The "med-comp" region is taken from Layden et al (1988). See text for details. By lining up the trypanosome and yeast U2 RNA (Tschudi et al, 1986), the sequence AAAUUAU corresponds to the branch-binding site (the overlined bulge in panel G). The invariant 5' (GU) and 3' (AG) eukaryotic mRNA intron boundary sequences, also found in trypanosome RNA, are indicated. The sequence YYRAY is a proposed consensus branch sequence from Borst (1986). Panel F is a representation of the U2 RNA (snR20) branch-binding site from yeast (Parker et al, 1987). Panel G presents the proposed interaction between U2 RNA and medRNA ("SL") (Tschudi et al, 1986). Note that, in contrast to other systems, the longest region of complementarity between medRNA ("SL") and U2 RNA occurs within the mini-exon, upstream of the cleavage site (nucleotide 35 in Panel G). The region corresponding to the branch binding site in yeast U2 RNA is overlined. The arrows in Panels A-E indicate the point of mini-exon, or pre-mRNA (Panel E), cleavage. See Figure 3 for the phylogenetic grouping of these species.



CHAPTER V

INITIAL ANALYSES OF THE CAP

The complete fingerprint analysis of the medRNA revealed which spot was the capped oligonucleotide. The mRNA cap was studied next for several reasons. First, since there are likely to be more mRNA molecules than medRNA molecules and there are less steps in the purification of mRNA, the cap from poly(A)⁺ RNA was expected to be obtained in higher yield. This would allow a detailed structural characterization. Second, because these RNAs are involved in discontinuous mRNA synthesis, it was of interest to study the mRNA cap. However, since it is not proven that every trypanosome mRNA contains the mini-exon, it was important to isolate the mini-exon specific cap. To do this, the same approach that was used to isolate the medRNA cap, hybrid-selection followed by fingerprinting, was used initially. One would predict that the two caps would be either identical or related to each other. The structure first obtained on the mRNA cap could then be compared with the less abundant medRNA cap.

This chapter details the identification and preliminary characterization of the mRNA cap, the initial analyses with nuclease P1, the comparison to the medRNA cap, and the detection of an artifactual derivatization of the cap.

RESULTS

USE OF RNASE T1 FINGERPRINTS TO FIND THE MRNA CAP

The mini-exon-selected RNA eluted from the high molecular weight region of 10% polyacrylamide gels (such as "area 1" in Fig. 14) was digested with RNase T1 and fingerprinted using standard techniques (Fig. 18A). A complex pattern, consistent with the expected pattern for mRNA (Robertson et al, 1977) was obtained. However, two of the mini-exon-specific oligonucleotides, spots 2 and 3 were easily identifiable. Similar results were obtained from fingerprints of poly(A)⁺ RNA (Fig. 18C). The fact that the mini-exon spots are more readily apparent in the 5' selected RNA (Fig. 18A) than in the 3' selected (Fig. 18C) is probably a reflection of two factors: there is likely to be a population of partially processed poly(A)⁺RNA that does not contain the mini-exon (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987) and second, if there is a slight amount of degradation, the poly(A)⁺ selected material will be depleted of 5' sequences. The absence of spot 1, from the 3' portion of medRNA (Fig. 11), indicates that spots 2 and 3 are from mRNA and not from contaminating medRNA. Spots 7 and 10 were not readily apparent. The mini-

exon spots were not visible in RNase A fingerprints, presumably due to co-migration with other oligonucleotides, so that an "mRNA RNase A spot D" was not isolated directly.

To confirm their identity as mini-exon oligonucleotides, spots 2 and 3 from the mRNA fingerprints were eluted and subjected to further digestions (Table V). Due to the complexity of the fingerprints, these spots contained contaminating oligonucleotides. However, the products obtained from both sources [poly(A)⁺ RNA or mini-exon selected] were quite similar to those obtained from medRNA (Tables III and V). When analyzed on DEAE HVE, pH 1.9, mRNA spot 3 co-migrated with medRNA spot 3 and produced the slow migrating doublet upon digestion by RNase A (Fig. 19). However, it was not possible to analyze the non-cap containing portion of this oligonucleotide, due to the products from the contaminating oligonucleotides (not visible in this exposure).

CYCLIZING RNASES

In order to study the mRNA cap structure in detail, the RNase T1 spot 3 was preparatively digested with RNase A (Fig. 19) and each of the two components purified. This accomplished two goals: the contaminating non-cap-containing oligonucleotides were removed and the two species comprising the cap-containing doublet were separately purified. These two species were called mRNA spots D (the faster component) and D'. When the non-cap-containing portions from D and D' were examined, products analogous to those obtained from medRNA spot D were found (Table V). C and AC were produced from each by RNase U2 and mononucleotides, likely to be A and C, were produced by RNase T2. These data suggested that the mRNA cap also ended with a 3' terminal A. The capped portions, migrating near the origin, were indistinguishable from the medRNA counterparts: D yielded the faster component of the U2 and T2 doublets, while D' yielded the slower component of each of the doublets (Fig.19). Aliquots of RNase T2 digests of bloodstream RNA and unfingerprinted medRNA are also shown (Fig. 19). They behaved similarly to the other RNase T2-resistant caps.

The identity of the patterns of products obtained from medRNA and mRNA spots 3 in the fingerprints and on DEAE HVE after no treatment, RNase A treatment, as well as between the faster component of medRNA spot 3 and mRNA spot D after RNase U2, RNase T2, RNase T2 followed by CIAP (Fig. 19) and nuclease P1 treatment (see below) indicated that the cap that was isolated from mRNA was the mini-exon specific cap. The fact that identical products were obtained after treatment with RNase T2 suggested that the two caps (medRNA and mRNA) did not differ with regard to ribose methylation.

INITIAL ANALYSIS WITH NUCLEASE P1

Nuclease P1 is a non-specific endonuclease that can cleave the phosphodiester bonds of modified nucleotides. It cannot cleave pyrophosphate bonds, so it can be used to isolate a structure referred to as the "core cap," consisting of $m^7GpppN_3\cdot OH$, where N is the first encoded nucleotide, from the other modified mononucleotides of a cap structure (Shatkin, 1976). A cap 0 structure would be expected to yield identical products when treated with nuclease P1 or RNase T2 and phosphatase.

When the medRNA and the mRNA (analyzed as the two separately purified components, D and D') capped oligonucleotides were digested with nuclease P1 [using "standard conditions" (Materials and Methods)], a complex pattern, consisting of a series of products, designated i', i, ii, iii, and v was obtained (Fig. 19). Note that none of the nuclease P1 products co-migrated with the RNase T2- and phosphatase-resistant cap, suggesting that the mini-exon cap cannot be a cap 0 structure. Some of these products appeared to migrate as oligonucleotides. The pattern of spots was highly reproducible, although their relative abundance differed between experiments. D and D' had identical patterns except that they differed in the slowest migrating component: D primarily had i while D' had primarily i' (Fig. 19).

ARTIFACT

Spots i' and i were separately eluted, digested with nucleotide pyrophosphatase and analyzed on TLC (Fig. 20). Each of the resulting products was eluted separately and identified on Whatman 540 paper HVE, pH 3.5 (data not shown). Spot i produced pA^* , pX , pm^7G and phosphate (not visible in this exposure), while spot i' produced pA^* , pX' and phosphate. The pA^* species, produced from each cap, migrated on Whatman 540 paper as an adenosine and on TLC in the position of a modified adenosine but the modification was not identified (see below). The presence of pm^7G and pA^* in spot i indicated that spot i contained the core cap, but subsequent analysis showed that it was not the core cap (see below; the production of pA^* was due to contaminating phosphodiesterase activity). 7-methylation of guanosine gives m^7G an extra positive charge that renders the nucleotide sensitive to ring-opening reactions at neutral or alkaline pH; ring opening results in the loss of positive charge (Shatkin, 1976; Furuichi et al, 1975a; Banerjee, 1980). This reaction is used in DNA sequencing (Maxam and Gilbert, 1977). pX and pX' are both artifactual derivatives of pm^7G ; pX was likely formed during the nucleotide pyrophosphatase incubation (pH 7.4), while pX' was formed during fingerprinting. Both

products: (i) were sensitive to phosphatase, (ii) but resistant to nucleotide pyrophosphatase (data not shown) and (iii) migrated, on Whatman 540 paper HVE, pH 3.5, in a position similar to a ring-opened pm^7G (Furuichi, 1975a and data not shown). The ring-opening reaction has several intermediates. It is likely that pX and pX' are different stages in the breakdown of pm^7G , but that they both have lost the positive charge, and therefore have, minimally, a ring-opened structure. Commercially obtained pm^7G , when treated with alkali, yielded two products, one of which migrated on TLC in a position similar to pX and the other, similarly to pX' (Fig. 20, lane e).

Since i and i' differ by the presence or absence of pm^7G , and since D and D' differ only with regard to i and i' , one can conclude that the difference between D and D' is the presence or absence of authentic pm^7G . Supporting this interpretation, alkaline treatment of spot i , but not spot i' , caused altered electrophoretic mobility (Fig. 19). The ring-opening reaction also explains why an apparently pure fingerprint spot (DEAE homochromatography "counts" phosphates) would contain two components whose behavior on DEAE HVE differs so greatly. No other differences between the two caps were detected.

These findings indicated that the mini-exon cap is homogeneous with regard to modification; therefore, the possible difference between two members of the capped doublets, mentioned earlier, that they are cap 0 and cap 1 versions of the same 5' end, is not correct.

SUMMARY

The medRNA RNase T1 fingerprint was used as a guide to identify the capped oligonucleotide from RNase T1 fingerprints of *in vivo* labeled $[^{32}\text{P}]\text{mRNA}$. The medRNA cap and the mRNA cap behaved indistinguishably in all subsequent assays, yielding the characteristic doublet upon treatment with A, T2 and U2. These data indicated that the medRNA and mRNA caps are identical with regard to both sequence and modification. By separation of the two components of the mRNA cap doublet, the slower-migrating member of the doublet was shown to be an artifactual derivative, created during the fingerprinting procedure. The faster-migrating member contained authentic m^7G . Such doublets have been observed in similar analyses by other workers (Ziff and Evans, 1978; Lai et al, 1982). This finding eliminated the possibility that the cap consisted of a mixture of cap 0 and cap 1. The initial analysis with nuclease P1 yielded a very complex pattern of products.

Figure 18
RNase T1 Fingerprints of Trypanosome mRNA.

High molecular weight, mini-exon selected RNA (Panel A) eluted from region 1 of a polyacrylamide gel (such as in Figs. 9 and 14) and oligo(dT) selected RNA (Panel C) were fingerprinted as described. A medRNA fingerprint (Panel B) is shown for comparison. The origin of the fingerprints is at the lower left. The mini-exon-specific oligonucleotides, 2, 3, 7 and 10, as well as medRNA spot 1, are indicated. Two of the mini-exon specific spots, 2 and 3, are easy to identify in the complex mRNA fingerprints. The sequence of the oligonucleotides was confirmed by analytical digests (Table V). Spots 7 and 10 were not readily apparent, presumably due to co-migration with other oligonucleotides. The alignment of these sequences with medRNA is shown in Figure 11. Note the absence (open arrows) of spot 1, from the 3' portion of medRNA, in Panels A and C, indicating that the mini-exon spots are from mRNA and not from contaminating medRNA. The isolated spot near the tops of Panels B and C is free G which is not visible in Panel A because the fingerprint was run slightly longer.

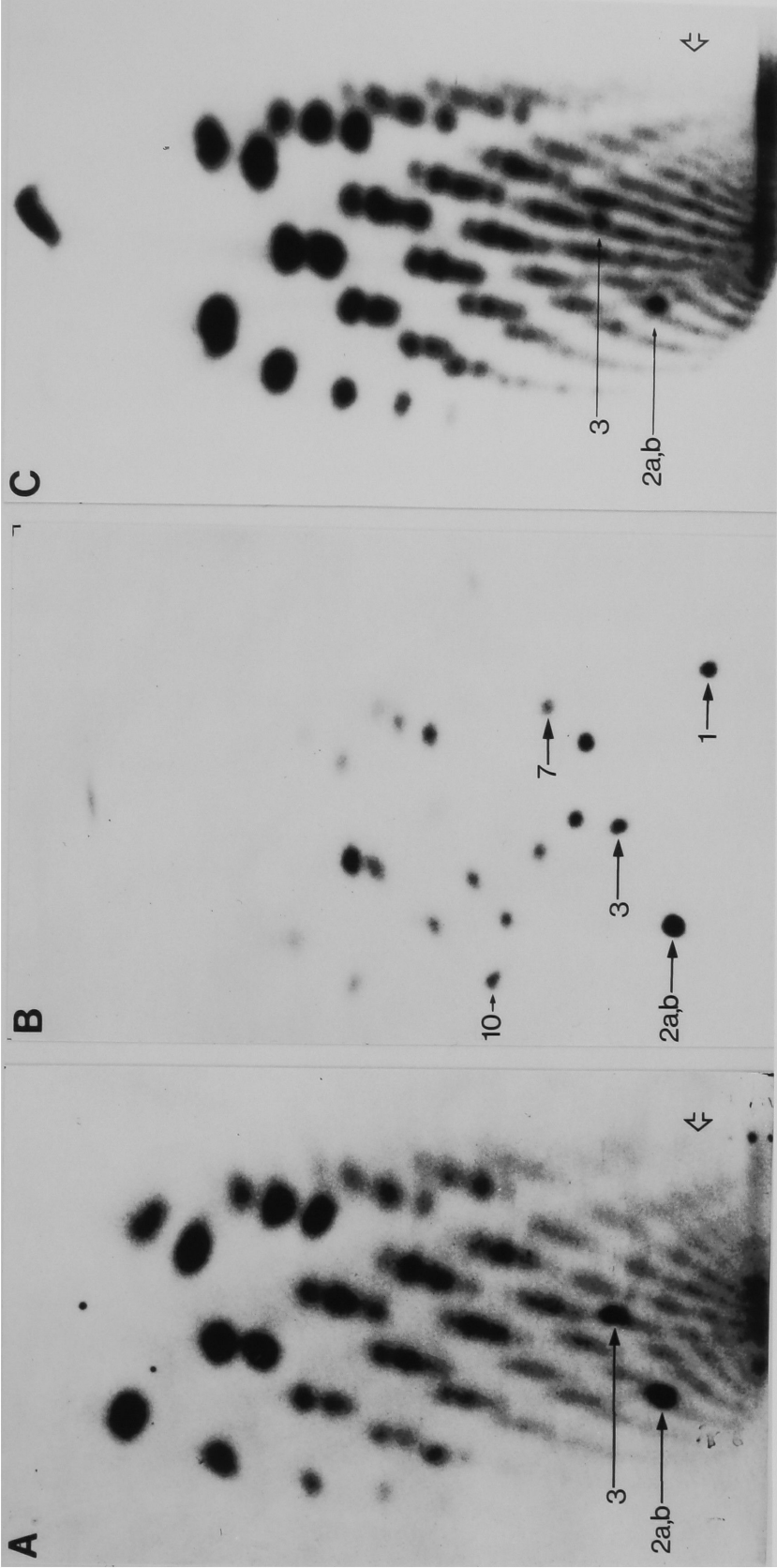


Table V
mRNA Analytical Digestion Products

Oligonucleotide (spot) designations refer to the mRNA fingerprints in Figures 18A [mini-exon-selected RNA high molecular weight (HMW) RNA] and 18C (poly (A)⁺ RNA). See also legend to Table III.

¹See Figure 19.

²A series of non-cap-containing oligonucleotides was also present; however, they are not visible in the exposure shown in Figure 19.

³Analyzed on DEAE paper, HVE, pH 1.9.

⁴The individual components of the RNase A-resistant cap-containing products, D and D' (see text and Fig. 19) were each digested with RNase T2 and U2 separately.

⁵The intensity of AC relative to C was slightly diminished, when compared to the same digest from medRNA, possibly indicating an increase in the modification of the AC bond in mRNA.

Table V. mRNA Analytical Digestion Products				
Spot	RNase T2	RNase A	RNase U2	Sequence
<u>Mini-exon Selected HMW RNA</u>				
2	C,A,G, <u>U</u>	ND	UUG,CUA,UUA,UA,G	CUAUUAUUAG(2a) UACUAUAUUG(2b)
3	C,A,K/G	K(origin)	K(origin doublet) ¹	CapAACG
<u>Oligo d(T) Selected RNA</u>				
2	C,A,G, <u>U</u>	ND	UUG,CUA,UUA,UA,G	CUAUUAUUAG(2a) UACUAUAUUG(2b)
3	C,A,K/G	K(origin doublet) ^{1,2}		CapAACG
D and D ⁴	K(origin doublet) ^{1,3} K ^{1,3} A/C		K ^{1,3} AC,C ⁵	CapAAC

Figure 19
Comparison of the medRNA and mRNA Cap

Different sources of the capped oligonucleotides were subjected to the indicated treatments and analyzed by HVE, pH 1.9, on DEAE paper. The sources of the capped oligonucleotides are:

- m-- spot 3 from RNase T1 fingerprints of medRNA (such as in Fig. 10);
- A⁺--spot 3 from RNase T1 fingerprints of oligo(dT) selected mRNA (such as in Fig. 18C);
- H--spot 3 from RNase T1 fingerprints of mRNA, isolated from the high molecular weight region of polyacrylamide gels of mini-exon selected RNA (such as in Fig. 18A);
- D and D' -- RNase A products from mRNA spot 3, as described below and in the text;
- "m(u)"--unfingerprinted medRNA;
- "b(u)"--unfingerprinted, *in vivo* labeled bloodstream form RNA;
- "m(D)"--medRNA spot D;
- i and i' --the slowest-migrating nuclease P1 products of D and D', respectively.

The third RNase A lane is a representative preparative experiment, from which each of the two products, D (the faster migrating) and D' was eluted and analyzed further. The slightly faster migration of the two unfingerprinted samples is due to a slightly longer electrophoresis time. Nuclease P1 produced a complex array of products, designated i', i, ii, iii and v. "Alk" refers to alkaline treatment. The origin (O) is at the bottom of the diagram; XC signifies the position of the xylene cyanol dye marker.

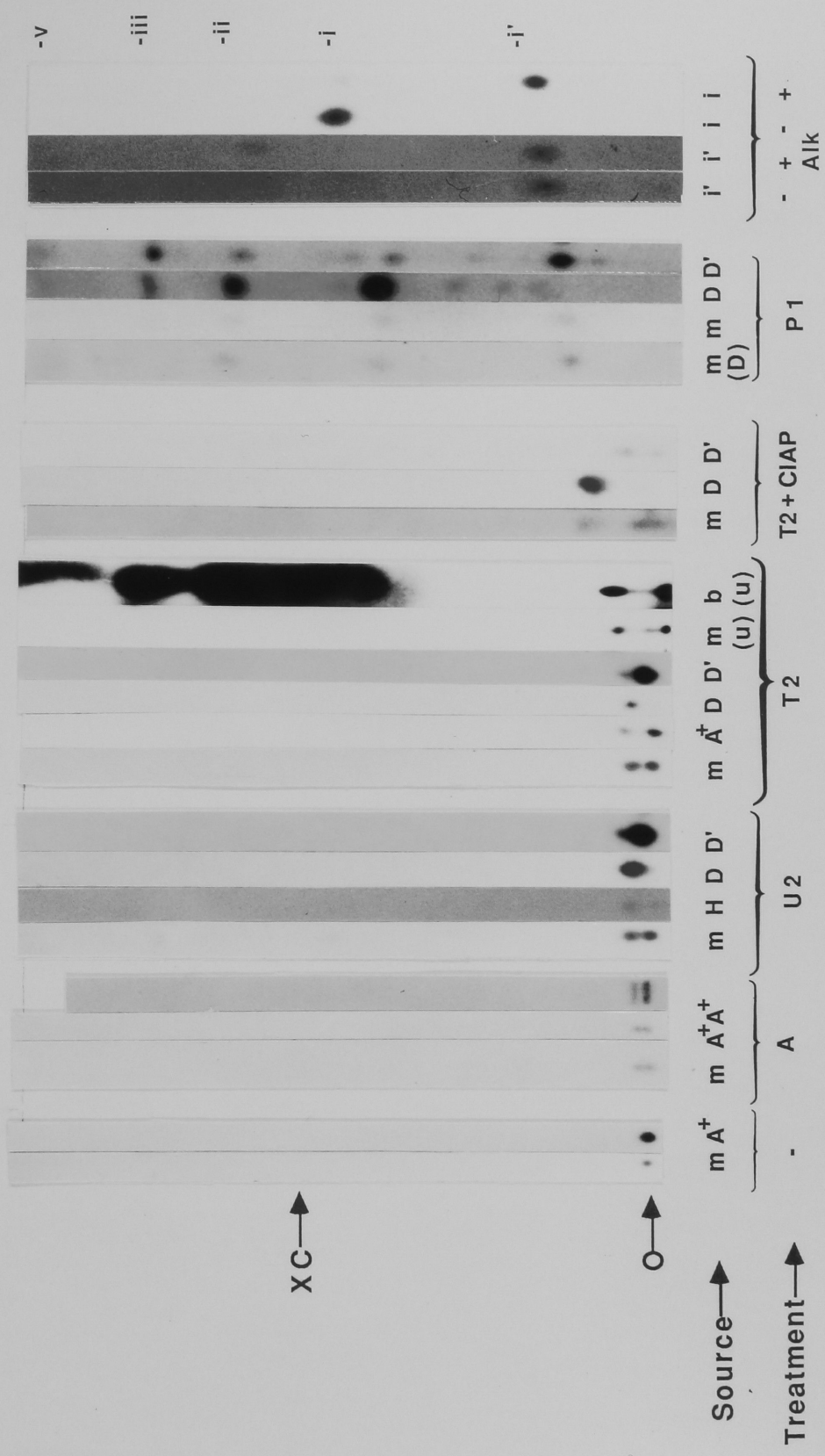
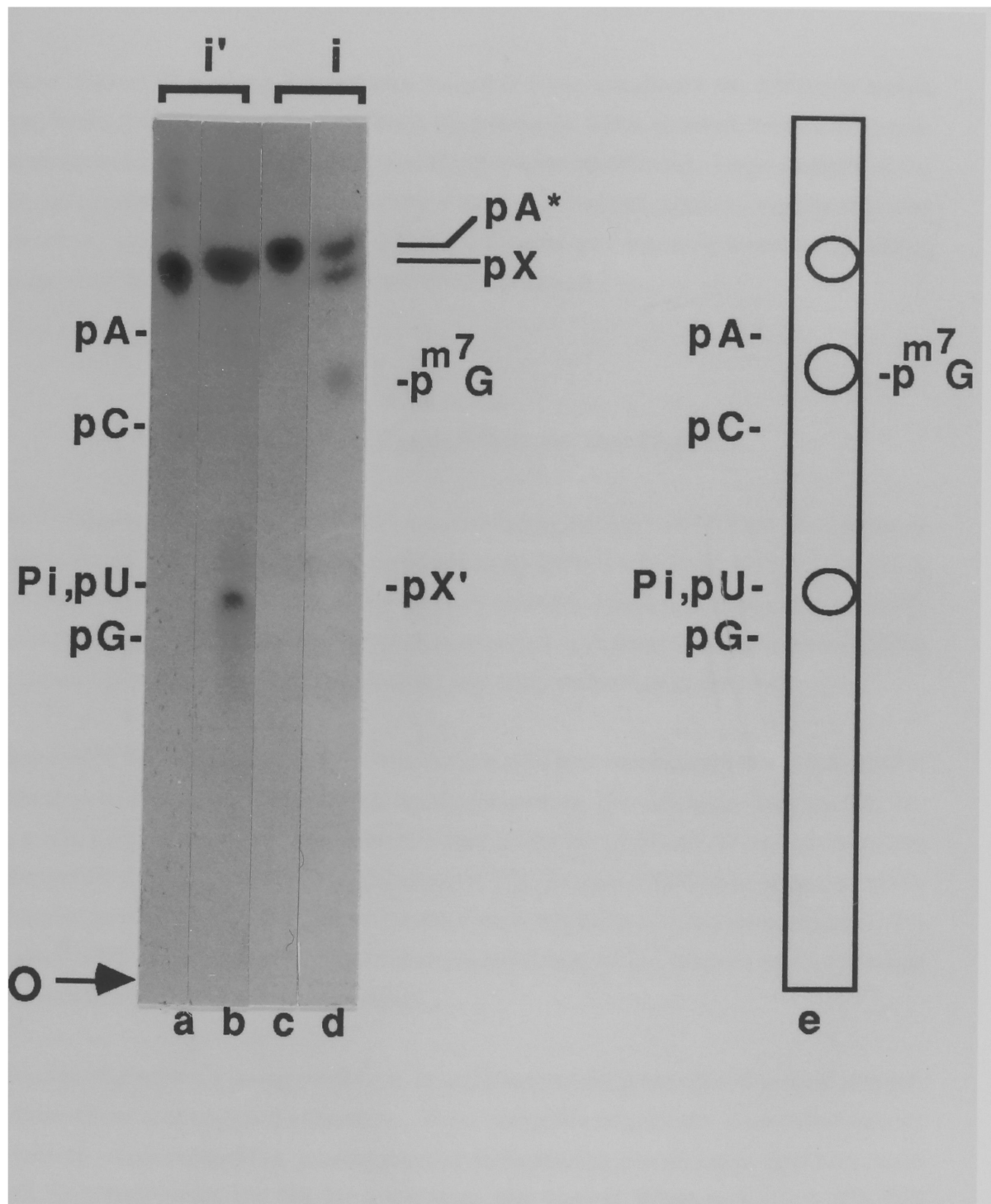


Figure 20
The Difference Between i' and i.

Spots i' and i (lanes a and c), from nuclease P1 digests of spots D' and D (such as shown in Fig. 19), respectively, were separately digested with nucleotide pyrophosphatase (lanes b and d) and analyzed on one-dimensional TLC. Each of the resulting products was eluted and identified by 540 paper HVE, pH 3.5. pX and pX' are likely to be ring opened forms of pm^7G . Lane e is a diagrammatic representation of the UV-absorbing spots that were detected after alkaline treatment of commercially obtained pm^7G . Two products, migrating approximately in the positions of pX and pX', in addition to unreacted pm^7G , were observed. The origin (O) is at the bottom of the figure and mononucleotide markers are indicated.



CHAPTER VI

UNFINGERPRINTED CAP

A simpler method of purifying the cap was sought in order to circumvent the artifactual losses incurred during fingerprinting and to accelerate the procedure. Once obtained, it was important to demonstrate that this unfingerprinted cap was the mini-exon-specific cap. Large amounts of the mRNA cap could then be used for fine structural analysis. The data obtained thus far indicated that medRNA cap and the mRNA cap were indistinguishable, so it was likely that results obtained in fine structural studies would apply to the medRNA cap as well.

RESULTS

ACCELERATED PURIFICATION OF THE CAP

In vivo ^{32}P -labeled procyclic poly(A)⁺ RNA was completely digested with RNase T2 followed by phosphatase treatment and preparative fractionation on DEAE paper HVE, pH 1.9 (Fig. 21). A predominant capped oligonucleotide was produced. In some experiments, the cap was eluted, redigested with RNase T2 and repurified on the same system in order to remove contaminating non-capped oligonucleotides. The artifactual cap was not produced using this technique.

To demonstrate that this oligonucleotide contained the mini-exon-specific cap, it was subjected to a series of comparisons to the cap from the fingerprint analysis. The unfingerprinted cap (Fig. 21, lanes a and d) co-migrated with the authentic caps produced by RNase T2 and phosphatase treatment of the medRNA (Fig. 21, lane b) and mRNA (Fig. 21, lane c) RNase T1 fingerprint spots, when analyzed on DEAE paper. When the caps from the three sources were digested with nuclease P1, the same, distinctive pattern of products (except for the absence of the artifactual spot i') was obtained (Fig. 21, lanes e, f and g).

The other products formed during the RNase T2 and phosphatase preparative digests of poly(A)⁺ RNA were eluted and analyzed separately. Three categories of products were detected (Fig. 22A, lane a). Mononucleotides, presumably due to incomplete phosphatase treatment, were present. Oligonucleotides that left the origin were also present (areas α , β and γ); the third category consisted of material that remained on the origin. The origin material, when eluted and analyzed, was resistant to DNase treatment (Fig. 22A, lane c). When it was digested with nuclease P1 (lane d), it yielded mononucleotides and residual material on the origin. When this procedure

was repeated (lane e and f) or if the origin material was redigested with RNase T2 (lane g), even after boiling (lane h), similar results were obtained, indicating that there is a core of relatively large-sized RNA that is resistant to RNase T2 and nuclease P1. The origin material is likely to be highly structured RNA with double stranded regions (Robertson et al, 1977) or highly or unusually modified RNA.

The oligonucleotides that left the origin comprised a diverse group of products; a full analysis of them is beyond the scope of this thesis. These products fell into 2 categories: products (area β and γ) that were fully digestible to mononucleotides by RNase T2 or P1 (Fig. 22B) and products (area α) that yielded some nonmononucleotide, nuclease-resistant structures (Fig. 22C). Products in the second category are candidates for other caps, branch point nucleotides (Wallace and Edmonds, 1983) or highly modified internal bonds.

COMPLETE NUCLEASE P1 DIGEST OF THE CAP

Nuclease P1 has classically been used to obtain core cap structures from adjacent modified nucleotides. It was necessary to obtain a core cap structure for the following reasons: (i) to verify that the 5' end of the mini-exon contained a conventional m⁷G in a pyrophosphate link, (ii) to determine the first encoded nucleotide and, therefore, the probable start site of transcription and (iii) the mononucleotides produced concurrently would give the base composition of the cap.

The complex pattern of nuclease P1 products (Figs. 19, 21 and 23) was obtained under conditions in which the phosphodiester bonds of conventional nucleotides were cleaved. However, nuclease P1 cleaves the phosphodiester bonds from modified nucleotides with reduced efficiency relative to unmodified nucleotides (Yamada and Ishikura, 1975; Silberklang et al, 1979). The complex pattern of oligonucleotides suggested that nuclease P1 was not completely cleaving all the bonds in the RNase T2-resistant cap. The differing ratio of nuclease P1 products from experiment to experiment was consistent with this interpretation. In order to obtain the core cap, increasingly harsh digestion conditions were tested (Fig. 23A). At the lower concentrations of P1, the partial products (Fig. 23A, lanes b-d; spots i, ii, iii and mononucleotides, seen earlier) were obtained. Upon increased treatment with P1 and omission of tRNA from the digest, spots iv, v and vi appeared, while spots i, ii and iii disappeared (Fig. 23A, lane e). Spot iv co-migrated with spot ii on this electrophoretic system; however, subsequent analysis demonstrated that they are distinct spots. Each of the spots was eluted from DEAE paper and further analyzed. A near complete P1 digest of the cap analyzed on 540 paper HVE, pH 3.5, is shown also (Fig. 23B).

Nuclease P1 products are expected to have 5' phosphate groups, which are sensitive to phosphatase treatment. Purified spot iv was resistant to phosphatase (data not shown), suggesting that it was a 5' terminal oligonucleotide. Spot iv was demonstrated to be the core cap by its sensitivity to three different enzymes: tobacco acid pyrophosphatase (TAPase, Shinshi et al, 1976), nucleotide pyrophosphatase and venom phosphodiesterase. Each of these digests yielded pm^7G , pA^* and phosphate (Fig. 24). The most likely way in which an entity that is itself resistant to phosphatase could contain three phosphates is if they are internal, such as in a conventional 5' to 5' pyrophosphate linkage. In addition, the latter two treatments produced a spot (X') that migrated near pU. This spot is likely to be a mixture of the ring-opened form of pm^7G and ADP. The artifactual form was likely produced during the incubation, because it was absent from the TAPase reaction. The pH of the latter two treatments (pH 9.0 for venom phosphodiesterase and pH 7.4 for nucleotide pyrophosphatase) is higher than that of the TAPase (pH 6.0). When the venom phosphodiesterase products were treated with alkali, pm^7G disappeared, but X' did not (Fig. 24). These combined data demonstrated that the mini-exon core cap contains m^7G in a conventional 5' to 5' triphosphate linkage and that the initiating nucleotide is an adenosine. The fact that spot iv migrated (on DEAE paper HVE, pH 1.9) close to, but not identically, with the either of the commercially obtained core cap markers, $\text{m}^7\text{GpppA}_{\text{OH}}$ or $\text{m}^7\text{GpppA}_{2'\text{m}}$ (Fig. 23A), indicated that the mini-exon core cap contained a modification in addition to the expected 2'-O-methylation. N⁶ methylation of adenosine is common in this position in caps; however, the nature of the extra modification was not determined.

The mononucleotides produced upon full digestion of the RNase T2- and phosphatase-resistant cap with nuclease P1 were identified by elution from DEAE paper followed by subsequent analysis on Whatman 540 paper HVE, pH 3.5, and TLC. Since these nucleotides were produced from an RNase T2-resistant entity, all but one should be 2'-O-methylated (Fig. 7). The 3' terminal nucleotide, which was originally recognized by RNase T2, should be unmodified. Because the Whatman 540 paper system is relatively insensitive to charge-neutral modifications (Cory and Adams, 1975), it was used to identify the nucleotides, while a different system was necessary to identify the nature of the modifications. The mononucleotides from the complete nuclease P1 digest were identified as pA^* , pA and pC^* (Figs. 23A, lane e, spot v; data not shown and Fig. 25A, lane a) and pU^* (Fig. 23A, lane e, spot vi and Fig. 25A, lane b). These modified mononucleotides migrated slightly faster, on DEAE paper, than their unmodified counterparts (Fig. 23A). When the pA/pC region of the DEAE paper from a complete P1 digest (Fig. 23A, lane e, spot v) was analyzed on Whatman 540 paper (Fig. 25A, lane b) both adenosine and cytidine were detected, indicating that spot v contains, minimally, two nucleotides. However, when spot vi was analyzed

on both one- and two-dimensional TLC, 2'-O-methyl cytidine, an unmodified adenosine as well as a modified adenosine were detected (data not shown), suggesting that the pA*, detected on Whatman 540 paper, was heterogeneous. However, the nature of the modification of the adenosine residue was not determined. These combined data indicated that spot v contained three nucleotides, pA*, pA and pC*.

Similarly, spot vi (from Fig. 23A, lane e) was identified as a uridine by its mobility on Whatman 540 paper HVE (Fig. 25A, lane b). However, it migrated much faster than a conventional pU on DEAE paper (Fig. 23A). It likely contains a modification in addition to 2'-O-methylation because it did not co-migrate with the commercial marker, 2'-O-methyl uridine 5' monophosphate, on two-dimensional TLC (Fig. 25B). The direction of shift is consistent with an additional methylation (Kuchino et al, 1987).

The production of an adenosine in addition to the core cap was not surprising because the analysis with the cyclizing nucleases indicated that the 3' terminus of the cap was an adenosine (Chapter III). However, the production of pC* and pU*, in addition to the core cap, from the RNase T2-resistant structure indicated that there were a minimum of three encoded nucleotides (the pA* from the core cap, pC* and pU*) upstream of the previously mapped 5' end (the 3' terminal adenosine of the cap, Fig. 26). Such a structure would be a "cap 3", since the 3' terminal adenosine would be unmodified. This result was inconsistent with the prediction that the mini-exon cap would be either a cap 1 or cap 0 (Chapter III), suggesting that the previously mapped 5' end was incorrect. The finding that spot v from the complete nuclease P1 digest contained two adenosine residues suggested the presence of 3 adenosines in the cap: two in addition to the pA* from the core cap. The combined data suggested that the cap could contain 5 nucleotides in addition to pm⁷G (3 adenosines, pC* and pU*). However, these data do not provide the sequence of these nucleotides. Conventional nucleotide sequencing methods (both chemical and enzymatic; for both DNA and RNA) are specific for unmodified nucleotides: arrays of modified nucleotides may be sequenced by a combination of partial and complete digests by nonspecific nucleases (Silberklang et al, 1979; Kuchino et al, 1987).

PARTIAL NUCLEASE P1 DIGESTS OF THE CAP

Partial digests can be used to obtain sequence information by redigesting the partial products (Lockard and RajBhandary, 1976, Silberklang et al, 1979, Kuchino et al, 1987). A mixture of all possible products from a cap 2 structure would yield 5 structures, 2 of which would be incomplete digestion products (Fig. 27, capN*N + N; capN* + N*N and capN* + *N + N). It was unexpected to

obtain 3 partial digestion products (i, ii and iii) from the mini-exon cap. These partial products were used to sequence the cap structure.

$$I = iv + pA^*$$

Spot i consists of the core cap and one adjacent nucleotide, adenosine. Spot i behaved similarly to spot iv in several assays: it was resistant to phosphatase (data not shown), it yielded pm^7G , pX'' , phosphate and pA^* when digested with nucleotide pyrophosphatase and venom phosphodiesterase; furthermore, m^7G disappeared when the venom phosphodiesterase products were treated with alkali (Fig. 28A, lanes c-e). As discussed previously (Chapter V), the nucleotide pyrophosphatase was contaminated with phosphodiesterase. However, treatment of spot i with TAPase (Fig. 28A, lane b), yielded phosphate, pm^7G and a novel spot, which replaced the pA^* produced from spot iv (Fig. 24, lane b). The presence of phosphate and pm^7G indicated that spot i did not differ from spot iv at the 5' end; the likely difference was at the 3' end. The novel spot was not analyzed further, but is likely to consist of pA^*pA^* , whose phosphodiester bond would be resistant to tobacco acid pyrophosphatase. When purified spot i was redigested with P1 (under the harsher conditions) it yielded a spot similar to spot iv and a mononucleotide, pA^* (Fig. 28B and C). The digest was analyzed by HVE on both Whatman 540 and DEAE paper; however, the mononucleotide was not visible in the latter case.

These combined data suggested that the 5' end of the cap is $m^7GpppA^*A^*$ (Fig. 29), implying that the core cap is directly linked to at least one pA^* . These two pA^* s are different from the two linked adenosine residues discussed in Chapter III because they come from an RNase T2-resistant structure that contains pC^* and pU^* . Note that spot i contains no cytidine. The two modified nucleotides, pC^* and pU^* , must be 5' of the 3' terminus of the cap, but 3' of spot i; therefore, pC^* and pU^* separate the two pairs of adenosines.

$$II = III + pC^*$$

When spot ii was purified and treated with nuclease P1 under mild digestion conditions, a spot similar to spot iii and a mononucleotide(s) migrating in the pA/pC region were produced (Fig. 30A). The mononucleotide was difficult to see, perhaps due to decreased hydrophilic interactions between the paper and the modified nucleotide, relative to an unmodified nucleotide. This result suggested that spot iii was contained within spot ii.

To determine the sequence of these two oligonucleotides, the following strategy was used (Fig. 30D). The base composition of each oligonucleotide was determined by digestion to completion by venom phosphodiesterase. In a separate reaction, the base composition was determined after pretreatment with phosphatase. Since these oligonucleotides are nuclease P1 products, they are expected to have 5' terminal phosphates. Phosphatase treatment will selectively remove the phosphate from the 5' nucleotide. If the 5' nucleotide is unique, it will disappear from the base composition after pretreatment with phosphatase.

Spot ii yielded the base composition, pC*, pA and pU* (Fig. 30B, lane b). Venom phosphodiesterase after pretreatment with phosphatase yielded pA and pU* (Fig. 30B, lane d), indicating that spot ii contained only one pC* and it was at the 5' end of spot ii. When spot iii was subjected to the same series of experiments, its base composition was shown to be pA and pU* (Fig. 30C, lane b). The disappearance of pU* after pretreatment with phosphatase (Fig. 30C, lane d) indicated that the 5' end of spot iii was pU*. Therefore, the sequence of spot iii is pU*pA or pU*pA*pA.

The combined data are consistent with spot iii being contained within spot ii, such that the sequence of spot ii is pC*pU*pA or pC*pU*pA*pA.

I + ii = Cap?

It was demonstrated above that spot i is at the 5' end of the cap structure and consists of m⁷GpppA*A*. Spot i contains no cytidine and spot ii begins with pC. Therefore, spots i and ii are not overlapping and their adenosines are separate (Fig. 31). These combined data support the earlier interpretation that the adenosines in the cap derive from, minimally, three separate adenosines. If spot i is directly bonded to spot ii in the cap structure, the cap contains 4 or 5 modified nucleotides (excluding m⁷G), in the sequence m⁷GpppA*A*C*U*A or m⁷GpppA*A*C*U*A*A.

VENOM PHOSPHODIESTERASE ANALYSIS OF THE CAP

In order to distinguish between these alternatives and to derive the sequence data by an independent method, the purified cap structure was subjected to partial and complete venom phosphodiesterase digests. This enzyme is a non-specific exonuclease that cleaves processively in a 3' to 5' direction such that partial products are related to one another by the loss of 3' nucleotides (Laskowski, 1971).

Partial Venom Phosphodiesterase Digests

When the purified cap structure was digested under conditions developed to obtain partial products, the digestion proceeded in two kinetic steps: first, the formation of oligonucleotide I and mononucleotides, spot II, followed by the production of oligonucleotide III and mononucleotide IV (Fig. 32A).

Each of the products was eluted and analyzed further. Spot I, when digested to completion by nuclease P1 and analyzed on Whatman 540 paper HVE, yielded pC*, pA*, and a spot likely to be a mixture of pU* and ADP (Fig. 33A, lane b). Although pU* and ADP do not resolve well on this system, when the nuclease P1 digest was analyzed on DEAE HVE (a system in which pU* and ADP are well separated), both spots were detected; however, pA* and pC* do not resolve well on the latter system (data not shown). Spot I was sensitive to phosphatase and tobacco acid pyrophosphatase (data not shown). These combined data suggested that the 5' end of spot I was a diphosphate and that its partial sequence was ppA*(pA*, pC*, pU*). The mononucleotides concurrently released from the cap structure (spot II) were identified by HVE, pH 3.5, (Fig. 33A, lane c) and by TLC (Fig. 33B). Two mononucleotides, pm⁷G and pA (unmodified) were observed. The composition of spot II did not change in the later time-points. Since pm⁷G was likely to have been released due to the presence of a free 3'-OH group, these results support the nuclease P1 data suggesting that the 5' end of the trypanosome cap contains a conventional m⁷G in a 5'-5' pyrophosphate linkage. They also confirm that the 3' terminal nucleotide of the cap structure is an unmodified adenosine.

The two products released later in the partial reaction, spots III and IV, were similarly analyzed. Spot III yielded ADP, pA* and pC* upon complete digestion with nuclease P1 (Fig. 33A, lane e). The difficulty in distinguishing pU* from ADP was resolved by analyzing the nuclease P1 digest on DEAE HVE, pH 1.9; ADP, but not pU*, was detected (data not shown). Spot IV was identified as a modified pU* by TLC (data not shown). These results indicated that spot III differs from spot I only by the absence of a pU* from its 3' end, suggesting the sequence, ppA*(pA*,pC*) for spot III.

In light of the nuclease P1 digestion of the cap, spot I is likely to be ppA*A*C*U* and spot III, ppA*A*C* (Fig. 32B). Since further partial products were not obtained, not all possible sequences were eliminated. For example, the sequences ppA*C*, ppA*C*A*, ppA*A*C*C* and ppA*A*A*C cannot be ruled out for spot III. However, the latter two structures are inconsistent with the mobility of spot III on DEAE paper (they would migrate slower) and the former two structures

contradict the nuclease P1 results (because it was demonstrated above that the core cap is directly linked to a pA*). The possibility, raised earlier, that the 3' end of the cap is pU*A*A, is unlikely for the following reasons. If the cap had two adenosines at its 3' end, either spot I would terminate in adenosine or the two adenosines would have been released to form spot I. Spot I does not terminate in adenosine because only a pU* was released from its 3' end. If two adenosines, in addition to pm⁷G, were released from the cap to form spot I, it is likely that spot I would not migrate so close to the starting material. Furthermore, spot II, from the early time-point, as well as from later time-points did not contain a modified adenosine. Table VI presents the reasons why other alternative sequences are unlikely; for example, A*A*C*C*U*A or A*A*C*A*U*A are consistent with the nuclease P1 data, but unlikely because of the mobility of the venom phosphodiesterase partial products. The combined data leave only one sequence consistent with all of the data, namely, that shown in Fig. 34.

SUMMARY

An RNase T2- and phosphatase-resistant cap structure was isolated directly from *in vivo* ³²P-labeled poly (A)⁺ RNA. This method of purification overcame the difficulties of losses and artifactual modifications that were encountered during the fingerprinting procedure. This cap was identified as the mini-exon specific cap by its migration on DEAE paper and by the production of characteristic nuclease P1 partial digestion products. The purified cap structure was sequenced by a series of partial and complete digests by nuclease P1 and venom phosphodiesterase and found to consist of m⁷GpppA*A*C(2'-O-m)U*A. The mini-exon is therefore 39 (excluding m⁷G), rather than 35, nucleotides long. Since each of the small RNAs analyzed in Chapter IV produced a spot 3 (Fig. 15) that had identical analytical digestion products to those of medRNA spot 3 (Tables III and IV), they likely have the same cap. The two small RNAs are thus 39 and 37, rather than 35 and 33, nucleotides long.

Figure 21
Isolation and Characterization of the Unfingerprinted Cap

In vivo labeled, oligo(dT) selected procyclic RNA was digested exhaustively with RNase T2 and phosphatase and fractionated on DEAE paper HVE, pH 1.9 (lane a). The cap structure that was isolated and used in subsequent analyses is indicated. An aliquot of the purified cap (lane d) was compared to the RNase T2- and phosphatase-resistant cap from medRNA (lane b, spot 3) and mRNA (lane c, spot D) fingerprints. The nuclease P1 products of the different cap preparations were compared: lane e, medRNA spot 3; lane f, mRNA spot D; and lane g, the unfingerprinted cap. The complex array of nuclease P1 products, spots i', i, ii, iii, v and vi, (similar to Fig. 19) were again obtained. Note the absence of the artifactual product, i', in the unfingerprinted cap. The origin (O) is at the bottom of the diagram; XC signifies the position of the xylene cyanol dye marker; the migration of mononucleotides and phosphate is indicated.

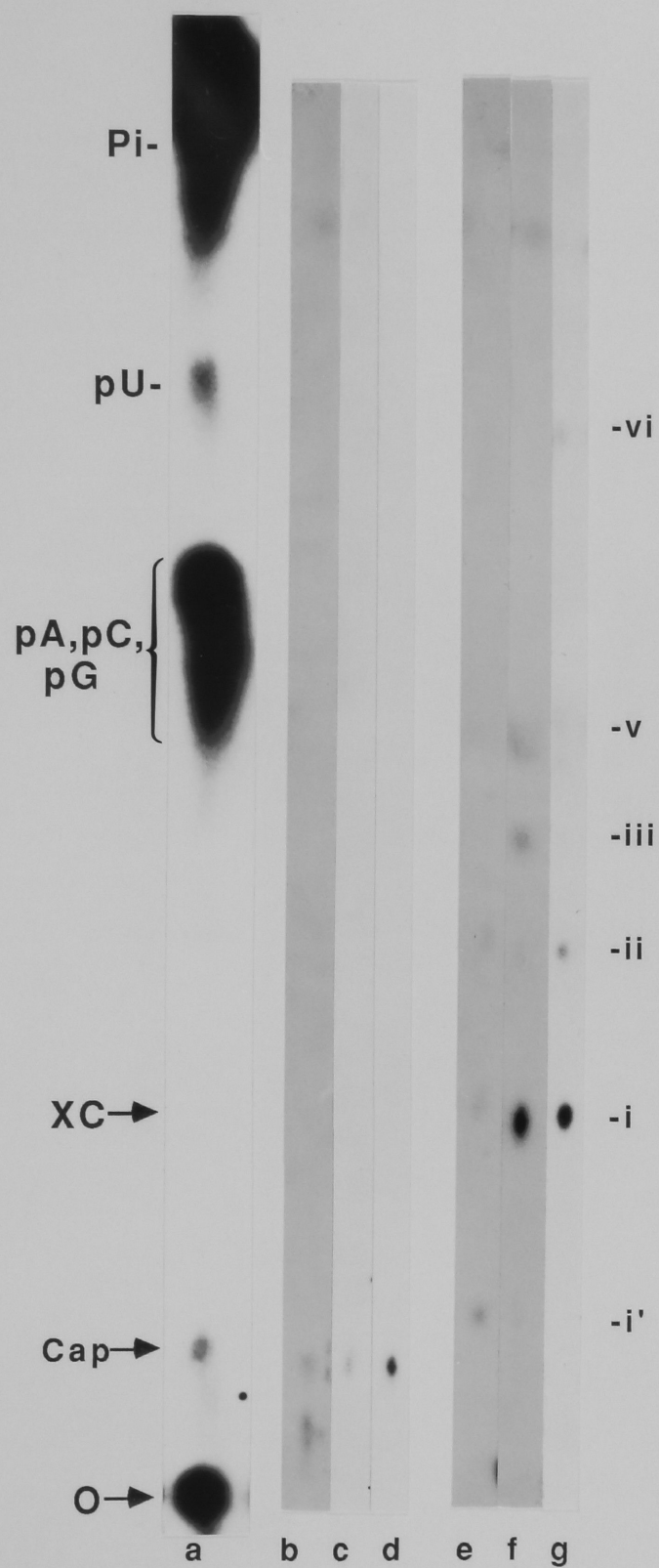


Figure 22
Other Products in the RNase T2 Digests

An RNase T2 and phosphatase digest of *in vivo* labeled RNA, similar to that shown in Figure 21, is shown in Panel A, lane a, after DEAE paper HVE, pH 1.9. Background products, designated α , β and γ are indicated. Spot α was probably on the origin in Figure 21, because a different batch of DEAE paper, one that generally moves oligonucleotides faster, was used in this experiment. RNA from each area was eluted and analyzed further. The material from the origin (O) was rerun after no treatment (lane b), DNase (lane c) or nuclease P1 treatment (lane d). The material remaining at the origin in lane c was again eluted and rerun untreated (lane e), after DNase (lane f), RNase T2 (lane g), or after boiling, followed by RNase T2 (lane h) treatment. Panels B and C depict experiments analyzed on 540 paper HVE, pH 3.5. RNA from area γ was analyzed after no treatment (Panel B, lane a), nuclease P1 (lane b) or nuclease P1 followed by phosphatase treatment (lane c). Similar results were obtained with the material from region β (data not shown). RNA from area α was analyzed untreated (Panel C, lane a), after nuclease P1 (lane b) or after nuclease P1 and phosphatase treatment (lane c). The " γ " material appears to be simple oligonucleotides, as indicated by the absence of residual material in lane c, while " α " appears to contain some P1-resistant material. The origin (O) is at the bottom of the diagrams; XC signifies the position of the xylene cyanol dye marker and the migration of mononucleotides and phosphate is indicated.

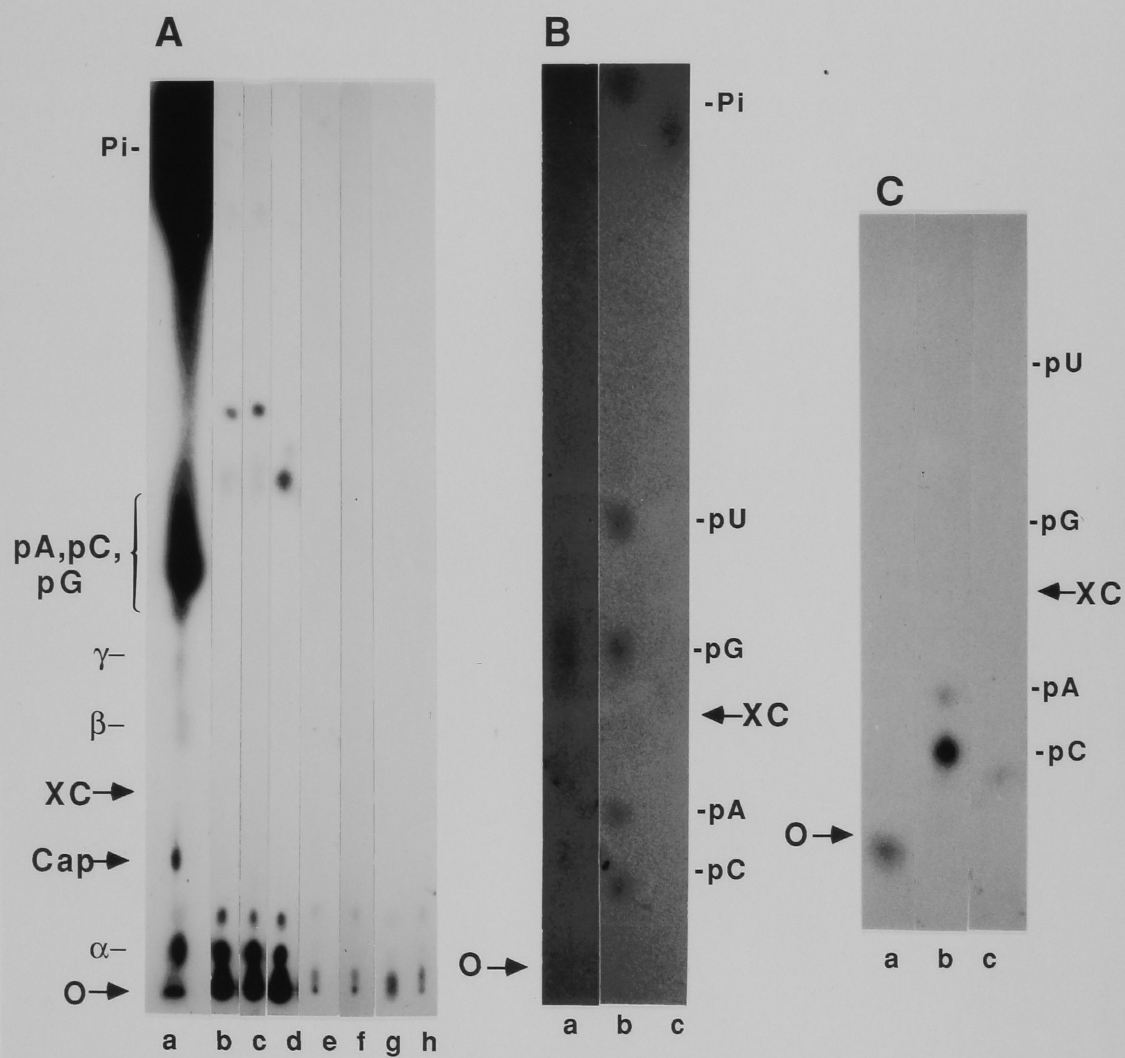


Figure 23
Nuclease P1 Analysis of the Cap

The unfingerprinted cap (Panel A, lane a) was treated with nuclease P1 under the specified conditions: 1.5 μ g, with 10 μ g tRNA for 5 minutes (lane b), 1.8 μ g of nuclease P1 with 10 μ g of tRNA for 2 hours (lane c), 10 μ g of nuclease P1 with no tRNA for 30 minutes (lane d) and 20 μ g of nuclease P1 with no tRNA for 2 hours (lane e) and analyzed on DEAE paper HVE, pH 1.9. A characteristic kinetic pattern of the complex P1 products, i, ii, iii, iv, v and vi emerged. pA and pC do not resolve well in this system. A near complete nuclease P1 digest (Panel B, lane b) of the RNase T2-resistant cap (lane a) analyzed on 540 paper HVE, pH 3.5, is also shown. Lane c is an aliquot of spot iv purified from DEAE paper. The origin (O) is at the bottom of the diagrams; XC signifies the position of the xylene cyanol dye marker; and the migration of cap, mononucleotides and phosphate is indicated.

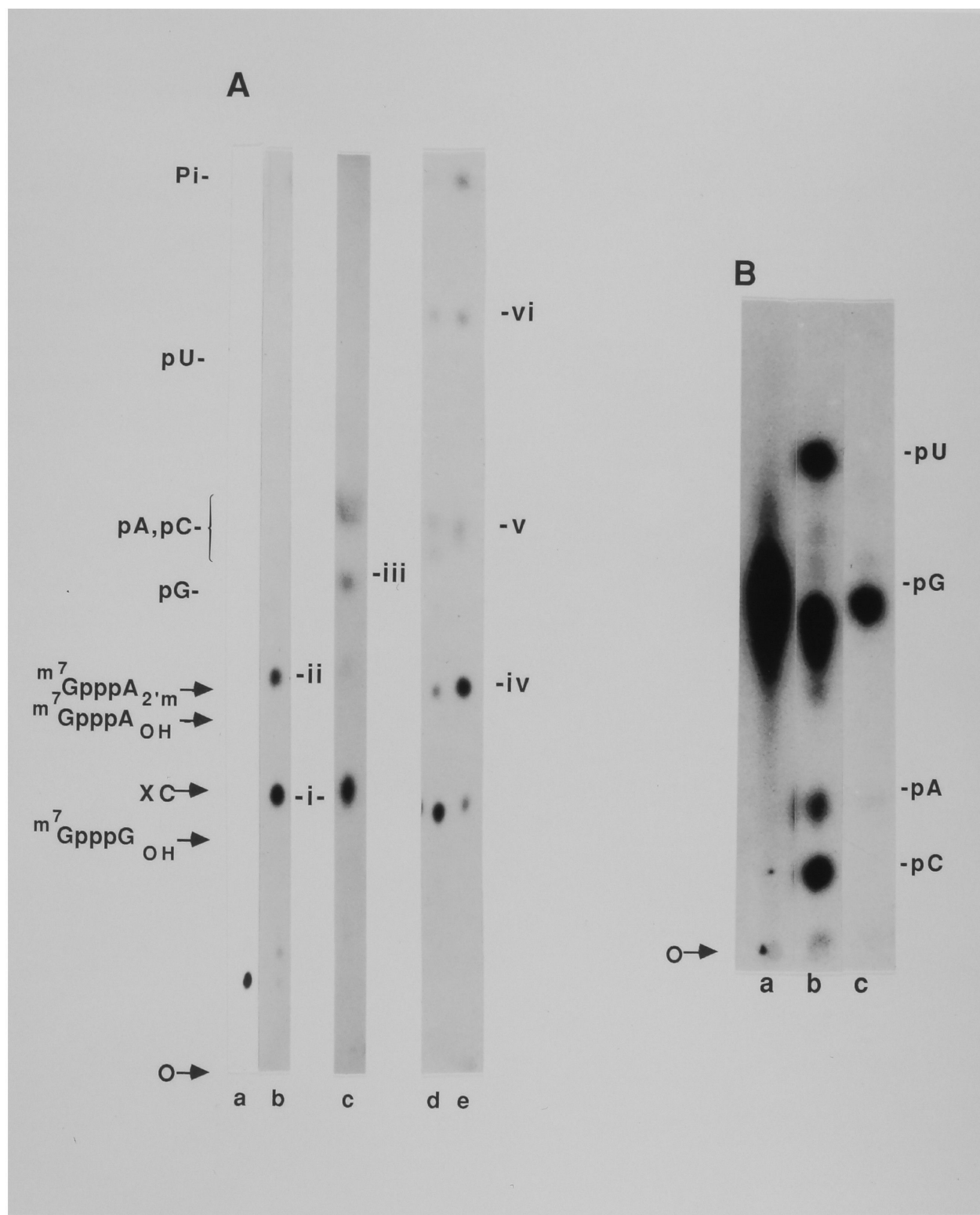


Figure 24
Identification of Spot Iv as the Core Cap

Purified spot iv (from Fig. 23A, lane e) (lane a) was digested with tobacco acid pyrophosphatase (lane b), nucleotide pyrophosphatase (lane c), venom phosphodiesterase (lane d) or venom phosphodiesterase followed by alkaline treatment (lane e) and analyzed on 540 paper HVE, pH 3.5. The high amount of nucleotide pyrophosphatase used (5 μ g) caused pm⁷G to diffuse from the origin in lane c. X' ' is likely to be the ring opened form of m⁷G, possibly mixed with the incomplete pyrophosphatase product, ADP. The origin (O) is at the bottom of the diagram; XC signifies the position of the xylene cyanol dye marker; and the migration of cap, mononucleotides and phosphate is indicated.

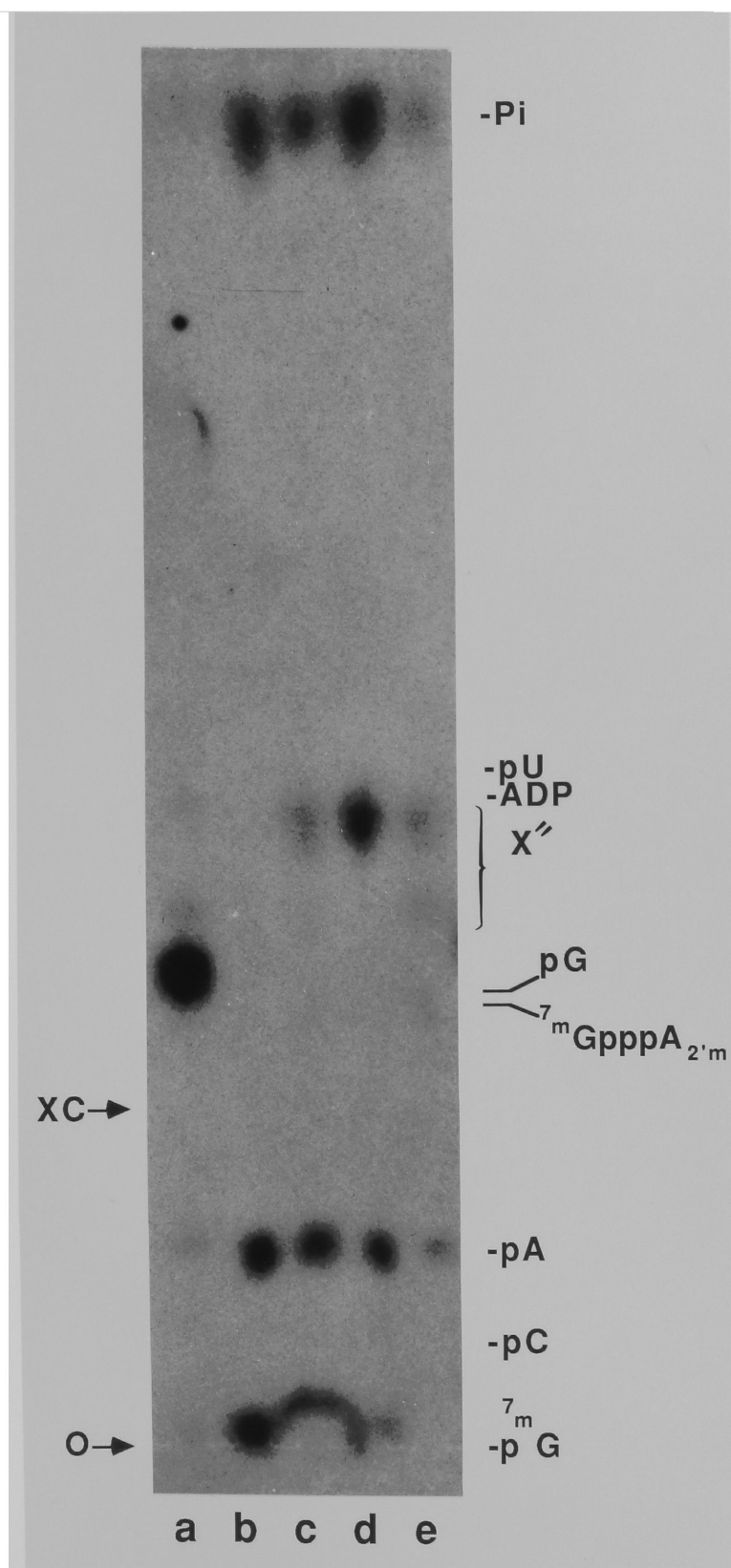


Figure 25
Identification of the Mononucleotides In the Cap

Spot v (Panel A, lane a) and spot vi (lane b) from Fig. 23A, lane e, were analyzed directly on 540 paper HVE, pH 3.5. Spot vi was also analyzed on two-dimensional TLC (Panel B). The origin (O) is at the bottom of Panel A and at the lower left in Panel B; XC signifies the position of the xylene cyanol dye marker; and the migration of mononucleotide markers is indicated.

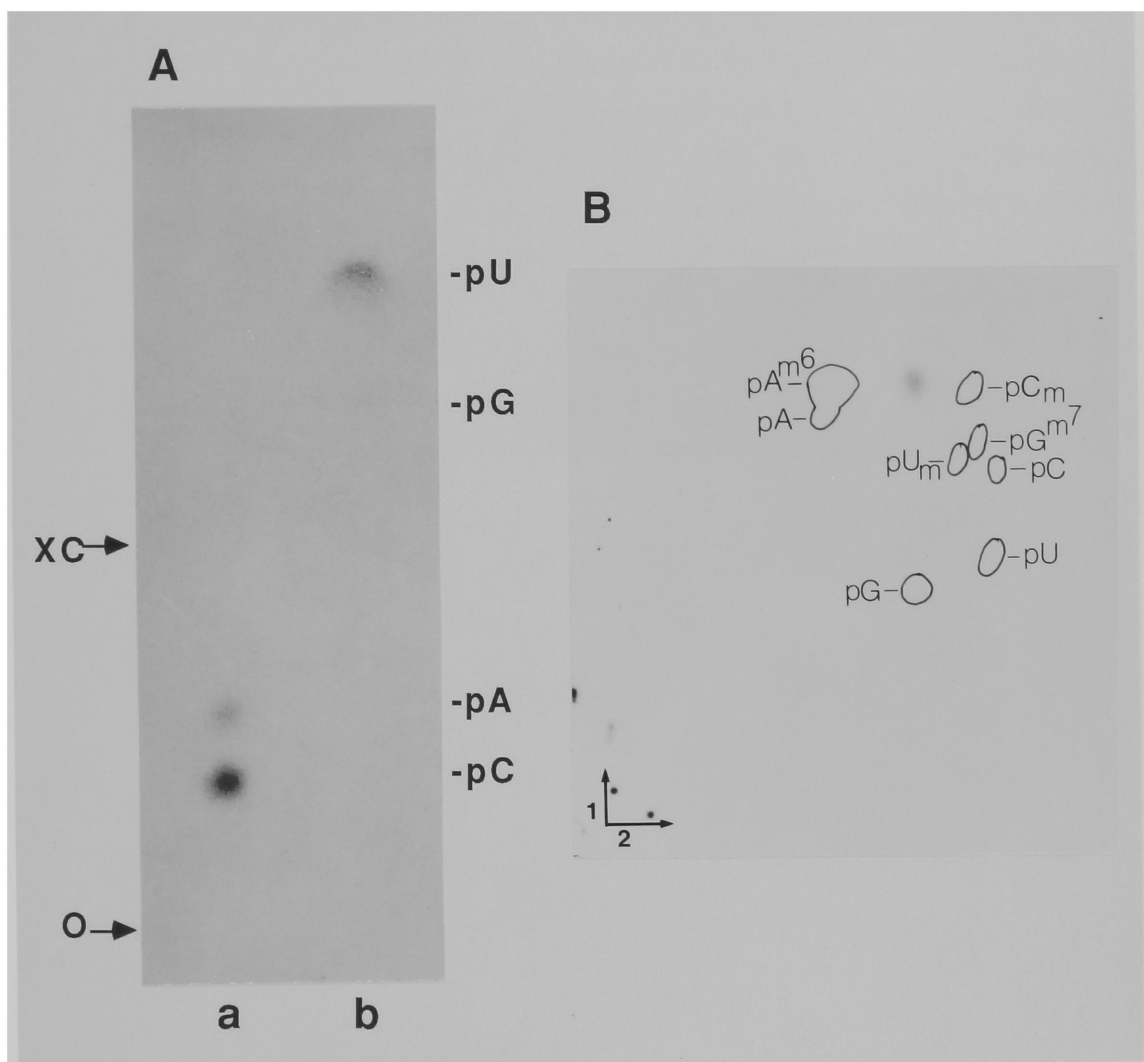


Figure 26
Preliminary Model of the Cap

The base composition of the mononucleotides (Fig. 25A), the identification of the core cap (Fig. 24) from the nuclease P1 complete digest and the earlier findings (Chapter V) that the 3' terminus of the cap is an adenosine, suggested the presence of at least 3 modified nucleotides, pC*, pU* and pA* (from the core cap, spot iv), upstream of the first unmodified nucleotide, the previously mapped 5' end. "T2 + CIAP" indicates the putative site where RNase T2 and phosphatase cleaved to yield the cap structure. 2' OH methylations were expected [confirmed for the cytidine residue only (see text)] to be present because the nucleotides were derived from an RNase T2-resistant structure. Since spot iv did not co-migrate with the core cap marker, m⁷GpppA_{2'}m, the presence of an unidentified modification, designated Y, was suggested. The presence of m⁶A, just downstream of the cap structure, was suggested by the differential resistance of the AC bond to RNases U2 and T2 (Table III).

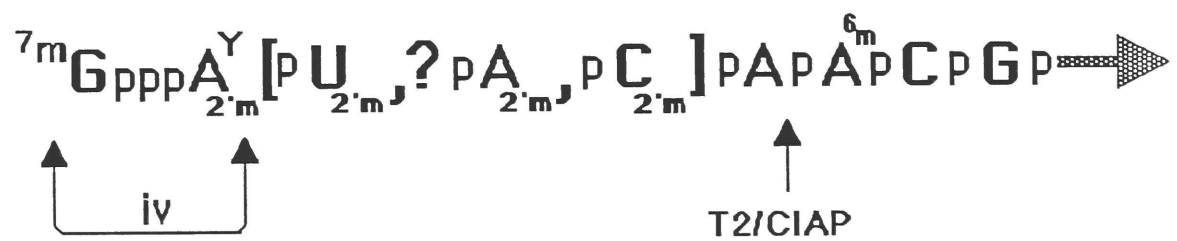


Figure 27
Hypothetical Nuclease P1 Partial Products

If all possible products of an incomplete nuclease P1 digestion of a cap 2 structure were present, five products (1, 2, 3, 4 and 5) would be detected. Two of these (2 and 3) would be partial digestion products.

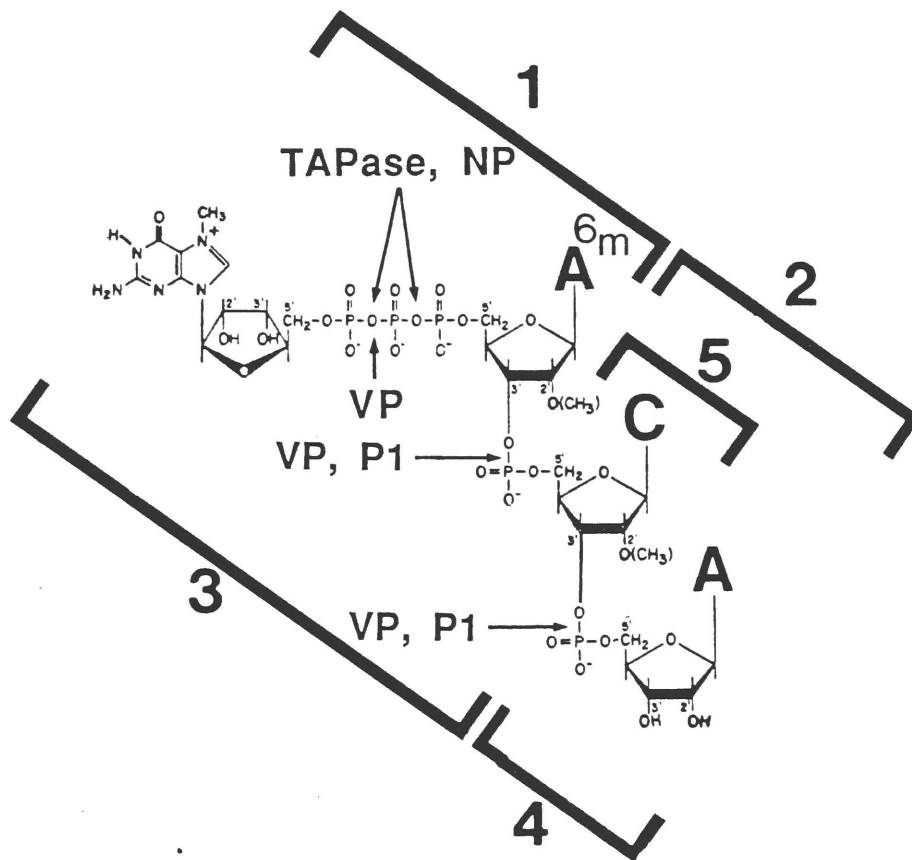


Figure 28
Spot i Contains Spot iv and pA*

Purified spot i (from lane e, Fig. 23) was analyzed on 540 paper HVE, pH 3.5, after no treatment (Panel A, lane a), tobacco acid pyrophosphatase (lane b), nucleotide pyrophosphatase (lane c), venom phosphodiesterase (lane d) and venom phosphodiesterase followed by alkaline treatment (lane e). Products quite similar to those produced by similar treatments of spot iv (Fig. 24) were obtained. However, the replacement of pA* by another spot (lane b) indicated that spots iv and i differ at their 3' ends. The material migrating near pG in lane c is likely to be the starting material. X' ' is probably a mixture of the ring-opened form of pm⁷G and the incomplete pyrophosphatase product, ADP. The production of pA* by nucleotide pyrophosphatase is due to a contaminating phosphodiesterase activity. The high amount of nucleotide pyrophosphatase used (5 µg) caused pm⁷G to diffuse around the origin in lane c. The unidentified spot in lane b is likely to be the dinucleotide, pA*pA*, because treatment of purified spot i (Panel B, lane a) with nuclease P1, under the harsher digestion conditions (lane b), released a spot that co-migrated (on 540 paper HVE, pH 3.5) with spot iv and the mononucleotide, pA*. The latter digest was also analyzed on DEAE paper HVE, pH 1.9 (Panel C) but the mononucleotide was not visible.

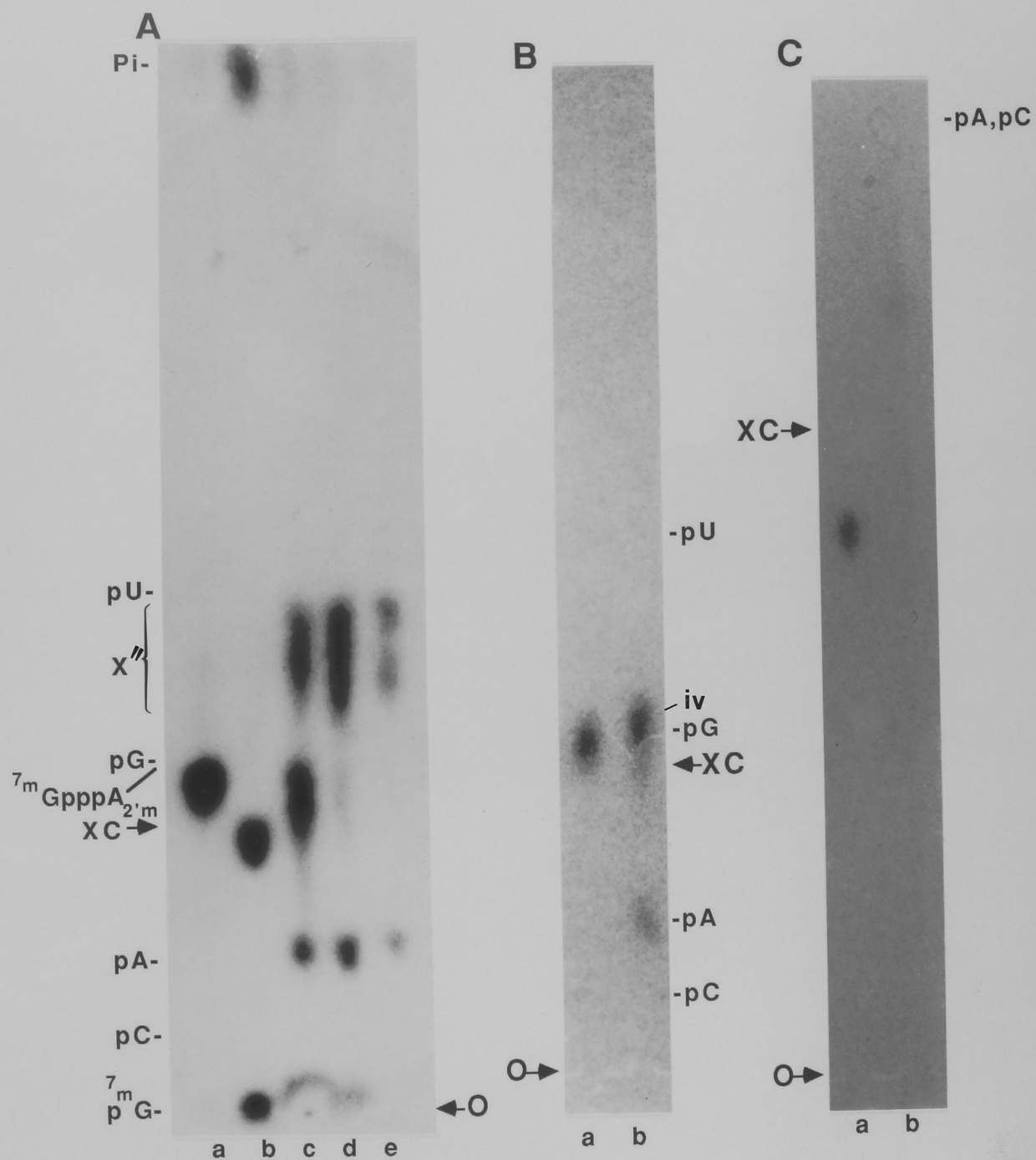


Figure 29
The Core Cap is Linked to pA*

The experiments presented in Fig. 28 demonstrated that spot iv, the core cap, is contained within the nuclease P1 partial digestion product, spot i. A minimum of one pA* is directly linked to the core cap. See text and legend to Fig. 26 also.

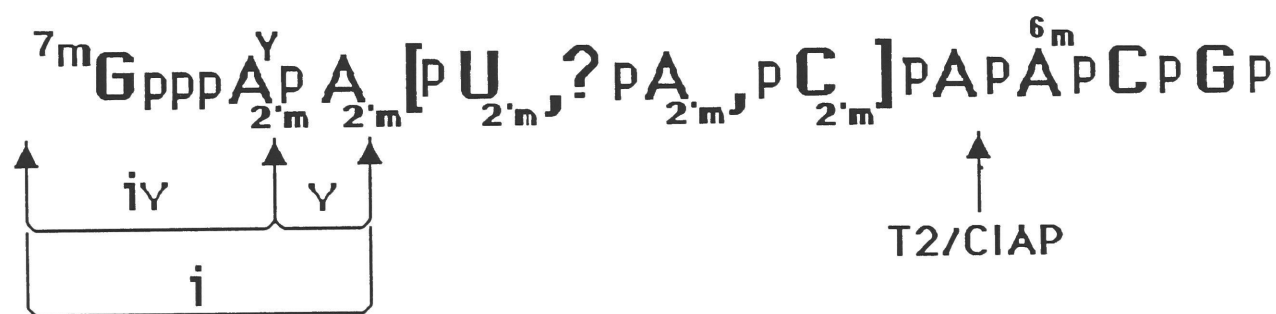


Figure 30**Spot ii Contains spot iii and a 5' pC***

Purified spot ii (from Fig. 23, Panel A, lane b) was redigested with nuclease P1 under standard conditions (Panel A, lane b) and analyzed on DEAE paper HVE, pH 1.9. The 5' ends of spots ii (Panel B) and iii (Panel C) were determined by the strategy schematized in Panel D. Each purified spot (lanes a) was treated with venom phosphodiesterase (VP, lanes b) to determine the base composition. In a separate reaction, the oligonucleotides were treated with phosphatase (CIAP) followed by VP (lanes d). The phosphatase alone reaction is shown in lanes c. Panels B and C are 540 paper HVE, pH 3.5. The origin (O) is at the bottom of the panels; XC signifies the position of the xylene cyanol dye marker; and the migration of mononucleotide markers is indicated.

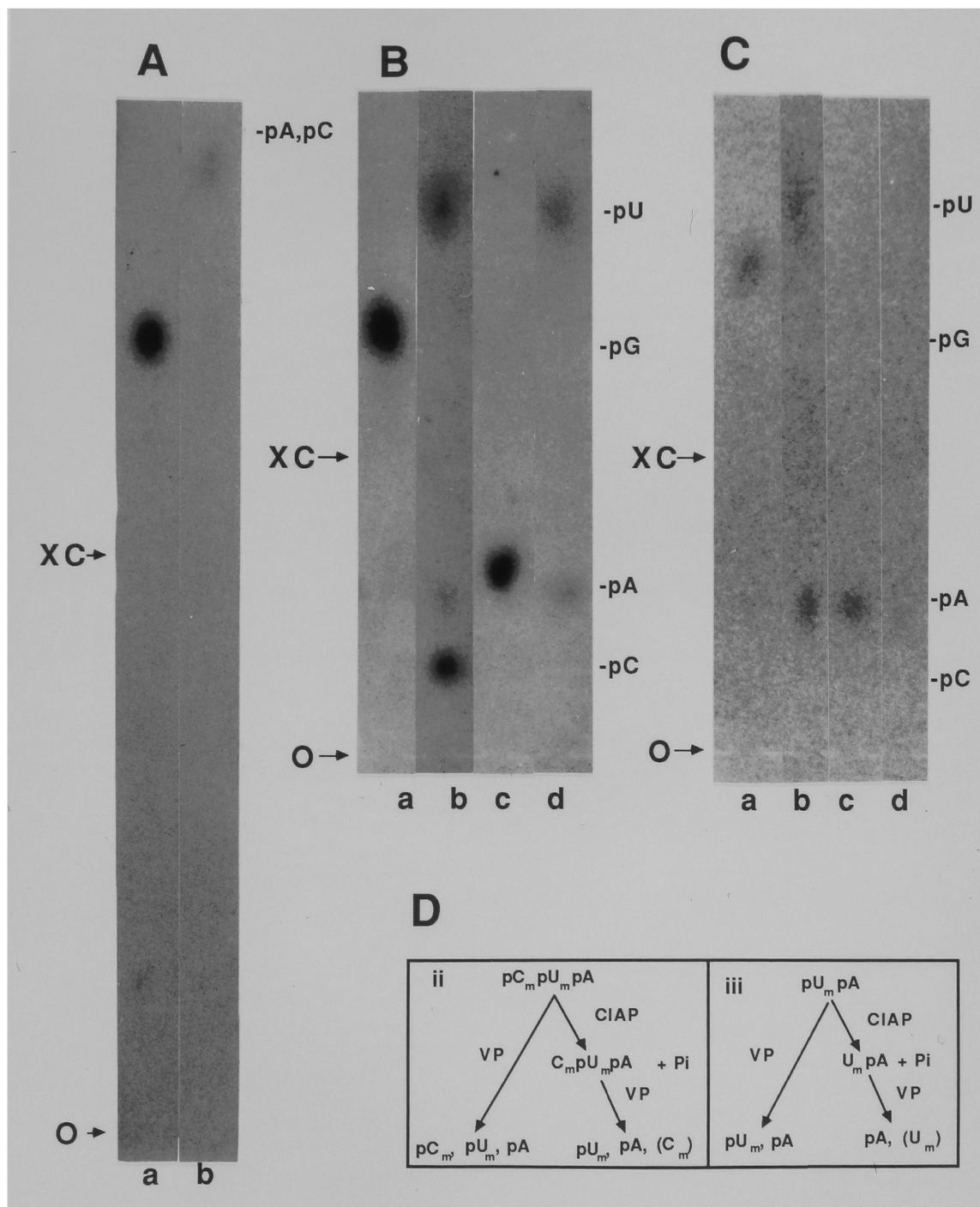


Figure 31
Partial Model of the Cap

Information about the nuclease P1 products, i-vi, yielded incomplete sequence information. Spot iii was shown to be contained within spot ii (see text). Spots i and ii were non-overlapping; however, it was not known if they were directly linked. In addition, the number of adenosine residues at the 3' end of the cap was still unknown. See also legend to Figure 26.

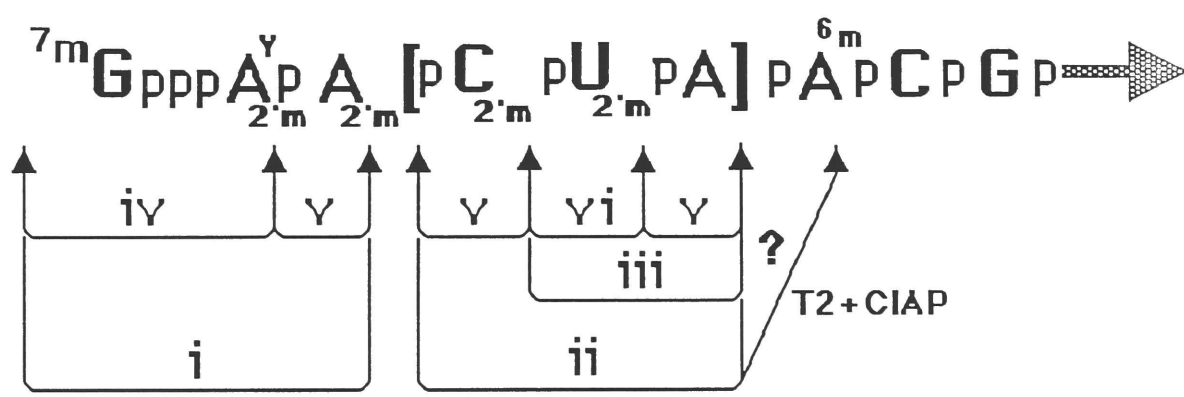


Figure 32
Venom Phosphodiesterase Partial Digestion Products of the Cap

The purified cap (Panel A, lane a) was digested with 2 μ g of venom phosphodiesterase in the presence of 20 μ g of tRNA in a 40 μ l volume. At the specified time-points, 5 μ l aliquots were removed and inactivated: 0 (lane b), 5 (lane c), 10 (lane d), 30 (lane e) 60 (lane f), 120 (lane g) and 180 (lane h) minutes. The products were fractionated on DEAE paper HVE, pH 1.9. The products are designated (in order of appearance) I, II, III and IV. The origin (O) is at the bottom of the panel; XC signifies the position of the xylene cyanol dye marker and the migration of mononucleotide markers is indicated. Panel B depicts the placement of the indicated partial products into the model of the cap structure.

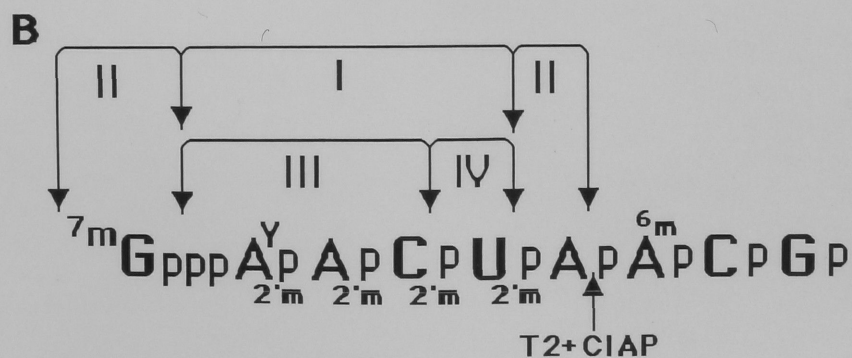
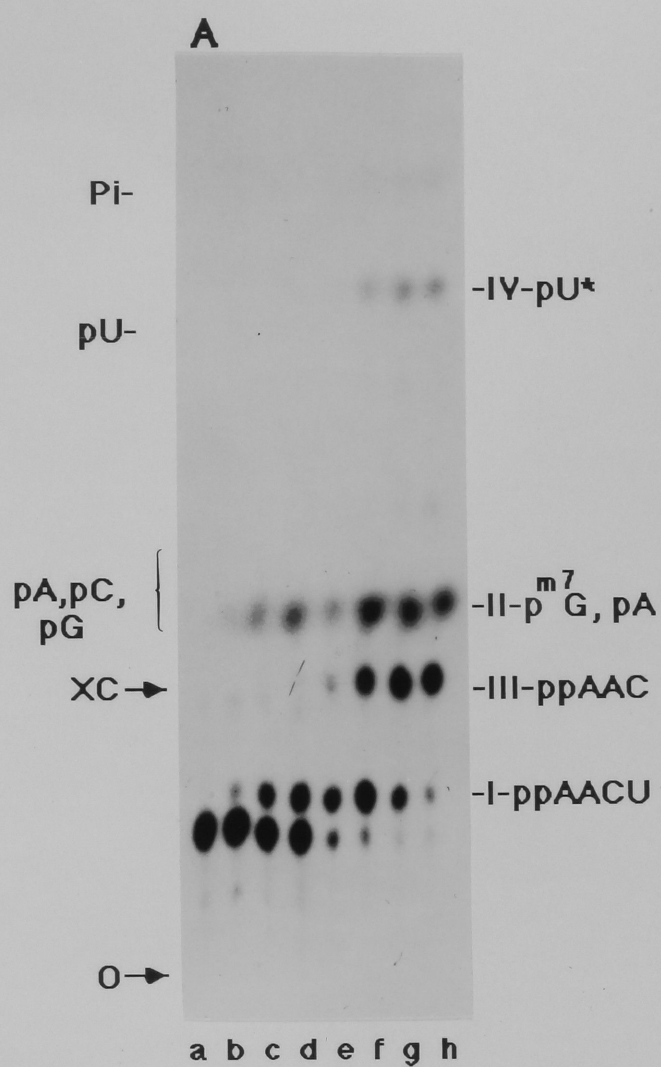


Figure 33
Analysis of the Venom Phosphodiesterase Partial Products

The venom phosphodiesterase products spot I (from the 5 minute time-point) (Panel A, lanes a and b), spot II (from the 10 minute time-point) (lane c) and spot III (from the 60 minute time-point) (lanes d and e) were either rechromatographed untreated (lanes a, c and d) or digested with nuclease P1 (lanes b and e) and fractionated on Whatman 540 paper HVE, pH 3.5. Panel B shows spot II analyzed on one-dimensional TLC. The origin (O) is at the bottom of the panels; XC signifies the position of the xylene cyanol dye marker; and the migration of mononucleotide markers is indicated.

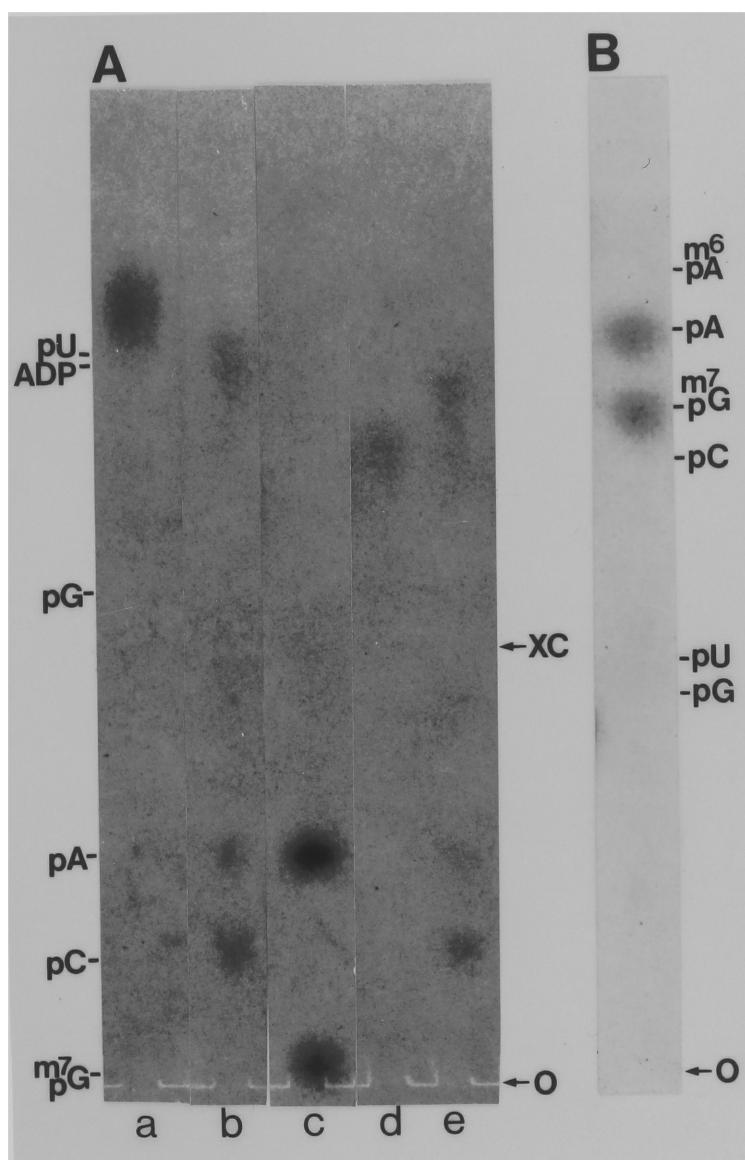


Figure 34
Model of the 5' End of *T. brucei* mRNA

The 5' end of the mini-exon, contained within RNase T1 spot 3, and the analytical digests used in the structural determination of the cap are indicated. Nuclease cleavages are designated by arrows. The cap structure extends from m⁷G to the first RNase T2 sensitive bond, as indicated. This bond was also sensitive to RNase U2. Venom phosphodiesterase products are indicated above the sequence; nuclease P1 products, below. Putative 2' O-methylations are marked below the phosphates (confirmed for the cytidine residue only); the detection of a modification, likely to be N⁶-methylation, indicated above the fourth adenosine, is described in Chapter III. "Y" and "Z" designate additional modifications that were detected but not fully characterized. Other analytical digests (Table III) used to sequence spot 3 are shown.

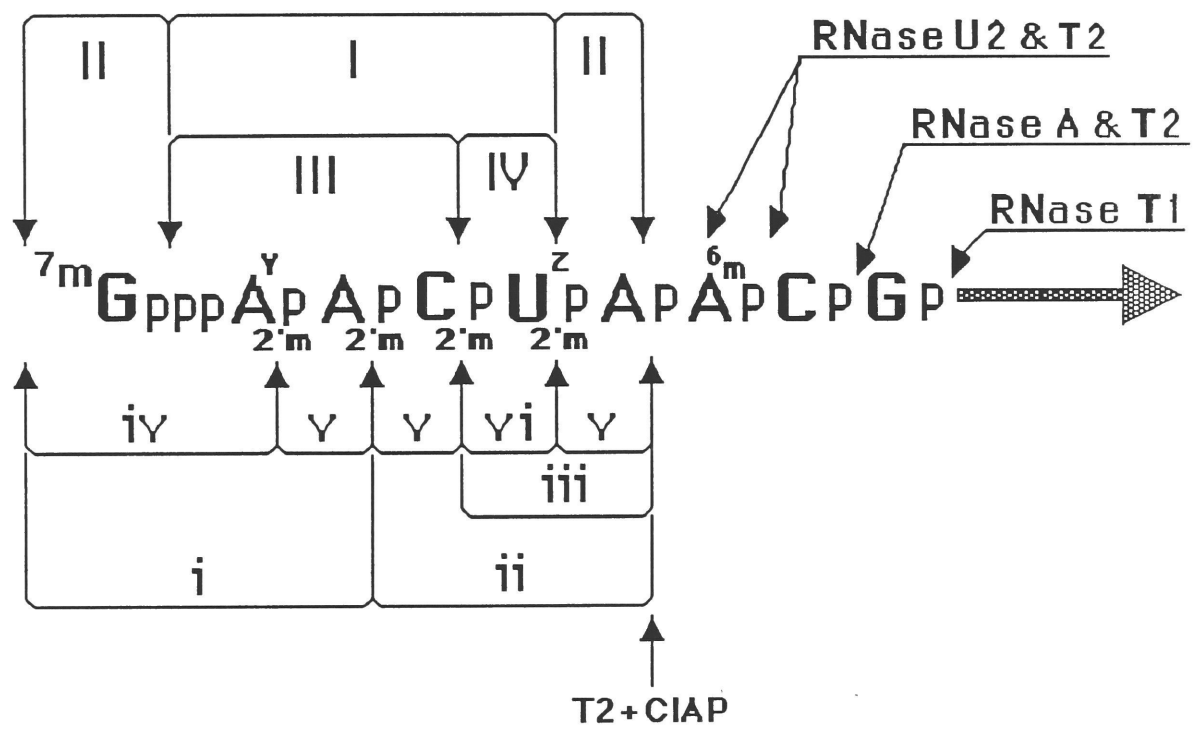


Table VI
Other Possible Cap Sequences

Possible sequences, considering the base composition, but excluding the 5' end, m⁷Gppp, for the cap are given in the first column. The reasons that each was eliminated are given in the second column. These reasons are based on the analysis of the partial digest products of nuclease P1 (small roman numerals; Figs. 22, 25 and 27) and venom phosphodiesterase (capital roman numerals; Figs. 28 and 29). For simplicity the indication of modifications is omitted in this table.

Table VI. Other Possible Cap Sequences

Possible Sequence	Reasons Against
<u>Extra A</u>	
AAACUA	<ol style="list-style-type: none"> 1) Mobility of spot i is too fast to be m7GpppAAA. 2) Mobility of spot III is too fast to be ppAAAC.
AACAUA	<ol style="list-style-type: none"> 1) Mobility of spot III is too fast to be ppAACA.
AAUACA	<ol style="list-style-type: none"> 1) Spot ii has a 5' terminal C and a U.
<u>Extra C</u>	
AACCUA	<ol style="list-style-type: none"> 1) Spot i contains no C. 2) Spot ii contains only one C. 3) Mobility of spot III is too fast to be ppAACC.
AACUCA	<ol style="list-style-type: none"> 1) Spot ii contains only one C and a U.
<u>Extra U</u>	
AAUCUA	<ol style="list-style-type: none"> 1) Spot i contains no U. 2) U was not released early in the nuclease P1 digests. 3) Spot III has no U.
AACUUA	<ol style="list-style-type: none"> 1) Spot III contains no U. 2) The mobility of spot III is too fast to be ppAAC. 3) The mobility of spot ii is too fast to be CUUA.

CHAPTER VII

DISCUSSION

This thesis presents experiments in which a direct analysis of *in vivo* ^{32}P -labeled RNA from *Trypanosoma brucei* was developed and utilized to analyze precursors, possible intermediates, products and a cap structure involved in discontinuous mRNA synthesis. Based on classical methods, this analysis will provide a standard for future studies, including *in vitro* work.

Eukaryotic parasites and the diseases they cause have been under-represented in the biomedical sciences because they occur in underdeveloped regions and because they are exceedingly complex diseases (Kuzoe, 1987). Unlike prokaryotic pathogens, these parasites have a genetic complexity that enables them to equal their hosts' complex survival strategies. Parasitologists have classically sought metabolic differences between host and parasite because these may be targets for specific chemotherapy. Paradoxically, several recent findings in parasites appear to be much more widely applicable than was initially recognized. Examples of this include the growth of telomeres (Bernards et al, 1981; Greider and Blackburn, 1987), bent DNA (Wells, 1988), the glycopospholipid membrane anchor (Ferguson and Williams, 1988) and discontinuous mRNA biosynthesis.

Before this work was initiated, the precursor in the trans-splicing reaction, medRNA, had been characterized by indirect methods (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984). The difficulties in phosphate labeling trypanosome RNA (Laird et al, 1987) were overcome in this work by a long pre-incubation in low phosphate medium. This direct analysis confirmed the existence of medRNA and yielded information concerning its homogeneity [despite clone to clone variation in the published DNA sequence literature (Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984)], a precise delineation of its 5' and 3' ends, the identification of an internal modification and the characterization of the cap structure. Having this analysis of medRNA gave a standard by which related RNAs in the same pathway could be analyzed.

RNAs related to medRNA were co-purified by the same scheme. This method would be expected to reveal intermediates in the mRNA biosynthetic pathway. An RNA slightly smaller than medRNA, called medRNA', was consistently present. Such an RNA has been detected by other workers (Kooter et al, 1984; Laird et al, 1985). Its functional significance is not clear. It could be an alternative transcription product, a premature termination product, an exonucleolytic degradation product or an intermediate in mRNA biosynthesis. A recent report (Laird et al, 1988) suggests that it may be a breakdown product.

Recent experiments suggest that an RNA corresponding to the 5' intran (that is, the 3' portion of medRNA) formed after metabolic debranching, should be detectable by these methods (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987). Such a molecule was not found in the course of the work presented here. Since this RNA is likely to be uncapped, it would be expected to be much less stable than the RNAs that were studied. Therefore, this result is not surprising. The 5' intran could be isolated by enzymatic debranching of trypanosome mRNA.

Several RNAs in the size range that would be expected for a free mini-exon were studied in detail. The RNA that was initially studied was a truncated version of the free mini-exon: it lacked two nucleotides (U and G) from its 3' end. As in the case of medRNA', the functional significance of this molecule is not clear. It is possible that it is an intermediate in mRNA biosynthesis. In prokaryotic tRNA biosynthesis, a nucleotidyl transferase specifically adds CCA to the 3' terminus (Zhu and Deutscher, 1987) even though some tRNAs have the CCA already encoded, suggesting that the CCA sequence may be continually removed and repaired. This 3' terminus, due to its important function in amino acid charging, is probably exposed to many nucleases in the cell, and the nucleotidyl transferase likely repairs damage as necessary. It is possible that, during mRNA biosynthesis in trypanosomes, the juncture between the extran (the mini-exon) and the intran (the downstream sequences) is exposed to nucleases. There may be repair enzymes to salvage the mini-exon during this critical time. In this regard, a mononucleotidyl transferase activity was recently detected in trypanosome whole cell extracts (White et al, 1987). The transferase activity utilized all four nucleotides as substrates, but appeared to be more active with UTP.

A more promising candidate for a biosynthetic intermediate was the molecule that corresponded precisely to the free mini-exon. It was found to terminate with a 3' OH. Such a terminus is consistent with the proposed trans-splicing mechanism. However, this RNA could be formed by any of a number of events such as premature termination, mRNA breakdown, or dephosphorylation of a 3' phosphorylated mini-exon RNA. An RNA primer would also be expected to have a 3' OH terminus. Whether this RNA is an authentic intermediate can be tested by number of experiments. For example, a successful *in vitro* trans-splicing reaction may yield such an intermediate. Such a system might incorporate the labeled mini-exon RNA into high molecular weight RNA. Other possible experiments include an *in vivo* pulse-chase experiment or heat shock, since temperature shock can inhibit splicing (Kamigo and Lizardi, 1986; Lindquist, 1986).

Computer-generated secondary structure models of several medRNA species suggested that a conserved structure is maintained to facilitate the joining reaction. A model of U2 RNA binding during trans-splicing was presented based on data from several different laboratories.

It was of interest to study the mini-exon cap from mRNA. Since it is not proven that every trypanosome mRNA contains the mini-exon, it was important to isolate the mini-exon specific cap from mRNA. The identification of the capped oligonucleotide in the fingerprints of medRNA suggested a method. It was fortunate that the capped oligonucleotide, spot 3, was well separated from other oligonucleotides in RNase T1 fingerprints. The identical mobility of the pairs of spots 2 and 3 on the fingerprints and of their analytical digestion products demonstrated that they were mini-exon-specific spots.

The cap from medRNA was indistinguishable from the mini-exon-specific mRNA cap; in particular, with regard to modification. This implies that capping and modification occur prior to the joining of the mini-exon to mRNA and that the cap is not altered during the transfer reaction. Partial confirmation of this finding was reported by Sutton and Boothroyd (1988). Slightly different results have been reported by Perry et al (1987). They reported that the medRNA cap was less highly modified than the mRNA cap and that the medRNA cap was heterogeneous with regard to methylation. It is possible that differences in methods can explain these different results. The medRNA that they were studying was apparently one of eight detected medRNA species, although the data were not shown. In addition, some of the heterogeneity in their work may have been caused by ring opening of m⁷G.

The mini-exon-specific mRNA cap appeared to contain five nucleotides in addition to m⁷G. Because these nucleotides were isolated from an RNase T2-resistant structure, the first four are likely to be modified by 2' OH methylation. Analysis of the nucleotides on a chromatographic system that was sensitive to modification supported this interpretation; however, due to the lack of commercial nucleotide markers, the precise nature of the modification was determined for only one of the nucleotides (C_{2'}-O-m). Other modifications, such as base modifications, may also be present. Perry et al (1987) report that the first encoded nucleotide (pA*) may contain a novel modification. There is no straightforward way to sequence modified nucleotides. Chemical and enzymatic methods for both RNA and DNA require unmodified nucleotides. Series of modified nucleotides, as in rRNA and tRNA, are conventionally sequenced by partial digests with non-specific nucleases that can cleave modified bonds (Silberklang et al, 1979; Kuchino et al, 1987). The strategies used to obtain the sequence in this work were adapted from such methods.

The trypanosome cap structure has several novel structural aspects. Since previously reported caps have zero, one or two 2'-O-methylated nucleotides (Shatkin, 1976; Banerjee, 1980), the detection of a cap with four such modifications, a "cap 4," is unique. Since lower eukaryotes, such as yeast and *Dictyostelium* tend to have less modified caps (Banerjee, 1980; Dottin et al, 1976), an overmethylated cap in a protozoan is particularly interesting. This observation is consistent with the early divergence of the Kinetoplastida in eukaryotic evolution (Sogin et al, 1986). Furthermore, capped mRNAs in other organisms tend to be heterogeneous with regard to modification, whether bulk mRNA (Banerjee, 1980; Cory and Adams, 1975) or individual messages are examined (Lockard and RajBhandary, 1976; Marcu et al, 1978). The appearance of a single predominant RNase T2- and phosphatase-resistant cap, in both the total digests and in digests of fingerprint spots (independent of the artifactual form), indicates that the mini-exon cap is unusually homogeneous in its extent of methylation. The second modification of uridine, likely a base modification, may be a novel modification. Although the detailed structural experiments were carried out with RNA from cultured procyclic forms, similar results were obtained with RNA from bloodstream forms, suggesting that there are not developmental changes with regard to the cap.

The detection of four extra nucleotides (excluding m⁷G) at the 5' end of the mini-exon indicates that the mini-exon is 39, rather than 35, nucleotides long. These 39 nucleotides are transferred from medRNA to mRNA. The modifications are the probable reason that the four 5' terminal nucleotides were not detected previously. It is likely that the progress of reverse transcriptase was impeded by the modifications, during the original primer extension experiments of medRNA and of individual mRNAs (Hagenbüchle et al, 1978). During the S-1 mapping, stable hybrids probably did not form.

Because the sequence of the ribonucleotides in the cap (excluding m⁷G) corresponds to the genomic sequence, it is likely that the mini-exon initiates four nucleotides upstream of the previously mapped initiation site. Since the sequence of the four upstream nucleotides is invariant among kinetoplastid mini-exon genomic sequences (Fig. 35; Cook and Donelson, 1987), it is likely that these related protozoa have a similar, overmethylated cap structure.

It is likely that trypanosomes utilize the same precursor of methyl groups in RNA, S-adenosyl methionine (SAM), that other organisms use. The Kinetoplastidae have a unique glutathione reductase co-factor, trypanothione (Fairlamb et al, 1985). Because the trypanothione biosynthetic pathway is closely linked to that for polyamines, intracellular levels of SAM may be higher in trypanosomes than in other organisms. Thus, the overmethylated cap structure may be related to

a pathway that is unique to the Kinetoplastida. This interpretation is supported by the experimental finding (data not shown) that Neplanocin A, an analog of S-adenosyl homocysteine that is an inhibitor of RNA methylation, was toxic to trypanosomes, but did not alter the cap structure.

The high number of methyl groups in the cap suggests the methyl labeling may provide a way to preferentially label medRNA. Such experiments may permit direct kinetic studies of trans-splicing *in vivo*. Whether the free mini-exon detected in this thesis is a genuine intermediate may be testable using this method. In addition, methyl labeling will be necessary to positively identify the nucleotide modifications in the cap. Preliminary results indicate that medRNA labels efficiently with ^3H methyl-labeled methionine (data not shown).

It will be important to determine what constitutes the 5' ends of primary transcripts of structural genes, such as VSG or tubulin, since the mature 5' ends of their mRNAs contain the mini-exon. If no cap is found, it would be the first known example of an uncapped pool of eukaryotic mRNA precursors. In that case, mini-exon addition would be the mRNA capping reaction. If a normal cap 0, 1 or 2 is found, then mini-exon joining would include a cap exchange, a novel reaction. If only the mini-exon cap is found, then it is likely that the mini-exon is trans-spliced very rapidly, perhaps in a manner analogous to conventional capping. However, such a finding would also be consistent with the priming model for trypanosome mRNA biosynthesis. If every trypanosome mRNA contains a mini-exon, then no other cap should be present on mature mRNA. Although only one mini-exon-specific cap was detected, the possibility that other caps are present could not be excluded in this work.

It is possible that a cap is required for some facet of the mini-exon's metabolism. The cleavage, branch formation (Murphy et al, 1986; Sutton and Boothroyd, 1986; Laird et al, 1987) and ligation reactions that certainly contribute to the mini-exon joining process may depend on the presence of a 5' cap. A conventional cap may be sufficient or the more highly modified cap may be required for the unusual trypanosome mRNA biosynthetic pathway. The absolute conservation among the Kinetoplastidae of the four nucleotides upstream of the mini-exon suggests a functional role for the base modifications. These hypotheses may be testable after an *in vitro* systems are available. One could compare the relative efficiency of trans-splicing using authentic, conventional or uncapped substrates.

What is the function of the mini-exon? If it functions during the biosynthesis of mRNA, it may be a unique level at which mRNA abundance is regulated. For example, the 221 VSG gene is

apparently co-transcribed with many other genes, perhaps as many as 10, including ESAG (Johnson et al, 1987; Kooter et al, 1987; Cully et al, 1985). However, ESAG and VSG mRNAs are present at 700-fold different levels (Cully et al, 1985). Since these transcripts are likely synthesized in an equimolar fashion (Johnson et al, 1987; Kooter et al, 1987), post-transcriptional mechanisms may regulate their differential abundance. Although there do not appear to be gross developmental changes in mini-exon usage, it is possible that the mini-exon joining is a regulatory event in mRNA biosynthesis. An important issue is whether every transcript covering a coding region receives a mini-exon. If not, for instance, if only one mRNA is made from every polycistronic transcript, as is the case for adenovirus (Ziff, 1980), then the choice of which gene to use would be a regulatory event. If every coding region is used, then other mechanisms, such as message stability or polyadenylation, would regulate differential abundance.

The mini-exon and its cap may also have a role in translation. The trypanosome translation machinery may have evolved to utilize mini-exon-containing mRNAs exclusively. The ability to test the role of the cap in translation has been deterred by the lack of a trypanosome cell-free translation system. Trypanosome ribosomes are unique in that they have nine rRNAs (Cross, 1970; Cordingley and Turner, 1980; Campbell et al, 1987). The presence of the mini-exon does not prevent translation in heterologous systems, although this question has not been studied quantitatively (Hoeijmakers et al, 1980a). The trypanosome analog of the Cap Binding Protein may have very high affinity for this highly methylated cap. Studies of such proteins may be feasible in trypanosomes.

The mini-exon pathway is a potential target for specific chemotherapy against trypanosomes and other Kinetoplastidae. If, for example, the methylation of the mini-exon cap is necessary for the trans-splicing reaction, specific inhibitors of RNA methylation, such as sinefungin or S-tubercidinylhomocysteine may inhibit the joining reaction, and therefore mRNA biosynthesis. The trypanosome methylases may be much more critical to the trypanosome's existence than its host's enzymes are to its survival. It may also be useful to look for inhibitors of trans-splicing: the mechanisms may be quite similar to cis-splicing, but the relative dependence of the trypanosome on a single mechanism may present a point of vulnerability.

It will be of interest to compare trypanosome RNA methylases to their eukaryotic counterparts. Specific enzymes are responsible for each of the methylations in other caps (Mizumoto and Kaziro, 1987). One issue will be whether specific or nonspecific methylases are responsible for these reactions in the mini-exon cap. Although medRNA has not been localized by cell fractionation, by comparison to other systems, it is likely to be a nuclear RNA because the joining

to mRNA is likely to be a nuclear event. Thus it is likely that the methylases are nuclear. Enzymes responsible for converting cap I into cap II structures in other eukaryotes should be cytoplasmic (Perry and Kelley, 1976; Banerjee, 1980). However, if medRNA spends part of its time in the cytoplasm, it is possible that some of the methylases are cytoplasmic. The trypanosome enzymes may provide a way to synthesize an authentic capped medRNA *in vitro*.

The presence of discontinuous RNA synthesis in plants, protozoa, nematodes and viruses, as well as the demonstration of trans-splicing with mammalian *in vitro* extracts (Fig. 2), suggests that similar pathways will be found with increasingly widespread distribution. Such findings invite evolutionary comparisons. Many current models of biotic evolution include a step in which unlinked exons, or exons, are joined to each other. The development of complex structures from simple ones requires a joining of pre-existing units. Because trans-splicing appears in different kingdoms, as well as in an endosymbiotic organelle, I would speculate that trans-splicing preceded cis-splicing, and that its presence today is a retention of an ancient mechanism. The recurrence of similar structures, such as the 3' OH end group of the extran supports the interpretation that similar trans-esterification mechanisms underlie cis- and trans-splicing. Caps are likely to be a eukaryotic embellishment. Because trypanosomes utilize discontinuous transcription so heavily, the cap structure may have co-evolved into a necessary aspect of mRNA biosynthesis in these protozoa. The fact that caps in other protozoa are less modified (Dottin et al, 1976) may reflect the early divergence of the Kinetoplastida.

Recent work (Krause and Hirsh, 1987) has shown that the nematode *C. elegans* has mRNAs both with and without a mini-exon-like sequence. It will be important to determine the cap structure of the *C. elegans* mini-exon. If a similar, over-methylated cap structure is found, its association with discontinuous mRNA synthesis across great phylogenetic distances may imply a functional role for such a cap. Since not every *C. elegans* mRNA has a mini-exon, different caps may serve to identify RNAs destined to be cis- or trans-spliced.

A search for similar caps in other organisms may be a way to identify additional occurrences of trans-splicing. If many more cases are detected, a new definition of a gene may be required. One might consider non-coding primers or exons as portable promoters that become incorporated into the message. However, if more cases (in addition to the chloroplast) in which transcripts for portions of coding regions are physically unlinked, the concept of a gene will not signify a distinct physical entity. It must be thought of as the sum of the parts that contribute to making an RNA.

The lack of suitable assay systems, either *in vivo* or *in vitro*, has prevented elucidation of mechanisms and regulatory signals in trypanosome mRNA biosynthesis. A detailed knowledge of the structure of medRNA, including its 5' end, may be necessary before such assay systems can be established. The direct analysis of trypanosome RNA presented in this work has led to the detection of several novel structures. Their functional significance may only become apparent when *in vitro* assay systems are available.

Figure 35
Comparison of Mini-exon Sequences

Mini-exon genomic DNA sequences from *T. brucei* ("Tbru", Campbell et al, 1984a; Kooter et al, 1984; Milhausen et al, 1984; Dorfman and Donelson, 1984), *Leptomonas seymori* ("Lsey", V. Bellofatto, unpublished observations), *Leishmania enrietti* ("Len", Miller et al, 1986), *Crithidia fasciculata* ("Cfas", Muhich et al, 1987) *Leptomonas collosoma* ("Lcol", Milhausen et al, 1984), *T. congolense* ("Tcong", Cook and Donelson, 1987), *T. cruzi* ("Tcruz", DeLange et al, 1984a) and *T. vivax* ("Tviv", DeLange et al, 1984a) are aligned with respect to the putative start site of transcription of medRNA. The phylogenetic relationships of these Kinetoplastidae are indicated in Figure 3. The position of the mini-exon (the first 39 nucleotides, except for *L. collosoma*) from these different sequences is indicated by the numbering above the sequences. The four nucleotides upstream (AACT) of the previously mapped site (AACG) are invariant.

Comparison of Mini-exon Sequences

	123456789	012345678901234567890123456789
Tbru	AACTAACGC	TATTATTAGAACAGTTTCTGTACTATATTGGTATGAGAAGCTC
Lsey	AACTAACGC	TATATAAGTATCAGTTTCTGTACTTTATTGGTATGAGAAGCTT
Len	AACTAACGC	TATATAAGTATCAGTTTCTGTACTTTATTGGTATGCGAAACCT
Cfas	AACTAACGC	TATATAAGTATCAGTTTCTGTACTTTATTGGTATAAGAAGCTT
Lcol	AACTAAAACA	TTTTTGAAGAACAGTTTCTGTACTTCATTGGTATGTAGAGACT
Tcong	AACTAAAGC	TTATAATAGAACAGTTTCTGTACTATATTGGTATGAGAAGCTC
Tgamb	AACTAACGC	TATTATTAGAATAGTTTCTGTACTATATTGGTATGAGAAGCTC
Tcruz	AACTAACGC	TATTATTGATACAGTTTCTGTACTATATTGGT ACGCGAAGCT
Tviv	AACTAAAGC	TTTTATTAGAACAGTTTCTGTACTATATTGGTATGAGAAGCTC

CHAPTER VIII

REFERENCES

- Abelson, J. (1979) *Ann. Rev. Biochem.* Vol. 48, 1035-1069.
- Aksoy, S., Lalor, T.M., Martin, J., Van der Ploeg, L.H.T. and Richards, F.R. (1987) *EMBO J.*, 6, 3819-3826.
- Arenas, J. and Hurwitz, J. (1987) *J. Biol. Chem.* 262, 4274-4279.
- Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412.
- Baldwin, A.S. and Sharp, P.A. (1988) *Proc. Natl. Acad. Sci. USA* 85, 723-727.
- Banerjee, A.K. (1980) in *Microbiological Reviews*, Vol. 44, pp. 175-205.
- Barrell, B.G. (1971) In *Procedures in Nucleic Acids Research*, Cantoni, G.L. and Davies, D.R., eds, Harper and Row, N.Y., Vol. 2, pp. 751-779.
- Beadle, G.W. and Tatum, E.L. (1941) *Proc. Natl. Acad. Sci. USA* 27, 499-506.
- Benzer, S. (1959) *Proc. Natl. Acad. Sci. USA* 45, 1607-1620.
- Bentley, D.L. and Groudine, M. (1986) *Nature* 321, 702-706.
- Berget, S.M., Moore, C. and Sharp, P.A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3171-3175.
- Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732.
- Bernards, A., Van der Ploeg, L.H.T., Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) *Cell* 27, 497-505.
- Bernards, A., Michels, P.A.M., Lincke, C.R. and Borst, P. (1983) *Nature* 303, 592-597.
- Bertholet, C., Van Meir, E., ten Heggeler-Bordie, B. and Wittek, R. (1987) *Cell* 50, 153-162.
- Birchmeier, C., Grosschedl, R. and Birnstiel, M.L. (1982) *Cell* 28, 739-745.
- Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
- Boothroyd, J.C. (1985) *Ann. Rev. Microb.* Vol. 39, pp. 475-502.
- Boothroyd, J.C. and Cross, G.A.M. (1982) *Gene* 20, 281-289.
- Borst, P. (1986) *Ann. Rev. Biochem.* Vol. 55, pp. 701-732.
- Borst, P. and Cross, G.A.M. (1982) *Cell* 29, 291-303.
- Bouloy, M., Morgan, M.A., Shatkin, A.J. and Krug, R.M. (1979) *J. Virol.* 32, 895-904.
- Bouloy, M., Plotch, S.J. and Krug, R.M. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3952-3956.
- Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* Vol. 50, 349-383.
- Breitbart, R.E., Nguyen, H.T., Medford, R.M., Destree, A.T., Mahdavi, V. and Nadal-Ginard, B. (1985) *Cell* 41, 67-82.
- Britton, R.J. and Kohne, D.E. (1968) *Science* 161, 529-540.
- Britton, R.J. and Davidson, E.H. (1969) *Science* 165, 349-357.
- Brown, D.D. and Dawid, I.B. (1968) *Science* 160, 272-280.
- Brun, R. and Schönenberger, M. (1979) *Acta Tropica* 36, 289-292.

- Brun, R. and Schönenberger, M. (1981) *Z. Parasitenkd.* 66, 17-24.
- Brownlee, G.G. (1972) *Determination of Sequences in RNA*, North Holland Publishing Co. Amsterdam.
- Budzilowicz, C.J., Wilczynski, S.P. and Weiss, S. (1985) *J. Virol.* 53,834-840.
- Campbell, D.A., Kubo, K., Clark, G.C. and Boothroyd, J.C. (1987) *J. Mol. Biol.* 196,113-124.
- Campbell, D.A., Thornton, D.A. and Boothroyd, J.C. (1984a) *Nature* 311, 350-355.
- Campbell, D.A., van Bree, M.P. and Boothroyd, J.C. (1984b) *Nucl. Acids Res.* 12, 2759-2774.
- Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) *Cell* 12,1-8.
- Chu, F.K., Maley, G.F., Maley, F. and Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* 81,3049-3053.
- Cech, T.R. and Bass, B.L. (1986) *Ann. Rev. Biochem.* 55,599-629.
- Clayton, C. (1985) *EMBO J.* 4, 2997-3003.
- Cordingley, J.S. and Turner, M.J. (1980) *Mol. Biochem. Parasit.* 1,91-96.
- Cook, G.A. and Donelson, J.E. (1987) *Mol. Biochem. Parasit.* 25,113-122.
- Cornelissen, A.W.C.A., Verspielen, M.P., Toulme, J.-J., Swinkels, B.W. and Borst, P. (1986) *Nucl. Acids Res.* 12, 5605-5614.
- Cory, S. and Adams, J.M. (1975) *J. Mol. Biol.* 99,519-547.
- Cross, G.A.M. (1970) *Biochem Biol. Acta* (1970) 204,470-477.
- Cross, G.A.M. (1975) *Parasit.* 71,393-417.
- Cross, G.A.M. (1978) *Proc. Roy Soc. London Ser B* 202,55-72.
- Cross, G.A.M. (1984) *Phil. Trans. R. Soc. Lond. B* 307,3-12.
- Cully, D.F., Ip, H.S. and Cross, G.A.M. (1985) *Cell* 42,173-182.
- Curtis, P.J., Mantei, N. and Weissman, C. (1978) In *Cold Spring Symposia on Quantitative Biology*. Vol. 42, Pt.2, pp.971-984.
- Darnell, J., Lodish, H. and Baltimore, D. (1986) *Molecular Cell Biology*. Scientific American Books, N.Y.
- DeLange, T., Berkrens, T.M., Veerman, H.J.G., Carlos, A., Frasch, C., Barry, J.D. and Borst, P. (1984a) *Nucl. Acids Res.* 12, 4431-4443
- DeLange, T., Liu, A.Y.C., Van der Ploeg, L.G.T. and Borst, P. (1983) *Cell* 34, 891-900.
- DeLange, T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, A.W.C.A. and Borst, P. (1984b) *Nucl. Acids Res.* 12,3777-3790.
- DeWachter, R. and Fiers, W. (1971) In *Methods in Enzymology*, Vol.21, pp.167-178.
- Dorfman, D.M. and Donelson, J.E. (1984) *Nucl. Acids Res.* 12, 4907-4920.
- Dorfman, D.M., Lenado, M.J., Reddy, L.V., Van der Ploeg, L.H.T. and Donelson, J.E. (1985) *Nucl. Acids Res.* 13,3533-3549.
- Donelson, J.E. and Rice-Ficht, A.C. (1985) In *Microbiological Reviews*, Vol. 49, 107-125.
- Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucl. Acids Res.* 8,2527-2538.

- Dottin, R.P., Weiner, A.M. and Lodish, H.F. (1976) *Cell* 8,233-244.
- Dunn, J.J. and Studier, F.W. (1973) *Proc. Natl. Acad. Sci. USA* 70,3296-3300.
- Earnshaw, D.L., Beebee, T.J.C. and Gutteridge, W.C. (1987) *Biochem. J.* 241,649-655.
- Edery, I. and Sonenbery, N. (1985) *Proc. Natl. Acad. Sci. USA* 82,7590-7594.
- Ehlers, B., Czichos, J., and Overath, P. (1987) *Mol. Cell. Biol* 7, 1242-1249.
- Engelke, D.R., Ng, S.-Y., Shastri, B.S. and Roeder, R.G. (1980) *Cell* 19,717-728.
- Englund, P.T., Hajduk, S.L. and Marini, J.C. (1982) *Ann. Rev. Biochem.* Vol. 51, 695-726.
- Feagin, J.E. and Stuart, K. (1988) *Mol. Cell. Biol.* 8,1259-1265.
- Filipowicz, W. and Shatkin, A.J. (1983) *Cell* 32,547-557.
- Fairlamb, A.H., Blackburn, P., Ulrich, P., Chait, B.T. and Cerami, A. (1985) *Science* 227,1485-1487.
- Ferguson, M.J. and Williams, A.F. (1988) *Ann. Rev. Biochem.* In Press.
- Franklin, R.M. (1966) *Proc. Natl. Acad. Sci. USA* 55,1504-1511.
- Freistadt, M.S., Cross, G.A.M., Branch, A.D. and Robertson, H.D. (1987) *Nucl. Acids Res.* 15,9861-9879.
- Freistadt, M.S., Cross, G.A.M. and Robertson, H.D. (1988) *J. Biol. Chem.*, submitted.
- Fujimoto, M., Fujiyama, K., Kuninaka, A. and Yoshino, H. (1974) *Agr. Biol. Chem.* 38,2141-2147.
- Furuichi, Y., Muthukrishnan, S. and Shatkin, A.J. (1975a) *Proc. Natl. Acad. Sci. USA* 72,363-366.
- Furuichi, Y., Morgan, M., Muthukrishnan, S. and Shatkin, A.J. (1975b) *Proc. Natl. Acad. Sci. USA* 72,742-745.
- Furuichi, Y., LaFiandra, A. and Shatkin, A.J. (1977) *Nature* 266,235-239.
- Gibson, W.C. and Miles, M.A. (1986) *EMBO J.* 5,1299-1305.
- Gall, J.G. (1981) *J. Cell Biol.* 91,3s-14s.
- Guyaux, M., Cornelissen, A.W.C.A., Pays, E., Steinert, M. and Borst, P. (1985) *EMBO J.* 4, 995-998.
- Glass, D.J., Polvere, R.I. and Van der Ploeg, L.H.T. (1986) *Mol. Cell. Biol.* 6,4657-4666.
- Gonzales, A., Lerner, T.J., Huecas, M., Sosa-Pineda, B., Nogueira, N. and Lizardi, P. (1985) *Nucl. Acids Res.* 13, 5789-5804.
- Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.* 2,1044-1051.
- Green, M.R., Maniatis, T. and Melton, D.A. (1983) *Cell* 32, 681-694.
- Greer, C.L., Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983) *Cell* 32, 537-546.
- Greider, C.W. and Blackburn, E.H. (1987) *Cell* 51,887-898.
- Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978). *Cell* 13, 551-563.
- Hart, R.P., McDevitt, M.A. and Nevins, J.R. (1985) *Cell* 43, 677-683.
- Hasan, G., Turner, M.J. and Cordingley, J.S. (1984) *Cell* 37, 333-341.

- Hay, N., Skolnik-David, H. and Aloni, Y. (1982) *Cell* 29,183-193.
- Hensgens, L.A.M., Brakenhoff, J., DeVries, B.F., Sloof, P. and Tromp, M. (1984) *Nucl. Acids Res.* 12,7327-7344.
- Hewlett, M.J., Rose, J.K. and Baltimore, D. (1976) *Proc. Natl. Acad. Sci. USA* 73,327-330.
- Hickey, E.D., Weber, L.A., Baglioni, C., Kim, C.H. and Sarma, R.H. (1977) *J. Mol. Biol.* 109,173-183.
- Hildebrand, M., Hallick, R.B., Passavant, C.W. and Bourque, D.P. (1988) *Proc. Natl. Acad. Sci. USA* 85,372-376.
- Hoeijmakers, J.H.J., Borst, P., van den Berg, J., Weissman, C. and Cross, G.A.M. (1980a) *Gene* 8, 391-417.
- Hoeijmakers, J.H.J., Frasch, A.C.C., Bernards, A., Borst, P. and Cross, G.A.M. (1980b) *Nature* 284, 78-80.
- Hofstetter, H., Kressman, A. and Birnstiel, M.L. (1981) *Cell* 24,573-585.
- Inoue, T., Sullivan, F.X. and Cech, T.R. (1985) *Cell* 43,431-437.
- Imboden, M.A., Laird, P.W., Affolter, M., and Seebeck, T. (1987) *Nucl. Acids Res.* 15, 7357-7368.
- Jacob, F. and Monod, J. (1961) *J. Mol. Biol.* 3,318-356.
- Jacobs, L. Spaan, W.J.M., Horzinek, M.C. and Van der Zeijst, B.A.M. (1981) *J. Virol.* 39,401-406.
- Jacquier, A. and Rosbach, M.M. (1986) *Science* 234, 1099-1104.
- Jefferson, D.M., Clayton, D.F., Darnell, J.E. and Reid, L.M. (1984) *Mol. Cell. Biol.* 4,1929-1934.
- Jelinek, W.R., Toomey, T.P., Lerinwand, L., Duncan, C.H., Biro, P.A., Chudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) *Proc. Natl. Acad. Sci. USA* 77,1398-1402.
- Jelinek, W.R. and Schmid, C.W. (1982) *Ann. Rev. Biochem.* Vol. 51,813-844.
- Jove, R. and Manley, J.L. (1984) *J. Biol. Chem.* 259, 8513-8521.
- Johnson, P.J., Kooter, J.M. and Borst, P. (1987) *Cell* 51,273-281.
- Kamigo, S. and Lizardi, P.M. (1986) *Mol Biol. Rep.* 11,77-80.
- Karnik, P., Gapalakrishna, Y. and Sarkar, N. (1986) *Gene* 49,161-165.
- Khoury, G. and Gruss (1983) *Cell* 33,313-314.
- Kitchen, P.A., Tyley, J.G. and Gutteridge, W.E. (1984) *Comp. Biochem. Physiol.*, 77,223-231.
- Koller, B., Fromm, H., Galum, E. and Edelman, M. (1987) *Cell* 48,111-119.
- Konarska, M.M., Padgett, P.A. and Sharp, P.A. (1984) *Cell* 38, 731-736.
- Konarska, M.M., Padgett, P.A. and Sharp, P.A. (1985) *Cell* 42, 165-171.
- Kooter, J.M. and Borst, P. (1984) *Nucl Acids Res.* 12, 9457-9472.
- Kooter, J.M., DeLange, T. and Borst, P. (1984) *EMBO* 3, 2387-2392.

- Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J. and Borst, P.J. (1987) *Cell* 51,261-272.
- Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) *Cell* 36, 993-1005.
- Krause, M. and Hirsh, D. (1987) *Cell* 49, 753-761.
- Kuchino, Y., Hanyu, N. and Nishimura, S. (1987) In *Methods in Enzymology*, Vol. 155, 379-396.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) *Cell* 31,147-157.
- Kuzoe, F.A.S. (1987) In *Tropical Disease Research; A Global Partnership*, John Maurice and Anna Marina Pearce. World Health Organization, Geneva.
- Lai, M.M.C. (1986) *Bioessays* 5,257-260.
- Lai, M.M.C., Patton, C.D. and Stohlman, S.A. (1982) *J. Virol.* 41,557-565.
- Laird, C.D., Chooi, W.Y., Cohen, E.H., Dickson, E., Hutchinson, N., Turner, S.H.,(1974) In *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 38,311-327.
- Laird, P.W., Kooter, J.M., Loosbroek, N. and Borst, P. (1985) *Nucl. Acids Res.* 13, 4253-4266.
- Laird, P.W., Zomerijk, J.C.B.M., de Korte, D. and Borst, P. (1987) *EMBO J.* 6, 1055-1062.
- Laird, P.W., ten Asbroek, A.L.M.A. and Borst, P. (1988) *Nucl. Acids Res.* In press.
- Landfear, S.M. and Wirth, D.F. (1985) *Mol. B. Parasit.* 15, 61-82.
- Langford, C.J. and Gallwitz, D (1983) *Cell* 33,519-527.
- Laski, F.A., Fire, A.Z., RajBhandary, U.L. and Sharp, P.A. (1983) *J. Biol. Chem.* 258,11974-11980.
- Laskowski, M., Sr. (1971) in *The Enzymes* (Boyer, P.D., ed) Third Edition, Vol. 4, pp. 313-328, Academic Press, N.Y.
- Layden, R.E. and Eisen, H. (1988) *Mol. Cell. Biol.* 8,1352-1360.
- Lenardo, M.J., Dorfman, D.M. and Donelson, J.E. (1985a) *Mol. Cell. Biol.* 9,2487-2490.
- Lenardo, M.J., Dorfman, D.M., Reddy, L.K. and Donelson, J.E. (1985b) *Gene* 35,131-141.
- Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) in *Ann. Rev. Biochem* Vol. 55, 1091-1117
- Lewin, Benjamin (1980) *Gene Expression 2*. John Wiley and Sons, N.Y.
- Lindquist, S. (1986) *Ann. Rev. Biochem.* Vol. 55,1151-1191.
- Lockard, R.E. and RajBhandary, U.L. (1976) *Cell* 9, 747-760.
- Longacre, S. and Eisen, H. (1986) *EMBO J.* 5, 1057-1063.
- McClure, W.R. (1985) *Ann. Rev. Biochem.* Vol. 54,171-204.
- McKnight, S.L. and Kingsbury, R. (1982) *Science* 217,316-324.
- Makino, S., Stohlman, S.A., and Lai, M.M.C. (1986) *Proc. Natl. Acad. Sci. USA* 83, 4202-4208.
- Manley, J.L., Sharp, P.A. and Gefter, M.L. (1979) *J. Mol. Biol.* 135,171-197.
- Marcu, K.B., Schibler, U. and Perry, R.P. (1978) *J. Mol. Biol.* 120,381-400.
- Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74,560-564.

- Meinkoth, J. and Wahl, G. (1984) *Anal. Biochem.* 138,267-284.
- Mendel, G. (1865) *Ver Naturforshung Ver. Brunn* 4,3-47. Reprinted in J.A. Peters, 1959, *Classic papers in Genetics*, Prentice-Hall.
- Michel, F. and Dujon, B. (1983) *EMBO J.* 2,33-38.
- Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M. and Agabian, N. (1984) *Cell* 38,721-729.
- Miller, S.F., Landfear, S.M. and Wirth, D.F. (1986) *Nucl. Acids Res.* 14,7341-7360.
- Mizumoto, K. and Kaziro, Y. (1987) *In Progress in Nucleic Acids Research and Molecular Biology*, Vol. 34, 1-28.
- Montoya, J., Gaines, G.L. and Attardi, G. (1983) *Cell* 34,151-159.
- Morgan, T.H. (1910) *Science* 32,120-122.
- Moore, C. L., Skolnick-David, H. and Sharp, P.A. (1988) *Mol. Cell. Biol.* 8,226-233.
- Mowry, K.L. and Steitz, J.A. (1987) *Mol. Cell. Biol.* 7, 1662-1672.
- Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell* 47, 517-525.
- Murphy, N.B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. and Pays, E. (1987) *J. Mol. Biol.* 195,855-871.
- Muhich, M.L., Hughes, D.E., Simpson, A.M. and Simpson, L. (1987) *Nucl. Acids Res.* 15, 3141-3153.
- Muthukrishnan, S., Both, G.W., Furuichi, Y. and Shatkin, A.J. (1975) *Nature* 255,33-37.
- Nelson, R.G., Parsons, M., Barr, P.J., Stuart, K., Selkirk, M. and Agabian N. (1983a) *Cell* 34,901-909.
- Nelson, R.G., Parsons, M., Selkirk, M., Newport, G. Barr, P.J. and Agabian N. (1983b) *Nature* 308,665-667.
- Nabel, G. and Baltimore, D (1987) *Nature* 326,711-713.
- Nomoto, A. , Lee, Y.F. and Wimmer, E. (1976) *Proc. Natl. Acad. Sci. USA* 73,375-380.
- Ohno, M., Sakamoto, H. and Shimura, Y. (1987) *Proc. Natl. Acad. Sci. USA* 84,5187-5191.
- Opperdoes, F.R., Baudhuin, P., Coppens, I., DeRoe, C., Edwards, S.W., Weijers, P. and Missett, O. (1984) *J. Cell Biol.* 98,1178-1184.
- Osinga, K.A., Swinkels, B.W., Gibson, W.C., Borst, P., Venneman, G.H., Van Boom, J.H., Michels, P.A.M. and Opperdoes, F.R. (1985) *EMBO J.* 4,3811-3817.
- Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S.I., and Sharp, P. (1986) *Ann. Rev. of Biochem.* Vol 55. 1119-1150.
- Palmiter, R.D. and Brinster, R.L. (1985) *Cell* 41, 343-345.
- Parker, R., Siliciano, P.G. and Guthrie, C.M. (1987) *Cell* 49,229-239.
- Parsons, M., Nelson, R.G., Watkins, K.P. and Agabian, N. (1984) *Cell* 38, 309-316.
- Parsons, M., Nelson, R.G., and Agabian, N. (1986) *Nucl. Acids Res.* 14,1703-1718.
- Patterson, J.L., Holloway, B. and Kolakofsky, D. (1984) *J. Virol.* 52,215-222.

- Patzelt, E., Thalman, E., Hartmuth, K., Blaas, D. and Kuechler, E.M. (1987) Nucl. Acids Res. 15, 1387-1399.
- Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983) Cell 32,525-536.
- Peebles, C.L., Mecklenburk, K.L., Perlman, P.S., Tabor, J. and Cheng, H.L. (1986) Cell 44, 213-223.
- Perry, K.L., Watkins, K.P. and Agabian, N. (1987) Proc. Natl. Acad. Sci. USA 84, 8190-8194.
- Perry, R.P., Kelley, D.E., Frederici, K.H. and Rottman, F.M. (1975) Cell 6,13-19.
- Perry, R.P. and Kelley, D.E. (1976) Cell 8, 433-442.
- Plotch, S.J., Bouloy, M and Krug, R.M. (1979) Proc. Natl. Acad. Sci. USA 76,1618-1622.
- Pribnow, D. (1975) Proc. Natl. Acad. Sci. USA 72,784-789.
- Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263,211-214.
- Ptashne, M. (1986) A genetic Switch: Gene control and Phage lambda. Cell Press and Blackwell Scientific Publications.
- Radic, M.Z., Lundgren, K. and Hamkalo, B.A. (1987) Cell 50, 1101-1108.
- Ransohoff, R.M., Narayan, P., Ayers, D.F., Rottman, F.M. and Nilsen, T.W. (1987) Antiviral Res. 7, 317-327.
- Reddy, R, Ro-Choi, T.S., Henning, D. and Busch, H. (1974) J. Biol. Chem. 249,6486-6494.
- Reddy, R, Ro-Choi, T.S., Henning, D., Shibata, H., Choi, Y.C. and Busch, H. (1972) J. Biol. Chem. 247,7245-7250.
- Reeder, R.H. (1984) Cell 38,349-351.
- Robertson, H.D. and Dickson, E. (1974) In Processing of RNA, Brookhaven Symposia in Biology; No. 26.,pp.240-266.
- Roberson, H.D., Dickson, E. and Jelinek, W. (1977) J. Mol. Biol. 115, 571-589.
- Robertson, H.D., Dickson, E., Plotch, S.J. and Krug, R.M. (1980) Nucl. Acids Res. 8,925-942.
- Roeder, R.G., Golomb, M.W., Joehning, J.A., Ng, S.Y., Parker, C.S., Schwartz, L.B., Sklar, V.E.R. and Weinmann, R. (1977) in Receptors and Hormone Action, O'Malley, B.W. and Birnbaumer, L., eds, Academic Press, Vol. I, 195-234.
- Roth, C.W., Longacre, S., Raibaud, A., Baltz, T. and Eisen, H. (1986) EMBO J. 5,1065-1070.
- Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980) Cell 19, 13-25.
- Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J.E., Jr. (1980) Cell 19,69-78.
- Sanger, F. (1981) Science 214,1205-1210.
- Sather, S. and Agabian, N. (1985) Proc. Natl. Acad. Sci. USA 82, 5695-5699.
- Schmidt, G.C. and Roberts, L.S. (1985) Foundations of Parasitology. College Publishing. St. Louis. pp.50-55.
- Schnare, M.N., Spencer, D.F. and Gray, M.W. (1983) Can. J. Biochem. 61,38-45.
- Schwer, B., Visca, P., Vos, J.C. and Stunnenberg, H.G. (1987) Cell 50,163-169.

- Shatkin, A.J. (1976) *Cell* 9,645-653.
- Shatkin, A.J. (1985) *Cell* 40,223-224.
- Sharp, P.A. (1985) *Cell* 42,397-400.
- Sharp, P.A. (1987) *Cell* 50,147-148.
- Shampay, J. Szostak, J.W. and Blackburn, E.H. (1984) *Nature* 310,154-157.
- Shea, C.I., Lee, M. G.-S. and Van der Ploeg, L.H.T. (1987) *Cell* 50,603-612.
- Shimotohno, K., Kodama, Y., Hashimoto, J. and Miura, K.-I. (1977) *Proc. Natl. Acad. Sci. USA* 74,2734-2738.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) *EMBO J.* 5, 2043-2049.
- Shinshi, H., Miwa, M., Kato, K., Noguchi, M., Matsushima, T. and Sugimura, T. (1976) *Biochemistry* 15,2185-2190.
- Sogin, M.L. Elwood, H.J. and Gunderson, J.H. (1986) *Proc. Natl. Acad. Sci. USA* 83,1383-1387.
- deSouza, W. (1984) *Intern. Rev.Cyt.* 86,197-283.
- Silberklang, M., Gillum, A.M. and RajBhandary, U.L. (1979) In *Methods in Enzymology*, Vol. LIX, pp. 58-109.
- Sollner-Webb, B. (1988) *Cell* 52,153-154.
- Solnick, D. (1985) *Cell* 42,157-164.
- Sonnenberg, N., Morgan, M.A., Merrick, W.C. and Shatkin, A.J. (1978) *Proc. Natl. Acad. Sci. USA* 75,4843-4847.
- Steitz, J.A. (1986) *The Harvey Lectures*, Ser 80, 39-47.
- Sutton, R.E. and Boothroyd, J.C. (1986) *Cell* 47, 527-535.
- Sutton, R.E. and Boothroyd, J.C. (1988a) *Mol. Cell. Biol.* 8, 494-496.
- Sutton, R.E. and Boothroyd, J.C. (1988b) *EMBO J.*, In press.
- Swinkels, B.W., Gibson, W.C., Osinga, K.A., Kramer, R., Veeneman, G.H., Van Boom, J.H. and Borst, P. (1986) *EMBO J.* 5,1291-1298.
- Taylor, J.M., Illmensee, R. and Summers, J. (1976) *Biochim. Biophys. Acta* 442,324-330.
- Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77,5201-5205.
- Tittawella, I. (1988) *FEBS Letters* 227,122-126.
- Trager, W. (1986) *Living Together: The Biology of Animal Parasitism*. Plenum Press, N.Y.
- Tschudi, C., Richards, F.F. and Ullu, E. (1986) *Nucl. Acids Res.* 14,8893-8903.
- Tschudi, C. and Ullu, E. (1988) *EMBO J.* 7,455-463.
- Tschudi, C., Young, A.S., Ruben, L., Patton, C.L., and Richards, F. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3998-4002.

- Uchida, T. and Egami, F (1971) In *The Enzymes*, Boyer, P.D., ed, Vol. IV, pp.205-250.
- Van der Ploeg, L.H.T., Cornelissen, A.W.C.A, Barry, J.D. and Borst, P. (1984a) *EMBO J.* 3,3109-3115.
- Van der Ploeg, L.H.T., Cornelissen, A.W.C.A, Michels, P.A.M. and Borst, P. (1984b) *Cell* 39,213-221.
- Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., DeLange, T., Borst, P., Majumder, H.K., Weber, H., Veenemna, G.H. and Van Boom, J. (1982) *Nucl. Acids Res.* 10,3591-3604.
- Van der Ploeg, L.H.T., Schwartz, D.C., Cantor, C.R. and Borst, P. (1984c) *Cell* 37,77-84.
- Vickerman, K. and Barry, J.D. (1984) In *Immunology of Parasitic Infections*, Cohen, S. and Warren, eds, K.S. Blackwell Scientific Publications, Oxford.
- Walder, J.A., Eder, P.S., Engman, D.M., Brenano, S.T., Walder, R.Y., Knutzon, D.S., Dorfman, D.M. and Donelson, J.E. (1986) *Science* 233,569-571.
- Walter, P. and Blobel, G. (1982) *Nature* 299,691-698.
- Wallace, J.C. and Edmonds, M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 950-954.
- Waring, R.B. and Davies, R.W. (1984) *Gene* 28,277-291.
- Wells, R.D. (1988) *J. Biol. Chem.* 263,1095-1098.
- White, T.C., Rudenko, G., and Borst, P. (1986) *Nucl. Acids Res.* 12,4907-4920.
- White, T.C. and Borst, P. (1987) *Nucl. Acids Res.* 15,3275-3290.
- Wigler, M., Pellicer, A., Silverstein, S. and Axel, R. (1978) *Cell* 14,725-731.
- Wu, C. (1984) *Nature* 309,229-234.
- Yamada, Y. and Ishikura, H. (1975) *Biochem. Biophys. Acta* 402,285-287.
- Yanofsky, C. (1981) *Nature* 289,751-758.
- Zaita, N., Torazawa, K., Shinozaki, K. and Sugiura, M. (1987) *FEBS Letters* 210,153-156.
- Zaug, A.J. and Cech, T.R. (1985) *Science* 229,1060-1064.
- Zaug, A.J. and Cech, T.R. (1986) *Science* 231,470-475.
- Zeitlin, S. and Efstradiadis, A. (1984) *Cell* 39,589-602.
- Zhu, L. and Deutcher, M.P. (1987) *EMBO J.* 6,2473-2477.
- Ziff, E.B. (1980) *Nature* 287,491-499.
- Ziff, E.B. and Evans, R.M. (1978). *Cell* 15,1463-1475.
- Zinn, K., DiMaio, D. and Maniatis, T. (1983) *Cell* 34,865-879.
- Zuker, M. and Stiegler, P. (1981) *Nucl. Acids Res.* 9,133-148.

End