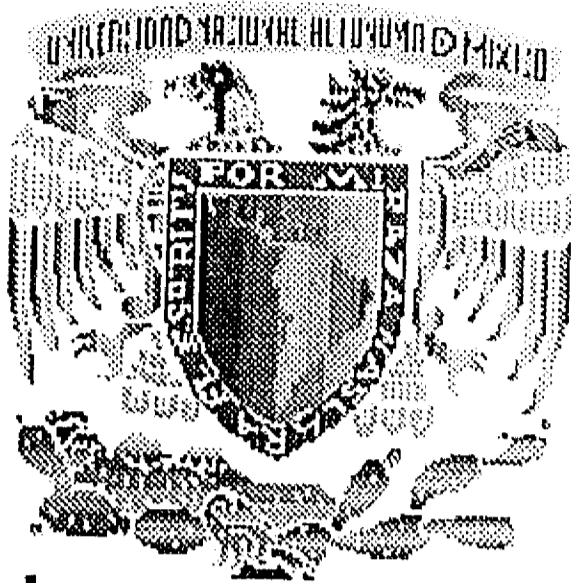


03088

2
24



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONALES
Y DE POSGRADO DEL COLEGIO DE CIENCIAS Y
HUMANIDADES

Identificación y análisis de la region promotora de los
genes de RNA ribosomal de *Entamoeba histolytica*

T E S I S

que para obtener el título de
Doctorado en Biotecnología

P R E S E N T A

Bertha Paula | Michel González

Cuernavaca, Mor.

1996

TESIS CON
FALLA DE ORIGEN

**TESIS CON
FALLA DE ORIGEN**



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

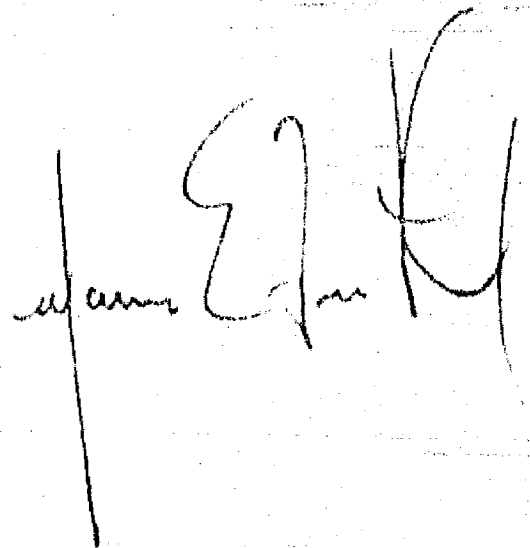
DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

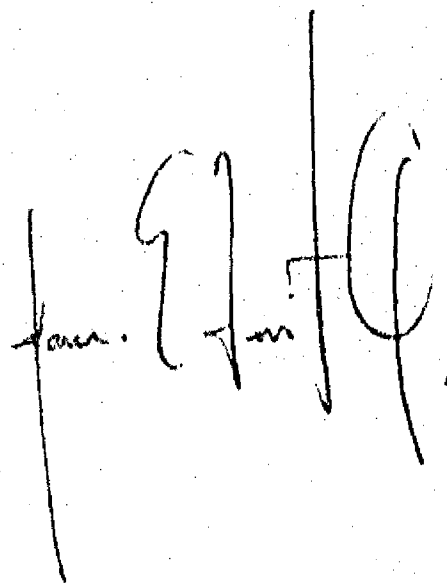
Resumen

En *Entamoeba histolytica*, los genes de RNA ribosomal se localizan en una molécula extracromosomal, conocida como el episoma ribosomal. Con el fin de caracterizar la región promotora de los genes de RNA ribosomal de *E. histolytica*, empezamos por clonar la región 5' del DNA ribosomal y definimos el extremo 5' de la unidad transcripcional usando ensayos de RNA nascente. La transcripción de los genes de RNA ribosomal empieza 2447 pb hacia arriba de la región codificante de la subunidad pequeña del RNA, en un residuo de adenosina. Este dato se apoya tanto en ensayos de mapeo por nucleasa S1 como en ensayos de extensión de primero. Demostramos que efectivamente el sitio identificado era un sitio de inicio de la transcripción haciendo un ensayo de protección del RNA marcado *in vitro* por la Guanidil Transferasa. Nuestros datos de secuencia muestran dos elementos repetidos que se localizan en la región transcrita, esto es, hacia abajo del sitio de inicio de la transcripción de los genes de RNA ribosomal, los cuales podrían estar involucrados en regulación transcripcional. También buscamos un consenso alrededor del sitio de inicio de la transcripción de los genes de RNA ribosomal de 14 especies diferentes utilizando un algoritmo de matrices de peso y encontramos un consenso débil. Utilizamos este consenso para buscar el valor de ajuste máximo en la secuencia del líder ribosomal, y este valor se localizó exactamente en el sitio de inicio de la transcripción que identificamos experimentalmente. Finalmente, en esta tesis reportamos un método para hacer secuencia genómica en el episoma ribosomal, y discutimos la posibilidad de hacer "footprinting" genómico para estudiar las posibles regiones involucradas en la regulación transcripcional y en la replicación del episoma.

A handwritten signature in black ink, appearing to be 'James E. H.', written in a cursive style. The signature is located in the lower right quadrant of the page.

Abstract

In the enteric parasite *Entamoeba histolytica*, ribosomal RNA genes are carried in an extrachromosomal DNA molecule, referred as the ribosomal episome. In order to start to characterize the rRNA promoter region of *E. histolytica*, we cloned the upstream region of the ribosomal RNA and we defined the 5' boundary of the transcription unit with nuclear run-on assays. Ribosomal transcription starts 2447 bp upstream the SSU ribosomal gene, at an adenosine residue. This data was supported both by S1 mapping and by primer extension analysis; that the mapped site was indeed the transcription initiation site was demonstrated by RNase protection of the *in vitro* capped RNA. Our sequence data around the transcription initiation site shows two different tandem repeat clusters immediately downstream of the transcription initiation site, which could be involved in transcriptional regulation. We also looked for a consensus sequence around the transcription initiation site of ribosomal RNA genes of 14 different species using a weight matrix algorithm and we found a weak consensus. This consensus was used to find the maximum fitting value in the ribosomal transcribed spacer, and this was found to be exactly the transcription initiation site that we identified by experimental data. We also report in this thesis a method to do genomic sequencing in the ribosomal episome, and we discuss the possibility of doing genomic footprinting to identify regions involved in transcriptional regulation and in replication.

A handwritten signature in black ink, appearing to read 'J. G. H. C.', is located at the bottom right of the page. The signature is written in a cursive style with a vertical line extending upwards from the top of the first letter.

Este trabajo se realizó en el Departamento de Genética y Fisiología Molecular y en el Departamento de Reconocimiento Molecular y Bioestructura del Instituto de Biotecnología de la UNAM, bajo la dirección del Dr Mario Zurita y del Dr. Paul Lizardi.

Durante este periodo fui becaria del Consejo Nacional de Ciencia y Tecnología.

Resumen

En *Entamoeba histolytica*, los genes de RNA ribosomal se localizan en una molécula extracromosomal, conocida como el episoma ribosomal. Con el fin de caracterizar la región promotora de los genes de RNA ribosomal de *E. histolytica*, empezamos por clonar la región 5' del DNA ribosomal y definimos el extremo 5' de la unidad transcripcional usando ensayos de RNA naciente. La transcripción de los genes de RNA ribosomal empieza 2447 pb hacia arriba de la región codificante de la subunidad pequeña del RNA, en un residuo de adenosina. Este dato se apoya tanto en ensayos de mapeo por nucleasa S1 como en ensayos de extensión de primero. Demostramos que efectivamente el sitio identificado era un sitio de inicio de la transcripción haciendo un ensayo de protección del RNA marcado *in vitro* por la Guanidil Transferasa. Nuestros datos de secuencia muestran dos elementos repetidos que se localizan en la región transcrita, esto es, hacia abajo del sitio de inicio de la transcripción de los genes de RNA ribosomal, los cuales podrían estar involucrados en regulación transcripcional. También buscamos un consenso alrededor del sitio de inicio de la transcripción de los genes de RNA ribosomal de 14 especies diferentes utilizando un algoritmo de matrices de peso y encontramos un consenso débil. Utilizamos este consenso para buscar el valor de ajuste máximo en la secuencia del líder ribosomal, y este valor se localizó exactamente en el sitio de inicio de la transcripción que identificamos experimentalmente. Finalmente, en esta tesis reportamos un método para hacer secuencia genómica en el episoma ribosomal, y discutimos la posibilidad de hacer "footprinting" genómico para estudiar las posibles regiones involucradas en la regulación transcripcional y en la replicación del episoma.

Abreviaturas

AdML	Middle Late Adenovirus
ARS	Secuencia de Replicación Autónoma
CAT	Cloranfenicol Acetil Transferasa
CHO	Chinese Hamster Ovaries
DMS	Dimetil Sulfato
DNA	Acido Desoxiribonucleico
Gal	Galactosa
GlcNac	N-acetil-glucosamina
GTF	General Transcription Factor
IBP	Iniciator-binding-protein
ICR	región de control interna
Inr	Secuencia iniciadora
NAcGal	N-acetil-Galactosamina
N-glicosilación	amino-glicosilación
PFGE	Pulse Field Gel Electrophoresis
PSE	elemento de secuencia proximal
RNA	Acido Ribonucleico
SL1	factor de selectividad 1
snRNAs	RNAs pequeños nucleares
SNAPc	complejo proteico activador de snRNAs
SREHP	Serine Rich E. histolytica Protein
SRP54	Signal Recognition Particle 54
TAF	TBP- Associated Factor
TBP	TATA-Binding Protein
TFII _X	Factor Transcripcional de la RNA polimerasa II
TFIII _X	Factor Transcripcional de la RNA polimerasa III
TTP	Timidina Trifosfato
UTP	Uridina Trifosfato
UCE	Elemento de control hacia arriba
UBF	Upstream Binding Factor

Indice

I- Introducción	
IA- <i>Entamoeba histolytica</i> es el agente causal de la amibiasis	1
IB-Ciclo de vida	1
IC-Biología de <i>E. histolytica</i>	2
IC1- Características morfológicas	2
IC2- Características bioquímicas	3
IC3- Estructura y organización genómica	4
IC4- Estatus filogénico de <i>Entamoeba</i>	6
ID- Transformación de <i>E. histolytica</i>	7
II- Antecedentes	9
IIA- Generalidades sobre la transcripción en eucariotes	9
IIB- Transcripción por la RNA polimerasa I	13
IIB1- Organización génica de los genes de RNA ribosomal	13
IIB1.1- Elementos regulatorios	14
IIB1.2- Promotores intergénicos	15
IIB1.3- Enhancers	15
IIB2- Factores transcripcionales	15
IIC- Organización extracromosomal de los rDNA	16
IID- Episoma ribosomal de <i>E. histolytica</i>	17
III- Objetivo	19
IV- Estrategia experimental	20
IVA- Análisis de extensión de primero y mapeo por S1	20
IVB- Capping <i>in vitro</i> y TAP-RLMPCR	21
IVC- Análisis de la secuencia del precursor del rRNA	22
IVD- "Footprint" <i>in vivo</i>	23
V- Materiales y métodos	24
VI- Resultados	30
VIA- Estrategia de secuenciación y mapa de restricción de la clona λ gt-10-6.8	30
VIB- Resumen de resultados	30
VIC- Secuencia genómica y reactividad de las guanosinas al DMS	33
VII- Discusión	34
VIII- Bibliografía	38

I-Introduccion

IA- Entamoeba histolytica: es el agente etiológico de la amibiasis

Entamoeba histolytica es un protozooario parásito del intestino humano. El trofozoito o amiba tiene alta capacidad lítica hacia todo tipo de células. La disentería amibiana y los abscesos hepáticos amibianos son problemas de salud importantes en países en vías de desarrollo. Walsh publica en 1986 una estimación de la magnitud mundial de la morbilidad y mortalidad asociadas a la infección por *E. histolytica*: anualmente, alrededor del 10 % de la población ha sido infectada por este parásito; las cifras muestran 40 millones de casos de disentería, y más de 40000 muertes anuales (1). En México, alrededor del 8.4% de la población ha sido infectada por *E. histolytica* y se reportan mil muertes al año (2). Sin embargo, el 90% de las infecciones son asintomáticas. De esta dicotomía con respecto al potencial patogénico de *E. histolytica* surgió la hipótesis de la posible existencia de dos especies, una patogénica, y la otra capaz de habitar como comensal inofensivo en el intestino humano; existen evidencias que sustentan esta hipótesis (3) (4) (5). El ser humano es el único reservorio y fuente de infección.

IB- Ciclo de vida

En el ciclo de vida de *E. histolytica* destacan dos estadios principales, que son la forma móvil y la forma inmóvil. La forma inmóvil, o el quiste, es la responsable de la transmisión y diseminación de la enfermedad, y la forma móvil, o el trofozoito es la forma agresora, responsable del cuadro clínico. Los quistes sobreviven varias semanas a la intemperie. Los quistes son ingeridos en alimentos y aguas contaminadas, y se desenquistan en el intestino medio, de donde migran para instalarse en el colon, y sobrevivir allí hasta por 12 meses. El ciclo se completa con el enquistamiento de los trofozoitos y la excreción de los quistes. La investigación se ha centrado en estudiar estas dos formas del ciclo de vida de la amiba.

Cabe enfatizar la extraordinaria capacidad histolítica del trofozoito agresor. *E. histolytica* produce tejido necrótico, tanto en su estado inicial de invasión (en el tracto intestinal) como en los lugares de invasión metastásica (que puede ser en el hígado, en el pulmón o en el cerebro). Su capacidad destructora involucra eventos secuenciales de adherencia, citólisis y fagocitosis (6), los cuales no son el tema a tratar de esta tesis.

IC- Biología de *E. histolytica*

IC1- Características morfológicas

E. histolytica es un eucariote unicelular estructuralmente simple. Las micrografías electrónicas muestran células que tienen un núcleo pero que carecen de mitocondrias, peroxisomas, microtúbulos citoplásmicos, de un retículo endoplásmico y de un aparato de Golgi desarrollado como en eucariotes superiores. Las micrografías de barrido muestran la extraordinaria plasticidad móvil de la amiba. Sobresalen especializaciones en la superficie de la célula, como evaginaciones (pseudópodos compuestos de filopodios y lobopodios) e invaginaciones (vacuolas pinocíticas y fagocíticas (7).

Destaca la presencia de vacuolas grandes (0.2 μm) en el citoplasma de la amiba, así como la presencia de un mismo tipo de vacuolas para realizar funciones digestivas y de secreción. La ausencia morfológica de un aparato de Golgi típico resulta paradójica dado que tanto la patogenicidad como otros fenómenos vitales de la amiba requieren de funciones de secreción y exocitosis que en otras especies están asociadas al Golgi. La membrana plasmática está cubierta en la superficie externa de glicoproteínas, las cuales juegan un papel central en la infección por este parásito. La lectina mejor caracterizada es la lectina que se une a Gal/GalNAc, involucrada en fenómenos de adherencia y citotoxicidad (8). La presencia de proteínas modificadas postraduccionalmente, y la reciente clonación en el laboratorio de genes involucrados en el tráfico intracelular asociado a retículo endoplásmico y a Golgi (ERD2 y SRP54, respectivamente), sugieren

fuertemente la existencia de organelos funcionalmente equivalentes en este organismo (Dr. A. Alagón, comunicación personal).

El núcleo tiene un diámetro de 4-7 μm . En secciones ópticas, sobresale la ausencia de un nucléolo típico, la presencia de cromatina periférica y de un endosoma en el centro del núcleo que parece ser un sitio de condensación del DNA. La división nuclear ocurre sin la disolución de la membrana nuclear, y sin la formación de cuerpos nucleares distintivos (7). Existen evidencias que indican que en la cromatina periférica hay síntesis de RNA. Albach y col. marcaron las células con UTP y con TTP radioactivos y observaron que el UTP se acumuló en la cromatina periférica, mientras que el TTP se distribuyó aleatoriamente en el núcleo (9). Por otro lado, por hibridaciones *in situ* con una sonda específica del episoma ribosomal, Zurita y col (10) demostraron que los genes de RNA ribosomal se localizan en la cromatina periférica. Estas observaciones sugieren fuertemente que la cromatina periférica es la contraparte del nucléolo eucariote, y el endosoma el sitio de condensación del DNA (o heterocromatina).

IC2- Características bioquímicas

En el intestino grueso las amibas encuentran un medio con baja tensión de oxígeno, el cual es un requisito para crecer las amibas en cultivo. Sin embargo, los trofozoitos de *E. histolytica* no son anaerobios absolutos, como se pensaba tradicionalmente; son capaces de consumir oxígeno pese a que carecen de mitocondrias (11). El metabolismo de este parásito es sorprendente ya que *E. histolytica* parece un aerobio facultativo. Este metabolismo puede presentar ventajas al parásito, permitiéndole cambiar de un medio ambiente con una baja tensión de oxígeno como existe en el intestino medio a tejidos con una irrigación sanguínea abundante, como el hígado.

Los carbohidratos son la fuente energética principal de este parásito. Los trofozoitos captan la glucosa por un sistema de transporte específico (12). El catabolismo de la glucosa es diferente, tiene enzimas glicolíticas poco

comunes que no se regulan alostéricamente, pero no tiene mitocondrias, citocromos, ni ciclo del ácido cítrico. Las amibas utilizan el pirofosfato inorgánico como fuente de energía en lugar de ATP en varias reacciones glicolíticas. *E. histolytica* tiene un metabolismo aeróbico; pese a la ausencia de mitocondrias y de un ciclo de ácido tricarboxílico, los electrones se transfieren de sustratos reducidos al oxígeno molecular a través de varios acarreadores como flavinas y hierro no hemo. No hay catalasas, peroxidasas u otras enzimas con grupos hemo. Carece de metabolismo de glutatión (13). En resumen, las rutas metabólicas de *E. histolytica* son diferentes a las de eucariotes que contienen mitocondrias. Por las características bioquímicas y morfológicas de este parásito, Reeves concluyó que "*E. histolytica*... podría ser uno de los organismos más primitivos, un fósil metabólico" (14). Más adelante se revisa brevemente el estatus filogenético de la amiba.

IC3- Estructura y organización genómica

La estructura y la organización genómica de *Entamoeba histolytica* es peculiar. Los genes de RNA ribosomal se localizan exclusivamente en una molécula de DNA circular, el episoma ribosomal (15). Estas moléculas se encuentran en copias múltiples en el núcleo de la amiba, y representan alrededor del 10% del DNA total de la amiba. Debido a la ausencia de un estado de condensación antes de la mitosis, no conocemos el cariotipo de la amiba. Los intentos por contestar esta pregunta (microscopía electrónica (16) y electroforesis de campo alterno, PFGE, (17)) generaron resultados poco consistentes. En un trabajo reciente, Lioutas y col hicieron un estudio sistemático comparando PFGE con DNA de *Entamoeba*, *leishmania* y levadura. Los resultados obtenidos con DNA de levadura y de *Leishmania* fueron claros y reproducibles y sin embargo, los geles PFGE con DNA de amiba mostraron bandas definidas de alto peso molecular y también un barrido de bandas, con tamaños que fluctuaron entre 225 kb a más de 2Mb; sorprendentemente, este patrón era diferente dependiendo de la preparación de DNA de amiba que utilizaron. Los autores examinaron las bandas de DNA

de alto peso molecular al microscopio electrónico y observaron moléculas de DNA circulares (18). Las moléculas circulares son sensibles a variaciones menores en la topología del DNA y a las condiciones electroforéticas, lo cual explica las diferencias observadas en diferentes preparaciones de DNA. Esta observación se reportó simultáneamente por Dhar y col (19). La tabla I muestra una compilación del tamaño de las moléculas circulares y de su abundancia relativa en diferentes cepas. Estas moléculas circulares no parecen estar relacionadas al episoma ribosomal. Existen evidencias de que estas moléculas circulares de DNA codifican para proteínas (18). Nuestro grupo caracterizó un elemento repetido abundante y discutimos su posible localización en una molécula circular diferente al episoma ribosomal (20). El cariotipo y la estructura de cromosomas queda aún por determinarse en *E. histolytica*.

La organización genómica muestra características diferentes a la de eucariotes superiores. Evidencias preliminares sugieren fuertemente que la transcripción es monocistrónica (21). La estructura de varios mRNA muestra regiones 5' y 3' no traducidas cortas (5-21 y de 14-61 nt, respectivamente). Las secuencias 5' de genes que codifican para proteínas muestran un motivo que podría ser equivalente a la caja TATA (A/GTATTTAAA) alrededor de la posición -30, diferente a la caja TATA típica de eucariotes superiores, y la transcripción inicia en una secuencia conservada (AAa/ta/cAa/tTCAA) localizada entre 5 y 21 nucleótidos antes de la primera metionina (ATG). Bruchaus y col (21) demostraron que el motivo "tipo TATA" se retarda en geles de acrilamida nativos en presencia de extractos nucleares, lo cual sugiere que esta secuencia es reconocida específicamente por una proteína, y que tenga una posible función en transcripción. Las secuencias 3' no traducidas de los mRNAs que se conocen son pequeñas (de 14 a 105 nt, a partir de codón de terminación a la cola de poli As). Hay una secuencia conservada TAA/TTT de 5-15 bases a partir de la cola de poli A, que se propone como probable señal de poliadenilación. Estas características muestran que *E. histolytica* tiene una organización genómica diferente.

En contraste con eucariotes superiores en donde la mayor parte de los

Tabla I

Abundancia relativa de las moléculas de DNA circulares en *E. histolytica*.

Cepa	Tamaño de los círculos (Kb ± D.S.) (abundancia relativa, %)				
<i>E. histolytica</i> HM1:IMSS	50.19 ±2.18 (4.8)	25.05 ± 1.37 (82.2)	12.18±0.81 (6.3)	6.02±0.24 (3.2)	3.95±0.17 (3.5)
200:NIH (clona 2)	30.30±1.3 (65.3)	24.93±0.55 (12.3)	12.1±0.52 (18.7)	7.06±0.24 (3.0)	4.03±0.13 (1.6)
HK-9 (clona 2)	32.9±1.7 (16.0)	15.06±0.45 (69.3)		5.94±0.22 (7.3)	4.07±0.22 (7.4)
<i>E. moshovski</i> Laredo	37.2±1.2 (2.0)	18.88±0.5 (71.1)	6.93±0.29 (14.1)	4.02±0.23 (9.2)	2.02±0.13 (3.6)

Dhar, S.K y col.(1995) Molec. Biochem. Parasitol. 70, 203-206

genes que codifican para proteínas tienen uno o más intrones, se han identificado sólo 2 genes que contienen un intrón, un homólogo del gen de levadura p34cdc2 (22) y un antígeno rico en glutamato (23). En nuestro grupo, se amplificó un fragmento de 50 pb que corresponde a la región interna de U6, y se clonó el gen que codifica para U6. Este gen existe en copia única en el genoma, y se expresa en *E. histolytica* (24). U6 juega un papel determinante en el "splicing" de intrones de mRNA, siendo la subunidad catalítica del "spliceosoma". La presencia de intrones y la existencia de U6 en *E. histolytica* sugiere que el sistema de procesamiento de los RNAs mensajeros podría ser similar al de otros eucariotes.

IC4- Estatus filogénico de *Entamoeba histolytica*

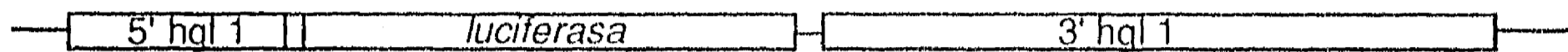
Basándose sólo en ultraestructura, Cavalier-Smith propuso en 1982 un subreino nuevo, el de los Archezoa, y acomoda en él los organismos cuyos ancestros se ramifican fuera de la línea de los eucariotes antes de la incorporación de la protomitochondria (25). Estos incluyen a *Entamoeba*, *Giardia* y *Varimorpha*. Para los dos últimos organismos, la temprana ramificación se confirmó con un árbol filogénico inferido a partir de la secuencia de la sub-unidad pequeña del rRNA. Para *Entamoeba*, los árboles inferidos a partir de secuencias moleculares han generado evidencias contradictorias con respecto a si son organismos amitocondriales primarios o secundarios; el análisis de la secuencia de la subunidad pequeña del rRNA apoya la hipótesis que propone que la ausencia de mitocondrias y de peroxisomas es una regresión, mientras que el árbol inferido a partir del factor de elongación 1 α sugiere que la ausencia de mitocondrias representa un rasgo original (26)(27), como también sugieren otros caracteres de *Entamoeba* (28). Sin embargo, contrariamente a lo que se dice sobre la ausencia de dictiosomas de Golgi en amiba por lo menos en un trabajo hay una micrografía que muestra algo que podría ser un pequeño dictiosoma de Golgi (29). La presencia de intrones (22) (23), de U6 snRNA (24) y por lo tanto de la maquinaria "spliceosomal", acercan a *E. histolytica* a los metacariontes y

la alejan de los Archezoa. La presencia en el genoma de *Entamoeba* de dos genes que se localizan en la mitocondria en los demás metacariontes, el de la enzima piridin nucleótido transhidrogenasa y el de la chaperonina cpn60, sugiere que existen reliquias mitocondriales en este organismo (30). El estatus filogénico de *E. histolytica* sigue generando polémica. Habría que analizar las secuencias de ubiquitina y del factor de elongación 1 α en organismos como *Giardia* y *Vairimorpha*, y saber si ellos sí llenan criterios de organismos primitivos prototípicos, y compararlos con *Entamoeba*. También, es importante considerar desde un punto de vista evolutivo, que *E. histolytica*, siendo un parásito, es un organismo particularmente susceptible de sufrir regresiones.

II- Transformación de *E. histolytica*

E. histolytica tiene un ciclo de vida que carece de ciclo sexual y por lo tanto, complica el estudio genético del parásito. La genética reversa es una herramienta poderosa para estudiar la biología en organismos que carecen de un ciclo sexual. Después de numerosos intentos frustrados, fue tan sólo en el año de 1994 que dos grupos independientes lograron introducir por electroporación, y expresar en forma transitoria un DNA heterólogo en *E. histolytica*, aclarando muchas de las incógnitas que existían sobre como introducir el DNA y sobre qué secuencias amibianas se requerían para la expresión de genes heterólogos (31) (32). Cabe señalar que ningún promotor amibiano ha sido caracterizado a la fecha. En estos trabajos, los autores mostraron que la expresión del gen heterólogo es dependiente de las secuencias de *E. histolytica*, y que una construcción control con un promotor, un enhancer y las señales de poliadenilación de SV40, capaz de expresarse en la mayoría de las células eucariotes, no se expresó en *E. histolytica*. Tampoco observaron actividad del gen reportero en construcciones que no tenían secuencias promotoras o de poliadenilación. Estos trabajos se hicieron utilizando dos tipos de genes reporteros, el gen de luciferasa y el gen de CAT, y las regiones que condujeron a la expresión funcional óptima de los

A- Construcción a partir de la cual se obtuvo expresión transitoria del gen de luciferasa en *E. histolytica*



B- Plásmido pTCV1 utilizado para transformación estable de *E. histolytica*.

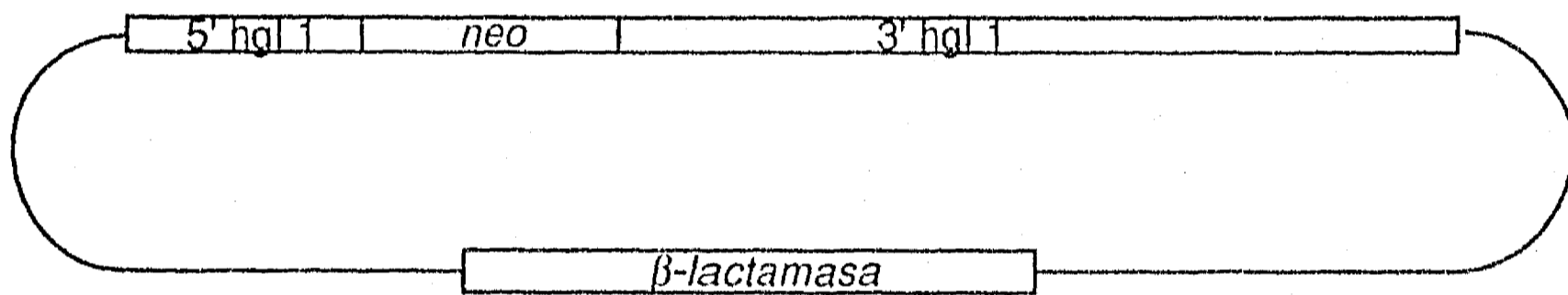
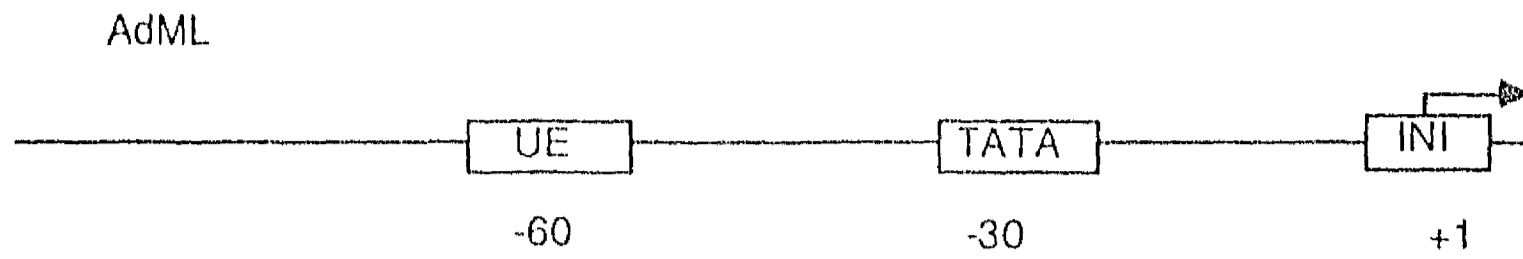


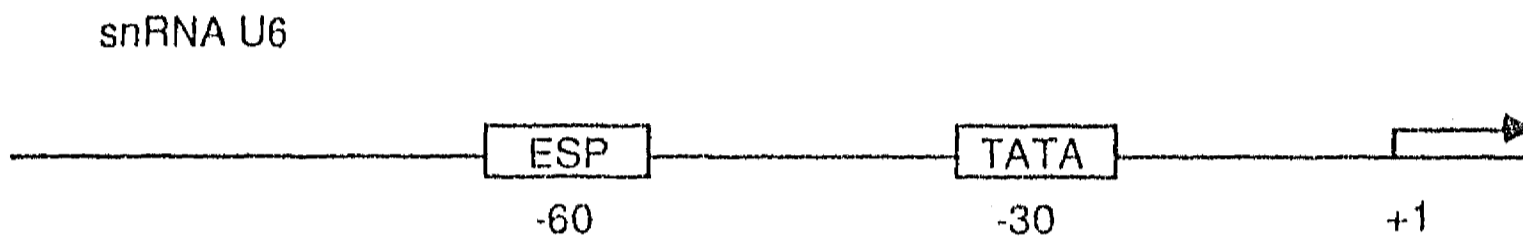
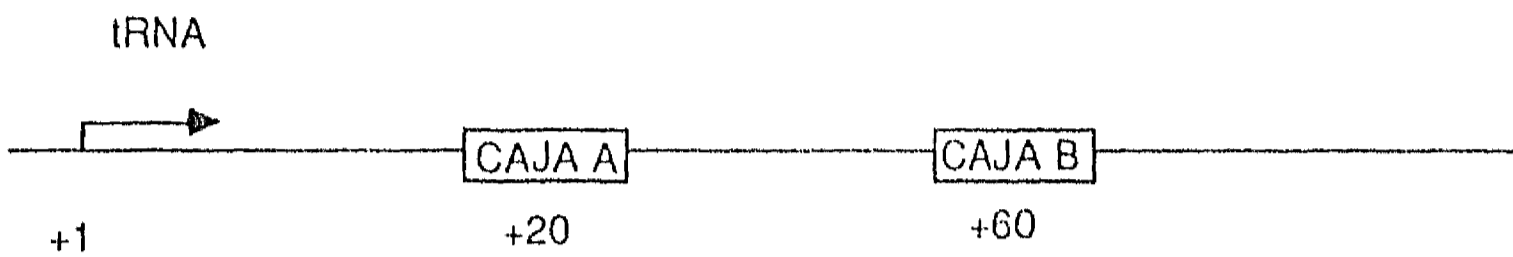
Fig. 1. Se presentan las construcciones que se utilizaron para transformar de forma transitoria (A) y de forma estable (B) *E. histolytica*. (Ver texto para referencias).

reporteros fueron los fragmentos 5' y 3' del gen hgl1, del gen hgl2 y del gen de actina (ver Fig. 1). Estos resultados confirman que la transcripción es monocistrónica en este organismo, e indican una dependencia absoluta de las regiones regulatorias amibianas. Otro resultado importante es la transformación estable de *E. histolytica* (33). La construcción que se introdujo dentro de la amiba que se mantuvo de manera estable y contiene un gen de resistencia a neomicina flanqueado por las regiones 5' y 3' del gen amibiano de la lectina hgl 1 (Fig. 1). No contiene las secuencias del episoma ribosomal que funcionan como ARS en levadura (75). Esta construcción se mantiene como molécula circular y se replica dentro de la amiba . Han obtenido amibas transformadas con esta construcción resistentes al antibiótico G418 por varios meses.

Promotor de la RNA polimerasa II



Promotor de la RNA polimerasa III



Promotor de la RNA polimerasa I

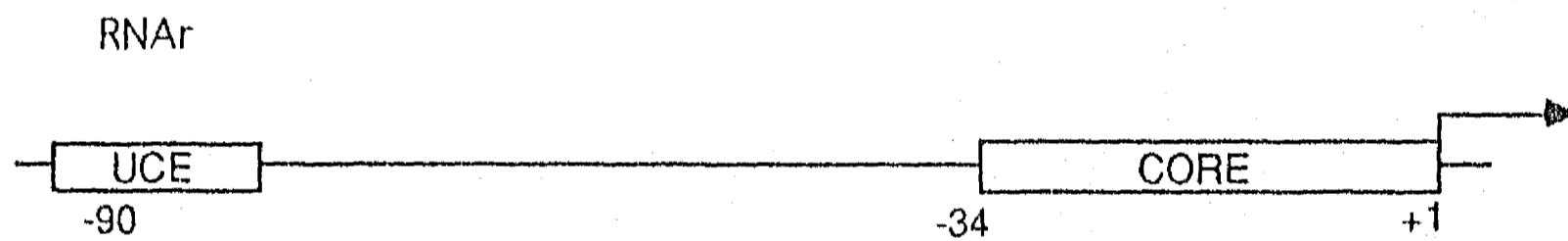


Fig. 2. Estructura de los promotores de la RNA polimerasa II, III y I. Los números indican la posición relativa de los elementos en cis en relación al sitio de inicio de la transcripción. El promotor de AdML indica un promotor prototípico de genes de clase II; un promotor de un tRNA y un promotor del snRNA U6 muestran dos tipos de promotores reconocidos por la RNA polimerasa III. El promotor de los RNAr de humano ilustra un promotor de la RNA polimerasa I. UE, elemento hacia arriba; UCE, elemento de control hacia arriba; ESP, elemento de secuencia proximal.

caja TATA en el promotor de U6, la cual es, paradójicamente, responsable de seleccionar la RNA polimerasa III, y no la II, para la transcripción de U6 (38), y que además requiere de TBP (39). Esta fué la primera sugerencia de que algunos de los factores transcripcionales, en este caso TBP, pudieran compartirse por las tres polimerasas. Por evidencias bioquímicas y genéticas, se mostró que TBP también se requiere para la transcripción de los genes transcritos por la RNA polimerasa II que no tienen caja TATA (40), para los genes transcritos por la RNA polimerasa III que tampoco tienen caja TATA (41) y para la transcripción por la RNA polimerasa I (42). Mutaciones en TBP afectan la transcripción por las tres RNAs polimerasas, lo cual muestra el papel general de TBP en la transcripción en organismos eucariotes (43). En resumen, TBP forma parte de los complejos de iniciación TFIID, SL1 y TFIIB formados en los promotores de las RNAs polimerasas II, I y III, respectivamente, con o sin caja TATA.

Los promotores de la RNA polimerasa II se pueden dividir en dos, los promotores de los genes de mRNA y los promotores de los snRNAs. Los promotores de los mRNA se pueden dividir a su vez en dos, dependiendo si contienen o no caja TATA (44). Los promotores de los mRNA con caja TATA pueden consistir solamente de una caja TATA en la posición -30 a partir del sitio de inicio de la transcripción, o de una caja TATA y de una secuencia iniciadora (Inr) que abarca el sitio de inicio de la transcripción. Los promotores de los mRNA sin caja TATA sólo consisten de la secuencia iniciadora. Tanto la caja TATA como la secuencia iniciadora reclutan el complejo TFIID que contiene TBP (45). Los promotores de los snRNAs consisten de un elemento de secuencia proximal (PSE), que también existe en el promotor de snRNA que se transcribe por la polimerasa III (Fig. 3). La Fig. 4 muestra un esquema de cómo se ensamblan los complejos de iniciación *in vitro* en los promotores de los mRNAs con o sin caja TATA, y en los promotores de las snRNAs transcritos por la polimerasa II. Para los promotores de los mRNAs con caja TATA (Fig. 4, A-D), el primer paso es la unión de TBP (o TFIID) a la caja TATA, la cual se facilita por TFIIA. Una vez que TBP se unió a la caja TATA, se unen al complejo de iniciación TFIIB, y la

U1, U2 (Pol II)



U6 (Pol III)



Fig. 3. Estructura de los promotores de U1 y de U2 transcritos por la RNA polimerasa II y del promotor de U6, transcrito por la RNA polimerasa III. Están indicados el elemento de secuencia distal (DSE), el elemento de secuencia proximal (PSE) y la caja TATA. Las flechas indican el sitio de inicio de la transcripción.

RNA polimerasa II, la cual está asociada con TFIIF. Una vez que la RNA polimerasa II se unió al complejo de iniciación, entran los factores TFIIE, TFIIH y TFIIJ (46), y la transcripción continúa al añadir ribonucleótidos trifosfatos. Para los promotores que no tienen caja TATA, la transcripción depende del Inr, el cual sobrelapa el sitio de inicio de la transcripción. Una de las secuencias Inr mejor caracterizadas es la del promotor TdT; estas secuencias existen en otros promotores sin caja TATA y en algunos otros con caja TATA (45). La secuencia Inr del promotor de TdT es suficiente para dirigir la transcripción basal del promotor (45). En la Fig. 4, E-G, se muestra la unión de IBP a la secuencia Inr, haciendo el mismo papel que TBP en dirigir el complejo de iniciación al sitio de inicio e la transcripción. Se han descrito varias IBPs; dos ejemplos son TFII-I, una proteína de 120 kd que reconoce la secuencia iniciadora del virus AD21Va21 (47), y YY1, que reconoce la secuencia iniciadora del virus p5 asociado a Ad2 (48). Sin embargo, TBP, TFIIB, la RNA polimerasa II y TFIIF se pueden asociar con la secuencia iniciadora en ausencia de IBP, lo cual sugiere que entran simultáneamente dentro del complejo de iniciación (49). Aún en promotores sin caja TATA, TBP se puede unir a la región -30 (50). Los factores TFIIE, TFIIH y TFIIJ entran al complejo de iniciación simultáneamente, igual que en los promotores con caja TATA. En el caso de los promotores de snRNAs, la presencia de una caja TATA determina la selección de la RNA polimerasa III, y no de la RNA polimerasa II para la transcripción (38); sin embargo, la transcripción de los genes de U1 y U2 requiere de TBP como parte de un complejo que se llama SNAPc, que a través de su unión específica al PSE, recluta algunos de los factores transcripcionales generales (40) (Fig. 4, H-I). SNAPc presenta dos propiedades que lo distinguen de otros complejos que contienen TBP (40). La primera es que, en contraste con TFIID, o SL1, que reconocen la caja TATA o que presenta afinidad baja o nula para secuencias de DNA, SNAPc se une específicamente al PSE. La segunda es que SNAPc se requiere para transcribir promotores de snRNAs reconocidos por la RNA polimerasa III y por la RNA polimerasa II, y por lo tanto es un factor transcripcional involucrado en la transcripción por dos diferentes RNAs polimerasas.

Promotores de los mRNA

Promotores de los snRNAs

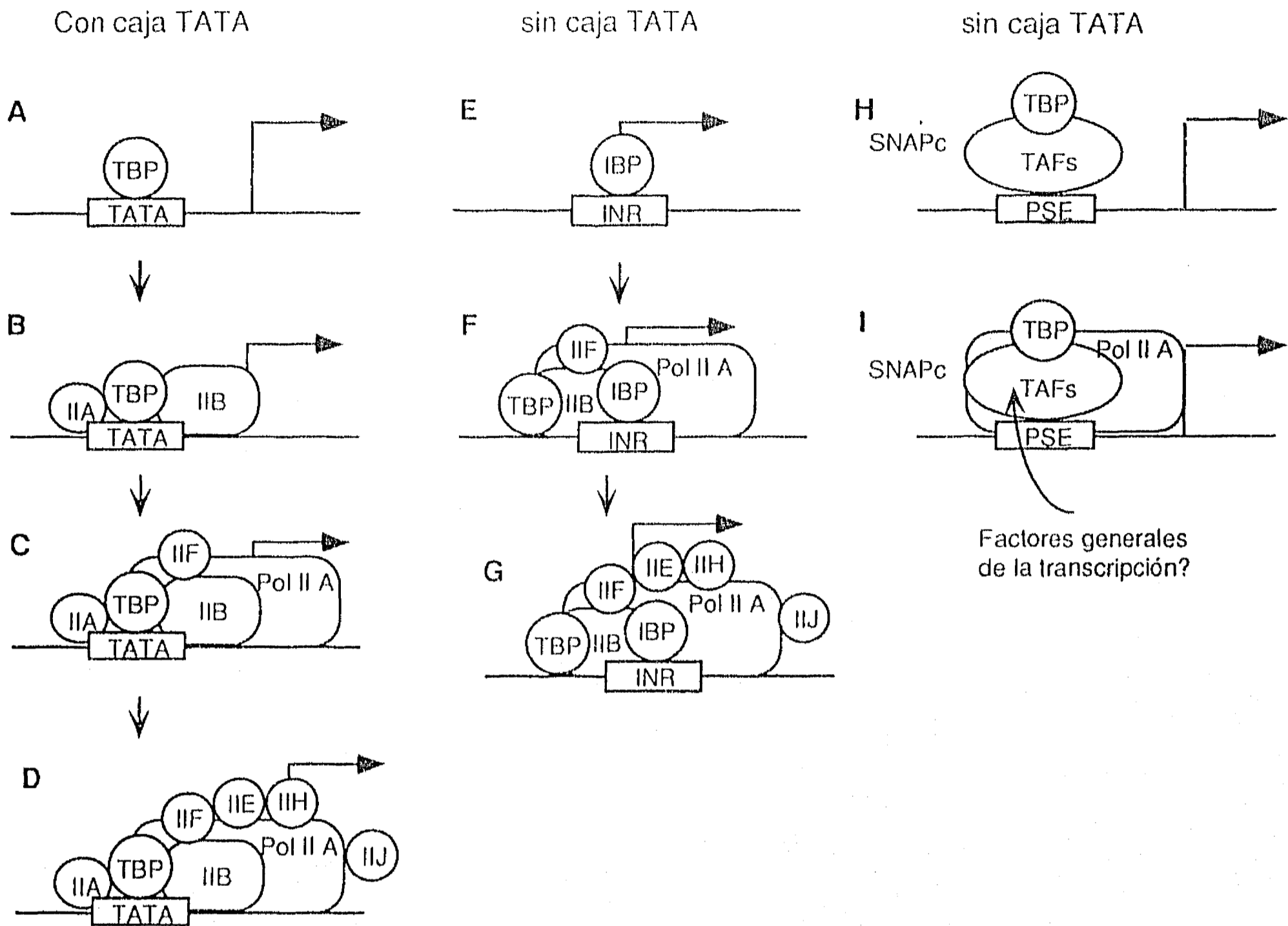


Fig. 4. Modelo para el ensamblaje de complejos de iniciación basales de la RNA polimerasa II en promotores de mRNA con o sin caja TATA y en promotores de snRNAs transcritos por la RNA polimerasa II. En promotores que contienen una caja TATA, TBP se une primero, con TFIIA seguido por TFIIB (B) y luego por el complejo RNA polimerasa II- TFIIF (C). TBP interacciona con estos tres componentes. Luego, entran TFIIE, TFIIH, y TFIIJ. En los promotores sin caja TATA, la formación del complejo de iniciación empieza con la unión de IBP a la secuencia Inr (E). Este evento permite el reclutamiento de TBP (o TFIID), TFIIB, y del complejo TFIIF-RNA polimerasa II (F), lo cual permite que entren el resto de los factores transcripcionales al complejo de iniciación. En los promotores de snRNAs, TBP se recluta como parte del complejo SNAPc (H). La hipótesis es que una vez unido al DNA, el complejo SNAPc recluta al resto de los factores generales de transcripción (I). Adaptado de Hernandez, N., 1993, Genes and Dev.

Los promotores de la RNA polimerasa III se han dividido en tres clases (51). Los promotores de clase 1 y 2 son promotores intergénicos sin caja TATA (Fig. 2) y son los de los genes que codifican para el gen 5S ribosomal, que consisten de la región de control interna (ICR), y los de los genes de tRNA y 7SL, de las secuencias Alu, y de los genes virales como Ad2 VAI y VAII que consisten de las cajas A y B, respectivamente (51). Los promotores clase 3 se localizan en los snRNA de vertebrados y en los genes que codifican para los RNA citoplásmicos, incluyendo U6, 7SK, hY1 y hY3, H1 y el RNA MRP/th (51). Estos promotores se localizan hacia arriba de la región codificante y son similares a los promotores de los snRNAs transcritos por la RNA polimerasa II pero contienen una caja TATA, que en este caso determina la especificidad por la RNA polimerasa III para la transcripción (39) (Fig. 2). La transcripción a partir de los promotores de la RNA polimerasa III con o sin caja TATA requiere TBP, en complejos diferentes. La transcripción del gen 5S requiere de TFIIIA, TFIIIB y de TFIIIC. TFIIIA se une primero al ICR, y esta unión permite la entrada subsecuente de TFIIIC y TFIIIB al complejo de iniciación (51). La transcripción de los promotores de clase 2 no requiere de TFIIIA, porque TFIIIC se une directamente a las cajas A y B y recluta entonces TFIIIB (Fig. 5) (51). TFIIIB es un complejo macromolecular que consiste de un polipéptido de 70kD, BRF, de TBP, y algunas veces también de un polipéptido de 90 kD (41). La Fig. 5 muestra el ensamblaje de TFIIIB en un promotor de tRNA. TFIIIC se une a las cajas A y B, y luego interacciona con TFIIIB. BRF interacciona con TFIIIC, y en presencia de TBP la asociación de BRF con el complejo de iniciación es más fuerte (41). Una vez que TBP se unió al complejo de iniciación, se recluta la proteína de 90kD.

La transcripción del gen U6 requiere también del SNAPc (40). Los experimentos de reconstitución muestran que la transcripción del gen U6 requiere además de SNAPc, TBP. Esto sugiere que el complejo de iniciación en estos promotores contiene por lo menos dos formas de TBP, una forma unida a PSE y que forma parte de SNAPc, y otra forma unida a la caja TATA (40).

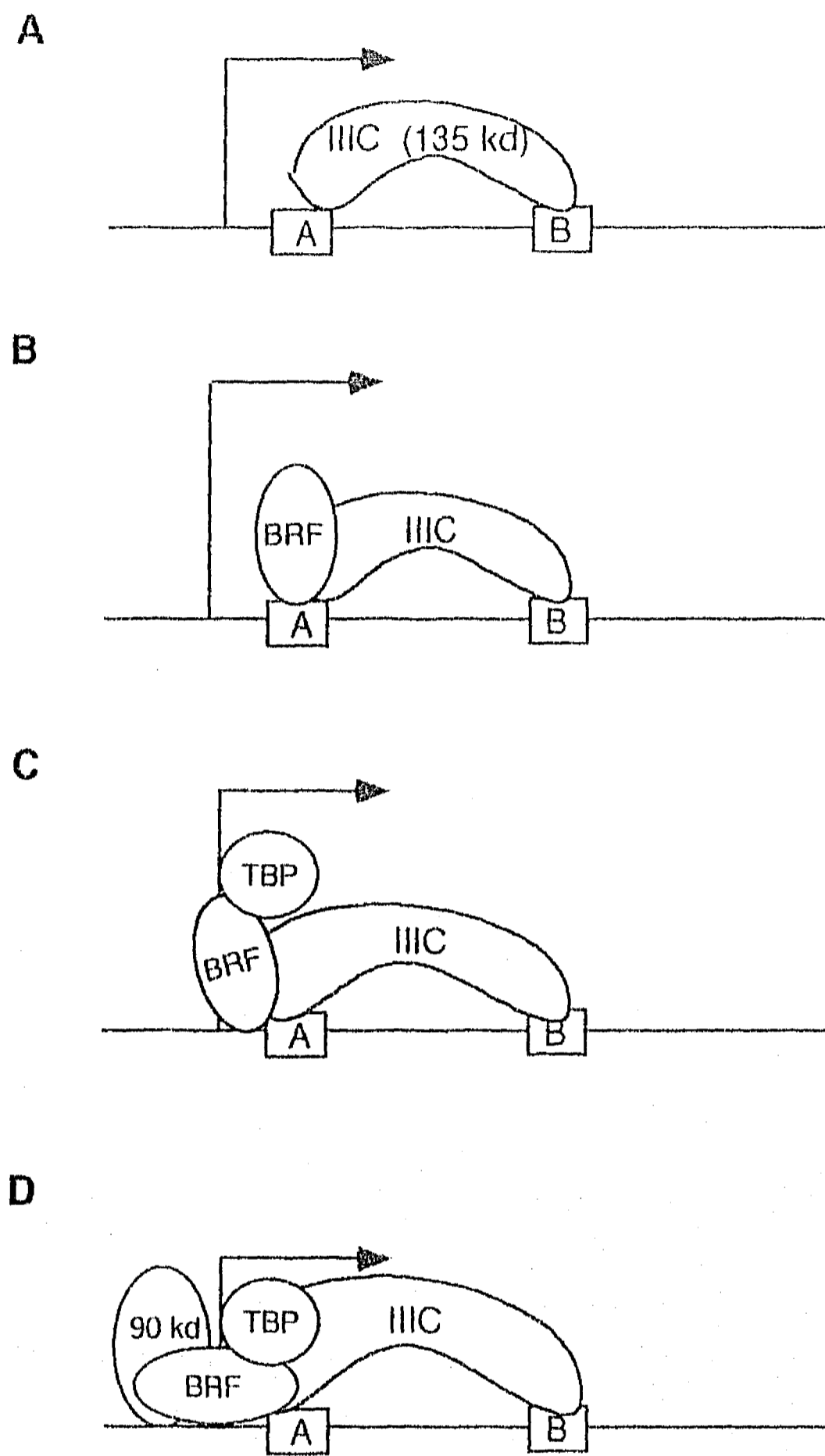


Fig. 5. Ensamblaje del complejo de iniciación de la RNA polimerasa III a un promotor de clase 2 de *S. cerevisiae* (adaptado de Kassavetis y col, 1992). TFIIC se une a las cajas A y B. Luego, TFIIB (compuesto de BRF, TBP y la subunidad de 90 kd) entran al complejo de iniciación. Primero, BRF interactúa con TFIIC (B), luego entra TBP, y induce un cambio conformacional en BRF que ahora entra en un contacto más estrecho con el DNA (C), y finalmente, la subunidad de 90 kd sella el complejo TFIIB sobre el DNA (D).

Los promotores de los genes de rRNA son reconocidos por la RNA polimerasa I y no contienen caja TATA. Los promotores están compuestos por un elemento "core" que sobrelapa el sitio de inicio de la transcripción y UCE, o elementos de control hacia arriba, localizados más allá de -100 y estimulan de 10 a 100 veces la transcripción (52). La transcripción requiere de dos factores, además de la RNA polimerasa I, que son UBF y SL1 (factor de selectividad 1). SL1 consiste de TBP y tres TAFs, de 48, 63 y 110kD, las cuales son esenciales para la transcripción (42).

Recapitulando, TBP forma parte de los complejos de iniciación TFIID, SL1 y TFIIB formados en los promotores de las RNA polimerasas II, I y III, respectivamente, que contienen cajas TATA y que no contienen caja TATA. También TBP se encuentra en el complejo SNAPc, que se forma en los promotores de las snRNAs de la RNA polimerasa III que contienen caja TATA y de la RNA polimerasa II sin caja TATA. La figura 6 ilustra la asociación de TBP con diferentes complejos transcripcionales (TAFs) y con las tres RNA polimerasas. Las TAFs asociadas a SL1 son diferentes a las que se asocian a TFIID. El modelo supone que la asociación de diferentes juegos de TAFs a las moléculas de TBP determinan el ensamblaje de diferentes complejos de transcripción.

IIB- La transcripción por la RNA polimerasa I

IIB1- Organización génica de los genes de RNA ribosomal

La RNA polimerasa I se encarga de transcribir los genes de DNA ribosomal. En casi todas las especies, estos genes están organizados de una manera similar (Fig. 7), y se transcriben en un compartimento nuclear específico, el nucléolo. Estos genes se transcriben en un precursor de varias kilobases, que contiene las secuencias que codifican para las regiones 18S, 5.8 S y 28 S, que se procesa hasta generar los componentes del ribosoma maduro (Fig. 7). El orden 5' a 3' de estas tres regiones es el mismo en todas las especies. La longitud de este precursor es particular a cada especie. La

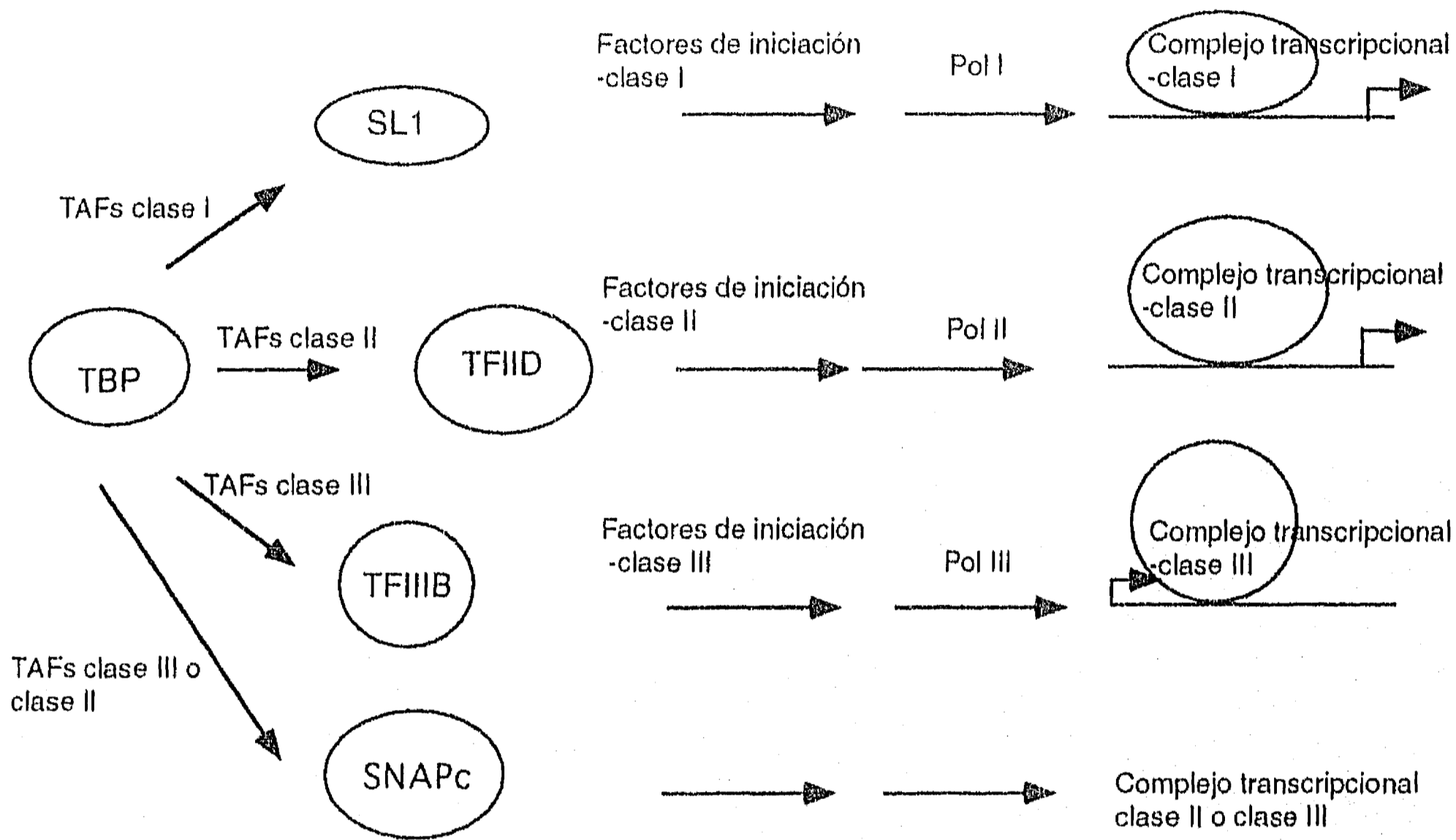


Fig. 6. Ilustración esquemática del ensamblaje de TBP en clases diferentes de complejos de transcripción mediante la asociación de TBP con diferentes grupos de TAFs y luego con diferentes factores de iniciación y RNA polimerasas.

tabla II muestra las longitudes de los transcritos primarios y de los rRNA citoplásmicos de varias especies.

IIB1- Elementos regulatorios

Los genes de RNA ribosomal tienen promotores localizados hacia arriba del sitio de inicio de la transcripción, y no tienen una caja TATA. En la mayoría de los organismos eucariotes, los genes ribosomales se localizan en cromosomas como unidades repetidas en tandem y cada gene que codifica para un precursor de RNA ribosomal se alterna con un espaciador intergénico (Fig. 7). La RNA polimerasa I reconoce dos promotores, el promotor de los genes de rRNA y los promotores intergénicos y este reconocimiento de la polimerasa por su promotor es especie específico. Dentro de la región génica, las secuencias que codifican para los RNAr 18S, 5.8S y 28S se han conservado fuertemente a lo largo de la escala filogenética mientras que las secuencias de los espaciadores intergénicos ha evolucionado rápidamente; aún entre especies relacionadas las secuencias muestran muy poca similitud. Existe, pese a esta divergencia, una conservación filogenética en el arreglo y en el tipo de elementos regulatorios dentro del espaciador intergénico. Se han identificado todos los elementos regulatorios de dos vertebrados, *Xenopus laevis* y de ratón, de *Drosophila*, y de varias especies de plantas y esta información sugiere que los elementos regulatorios están organizados de forma similar (52). La conservación en el arreglo de los elementos regulatorios sugiere fuertemente que los mecanismos de control también están conservados. El examen detallado de los promotores de los genes ribosomales de 4 vertebrados (humano, ratón, rata y rana) muestra que comparten una arquitectura básica similar: el promotor se extiende en 150 pares de bases, con el extremo 3' en los primeros 4-6 nucleótidos del transcrito primario (52). El promotor se divide en dos dominios denominados la región "core" o medular, que comprende el sitio de inicio de la transcripción y un dominio hacia arriba que define el extremo 5' del promotor (Fig. 2). La región "core" es suficiente para dirigir la iniciación específica *in vitro* pero la región hacia arriba es necesaria para iniciar *in vivo* (52).

Tabla II

Longitud de los transcritos primarios y de los RNAr maduros de varias especies

	Transcritos primarios		Longitud de los RNAr			
	Coef. Sed. (S)	Longitud (Kb)	26S-28S (Kb)	16S-18S (Kb)	% del transcrito primario	Ref.
<i>E. coli</i>	30	6.0	3.0	1.5	75	50
<i>S. cerevisiae</i>	37	6.6	3.8	1.7	77	51
<i>D. discoideum</i>	37	7.4	4.1	1.8	80	55
<i>D. melanogaster</i>	34	7.7	4.1	1.8	76	52
<i>X. laevis</i>	40	7.9	4.5	1.9	81	53
<i>M. musculus</i>	45	13.7	5.1	1.9	51	52,53
<i>H. sapiens</i>	45	13.7	5.1	1.9	51	53
<i>E. histolytica</i>	ND	8.25	3.46	1.94	65	59

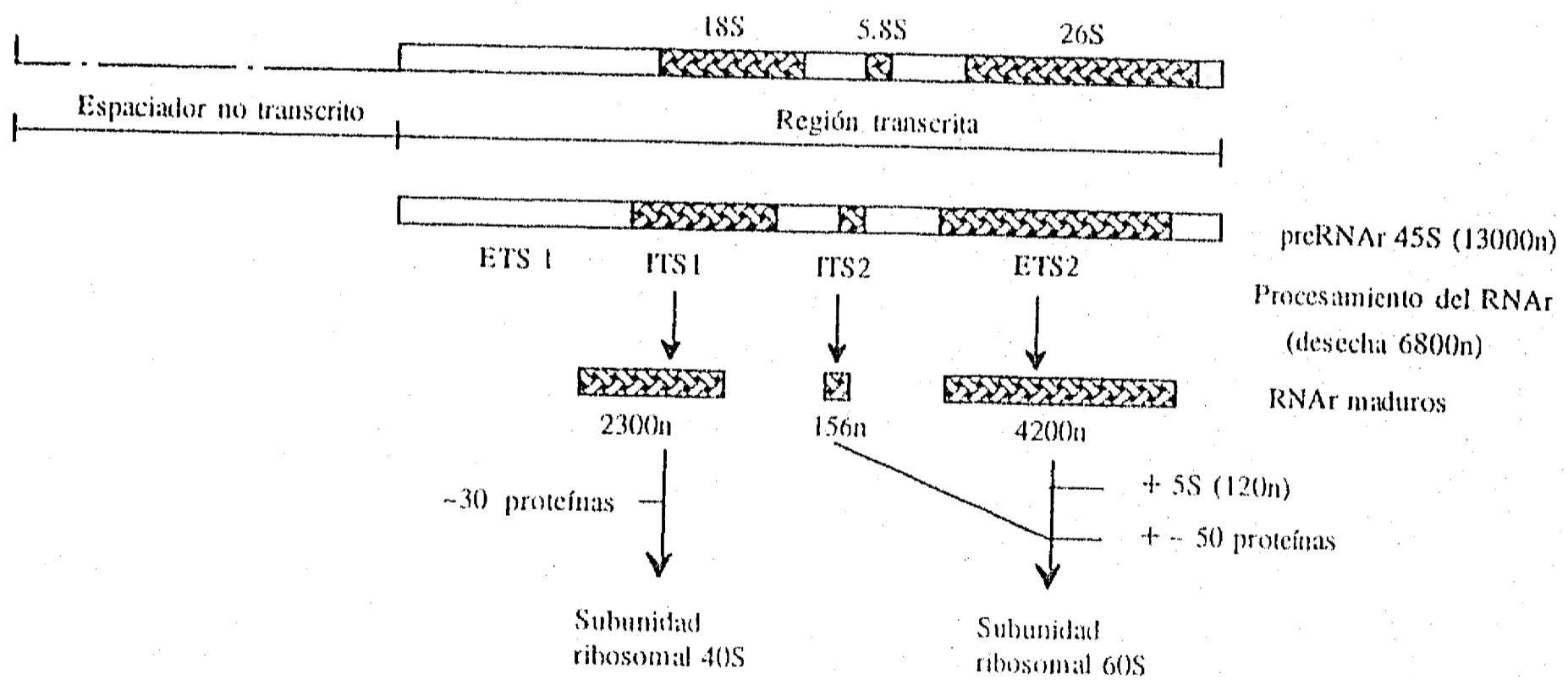
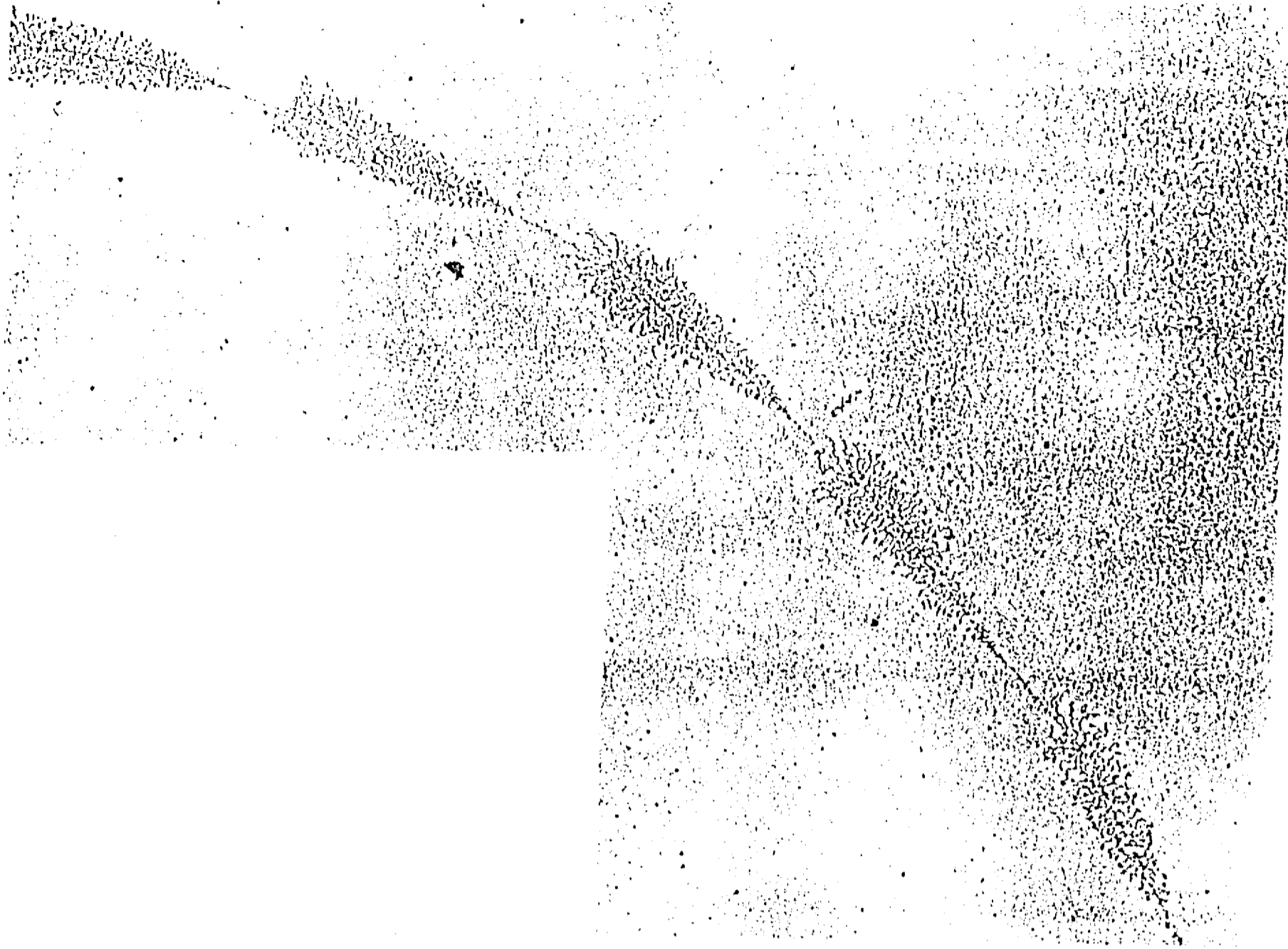


Fig. 7. Organización de la unidad transcripcional de los RNAr. La micrografía electrónica muestra la cromatina ribosomal transcribiéndose. Se ven estructuras parecidas a árboles de navidad, que corresponden a unidades transcripcionales individuales. En el esquema se ilustran la región espaciadora no transcrita y la región transcrita, dentro de la cual se muestra la organización de los genes 18S, 5.8S y 26S. El esquema también muestra los pasos involucrados en la transformación de la molécula precursora 45S a la molécula madura. ETS, espaciadores externos transcritos, ITS, espaciadores internos transcritos (Watson y col., 1987).

IIB2- Promotores intergénicos

Entre las unidades transcripcionales de los genes de DNA ribosomal existen promotores intergénicos, los cuales son responsables de la transcripción de la región espaciadora que algunas veces se observa en micrografías electrónicas. Se han descrito promotores intergénicos en dos especies de rana, en ratón, en rata, en hamsters chinos y en *Drosophila* (52). En humano no se han identificado los promotores intergénicos pero sí hay transcripción intergénica. Los promotores intergénicos son los otros promotores que también son reconocidos por la RNA polimerasa I. Estos promotores, en rana, son solamente diferentes en un 10% al promotor génico (60). En ratón, sin embargo, estos sólo comparten 12 pb con el promotor génico (61). Estos promotores se localizan hacia arriba de los elementos repetidos que actúan como enhancers de la transcripción. Los transcritos de estos promotores terminan en un terminador que se encuentra inmediatamente hacia arriba del promotor génico (52). La actividad de estos promotores intergénicos está correlacionada con el desarrollo. Tanto en *Xenopus laevis* (52) como en *Drosophila* (62), la actividad transcripcional de la región espaciadora es mayor durante el desarrollo temprano que durante el estado adulto del organismo.

IIB3- "Enhancers"

La región espaciadora de los genes de DNA ribosomal se caracteriza por la presencia de elementos repetidos que funcionan como "enhancers" (52). Estos "enhancers" actúan en *cis* para estimular promotores adyacentes, y son capaces de competir con promotores localizados en *trans* (108).

IIB4- Factores transcripcionales

La transcripción de los genes de rDNA, requiere, además de la RNA polimerasa, dos factores transcripcionales, UBF y SL1. El factor UBF, es un factor que contiene múltiples dominios tipo HMG (63), es esencial para la

activación eficiente del promotor, y se une a las secuencias de DNA que conforman el arreglo complejo de los espaciadores intergénicos (enhancers, promotores duplicados y terminadores). UBF también se une al promotor "core" y a los UCE, de forma secuencia específica, y así recluta el complejo SL1-RNA polimerasa I, y por lo tanto funciona como un factor de transcripción general y como un enhancer. SL1, por sí solo, tiene afinidad muy baja o nula para el promotor de los rRNAs. Con base en estas observaciones, el complejo de iniciación de la RNA polimerasa I podría ensamblarse como lo muestra la Fig. 8. UBF se une primero al DNA, y esto permite reclutar SL1, lo cual estabiliza el complejo de iniciación favoreciendo la entrada de la RNA polimerasa I. En humanos, SL1 es un complejo macromolecular que contiene TBP y tres TAFs, de 48, 63 y 110kD. El modelo propone que las TAFs están involucradas en reclutar a TBP al promotor de la RNA polimerasa I a través de interacciones proteína-proteína con UBF, y de interacciones con el DNA. La especificidad de especie está otorgada por las TAFs (42).

IIC- Organización extracromosomal de los rDNA

En anfibios, peces y en algunos insectos, además de los genes de DNAr cromosomales, existen copias extracromosomales asociadas a estadios de desarrollo, en este caso a la ovogénesis (64)(65). En protozoarios, los rDNA extracromosomales son relativamente comunes, tanto en moléculas lineales palindrómicas (*Physarum* y *tetrahymena*)(66)(67), en unidades repetidas en moléculas lineales y circulares (*Paramecium*) (68), o exclusivamente en moléculas circulares (*Euglena*, *Naegleria*, *Entamoeba*) (69)(70)(15)(71). En algunos de ellos, existen también copias cromosomales de los rDNA. Pero en *Naegleria* y en *E. histolytica* los rDNA existen solamente en moléculas circulares extracromosomales. En estos organismos no se conocen los elementos involucrados en regulación transcripcional.

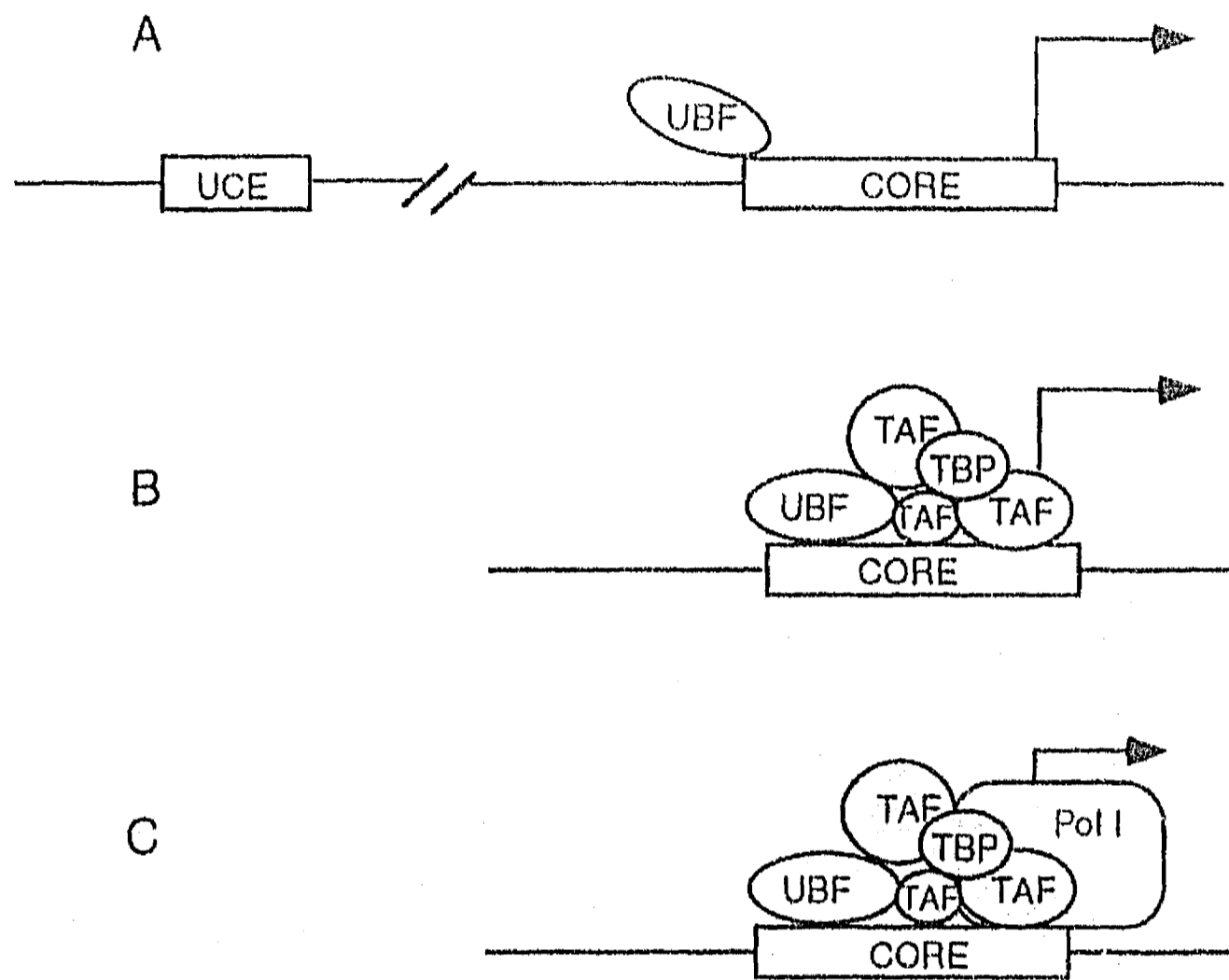


Fig. 8. El esquema muestra el ensamblaje del complejo de iniciación sobre un promotor ribosomal de mamífero. La unión de UBF al promotor se estabiliza con la unión de SL1 (B), que consiste de TBP y tres TAFs. Están señaladas en amarillo porque forman parte de SL1. No se sabe cual de las TAFs se une al DNA. Luego, UBF y SL1 reclutan a la RNA polimerasa I (C).

IID- Episoma ribosomal de *E. histolytica*

El episoma ribosomal parece ser el recinto único de los genes de RNA ribosomal (15), puesto que no se han logrado identificar copias cromosomales de los rDNA. El episoma ribosomal que mejor se conoce es el de la cepa patógena HM1 (Fig. 9). Tiene un tamaño de 24.5 kb, existen alrededor de 200 copias por núcleo, y constituye el 10% del DNA total de cada célula. Seghal y col (72) secuenciaron y analizaron la organización del episoma ribosomal de la cepa patógena HM1.

En HM1 el episoma ribosomal contiene dos copias de los genes que codifican para la subunidad pequeña (ssu, 16S) y para la subunidad grande (lsu, 5.8 S y 25 S) de los RNA ribosomales, y los tres genes se transcriben en una misma unidad transcripcional como en todos los organismos eucariotes. Las dos copias de rDNA están arregladas en la molécula circular como repetidos inversos, pero sin embargo no se sabe si ambas copias conducen a un transcrito funcional. Existen episomas ribosomales en otras cepas de *E. histolytica* y en otras *Entamoebas*, aunque estos episomas no son iguales, y no siempre contienen dos copias del cistrón ribosomal (73)(74)(72). Por ejemplo, en las cepas Rahman, Laredo y HK9 los episomas son más pequeños (18.3 Kb, 19.3 Kb y 15.3 Kb respectivamente) y contienen una sola copia de los genes de rDNA (72). La cepa 200:NIH tiene dos copias de los rDNA (72). La Fig. 10 muestra el episoma ribosomal de la cepa HK9, que es la cepa que utilizamos para realizar este trabajo. La presencia de dos unidades de los rDNA no parece ser obligatoria.

Los episomas ribosomales presentan una organización a nivel de estructura primaria característica. La región río-arriba y río-abajo (region 5' y 3' con respecto a los genes de DNA ribosomal, respectivamente) consiste de varias familias de elementos repetidos (72). Estas familias de elementos repetidos se han secuenciado y analizado detenidamente en la cepa HM1. Algunas de estas familias tienen secuencias relacionadas a ARS de levaduras, y por lo tanto se les implica en replicación del episoma (75). Otras de ellas, por su alto contenido A-T, y por la presencia de tractos de pirimidinas

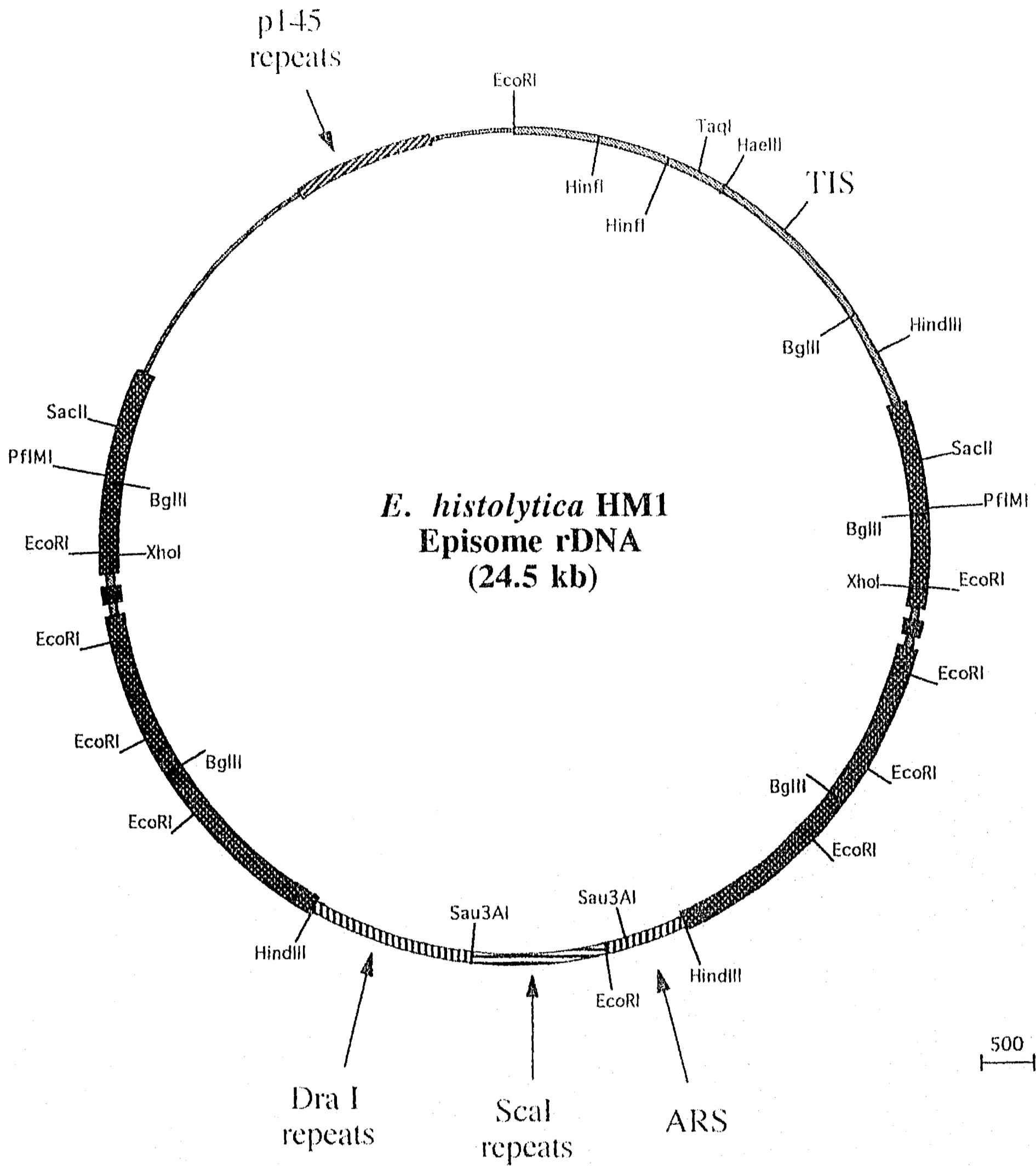


Fig. 9. Episoma ribosomal de la cepa histolítica HM1.

Se muestran los genes de DNAr organizados como inversos repetidos en una molécula circular y un mapa de restricción de la molécula. En verde, azul y en rojo, están indicadas las regiones que codifican para los RNAs 17S, 5.8S y 25S, respectivamente. Se muestran los repetidos *Dra* I, *Sca* I, p145 y la secuencia que funciona como ARS en levadura.

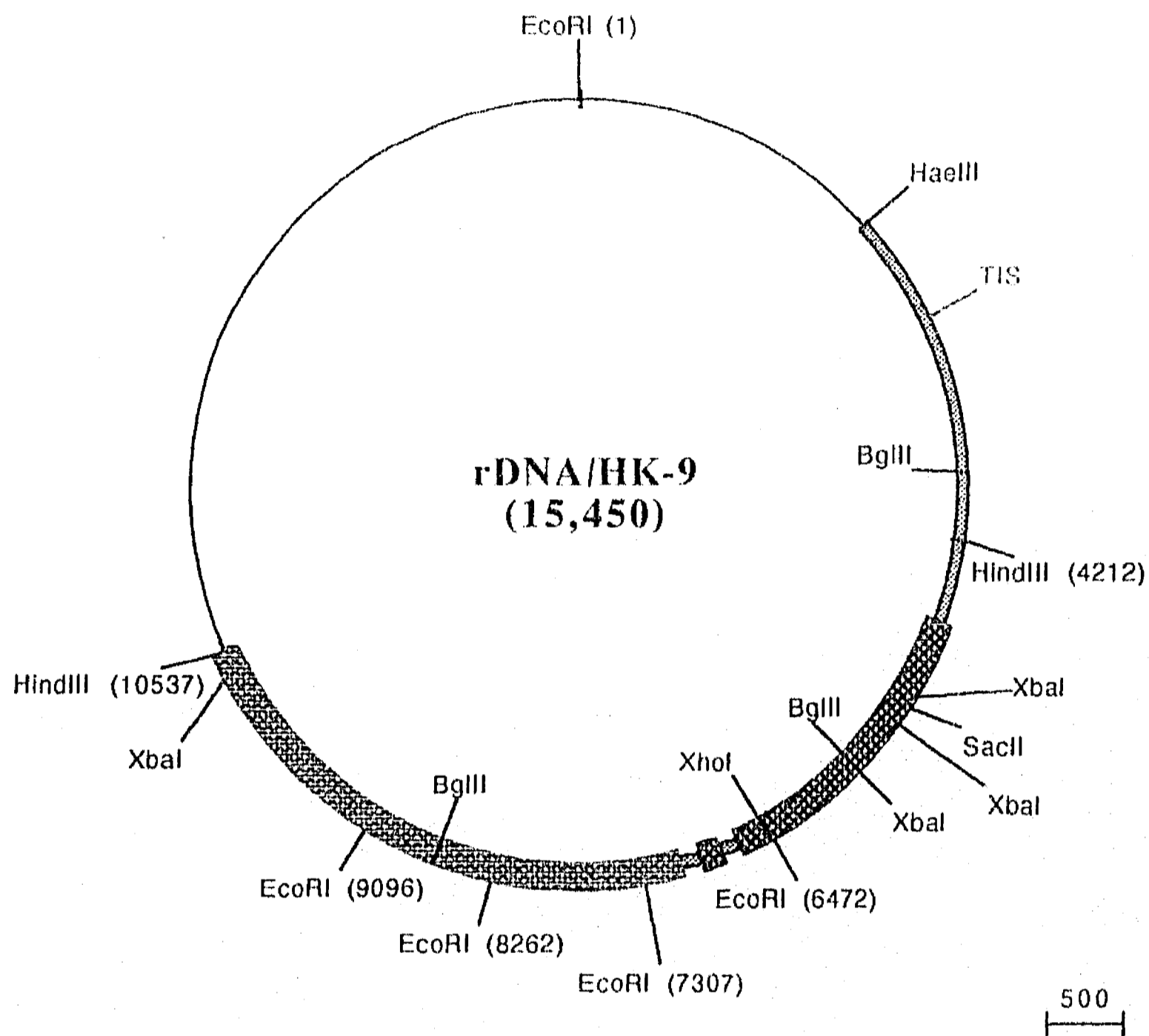


Fig. 10. Mapa de restricción detallado de la cepa histolítica HK9. En verde, azul, y rojo, respectivamente, están indicadas las regiones codificantes de los RNAs 17S, 5.8S y 25S. En negro, se muestra la región que se secuenció en este trabajo. TIS, sitio de inicio de la transcripción.

sugieren que puedan participar en recombinación (73). En particular, los repetidos *Dra* I tienen un tracto de pirimidinas, y muestran polimorfismo en los fragmentos de restricción con enzimas de restricción de corte frecuente, probablemente como resultado de pérdidas de unidades repetidas por recombinación (Fig. 9).

Aparte de ser el depósito único de los RNA ribosomales, existe un fragmento de DNA en la región río arriba del episoma que hibrida en experimentos tipo "northern" con un transcrito de 0.7 kb (denominado Tr) en la cepa HM1 (76). La función de este transcrito se desconoce; no se conocen los límites de este transcrito, ni tampoco se localizaron marcos abiertos de lectura en las regiones próximas a este transcrito. En un trabajo reciente describen la presencia de marcos abiertos de lectura en el rDNA con secuencias que codifican para hemolisinas (77), aunque éstos son anormalmente cortos, y no se han identificado ni los mensajeros ni los productos de estos genes putativos. Esta observación podría relacionar el potencial citolítico de *E. histolytica* al episoma ribosomal. En otras cepas no se ha detectado el transcrito Tr (p.e. en las histolyticas Rahmann y HK9, en la Moshowski, o en Laredo (76)) pero, sin embargo, no se puede descartar que en estas cepas exista un transcrito diferente con una función equivalente.

III- Objetivo

El objetivo general de este trabajo es identificar y analizar el promotor de los genes de RNA ribosomal de *E. histolytica*.

Los objetivos particulares son:

- identificar el sitio de inicio de la transcripción de los genes de RNA ribosomal de *E. histolytica*
- analizar la secuencia que rodea el sitio de inicio de la transcripción, con la idea de encontrar los elementos típicos involucrados en la regulación de la transcripción de los genes ribosomales que existen en otros organismos eucariotes
- identificar probables regiones involucradas en regulación transcripcional en el episoma ribosomal mediante "footprinting" *in vivo*.

IV- Estrategia experimental

Como ya lo mencioné anteriormente, la transcripción de los genes de rRNA da origen a un transcrito primario de varias kilobases que sufre múltiples pasos de procesamiento antes de dar origen a las especies ribosomales maduras. La identificación del extremo 5' y del extremo 3' del transcrito primario definen el tamaño de la unidad transcripcional, y en el caso de los genes de DNAr, además, el sitio de inicio de la transcripción forma parte del promotor de los genes ribosomales (52). Los avances en la última década en la Biología Molecular ofrecen herramientas que permiten diferenciar un transcrito primario de un precursor procesado, con base en sus características moleculares. Durante mi trabajo de maestría, cloné la región que contenía el extremo 5' del cistron de RNA ribosomal de *Entamoeba histolytica* y limité el extremo 5' de la unidad transcripcional mediante un ensayo de RNA naciente. En esta tesis utilice varias técnicas complementarias, que describo a continuación, que permitieron identificar el punto de inicio de la transcripción de los genes de RNA ribosomal de *Entamoeba histolytica*.

IVA- Análisis de extensión de primero y mapeo por S1

Los dos métodos más comunes para identificar el sitio de inicio de la transcripción de un gen transcrito por la RNA polimerasa II son el análisis de extensión de primero para mapear extremos 5' y el mapeo por S1 para mapear tanto extremos 5' como extremos 3' de los RNA mensajeros. El ensayo de extensión de primero consiste en extender una molécula de cDNA sobre el RNA templado utilizando un oligonucleótido o un fragmento de restricción como primeros localizados cerca pero hacia abajo del extremo 5' del RNA, con la enzima Transcriptasa Reversa. El tamaño de la extensión del cDNA indica el límite 5' del transcrito (Fig. 11). Es indispensable utilizar como marcadores finos de peso molecular una reacción de secuencia iniciada con el mismo oligonucleótido que se utilizó para sintetizar la molécula de cDNA.

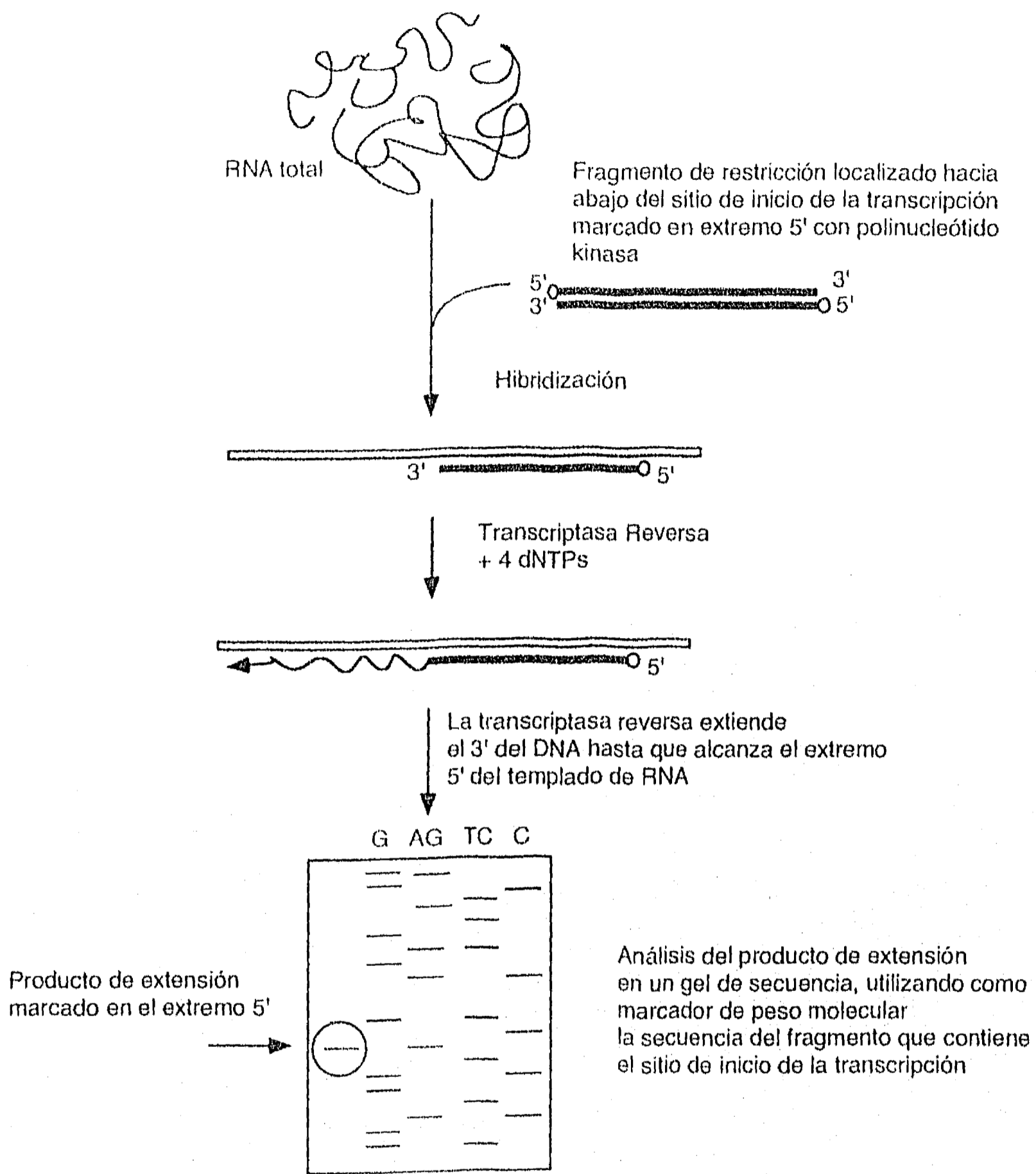


Fig. 11. **Ensayo de extensión de primero.** Es frecuente utilizar oligonucleótidos marcados en el extremo 5' como primeros complementarios al RNA, en lugar de fragmentos de restricción.

El mapeo por nucleasa S1 involucra la formación de un híbrido de RNA:DNA resistente a la digestión por la nucleasa S1, que corta DNA o RNA de cadena sencilla. La sonda es un fragmento que incluye la secuencia del extremo que se quiere determinar. Dependiendo de como se marca la sonda, se pueden identificar los límites 5' o 3' de la molécula de RNA de acuerdo al tamaño del híbrido resistente a la digestión por la nucleasa (Fig. 12). Estas dos técnicas han sido sumamente útiles para identificar extremos 5' y 3' de RNA mensajeros.

El principal inconveniente que presentan estas dos técnicas es que no permiten diferenciar entre un transcrito primario y un transcrito procesado.

IV B- Capping *in vitro* y TAP-RLMPCR

Todos los transcritos primarios que se encuentran dentro de una célula eucariote tienen un extremo 5' característico. Los genes transcritos por la RNA polimerasa II tienen en el extremo 5' una estructura "cap" (Fig. 13), y los transcritos por la RNA polimerasa I, carecen de estructura cap, pero tienen en el extremo 5' un nucleótido trifosfato. En contraste, los transcritos procesados tienen en el extremo 5' un nucleótido monofosfato. La enzima guanidil transferasa (capping enzyme) incorpora específicamente una molécula de GTP a extremos 5' tri o difosfato, y por lo tanto se utiliza para marcar específicamente el extremo 5' de los transcritos primarios. La reacción catalizada por la guanidyl transferasa se muestra en la Fig. 13. El marcaje específico en el extremo 5' del transcrito primario, y la formación de un híbrido RNA:DNA, o RNA:RNA antisentido resistente a la digestión por una nucleasa permite identificar transcritos primarios y permite identificar el sitio de inicio de la transcripción (Fig. 14).

Un trabajo reciente reporta una metodología que ofrece una manera alternativa para identificar inicios de la transcripción, que llamaron **TAP-RLMPCR** (Tobacco Acid Pyrophosphatase- Reverse Ligation mediated PCR) (78). Esta metodología es especialmente útil para identificar extremos 5' de transcritos poco abundantes, porque contiene un paso de amplificación. El

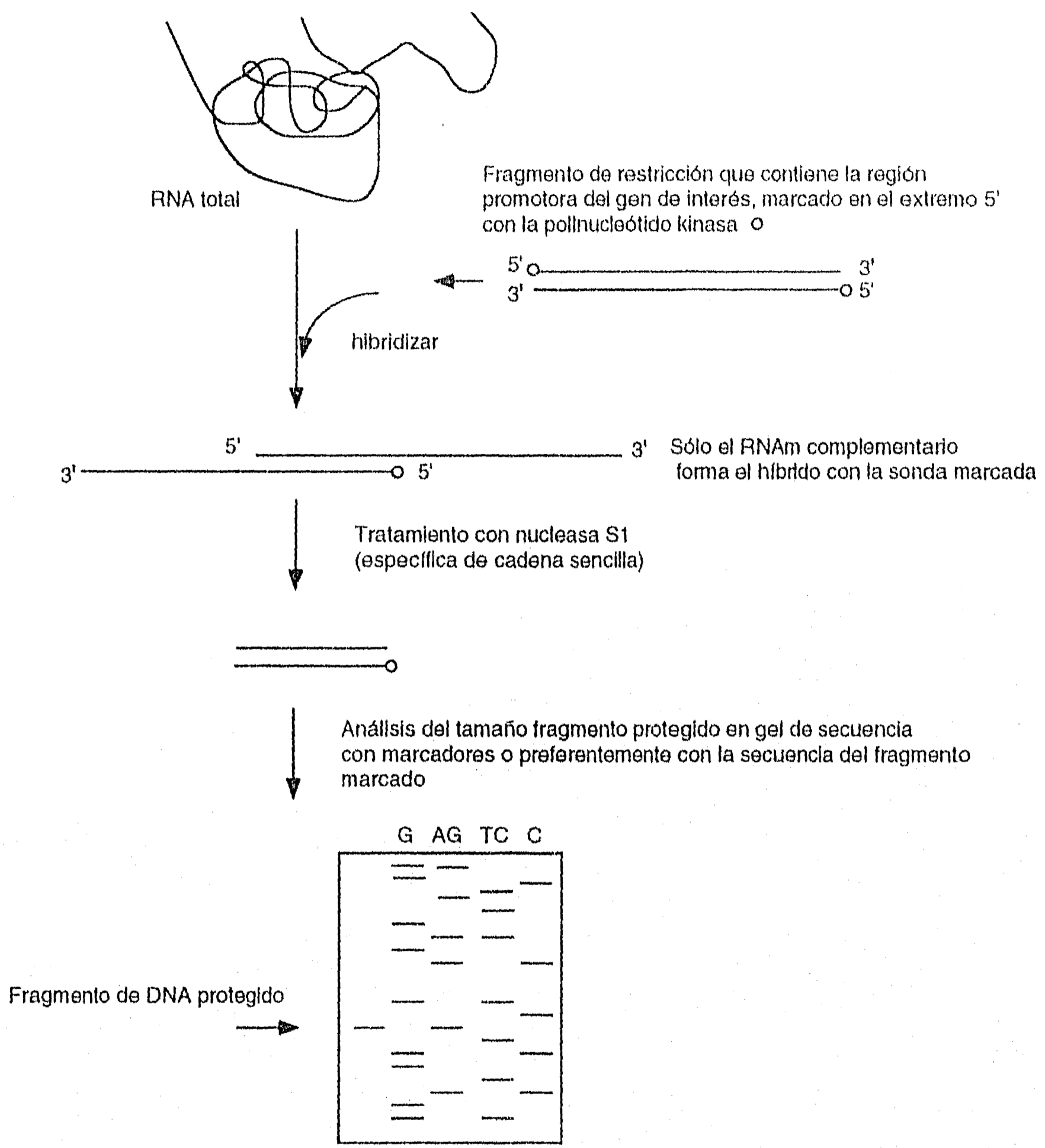


Fig. 12. Mapeo por nucleasa S1.

El mapeo por nucleasa S1 puede utilizarse también para identificar el extremo 3' de un RNA marcando la sonda en el extremo 3', en lugar de marcar la sonda en el extremo 5', como está indicado en el esquema.

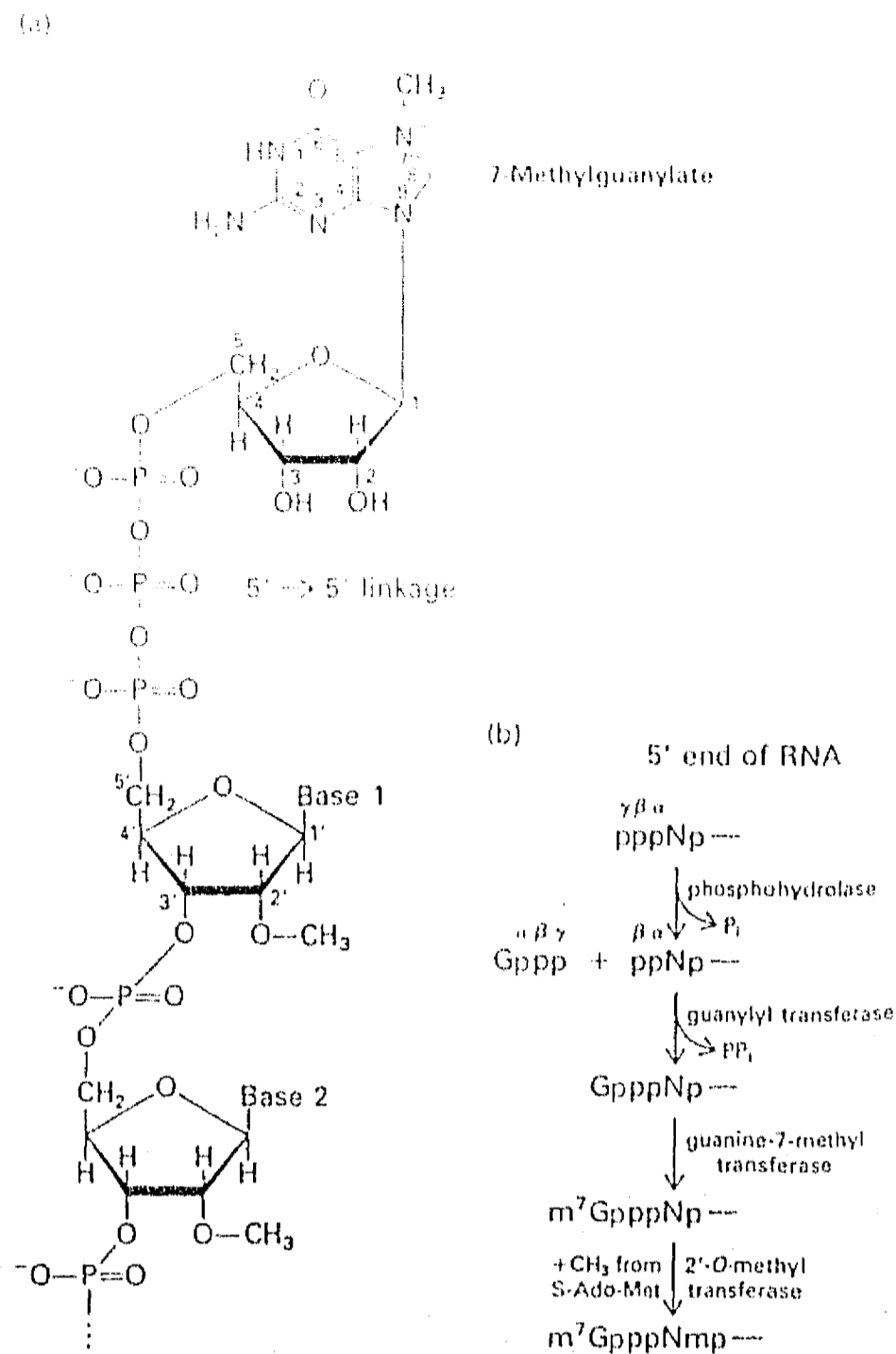
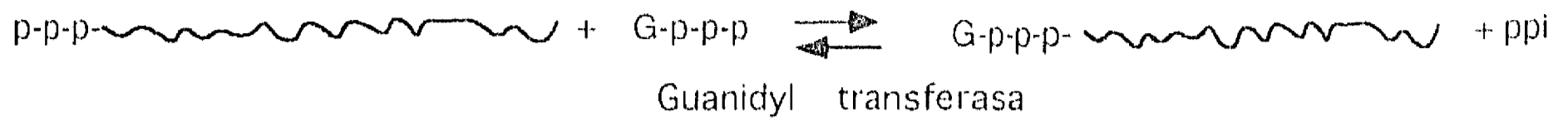


Fig. 13. (a) Estructura del cap 5' metilado de los mRNAs de eucariotes. Los rasgos químicos característicos son la unión 5'-5' del 7-metilguanilato al nucleótido inicial de la molécula de mRNA y el grupo metilo en el hidroxilo 2' de la ribosa del primer nucleótido (base 1). Ambas características ocurren en la mayoría de las células animales y en las células de plantas superiores. (b) Pasos involucrados en la biosíntesis de la estructura cap en los núcleos de las células de mamíferos. La Guanidyl Transferasa cataliza la adición de guanilato (Gppp) sólo a extremos polifosfatos (ppN o pppN). El donador del metilo (CH₃) es la S-adenosyl-metionina.

Venkatesan, S. and Moss, B. (1982) Proc. Natl. Acad. Sci. USA 79:304

Marcaje del transcrito primario:



Ensayo de protección:

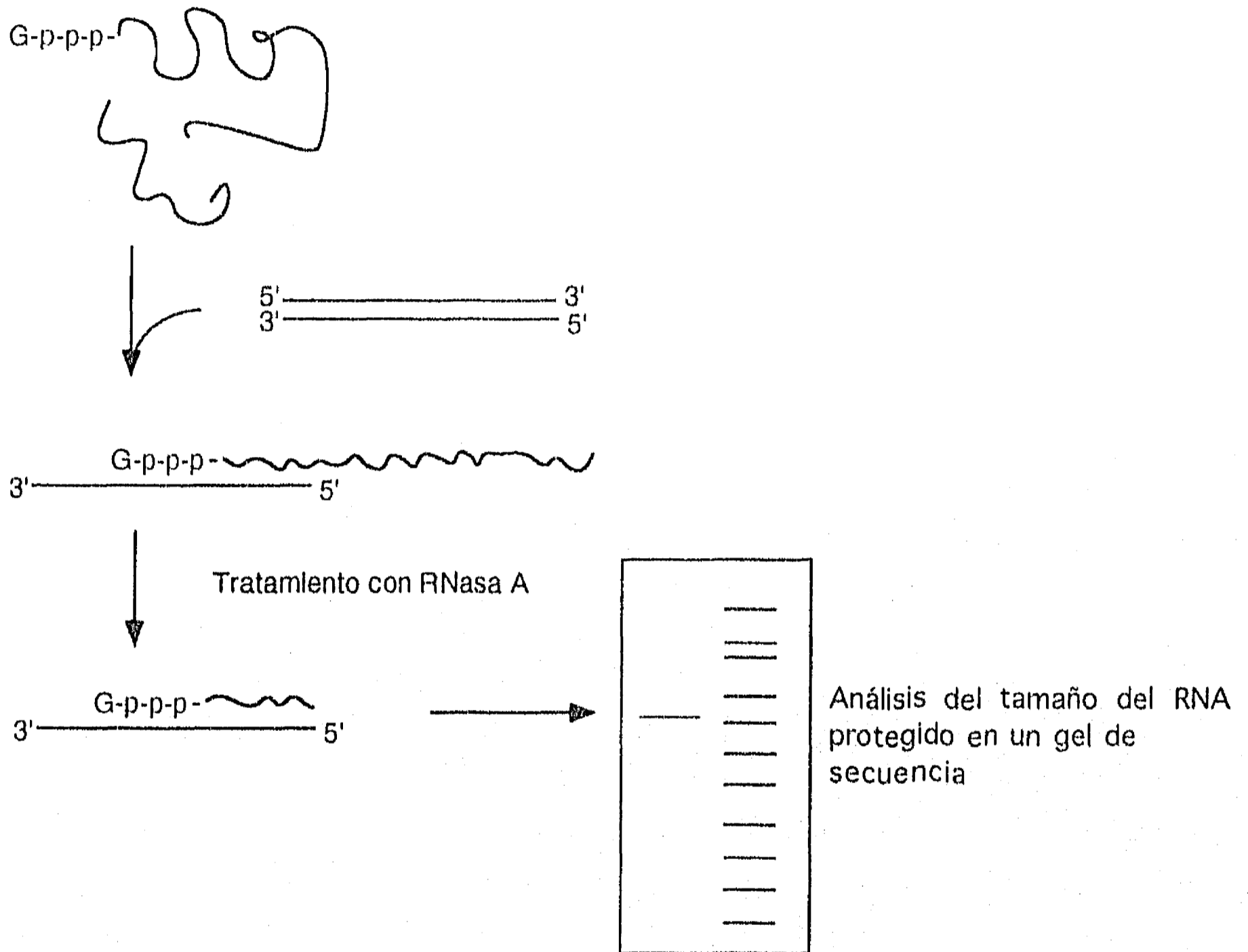


Fig. 14. Capping *in vitro*

A- Marcaje del transcrito primario con Guanidyl Transferasa
B- Protección del RNA marcado con un fragmento de restricción que contiene el sitio de inicio de la transcripción.

método consiste básicamente en hacer a los transcritos primarios susceptibles de ser amplificados. Para ello, el RNA se trata inicialmente con la enzima TAP (por sus siglas en inglés, Tobacco Acid Pyrophosphatase) que genera extremos 5' monofosfato a partir de moléculas de RNA que contienen extremos 5' trifosfatos o con una estructura "cap". Los extremos 5' monofosfato son sustratos de la RNA ligasa para incorporar un linker de RNA. Una vez incorporado el RNA linker en el extremo 5' de las moléculas de RNA, se sintetiza cDNA con un oligonucleótido específico, y éste se amplifica por PCR, utilizando como primeros de amplificación al linker y de preferencia otro oligonucleótido específico, interno al cDNA generado en el paso anterior (Fig. 15).

IVC- Análisis de la secuencia del precursor de los RNA ribosomales

Analizamos la secuencia de los 3000 nucleótidos que se localizan hacia arriba de la región codificadora del 17S ribosomal utilizando herramientas computacionales. Buscamos elementos repetidos en la secuencia que rodea el sitio de inicio de la transcripción haciendo un análisis tipo "dot matrix" (Geneworks, Intelligenetics). Paralelamente, con la idea de encontrar un "consenso" entre las secuencias que constituyen una secuencia iniciadora en diferentes especies, utilizamos el algoritmo de Staden (79) que permite localizar secuencias de reconocimiento poco definidas, mediante la asignación de valores individuales a cada base en cada posición y que por lo tanto indica la importancia relativa de estas. En términos generales, el algoritmo del programa toma las secuencias iniciadoras de las diferentes especies y mediante una matriz de peso, cuenta las veces en que una base se encuentra en una posición determinada. La base de datos para la matriz de peso consistió de las secuencias iniciadoras de 14 especies, las cuales escogimos como representantes de diferentes fila.

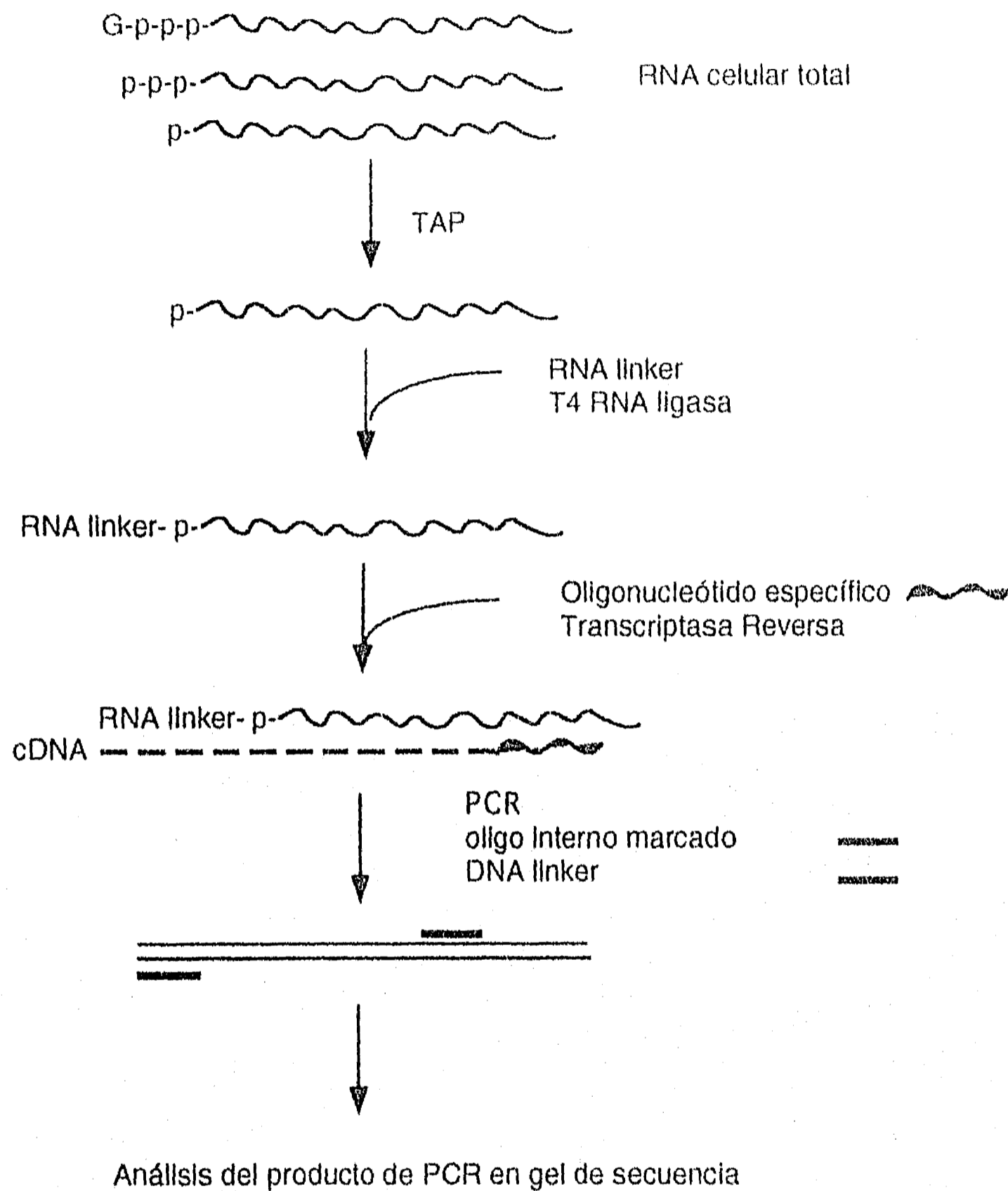


Fig. 15. TAP-RLMPCR

El RNA celular total se trata primero con la enzima TAP, para luego ligar a los extremos 5' monofosfatos generados un "linker de RNA". Posteriormente, se sintetiza cDNA a partir de un oligonucleótido específico, y finalmente, se hace una reacción de PCR, utilizando un oligonucleótido interno marcado radioactivamente, y un primer idéntico al linker de RNA.

IV D- Footprint *in vivo*

Los experimentos de "footprint" han sido sumamente útiles para determinar los sitios de interacción de una proteína con el DNA, ya que el "footprint" (o huella) se obtiene cuando se trata un complejo DNA-proteína con agentes químicos y/o enzimáticos, de manera que solamente las regiones del DNA no protegidas por la proteína son susceptibles de reaccionar. Las regiones protegidas se interpretan como "huellas" que dejan las proteínas al secuenciar el DNA. Estos experimentos se han realizado comunmente *in vitro*. Sin embargo, en las células eucariotes el DNA genómico forma una estructura compacta y organizada con histonas y otras proteínas, llamada cromatina, de forma que cambios en su estructura (p.e. compactación) o modificaciones en el DNA (p.e. metilaciones) pueden modular el acceso de factores de transcripción a sus secuencias blanco; por lo tanto, las interacciones DNA-proteína que se observan *in vitro* no necesariamente ocurren *in vivo* (80). El footprint *in vivo* involucra los siguientes pasos: primero, modificar el DNA químicamente (o enzimáticamente) en su estado *in vivo* con un reactivo capaz de penetrar dentro de las células; segundo, extraer el DNA y finalmente, detectar los sitios que se modificaron químicamente. De los reactivos químicos que se utilizan para hacer footprint *in vivo*, nos referiremos particularmente al DMS. El DMS metila la posición N7 de las guanosinas y la posición N3 de las adenosinas del surco mayor (81). Un tratamiento químico posterior con piperidina rompe específicamente las guanosinas metiladas por el DMS (81). La detección de los sitios modificados es el paso más difícil de lograr, sobre todo para genes que están poco representados dentro de un genoma complejo. Para detectar las modificaciones químicas sobre el DNA, es indispensable poder hacer secuenciación genómica dado que es la única herramienta que nos permite detectar diferencias químicas entre un nucleótido y otro.

Con el fin de identificar posibles regiones regulatorias en la región que rodea el sitio de inicio de la transcripción, empezamos por implementar una técnica que nos permitiera detectar las bases modificadas en el DNA haciendo secuenciación genómica sobre el episoma ribosomal, y la aplicamos para detectar la reactividad del DNA ambiente al DMS.

V- Materiales y Métodos

Materiales

Los trofozoitos de *E. histolytica*, cepa HK9, se crecieron axénicamente en el medio TYI-S-33 (82). Obtuvimos otros materiales como enzimas de restricción, DNAsa I y la enzima Transcriptasa Reversa de Boehringer-Mannheim. Obtuvimos los nucleótidos radioactivos de New England Nuclear. Las reacciones de secuencia se hicieron usando un kit de Secuensa 2 de United States Biochemical Corporation, y las manipulaciones clásicas del DNA de acuerdo a los protocolos descritos por Sambrook y col (83).

Aislamiento de la region 5' de los genes que codifican para los RNA ribosomales de E. histolytica

Se construyó una librería genómica de la cepa HK9 clonando digestiones parciales y totales del DNA de la cepa HK9 cortado con *Eco* RI en el fago λ gt-10. La biblioteca genómica contiene 5×10^5 clonas recombinantes independientes, y el tamaño promedio de los insertos es de 3Kb. Tamizamos la biblioteca genómica con un oligonucleótido sintético específico del gen 17S ribosomal de *E. histolytica* (Oligo 17S; 5'-GTGTTCTACAATTGTCGTGG-3'). Este oligonucleótido es complementario a los nucleótidos 251-273 de la secuencia del gen 17S de *E. histolytica* (84). Se utilizó el protocolo de hibridización descrito en (83).

Purificación del RNA

RNA celular total se purificó de acuerdo al protocolo descrito en (85). Brevemente, los trofozoitos en fase logarítmica de crecimiento se cosecharon y lavaron 2 veces en PBS, y se resuspendieron en isotiocianato de guanidina 6M. Después de homogenizar vigorosamente, se eliminaron los restos celulares. El sobrenadante se precipitó dos veces con 0.3 volúmenes de etanol. El pellet se resuspendió en solución TE (10mM Tris/HCl -pH8, 1 mM

EDTA), y se extrajo dos veces con 1 volumen de fenol/cloroformo y se precipitó con 1/10 de acetato de sodio 3M y con 2.5 volúmenes de etanol.

Análisis del RNA y del DNA en ensayos tipo "southern" y "northern"

El RNA se separó por electroforesis en geles de agarosa/formaldehído de acuerdo al protocolo descrito en (86), y se transfirió a filtros de nitrocelulosa. Cuando se usaron fragmentos de DNA como sondas, se hibridó toda la noche a 42 °C en 50% formamida, 5XSSC, 1X solución de Denhart's, y 100 µg/µl de DNA de esperma de salmón. Se lavaron los filtros a 55° C en 0.1X SSC y 0.1 % SDS. Las hibridaciones con el oligonucleótido 17S se hicieron de acuerdo a Grosschandel y col (87). Los ensayos tipo "northern" se hicieron de acuerdo a Goldberg y col (88). Los transcritos de RNA naciente se hibridaron de acuerdo a Kooter y col (89).

Análisis de la secuencia de DNA

Se utilizó el método de los didesoxinucleótidos descrito por Sanger para hacer las reacciones de secuencia (90). Los fragmentos **a** (*Hind* III- *Xba* I) y **b** (*Bgl* II- *Hind* III) se insertaron entre los sitios *Hind* III y *Xba* I y entre los sitios *Bam* HI y *Hind* III del M13 mp18 y M13mp19, respectivamente para generar cuatro subclonas independientes. El oligonucleótido 17S se utilizó como primero para empezar a secuenciar el fragmento **a**. Se generó la clona p3.8 insertando al fragmento **c** (*Eco* RI- *Bgl* II) entre los sitios *Eco* RI y *Bam* HI del vector pbluescriptKS, y se usó como fásmido para producir templado de cadena sencilla (ver Fig. 16 para localizar los fragmentos **a**, **b** y **c**). Se usaron tres oligonucleótidos sintéticos (oligo 2: 5'-CTCTGATGGTCCTAGATTAAC-3', oligo 3: 5'-GTTATACTCCTCCCCCTGGTATT-3' y el oligo C : 5' TAGAAAACCAAGAAAATT- 3', complementarios a los nucleótidos 1556-1576, 1722-1745 y 340-358, de la secuencia que se presenta en la Fig. 16 respectivamente) para las reacciones de secuencia de la clona p3.8. Utilizamos las enzimas *Hae* III y *Alu* I para subclonar fragmentos de la clona

p3.8. Por ensayos tipo "southern", utilizando el RNA naciente como sonda, se identificaron los fragmentos *Hae* III, de 1.7 kb, y *Alu* I, de 1.4 kb, como parte de la unidad transcripcional de los genes de RNA ribosomal. Estos fragmentos se subclonaron en el m13mp18. Se utilizó un oligonucleótido sintético (oligo B1783: 5'-ATGAACATAACCTTTGTA-3') idéntico a los nucleótidos 1683-1700 de la secuencia *Eco*RI-*Taq* I (91) para secuenciar la hebra opuesta de la región entre el sitio *Taq* I-*Hae* III (ver mapa de restricción de la Fig. 1).

Ensayo de RNA naciente

Se prepararon los núcleos rompiendo las células en una solución de lisis (15 mM Tris/HCl, pH 8, 15 mM NaCl, 0.25 mM Cl_2Mg y 0.5% NP40) en un homogenizador manual. El lisado se colocó en un colchón de sacarosa 1M (10 mM Tris/HCL pH8, 1 mM EDTA) y se centrifugó a 5000g 10 min a 4°C. Los núcleos se congelaron en 50 mM Tris HCl pH 8, 75 mM NaCl, 0.5 mM EDTA pH 8, 200 mM PMSF, 5 mM DTT, y 50% glicerol y se guardaron en nitrógeno líquido. El RNA naciente se marcó siguiendo el protocolo de Kooter y col (89). Brevemente, los núcleos se descongelaron en hielo, se centrifugaron para eliminar la solución de congelación, y se resuspendieron en 100 mM Tris/HCl, pH 8, 0.5 mM NaCl, 50 mM KCl, 2 mM Cl_2Mg , 4 mM Cl_2Mn , 0.15 mM espermidina, 2 mM DTT y glicerol al 25%. La reacción de elongación se inició con la adición de 2 mM ATP, 2 mM GTP, 2 mM CTP, 10 μ M UTP y 0.2-0.5 μ Ci de [^{32}P]- UTP (3000 Ci/mmol) en un volumen final de 100 μ l. La reacción se incubó a 37 °C al cabo de los cuales se añadieron 27 U de DNasa I y se prosiguió la incubación a 37°C por 5 min más. Se paró la reacción incubando 15 min a 37°C y añadiendo un volumen equivalente de 10 mM Tris/HCl, pH 8, 1 mM EDTA, 1% SDS y 0.5 mg/ml de proteinasa K. Finalmente, se extrajo con fenol/cloroformo y el RNA se separó de los radionucleótidos libres pasándolo a través de una columna de Biogel P-10. El RNA naciente marcado se utilizó como sonda en ensayos tipo "southern" de acuerdo a Kooter y col.

Capping in vitro del precursor de RNA ribosomal

El RNA total de *E. histolytica* se fraccionó utilizando una columna de Sepharosa 4B. Se precipitó la fracción de alto peso molecular. La reacción de capping se hizo de acuerdo a Mulligan y col. (92), con algunas modificaciones. Las reacciones de marcaje se hicieron en 20 μ l con 13 μ g de RNA de alto peso molecular, 50 mM Tris HCl pH8, 1.25 mM Cl_2Mg , 2.5 mM DTT, 0.5 mM SAM, 10 U de guanilyl transferasa (GIBCO BRL) y 120 μ Ci de α - ^{32}P -GTP. Después de incubar 30 min a 37 °C, se paró la reacción usando 20 μ g de proteinasa K por 15 min a 37 °C. El RNA marcado se extrajo una vez con fenol y una vez con cloroformo. El α - ^{32}P -GTP no incorporado se eliminó con 2 precipitaciones con acetato de amonio 2.8 M y 75% de etanol. La incorporación del isótopo al RNA dió una actividad específica de 1.2×10^6 cpm/ μ g.

Protección de la digestión por RNAsa del RNA marcado in vitro por la Guanidil transferasa

Los ensayos de protección por RNAsa se hicieron de acuerdo a Mulligan y col. (93), con modificaciones menores. Se desnaturalizaron 400 ng de DNA de doble cadena y 1.5 mg de RNA marcado *in vitro* a 75 °C por 5 min., y se hibridaron por 5 horas a 43 °C en presencia de 80 % formamida. Se añadieron 0.3 ml del buffer de digestión de la RNAsa A (10 mM Tris HCl pH 7.7, 300 mM NaCl, 5 mM EDTA pH 8.0) con 100 ng de RNAsa A. Se digirió el heteroduplex 1 hora a 37 °C. la reacción se paró con 20 μ l de SDS 10% y 40 μ g de proteinasa K 15 min a 37 °C. Se extrajo con fenol-cloroformo y se precipitó con etanol. Los fragmentos protegidos se analizaron en un gel de secuencia al 5% y se expuso en un fosforimager (Molecular Dynamics).

Mapeo por nucleasa S1

Los ensayos de protección por nucleasa S1 se hicieron de acuerdo a Ausubel y col (86). Los fragmentos resistentes a la digestión por la nucleasa S1 se analizaron en un gel de secuencia al 5%.

Análisis de extensión de primero

Los ensayos de extensión de primero se hicieron de acuerdo a Mulligan y col (93). Un oligonucleótido sintético complementario a los nucleótidos 607-628 (ver Fig 16, oligo V: 5'- TTCGTTTTCTCATTCTATTCA-3' se marcó en el extremo 5' por fosforilación y se hibridó 3 horas a 42 °C a 15 µg de RNA celular total. La extensión a partir del primero se hizo 60 min a 43 °C por 60 min con 50U de transcriptasa reversa (GIBCO BRL). La reacción se paró añadiendo 20 mM EDTA y 2 mg/ml de RNasa A. El cDNA se extrajo con fenol/cloroformo, se precipitó, y se analizó en un gel de secuencia al 5%, utilizando como marcador de peso molecular la secuencia p3.8 iniciada con el mismo primero.

TAP-RLMPCR

Se siguió el protocolo de Fromont-Racine y col (78). Brevemente, el RNA total de *E. histolytica* se trató 30 min a 37 °C con 1 U de TAP (Epicentre) en 10 µl del buffer TAP (Epicenter) con 2.5 µM de ATP. Se tomó 1/10 de la reacción como sustrato en una reacción de ligasa con 36 pmolas del linker de RNA (Cybersyn, Inc) en presencia de 6U de T4 RNA ligasa, 45 min a 37°C. Se generó cDNA utilizando 1/10 de la reacción de ligasa a partir del oligo 3: 5'- GTTATACTCCTCCCCCTGGTATT-3', complementario a los nucleótidos 1722-1745 de la secuencia que se muestra en la Fig. 16. Se amplificó el cDNA usando condiciones estandar de PCR con 3 pmolas del primer V marcado en el extremo 5', 6.25 pmolas del linker primer en 20 µl de reacción. Después de desnaturalizar 5 min a 95 °C, la reacción se repitió 30 veces a 94 °C por 1 min, 48 °C, 1 min y 72 °C, 45 sec. Finalmente, la reacción se incubó 7 min a 72 °C y 7 min a 52 °C. La secuencia del linker primer es 5'- d(GTTTTCCCAGTCAGGAC)-3', y es idéntica a la del RNA linker.

Secuencia genómica y metilación in vitro del episoma ribosomal

Se siguió el protocolo de secuenciación del kit de secuencia ampli-cycle (Perkin-Elmer). Para la reacción de secuencia se utilizaron 4 μg de DNA total de *E. histolytica* cortado con *Eco* RI y 1 pmol de oligonucleótidos fosforilados (oligonucleótido V, y oligonucleótido up325, 5'- GGC GTC ATC TTT TAA ATT TTC T-3', el cual es idéntico a los nucleótidos 325-345 de la secuencia que se muestra en la Fig. 16), con una actividad específica de 1×10^8 cpm/ μg . La reacción se repitió 25 veces 1 min a 95 °C, 1 min a 48 °C, y 1 min a 72 °C.

Como control del patrón de metilación en ausencia de proteínas, el DNA de la cepa HK9 se metiló *in vitro* siguiendo el protocolo convencional de Maxam y Gilbert (81), modificado por Morett, y col. (94). Para el caso de *E. histolytica*, tratamos 5 μg de DNA de la cepa HK9 con 0.05% de DMS diluido en buffer de cacodilato (100 mM Cacodilato, pH 7.0) por 1 min a temperatura ambiente. La reacción se paró con 0.5 volúmenes de DMS "stop solution" (acetato de sodio 1.5M, pH 7.0, β -mercaptoetanol 1M), y el DNA se precipitó con etanol, y se redisolvió en 0.3 M de acetato de sodio y se precipitó nuevamente. Después de lavar y secar el DNA, se resuspendió en H₂O a una concentración de 0.5 $\mu\text{g}/\mu\text{l}$. Posteriormente, el DNA tratado con DMS se rompió en los residuos metilados de guanosinas usando la reacción única de las Gs de Maxam y Gilbert (81). Se añadió la piperidina al DNA a una concentración final de 1M y se calentó la mezcla a 90°C por 30 min en un tubo cerrado. La piperidina se eliminó mediante tres liofilizaciones consecutivas añadiendo cada vez 100 μl de H₂O. El DNA se resuspendió finalmente en H₂O a una concentración de 0.7 $\mu\text{g}/\mu\text{l}$. Las reacciones de extensión se hicieron con 5 μg de DNA y 1pmol de los oligonucleótidos fosforilados (oligo V y oligo up325) con la misma actividad específica que la que se usó para la reacción de secuencia. La reacción de extensión se cicló 25 veces 1min a 94°C, 1min a 48°C y 1min a 72°C. Los productos de extensión se resolvieron paralelamente a la reacción de secuencia en un gel de poliacrilamida al 6%.

VI- Resultados

VIA- Mapa de restricción y estrategia de secuenciación de la clona genómica que contiene el extremo 5' de la unidad transcripcional de los genes de RNA ribosomal de *E. histolytica*

Durante mi trabajo de maestría, aislé la clona genómica λ gt10-6.8 de un banco de la cepa HK9 que contiene el extremo 5' de la unidad transcripcional ribosomal de *E. histolytica* y por ensayos de RNA naciente, limité el extremo 5' de la unidad transcripcional. Secuencié en ese periodo la región comprendida entre el oligonucleótido 17S ribosomal y el sitio de restricción *Bgl* II de la clona λ gt10-6.8 (Fig. 16). En el doctorado, secuencié la región comprendida del sitio de restricción *Bgl* II al sitio de restricción *Taq* I. El mapa de restricción, la estrategia de secuenciación y la secuencia de 2678 pares de bases hacia arriba del gen estructural 17S de *E. histolytica* se muestran en la Fig. 16.

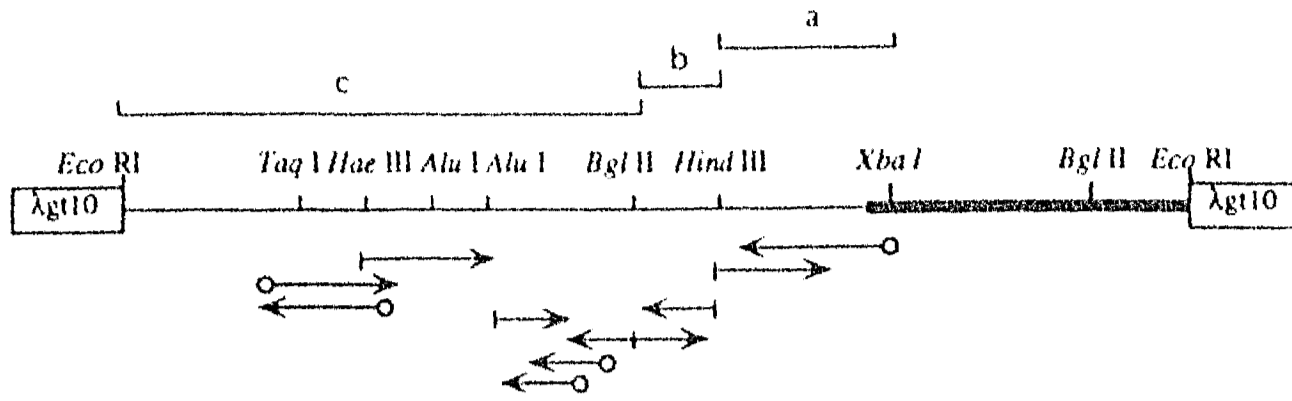
VIB-Resumen de resultados:

Los resultados de esta tesis están publicados en el trabajo : "Identification and analysis of the transcription initiation site of ribosomal RNA transcription of *Entamoeba histolytica*" que apareció en la revista Molecular and Biochemical Parasitology, Vol 73, 19-30 (1995), y esta anexado en esta tesis.

(El número de las figuras en esta sección corresponde al número de las figuras del artículo.)

VIB.1- Mapeo tipo run-on de la unidad transcripcional ribosomal.

En la Fig. 1 se muestra la hibridación del RNA naciente marcado con [³²P]-UTP con diferentes fragmentos de restricción de la clona genómica λ gt10-6.8. El panel A de la figura indica que el RNA naciente marcado hibrida con los fragmentos de restricción que contienen parte de la región codificante del gen estructural 17S: los fragmentos a y b (línea 4) y con el fragmento d (línea 5). Se observa también una fuerte señal con los fragmentos e y c, que se

A**B**

1 CGACACGAGA AAAAGTTCAA CAAAAAGTGC AAAAACCCCC CCCCTACTCA AACTACGAAT TTTGAAAATG
 71 CCATTTCTAC GAATTTTACC ATTTTTTGT TGA AATTGGA AGGAATGGAA TAAAAAATT GAAATTTCAA
 141 TGAAAAGTGA ACCTTTATTA CACAAAGTCT GGGAAAAACC ATCAGATTAA CACTAGACCA AGGATTTAAG
 211 TTGGTAAAG CTCTATTTGC CGAGAAAAGA TGACGTGTGA ACAACTTTTT TTCGGCATT TCGGCCGATA
 281 TATTTGTATG GATATCTGA ATATATTCT ATTTTAGAA ATATGGCGTC ATCTTTTAAA TTTCTTGGT
 351 TTTCTAAAA TAAATAACCG TTTATGAATA AAAAAATGTT AATAGTAAA TTATAGTAAA TTA ACTATAA
 421 AATTATAGTT TGATGTGGAT TTTTCAAAAT TTTCGTTAGC TACAAATAGA CCAATAGAAT ACTTGTTAGT
 491 GAAAATATAC TATACAGGAG AAGTCGGTTT GAGGTCAGTT GAACAGTTTT TAAAGAAAAG CAAAAAACA
 561 TCTTCAATTG AAATCAGCTA TAAATAACGC CAACGAAAGA TGG AATGAAT AGAATGAGGA AAACGAAAAC
 631 GAAAGATGGT TGAGTTAGAA CGAAAGATGG TTGAGTTAGA ACGAAAGATG GTTGAGTTAG AACGAAAGAT
 701 GGTGAGTTA ACACGGAGGC ATTAAGTAAT TGAGAATAAA AAATGATTAA TTTAAAGATT AAGCATATTA
 771 GAGATTGAAA AGAAAATCAA ACAAAAAAAG AATTTAAGA TTAAGCAGAT TAGAATAGA AAAAAAAT
 841 GAAACAAAA AACAAATGAT ATTTAGTAAT GAAGAAATGA AATTGATAAA AGATAGAAGA TTGTTCACGA
 911 GATTAAGTA TTGTGTTTCT GGACCGATTG TAATGCAAAG ATTTCAAGCG TCAGAAAAAT GATAGTGAAT
 981 TTAGAATATC TGAAAATGAC CAAATAATTA AGAAGAAAT AATGGAGAAA TGAAAATTAAG AAATTTATATA
 1051 AGATTTTCA CGAGATTA AAATTTGTGT TCTGGACCGA TTCTAATGCA AAGATTTTAA GCGTAGCAAA
 1121 AATGATAGTG AATTAGAAT ATCTGAAAAT GACCAAAATA TTAAGAAGAA ATTAATGGAG AAATGAAAT
 1191 AAGAAATTAT ATAAGATTTT TCACGAGATT AAAATATTGT GTTCTGGAC CGATTCTAAT GCAAAGATT
 1261 CAAGCCTAGC AAAAATGATA GTGAATTTAG AATATCTGAA AATGACCAAA TAATTAAGAA GAAATTAATG
 1331 GAGAAATGAA ATTAAGAAAT TATATAAGAT TTTTACCAG ATTAATAAT TGTGTTTCTG GACCGATTCT
 1401 AATGCAAAGA TTTAAGCGT AGCAAAAATG ATATAGATTT TGGATGATCA TGGACAATGG ATAAAAAAA
 1471 GAAAATCTAA AAATAAAAA TGAAAGGGAA ATTAAGGATG AGAATTGACC AAAAAATGTC TCATTTTCA
 1541 CGAGATTA AAATGTTAAT CTAGGACCAT CAGAGATGCA AAGATTTTAA TTTGTTGAAA AATGACGAGG
 1611 TTTTGGACC ATCAAAAAA AAGAGAAAT CTAAAAAAA AAATGAAAGG GAAATTAAGG ATGAAAATTG
 1681 ACCAAAAAT ATCTCATTTT TCACGAAAT CAAGTATGTT AATACCAGG GGGAGGAGTA TAACATATTT
 1751 GAATTTCTG AAGAATCATG ACATTTGAT CAAATAGCAT CTGACAATT AAAAAATAGAA CAAAAATGAA
 1821 ATAGGTA AAA TAGGGTAAA TAAAAAAT AAAAAAAATAG TAAAAAAT AATAATAAAA AAAAAAGAGA
 1891 AAAAAATAAA AAAAAAGACT GAAGAATAA AATAATATCA AATTAAGGAG ACAATAGGA TGAACAACA
 1961 AGTAAAAATA AAAAAAGAAA AAAATAAAA AGAAGA ACT AAGATCTAAA ATAATATCAA ATCAAGGGGA
 2031 CAAATAGGAT GAACAACAAA ATAAGGAGTA AAATGAAAT GAAAAAATAA TGAAAGCTAG GAAGAGACCA
 2101 CTAGAATAGT TGACATTA AAATATATTA TATGTACTAG TGTACAAGAA GAGTAAAAA AATAATATGT
 2171 TTGTTTATC ATTCTAAAC AAAATACTAA CAATTAGATG AACTACTGTT CTCAAAGAAC CCACGAGAAA
 2241 CAAGTCATGT AATTAGTATT AGTATGTTG TAAGTAGTAA TGAGTAAGTT TGTAGCTAGT ACTGAGTATG
 2311 GTTGTAAGTA GTAATGAGTA GTTAACTAC TCTTAAGGTA ATATTGATTA ATGTATGTGA GTAGTATTAG
 2381 TTATGATAAT CATCTTAAG TAGTATATTT ATTAGACTGA TGTGTAGGA AGCTTTAGTT TTATGGACCA
 2451 TAAGAGCAAT TAATGAATAA GGATGATTA TAATAATTA TGCTATGATG AGAAGAGTTG ATAAGTTTCC
 2521 TTTATAATTG TGCAGATAAT ATATAAAGG TTTGTTGTTG TATTTTAATA AGTAAAAATA AGACTAACAA
 2591 TAGAGTTAAT AAATAATAAT AGGTGAGTAT AAATAAGTAA TCTCATAGAG GAGATTAAT AAATAAGAA
 2661 GAAGTTCTTA TTAAGAAGAT TTAATAGGAA CATAATAAAT AGATGTTGTT AAGAAAAGAGA GAAATAATA
 2731 AAGAAAGATA AGACTGAATA AGAAGAAAAT TATACTTATG CTGATTATAT AAAATGTAAT TAGATGACAG
 2801 AACAGATTA TATTATATAA GTTTGTTTT AGTATAATTA GTAATAAGA AGAAGTCAA AGTAGGAGAA
 2871 TAAAGAAAGG AAAAATCGAT ATAAATAATA AATACAAATA AGAATATTA AATCTTCTT TAAGAAGGAA
 2941 GAATAATAAA ATATCTGTT CATCTGCTCA GTATATATG CTGATGTTAA AGAATAAGCC ATGCAATGTT
 3011 AAGTAAAG ACCAAGTAGG ATGAAACTGC GGACGGCTCA TTAATACAGT AATAGTTTT TGGTATAGTA
 3081 AAATACAAGC ATAGCTTTGT GAATGATAAA GATAATACTT GAGACGATCC AGTTGTATTT AGTACAAAA
 3151 GGCCAATCA TTCAATGAAT TGAGAAATA CATTCTA

Fig. 16. Estrategia de secuenciación y secuencia del fragmento *Xba*-*Taq*1.

En A, se muestra la estrategia de secuenciación. Los círculos abiertos indica que diferentes oligonucleótidos se utilizaron como primeros de secuencia. Las barras verticales indican que los fragmentos se subclonaron y se secuenciaron. Para más detalles, ver Materiales y Métodos.

En B se muestra la secuencia del fragmento *Taq* 1- *Xba* 1. La región codificadora está subrayada.

encuentran lejos hacia el 5' del inicio de la región codificante del gen 17S. El panel B muestra la hibridación del RNA naciente con delecciones del fragmento c con Bal 31. Se observan fragmentos generados por la digestión con Bal 31 de 3.1 kb, 2.7 kb, 2.3 kb y de 1.6 kb. El RNA naciente hibridó con el fragmento c no digerido con Bal 31, y con los fragmentos de 3.1Kb, 2.7 Kb, y 2.3 Kb, pero no hibridó con el fragmento de 1.6 Kb. Este resultado localiza el extremo 5' de la unidad transcripcional ribosomal en la región que se localiza entre los productos de 1.6Kb y 2.3 Kb obtenidos por la digestión con Bal 31.

VIB.2- Mapeo del sitio de inicio de la transcripción por ensayos de protección de nucleasas: mapeo por S1 y capping in vitro.

De acuerdo a los resultados de la Fig. 1, utilizamos un fragmento de restricción *Hae* III- *Ase* I de 474-pb como sonda para determinar el extremo 5' de la unidad transcripcional ribosomal en ensayos de protección con nucleasa S1. El panel A de la fig. 2 muestra el mapeo por S1. Se observan tres híbridos resistentes a la digestión por la nucleasa S1, con tamaños de 240, 225 y 215 pb (Fig. 2, panel A, líneas 4-6). Estos sitios mapean en los nucleótidos 509 ± 5 , 523 ± 5 y 533 ± 5 , en la secuencia que se muestra en la Fig. 4. El panel B de la Fig. 2, muestra el capping *in vitro*. En la línea 2 del panel B se observa el RNA marcado *in vitro* con la Guanidil Transferasa. Las líneas 3 y 4 muestran la presencia de un híbrido único resistente a la digestión por RNasa con un tamaño de 240 nt formado entre el RNA marcado y el fragmento de restricción de 474 pb. La protección de un fragmento de 240 nt en los dos ensayos de protección indica que este fragmento es el transcrito primario y lo sitúa cerca de la posición 509 ± 5 en la secuencia que mostramos en la Fig. 4.

VIB.3- Mapeo de alta resolución del sitio de inicio de la transcripción: ensayo de extensión de primero y TAP-RLMPCR

En el panel A de la Fig. 3, se muestra el ensayo de extensión de primero iniciando con el oligonucleótido V. Están señalados en la figura los

dos productos de extensión más largos, cuyos tamaños son 123 y 124 nt (línea 1). El tratamiento del RNA con RNAsa (línea 2) no generó producto de extensión alguno, lo cual indica que los productos de extensión generados en la línea 1 son específicos del RNA. El producto de extensión de 124 nucleótidos localiza un sitio de inicio de la transcripción en la posición 505 de la secuencia que mostramos en la Fig. 4.

Para confirmar los datos que obtuvimos con los ensayos de protección de las nucleasas y con el ensayo de extensión del primer, hicimos un experimento adicional que es el TAP-RLMPCR. Utilizamos para amplificar el transcrito primario un oligonucleótido idéntico al linker primer y el oligonucleótido V, que fué el que utilizamos para el ensayo de extensión de primer. El producto amplificado por PCR sobre el cDNA es de 141 nt. Tomando en cuenta la longitud del linker primer más dos adenosinas añadidas por la actividad de transferasa terminal de la Taq polimerasa en cada extremo 3' del fragmento de doble cadena (95), el extremo 5' mapea en la misma posición que el producto extendido en el ensayo de extensión de primer, en la posición 505 de la Fig. 4.

VIB.4- Análisis de la secuencia de la región promotora del precursor del RNA ribosomal.

En la Fig. 4 se muestra la localización de dos familias de elementos repetidos que se encontraron utilizando el Geneworks en la secuencia del líder de los rRNA. Ambas familias de elementos repetidos están hacia abajo del sitio de inicio de la transcripción; una de ellas consiste de 4 repetidos perfectos de un elemento de 21pb localizados 123 nt hacia abajo del sitio de inicio de la transcripción, y la otra consiste de 4 repetidos imperfectos de 118, 149, 149 y 148, localizados 376 nt hacia abajo del sitio de inicio de la transcripción.

Analizamos los 3000 nt que se localizan hacia arriba del gen 17S ribosomal utilizando el programa PatternScan, que tamizó toda la secuencia utilizando una matriz de peso, generada a partir de un consenso obtenido del análisis de secuencias iniciadoras de 14 diferentes especies. Los resultados

se muestran en la Fig. 5. El panel A muestra el consenso y el orden que el programa da a las diferentes secuencias de acuerdo a su similitud con dicho consenso. De todas éstas, solamente la secuencia iniciadora de *Crithidia fasciculata* no fue agrupada junto a las 14 especies restantes, por diferir de manera considerable con el consenso generado por el programa. El panel B muestra el tamizado de los primeros 1600 nt que aparecen en la Fig. 16, indicando los valores de similitud de acuerdo a lo mostrado en el panel A. El resultado del programa muestra un valor de similitud máximo en la posición 505 de los primeros 1600nt de la secuencia que se muestran en la Fig. 16, y es notable que esta posición es exactamente el sitio inicio de la transcripción que identificamos experimentalmente. Finalmente, introdujimos la secuencia iniciadora ribosomal de *E. histolytica* a la base de datos junto con las 14 especies restantes. Esta última quedó localizada en décimo lugar de acuerdo a las 14 especies restantes, y nuevamente, la secuencia iniciadora de *C. fasciculata* quedó fuera del grupo.

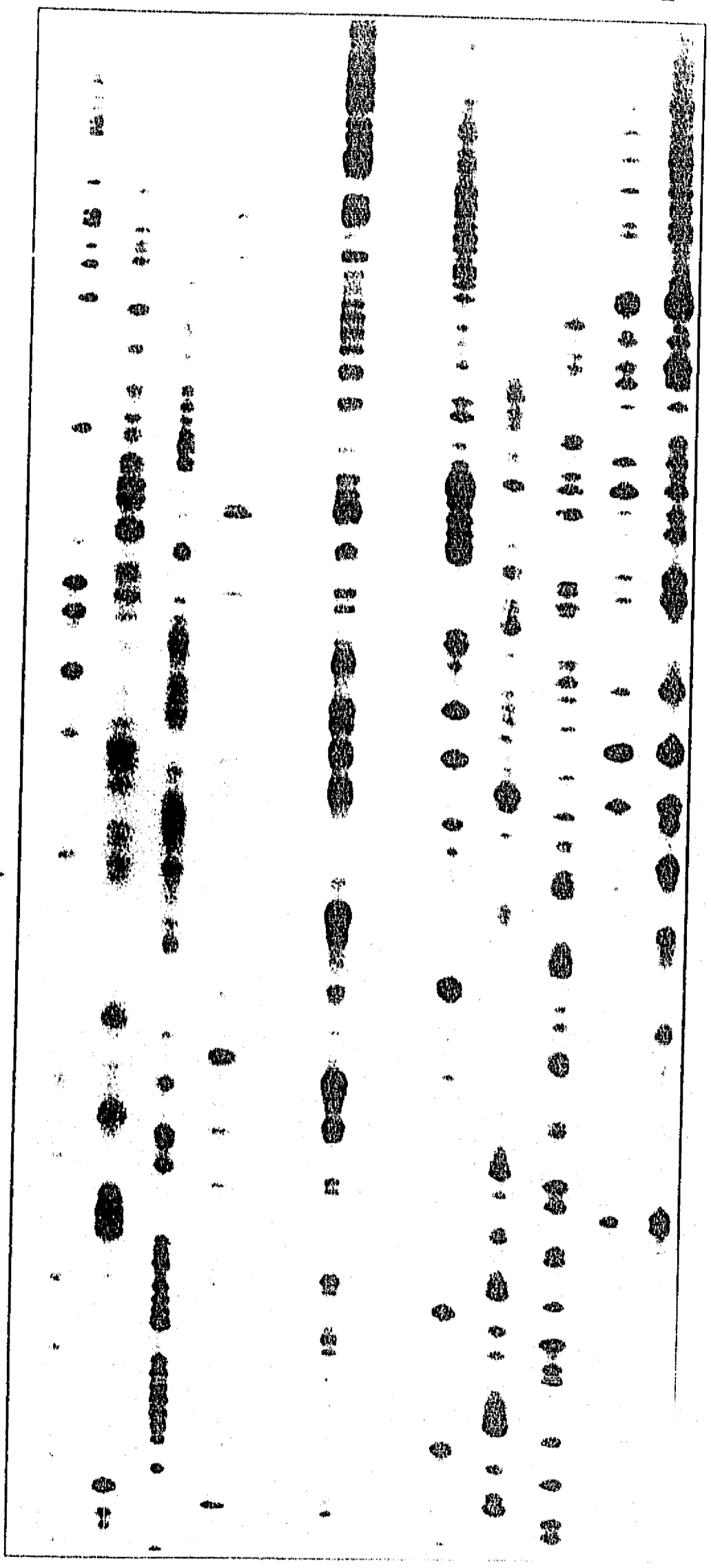
VIC- Secuencia genómica y reactividad de las guanosinas al DMS

Dado que el episoma ribosomal es una molécula abundante dentro del genoma de la amiba, pensamos en secuenciar directamente la región que rodea el sitio de inicio de la transcripción para empezar a hacer ensayos de footprinting *in vivo*. En la Fig. 17 mostramos la secuencia genómica de la hebra codificante y de la hebra no codificante de la región que rodea al sitio de inicio de la transcripción y resultados preliminares de la metilación *in vitro* de esta región. El oligonucleótido que utilizamos para secuenciar la hebra codificante es el oligonucleótido V y para secuenciar la hebra no codificante utilizamos el oligonucleótido UP325. Los carriles 1 y 2 muestran las extensiones de la Taq polimerasa sobre el DNA de amiba tratado *in vitro* con DMS. En el carril 1 se muestran las extensiones de la Taq polimerasa a partir del oligo V, y en el carril 2 las extensiones a partir del oligo UP325. En estas extensiones se observa la reactividad de las guanosinas que reaccionaron con el DMS. Este resultado muestra que tenemos la sensibilidad que requerimos para realizar experimentos de footprinting *in vivo*.

G A T C

1

G A T C 2



Hebra no codificante

Hebra codificante

Fig. 17. Secuencia genómica y metilación *in vitro*.
Utilizamos los oligonucleótidos UP325 y V para secuenciar la hebra no codificante y la hebra codificante, respectivamente. Los carriles 1 y 2 muestran la extensiones de la Taq polimerasa sobre el DNA tratado *in vitro* con DMS, con los oligonucleótidos UP325 y V, respectivamente.

VII- Discusion

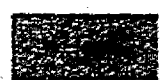
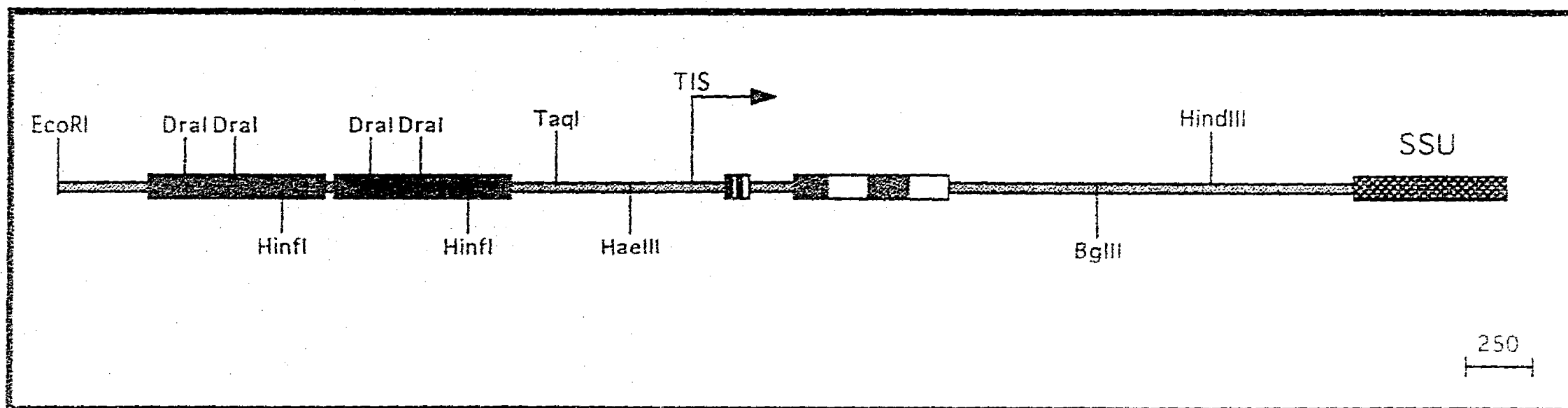
Aún cuando los genes 17S, 5.8S y 25S de los RNA ribosomales se han conservado a lo largo de la escala de la evolución, tanto las secuencias como el tamaño de los líderes o espaciadores externos transcritos han divergido considerablemente (ver Tabla II en sección de antecedentes). En este trabajo reportamos el tamaño del líder de los RNA ribosomales, que corresponde a 2447 nucleótidos. En ensayos tipo "northern" no logramos detectar parte de la región transcrita, lo cual sugiere que el transcrito primario se procesa rápidamente como ocurre para los transcritos primarios ribosomales de la mayoría de los organismos eucariotes. El transcrito primario se procesa en varios pasos antes de generar las especies ribosomales maduras; el primer paso de procesamiento ocurre muy rápido dentro de la célula (96), y frecuentemente el transcrito primario no se detecta dentro de la población de los RNAs totales. Esto ha provocado en otras especies, por ejemplo en ratón, la identificación equivocada de un sitio de inicio de la transcripción de los genes de DNA ribosomal interpretando un dato de mapeo por S1 (97) (98). En esta tesis utilicé varios enfoques metodológicos para identificar precisamente el sitio de inicio de la transcripción de los genes de RNA ribosomal de *E. histolytica*. Identifiqué inicialmente la región activa transcripcionalmente por ensayos de RNA naciente; esta metodología permitió identificar regiones transcritas que no logramos detectar en ensayos tipo "northern". Por experimentos de mapeo con S1 identificamos un probable sitio de inicio de la transcripción, y por un ensayo de "capping *in vitro*" confirmamos que el sitio identificado por mapeo por S1 contiene un extremo 5' trifosfato característico de un transcrito primario. En este ensayo, la enzima Guanidil Transferasa incorpora un GTP únicamente a transcritos que tienen extremos 5' tri o di fosfatos. Finalmente, identificamos el punto de inicio de la transcripción por un ensayo de extensión de primero y por TAP-RLMPCR. Estos resultados, en resumen, localizan el sitio de inicio de la transcripción en la posición 505 de la secuencia que presento en la Fig. 4. Cabe señalar que si al hacer el TAP-RLMPCR se demuestra que el producto amplificado es

dependiente estrictamente del tratamiento con TAP, el método sirve no sólo para generar extremos 5' trifosfatos susceptibles de ser amplificados, sino también puede utilizarse para diferenciar entre un transcrito primario y un transcrito procesado.

No se han descrito consensos entre las secuencias que conforman el promotor de los genes ribosomales, pero los promotores de varios vertebrados han sido estudiados con suficiente detenimiento como para concluir que comparten una arquitectura similar. Se ha demostrado que los nucleótidos localizados entre la posición -45 hasta la +18, con respecto al sitio de inicio de la transcripción, son esenciales para la transcripción por la RNA polimerasa I y por esa razón constituyen el promotor "core" (99)(100)(101)(102). En este trabajo, buscamos un consenso en 26 bases que rodean al sitio de inicio de la transcripción, pero utilizamos una herramienta que considera la frecuencia de las bases en cada posición. A partir de este análisis, obtuvimos un consenso, aunque débil. El tamizado de los primeros 1600 identificó un valor de ajuste máximo en el nucleótido 505. Es notable que experimentalmente identificamos el sitio de inicio de la transcripción precisamente en esta posición. Es importante señalar que el tamizado se hizo con un consenso que no contenía la secuencia iniciadora de *E. histolytica*. La secuencia iniciadora de *C. fasciculata*, por diferir de manera considerable con el resto de las secuencias, no entró en el mismo grupo que las 14 especies restantes. Es importante señalar que la identificación del sitio de inicio de la transcripción de los genes ribosomales de *C. fasciculata* se realizó con base a la sola interpretación de un dato de mapeo por S1 (103); existe la posibilidad de que el sitio que identificaron como sitio de inicio de la transcripción sea un sitio de procesamiento. Este resultado indica que este programa puede utilizarse para buscar probables sitios de inicio de la transcripción de genes de RNA ribosomal en otras especies. Aún cuando en a ventana siguiente aparecen tres picos con un valor de ajuste similar, el programa no deja de tener un valor predictivo en la primera ventana (en los primeros 1600 nt).

La presencia de elementos repetidos en los espaciadores intergénicos es una característica general de la organización de los genes ribosomales, y se ha demostrado que estos elementos repetidos están íntimamente relacionados con regulación de la transcripción de los genes ribosomales, en particular funcionando como enhancers. En el episoma ribosomal de *E. histolytica* existen varios elementos repetidos cuya función no se conoce (Fig. 9). Dentro de la clona genómica λ gt10-6.8 existen dos copias de un elemento repetido de 644 pb localizados 669 nucleótidos hacia arriba del sitio de inicio de la transcripción de los genes ribosomales (104); la Fig. 18 muestra la localización de estos repetidos con respecto al sitio de inicio de la transcripción. Existe también otro elemento repetido (p145) en el episoma ribosomal. Aún cuando la función de este elemento repetido tampoco se conoce, nuestro grupo mostró que esta secuencia repetida es un sitio de hipersensibilidad a DNasa I, y por geles de retardamiento y experimentos de footprinting *in vitro*, se mostró que interactúa de manera específica con proteínas nucleares; estos datos sugieren una función específica de este elemento dentro del episoma ribosomal (10). En la secuencia que presento en la Fig. 4 del artículo, están indicados dos elementos repetidos; el más corto consta de cuatro copias idénticas de un elemento de 21 pares de bases y el más largo consta de cuatro repetidos imperfectos de un elemento de 149 pb. Ambos elementos repetidos forman parte del transcrito primario. Pese a que la localización de estos elementos repetidos es atípica si la comparamos con la organización de los elementos regulatorios de los rDNA de eucariotes superiores, su localización cerca del sitio de inicio de la transcripción sugiere que puedan estar involucrados en regulación transcripcional. Se ha mostrado que los "enhancers" de la transcripción por la RNA polimerasa I son independientes de la orientación y de la distancia, y que funcionan aún localizados hacia abajo del promotor sobre el cual ejercen su función (105). Si la RNA polimerasa I de *E. histolytica* es bioquímicamente similar a la de otros eucariotes, estos elementos podrían ser "enhancers". Sabremos si estos elementos repetidos tienen un papel en la transcripción de los genes de rDNA por ensayos funcionales. El futuro desarrollo de un ensayo de transcripción *in*

Figura 18. ELEMENTOS REPETIDOS ALREDEDOR DEL SITIO DE INICIO DE LA TRANSCRIPCION DE LOS rDNA DE *Entamoeba histolytica*



repetidos de 644 pb (repetidos HinfI)



repetidos de 21 pb



repetidos 149 pb

vitro, junto con la utilización de ensayos de transformación transitoria para *E. histolytica* permitirán definir el o los dominios funcionales del promotor de los genes de rDNA y aclarar el probable papel regulatorio de los elementos repetidos en la regulación transcripcional de estos genes. Otra forma de saber si una región del DNA tiene una función regulatoria en la transcripción es mostrando que es reconocida específicamente por una proteína; esto se logra buscando la huella (footprint) que deja la proteína cuando se une al DNA. Existen agentes químicos que reaccionan de manera específica con el DNA, y que pueden mostrar qué bases y qué residuos de las bases están involucradas en el pegado con la proteína. Los más comunes son los agentes metilantes como el DMS y los agentes que oxidan los dobles enlaces de las pirimidinas, como el tetróxido de osmio y el permanganato de potasio. La unión de una proteína impide la metilación de estos residuos por el DMS. El permanganato de potasio se utiliza como reactivo específico de las pirimidinas que están en cadena sencilla y permite detectar complejos de iniciación de la transcripción (complejos abiertos) (94) (106) así como las distorsiones en la helice que ocurren cuando el DNA se dobla ("bending") (107). En este trabajo implementamos la secuencia genómica sobre el episoma ribosomal de *E. histolytica*, y presentamos la reactividad de las guanosinas al DMS, utilizando un oligonucleótido específico y varias reacciones de extensión (Fig. 17). Quedan por determinarse las condiciones fisiológicas óptimas para poder detectar la huella de la proteína en esa región. Pensamos comparar "footprints" de amibas en fase logarítmica y en fase estacionaria de crecimiento, tratadas con el químico. El control más adecuado en este caso es la metilación del DNA de amiba en ausencia de proteínas. Estos experimentos ayudarán a identificar las regiones del episoma que por su interacción con una proteína estarían involucradas en regular la transcripción de los genes de rDNA y a entender cómo ocurren los procesos de replicación y de recombinación dentro del episoma ribosomal.

VII- Bibliografía

1. Walsh JA. Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Rev. Infect. Dis.* 1986;8:228-238.
2. Caballero-Salcedo, A., Viveros-Rogel, M., Salvatierra, B., Tapia-Conyer, R., Sepúlveda-Amor, J., Gutierrez, G., Ortiz-Ortiz, L. Seroepidemiology of amebiasis in Mexico. *Am. J. Trop. Med. Hyg.* 1994;50:412-419.
3. Tannich, E., Horstmann, R.D., Knobloch, J. and Arnold, H.H. Genomic DNA differences between pathogenic and non pathogenic *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. USA* 1989;86:5118-5122.
4. Diamond, L.S., and Clark, C. G. A redescription of *Entamoeba histolytica* Schaudinn, 1933 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J-Eukaryot-Microbiol.* 1993;40(3):340-343.
5. Clark CG and Diamond, L.S. *Entamoeba histolytica*: an explanation for the reported conversion of "nonpathogenic" amebae to the "pathogenic" form. *Exp. Parasitol.* 1993;77(4):456-460.
6. Petri, W. Invasive amebiasis and the galactose-specific lectin of *Entamoeba histolytica*. *ASM news* 1991;57(6):299-306.
7. Martínez-Palomo, A. The biology of *Entamoeba histolytica*. Chichester: 1982(Press RS, ed.
8. Saffer, L.D. and Petri, W.A. Role of the galactose lectin of *Entamoeba histolytica* in adherence-dependent killing of mammalian cells. *Infect. Immun.* 1991;59:4681-4683.
9. Albach, R.A., Booden, T., Boonlayangoor, P. and Downing, S. Concepts of function of peripheral non-chromatin and endosome in *Entamoeba histolytica*. *Arch. Invest. Med. (México)* 1980;11(1):63-74.
10. Zurita, M., Alagon, A., Vargas, V.J., Lizardi, P.M.. The *Entamoeba histolytica* rDNA episome: nuclear localization, DNAase I sensitivity map, and specific DNA-protein interactions. *Mol Microbiol* 1991;5(8):1843-51.
11. Weinback, E.C. and Diamond, L.S. *Entamoeba histolytica*. I. Aerobic metabolism. *Exp. Parasitol.* 1974;35:232-243.

12. Serrano, R. and Reeves, R. E. Glucose transport in *Entamoeba histolytica*. *Biochem. J.* 1975;144:43-48.
13. Fahey, R.C., Newton, G. L., Arrick, B., Overdank-Bogart, T. and Aley, S. B. *Entamoeba histolytica*: a eukaryote without glutathion metabolism. *Science* 1984;224:70-72.
14. Reeves, R.E. *Adv. Parasitol.* 1984;23:105-142.
15. Bhattacharya, S., Bhattacharya, A., Diamond, L.S., Soldo, A.T. Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. *J Protozool* 1989;36(5):455-8.
16. Orozco, E. Solis, F. J., Domínguez, J., Chávez, B., and Hernández, F. *Entamoeba histolytica*: cell cycle and nuclear division. *Exp. Parasitol.* 1988;67:85-95.
17. Valdes, J., Hernández, D.A., Ocadiz, R., and Orozco, E. Molecular karyotype of *Entamoeba histolytica* and *Entamoeba invadens*. *Trans. Roy. Soc. Trop. Med, and Hyg.* 1990;84:537-541.
18. Lioutas, C., Christel, S. and Tannich, E. Identification of various circular DNA molecules in *Entamoeba histolytica*. *Exp. Parasitol.* 1995;80:349-352.
19. Dhar, S.K., Choudhoury, N. R., Bhattacharya, A. and Bhattacharya, S. A multitude of circular DNAs exist in the nucleus of *Entamoeba histolytica*. *Molec. Biochem. Parasitol.* 1995;70:203-206.
20. Michel, B., Alagón, A., Lizardi, P.M. and Zurita, M. Characterization of a repetitive DNA element from *Entamoeba histolytica*. *Molec. biochem. Parasitol.* 1992;51:165-168.
21. Bruchhaus, I., Leippe, M., Lioutas, C. and Tannich, E. Unusual gene organization in the protozoan parasite *E. histolytica*. *DNA Cell Biol* 1993;12:925-933.
22. Lohia, A., and Samuelson, J. Cloning of the EH CDC2 gene from *Entamoeba histolytica* encoding a protein kinase p34 CDC2 homolog. *Gene* 1994;127:203-207.
23. Plaimauer, B., Ortner, S., Weindermann, G., Scheiner, O. and Duchene, M. An intron containing gene coding for a novel 39-kilodalton antigen of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 1994;66:181-185.

24. Miranda R. Clonación, secuenciación y caracterización del gen U6 snRNA de *Entamoeba histolytica*. [Tesis de Licenciatura]. UNAM, 1995.
25. Cavalier-Smith T. A 6-kingdom classification and a unified phylogeny. In: Schenk HEAaS W., eds., ed. Endocytobiology. Berlin: de Gruyter, 1983: 1027-1034.
26. Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N., and Miyata, T. Early divergences in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. J. Mol. Evol. 1992;36:380-388.
27. Wostmann, C., Tannich, E., and Bakker-Grunwald, T. Ubiquitin of *Entamoeba histolytica* deviates in six amino acid residues from the consensus of all other known ubiquitins. FEBS lett 1992;308(1):54-58.
28. Muller, M. Energy metabolism of ancestral eukaryotes: a hypothesis based on the biochemistry of amitochondriate parasitic protists. BioSystems 1992;28:33-40.
29. Gicquaud CR. Etude de l'ultrastructure de noyau et de la mitose de *Entamoeba histolytica*. Biol. Cell. 1979;35:305-312.
30. Clark, G.C. and Roger, A. J. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc. Natl. Acad. Sci. 1995;92:6518-6521.
31. Purdy, J.E., Mann BJ, Pho LT, Petri WJ. Transient transfection of the enteric parasite *Entamoeba histolytica* and expression of firefly luciferase. Proc Natl Acad Sci U S A 1994;91(15):7099-7103.
32. Nickel, R., Tannich, E. Transfection and transient expression of chloramphenicol acetyltransferase gene in the protozoan parasite *Entamoeba histolytica*. Proc Natl Acad Sci U S A 1994;91(15):7095-7098.
33. Vines, R.R., Purdy, J.E., Ragland, B.D., Samuelson, J., Mann, B.J. and Petri, W.A. Stable episomal transfection of *Entamoeba histolytica*. Molec. Biochem. Parasitol. 1995;71:265-267.
34. Chambon, P. Ann. Rev. Biochem. 1974;43:613-638.
35. Conaway, R.C. and Conaway, J. W. Annu. Rev. Biochem. 1993;62:161-190.

36. Matsui, T., Segall, J., Weil, P. A. and Roeder, R. G. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* 1980;255:11992-11996.
37. Buratowski, S. and Sharp, P.A. Initiation of transcription by RNA polymerase II. In: McKnight SaY K.R., ed. *Transcriptional regulation*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1992: 227-246.
38. Lobo, S.M., and Hernandez, N. A 7bp mutation converts a human RNA polymerase II snRNA promoter into a RNA polymerase III promoter. *Cell* 1989;58:55-67.
39. Lobo, S.M. Tanaka, M., Sullivan, M.L., and Hernandez, N. A TBP complex essential for transcription from TATA-less but not TATA containing RNA polymerase III promoters is part of the TFIIIB fraction. *Cell* 1992;71:1029-1041.
40. Sadowski, C.L., Henry, R.W., Lobo, S.M. and Hernandez, N. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates RNA polymerase II and III transcription from snRNA promoters through the PSE. *Genes and Dev.* 1993;
41. Kassavetis, G.A., Joazeiro, C.A.P., Pisano, M., Geiduschek, E.P., Colbert, S., Hahn, S. and Blanco, J.A. The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell* 1992;71:1055-1064.
42. Comai, L., Tanese, N. and Tjian, R. The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* 1992;68:965-976.
43. Cormack, B.P., and Struhl, K. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* 1992;69:685-696.
44. Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev.* 1993;7:1291-1308.
45. Smale, S.T. and Baltimore, D. The "initiator" as a transcriptional element. *Cell* 1989;57:103-113.
46. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 1989;56:549-561.

58. Batts-Young, B., Maizels, N., and Lodish, H.F. Precursors of rRNA in the cellular slime mold *D. discoideum*. *J. Biol. Chem.* 1977;252:3952-3960.
59. Michel, B., Lizardi, P., Alagón, A., and Zurita, M. Identification and analysis of the start site of ribosomal RNA transcription of *Entamoeba histolytica*. *Molec. Biochem. Parasitol.* 1995;73:19-30.
60. Morgan, G.T., Roan, J.G., Bakken, A.H., and Reeder, R. Variations in transcriptional activity of rDNA spacer promoters. *Nucleic Acids Res.* 1984;12:6043-6052.
61. Kuhn, A., and Grummt, I. A novel promoter in the mouse rDNA spacer is active *in vivo* and *in vitro*. *EMBO J* 1987;6:3487-3492.
62. Miller, K.G., Hayward, D.C., and Glover, D.M. Transcription of the non-transcribed spacer of *Drosophila melanogaster*. *Nucleic Acids Res* 1983;11:11-19.
63. Jantzen, H.M., Admon, A., Bell, S.P., and Tjian, R. Nucleolar transcription factor hUBF contains a DNA binding motif with similarity to HMG proteins. *Nature* 1990;344:830-833.
64. Mandal, R.K. The organization and transcription of eucaryotic ribosomal RNA genes. *Prog. Nucleic Acids Res. Mol. Biol.* 1984;31:115-160.
65. Long, E.O., and David, I.B. Repeated genes in eucariotes. *Annu. Rev. Biochem.* 1980;49:727-764.
66. Molgaard, H.V., Matthews, H.R., and Bradbury, E. M. Organization of genes for ribosomal RNA in *Physarum polycephalum*. *Eur. J. Biochem.* 1986;68:541-549.
67. Yao, M.C., Zhu, S.G. and Yao, C.H. Gene amplification in *Tetrahymena Thermophila*: formation of extrachromosomal palindromic genes coding for rRNA. *Mol. Cell. Biol* 1985;5:1260-1267.
68. Findly, R.C., and Gall, G.G. Free ribosomal RNA genes in *Paramecium* are tandemly repeated. *Proc. Natl. Acad. Sci. USA* 1978;75:3312-3316.
69. Ravel, C.P., Nicolas P, Nigon V, Neyret O, Freyssinet G. Extrachromosomal circular nuclear rDNA in *Euglena gracilis*. *Nucleic Acids Res* 1985;13(20):7529-37.

70. Clark, C.G., Cross GA. rRNA genes of *Naegleria gruberi* are carried exclusively on a 14- kilobase-pair plasmid. *Mol Cell Biol* 1987;7(9):3027-31.
71. Huber, M., Koller B, Gitler C, et al. *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules. *Mol Biochem Parasitol* 1989;32(2-3):285-96.
72. Sehgal, D., Mittal, V., Ramachandran, S., Dhar, S. K., Bhattacharya, A., Bhattacharya, S. Nucleotide sequence organisation and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* 1994;67:205-214.
73. Cázares, F., Manning-Cela, R., and Meza, I. Heterogeneity of the ribosomal DNA episome in strains and species of *Entamoeba*. *Molec. Microbiol.* 1994;12(4):607-612.
74. Michel, B. Identificación y análisis de la región promotora de los genes de RNA ribosomal de *Entamoeba histolytica*. [Maestría]. UNAM, 1992.
75. Mittal, V., Ramachandran, S., Sehgal, D., Bhattacharya, A. and Bhattacharya, S. Sequence analysis of a DNA fragment with yeast autonomously replicating sequence activity from the extrachromosomal ribosomal DNA circle of *Entamoeba histolytica*. *Nucleic Acids Res.* 1991;19:2777.
76. Burch, D.J., Li, E., Reed, S., Jackson, T.F.H.G., and Stanley, S.L., Jr. Isolation of a strain specific *Entamoeba histolytica* cDNA clone. *J. Clin. Microbiol.* 1991;29:696-701.
77. Jansson, A., Gillin F, Kagardt U, Hagblom P. Coding of hemolysins within the ribosomal RNA repeat on a plasmid in *Entamoeba histolytica*. *Science* 1994;263(5152):1440-3.
78. Fromont, R.M., Bertrand E, Pictet R, Grange T. A highly sensitive method for mapping the 5' termini of mRNAs. *Nucleic Acids Res* 1993;21(7):1683-4.
79. Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* 1984; 505-19.
80. Becker, P.B., Ruppert, S., and Shultz, G. Genomic footprint reveals cell type-specific DNA binding of ubiquitous factors. *Cell* 1987;51:435- 443.

81. Maxam, A. M. and Gilbert, W. Sequencing end-labelled DNA with base specific chemical cleavage. *Methods in Enzymol.* 1980;65:499-560.
82. Diamond, L.S., Harlow, D. R. and Cunnick, C.C. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans. Roy. Soc. Trop. Med. Hyg.* 1978;72:431-432.
83. Sambrook, J., Fritsch, E., F. and Maniatis, T. *Molecular Cloning. A laboratory manual.* Cold Spring Harbor Laboratory Press, 1989
84. Que, X., Reed SL. Nucleotide sequence of a small subunit ribosomal RNA (16S-like rRNA) gene from *Entamoeba histolytica*: differentiation of pathogenic from nonpathogenic isolates. *Nucleic Acids Res* 1991;19(19):5438.
85. Zurita, M., Bieber D, Ringold G, Mansour TE. Cloning and characterization of a female genital complex cDNA from the liver fluke *Fasciola hepatica*. *Proc Natl Acad Sci U S A* 1987;84(8):2340-4.
86. Ausubel, F.M., Brent, R., Kingston, R. E., Moore, D. R., Seidman, J. A., Struhl, K. *Current protocols in molecular biology.* John Wiley & sons, Inc., 1994
87. Grosschedl, R., Weaver D, Baltimore D, Costantini F. Introduction of a mu immunoglobulin gene into the mouse germ line: specific expression in lymphoid cells and synthesis of functional antibody. *Cell* 1984;38(3):647-58.
88. Goldberg, D.A. Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc Natl Acad Sci U S A* 1980;77(10):5794-8.
89. Kooter, J. M., van der Spek, HJ, Wagter, R., d'Oliveira, C. E., van der Hoeven, F, Johnson, P. J., and Borst, P. The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. *Cell* 1987;51(2):261-72.
90. Sanger, F., Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74(12):5463-7.
91. Sehgal, D., Mittal, V., Ramachandran, S., Dhar, S. K., Bhattacharya, A., Bhattacharya, S. Nucleotide sequence organization and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 1994;67:205-214.
92. Mulligan, R.M., Leon P, Walbot V. Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. *Mol Cell Biol* 1991;11(1):533-43.

93. Mulligan, R.M., Maloney AP, Walbot V. RNA processing and multiple transcription initiation sites result in transcript size heterogeneity in maize mitochondria. *Mol Gen Genet* 1988;211(3):373-80.
94. Morett, E. and Buck, M. In vivo studies on the interaction of the RNA Polymerase-s with the *Klebsiella pneumoniae* and *Rhizobium meliloti* nifH promoters: the role of NifA in the formation of an open promoter complex. *J. Mol. Biol.* 1989;210:65-77.
95. Clark, J.M. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* 1988;16(20):9677-86.
96. Miller, K. and Sollner-Webb, B. Transcription of mouse rRNA genes by RNA polymerase I: in vitro and in vivo initiation and processing sites. *Cell* 1981;27:165-177.
97. Grummt, I. Mapping of a mouse ribosomal DNA promoter by in vitro transcription. *Nucleic Acids Res* 1981;9(22):6093-102.
98. Urano, Y., Kominami R, Mishima Y, Muramatsu M. The nucleotide sequence of the putative transcription initiation site of a cloned ribosomal RNA gene of the mouse. *Nucleic Acids Res* 1980;8(24):6043-58.
99. Yamamoto, O., Takakusa, N., Mishima, Y., Kominami, R. and Muramatsu, M. Determination of the promoter region of mouse ribosomal RNA gene by an in vitro transcription system. *Proc. Natl. Acad. Sci.* 1984;81:299-303.
100. Haltiner, M.M., Smale, S. T., and Tjian, R. Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. *Mol. Cell: Biol.* 1986;6:227-235.
101. Windle, J. and Sollner-Webb, B. Two distant and precisely positioned domains promote transcription of *Xenopus laevis* rRNA genes: analysis with linker-scanning mutants. *Mol. Cell. Biol.* 1986;6:4585-4593.
102. Kohorn, B.D., Rae PM. Localization of DNA sequences promoting RNA polymerase I activity in *Drosophila*. *Proc Natl Acad Sci U S A* 1983;80(11):3265-8.

103. Grondal, J., Evers R, Cornelissen AW. Identification and sequence analysis of the ribosomal DNA promoter region of *Crithidia fasciculata*. *Nucleic Acids Res* 1990;18(6):1333-8.
104. Sehgal, D., Bhattacharya A, Bhattacharya S. Analysis of a polymorphic locus present upstream of rDNA transcription units in the extrachromosomal circle of *Entamoeba histolytica*. *Mol Biochem Parasitol* 1993;62(1):129-30.
105. Labhart, P. and Reeder, R. H. *Xenopus* ribosomal genes enhancers function when inserted inside the gene they enhance. *Nucleic Acids Res.* 1985;13:8999-9009.
106. Mirkovitch, J. and Darnell, J. E. Jr. Mapping of RNA polymerase on mammalian genes in cells and nuclei. *Mol. Biol. Cell* 1992;3:1085-1094.
107. Boroweic, J.A., Zhang, L., Sasse-Dwight, S., and Gralla, J. D. DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. *J. Mol. Biol.* 1987;196:101-111.
108. Pape, L. K., Windle, J. J., Mougey, E. B., Sollner, Webb B. The *Xenopus* ribosomal DNA 60- and 81-base-pair repeats are position- dependent enhancers that function at the establishment of the preinitiation complex: analysis in vivo and in an enhancer-responsive in vitro system *Mol Cell Biol.* 1989; 11, 5093-104

Identification and analysis of the start site of ribosomal RNA transcription of *Entamoeba histolytica*

Bertha Michel^{a,b}, Paul M. Lizardi^b, Alejandro Alagón^b, Mario Zurita^{a,*}

^a Departamento de Genética y Fisiología Molecular Instituto de Biotecnología, Universidad Nacional Autónoma de México, A.P. 510-3, 62271, Cuernavaca, Morelos, México

^b Departamento de Reconocimiento Molecular y Bioestructura, Instituto de Biotecnología, Universidad Nacional Autónoma de México, A.P. 510-3, 62271 Cuernavaca, Morelos, México

Received 24 January 1995; accepted 9 May 1995

Abstract

In this article we report the identification of the start site of ribosomal RNA transcription unit of the enteric parasite *E. histolytica*. We cloned the upstream region of the ribosomal RNA and we defined the 5' boundary of the transcription unit with nuclear run-on assays. We report that ribosomal transcription starts 2447 bp upstream the SSU ribosomal gene, at an adenosine residue. This data was supported both by S1 mapping and by primer extension analysis; that the mapped site was indeed the transcription start point was demonstrated by RNase protection of the in vitro capped RNA. Our sequence data around the transcription start point shows two different tandem repeat clusters immediately downstream from the transcription start point.

Keywords: *Entamoeba histolytica*; rRNA transcription start site; Tandem repeat element

1. Introduction

The enteric protozoan parasite *Entamoeba histolytica* is a major cause of debilitating illness and

death, infecting 500 million people and causing an estimated 50 million cases of diarrhea and 50000 deaths yearly [1]. Despite its medical importance, little is known about the genomic organization and control of gene expression in this parasite.

Huber et al. [2] reported that the rDNA of the human parasite *Entamoeba histolytica* is located in a circular extrachromosomal 24.5-kb molecule present in 200 copies per cell, and suggested that this molecule can replicate autonomously [3,4]. The 200 or so copies of the rDNA episome are located all along the inner nuclear membrane, appearing to be part of the peripheral chromatin [5]. These observations, as well as previous work by Albach et al. [6], strongly suggest that ribosomal RNA (rRNA) syn-

Abbreviations: EGTA, [ethylenebis(oxyethylenitrilo)]tetraacetic acid; NTS, non-transcribed spacer; rDNA, ribosomal DNA; rRNA, ribosomal RNA; TAP-RLMPCR, tobacco acid pyrophosphate-reverse ligation mediated polymerase chain reaction; UCE, upstream control elements.

Note: Nucleotide sequence data reported in this paper are available in the GenBank data base under the accession No. L36807.

* Corresponding author. Tel.: (52-73) 114-900; Fax: (52-73) 114-900.

thesis occurs in the peripheral chromatin, and that the episome can be considered as a nucleolar organizer, since there is no evidence for a typical nucleolus in *E. histolytica*.

A transient transfection system for *E. histolytica* has just been established [7,8], thus creating the possibility of doing genetic analysis in this organism. Nevertheless, no definitive information about regulatory sequences has been documented for *E. histolytica*. The characterization of the ribosomal promoter is relevant to ongoing efforts toward the construction of an episomal DNA vector for both transient and stable transformation of this important human parasite. To identify the site of rRNA transcriptional initiation in *E. histolytica*, we first mapped transcriptionally active regions using nuclear nascent labeled RNA. The start site was identified with different methods, including S1 mapping, RNase protection of in vitro capped RNA and primer extension analysis. The results, which complement each other, show that transcriptional initiation occurs at an adenosine residue, within the sequence ATATACTATACAGG. We performed a comparison of the sequences of the transcription start points of 15 ribosomal genes of different organisms, including *E. histolytica*, and obtained a weak consensus. Aside from a number of repetitive elements located upstream from the transcription start site, our sequence analysis also identified tandem repeats downstream from the putative core promoter region.

2. Materials and methods

2.1. Materials

Growth and culture of trophozoites from *E. histolytica* strain HK9, was done axenically in TYI-S-33 medium [9]. Other materials were obtained from the following sources: restriction enzymes, DNase I and reverse transcriptase were from Boehringer-Mannheim. Radiolabeled nucleotides were purchased from New England Nuclear. Sequencing reactions were performed with a Sequenase-2 kit (US Biochemical). Standard DNA manipulations were performed according to protocols of Sambrook et al. [10].

2.2. RNA analysis

Total cellular RNA was prepared as described previously [12]. RNA was separated by electrophoresis in formaldehyde-agarose gels as described [13], and transferred to nitrocellulose filters. In experiments involving the 17S oligonucleotide probe, hybridization was performed according to Grosschedl et al. [14]. DNA blots were done according to Goldberg [15]. Nascent-labeled transcripts (see below) were hybridized according to Kooter et al. [16]. Sequencing reactions were performed using the dideoxy method of Sanger et al. [11].

The in vitro capping reaction was done according to Mulligan et al. [17], using high-molecular-mass *Entamoeba histolytica* RNA fractionated by Sepharose 4B gel filtration. Incorporation of the isotope into RNA yielded a specific activity of 1.2×10^6 cpm μg^{-1} . RNase protection assays were performed according to Mulligan et al. [13]. Protected fragments were analyzed on a 5% denaturing polyacrylamide sequencing gel and analyzed using a phosphorimager (Molecular Dynamics). S1 protection assays were performed according to Ausubel et al. [18]. Primer extension analysis was done according to Mulligan et al. [13]. A synthetic oligonucleotide complementary to nt 607–628 on the sequence shown in Fig. 4 was used as primer in the extension reaction.

For TAP-RLPCR, we followed the protocol reported by Fromont-Racine et al. [19]. The nucleotide sequence of the linker primer is 5'-d(GTTTTCC-CAGTCAGGAC)-3'. The sequence of the RNA linker is the same as that of the linker primer.

2.3. Nascent RNA assay

Nuclei from exponential phase trophozoites were prepared by disrupting the cells in a manual homogenizer. The lysate was layered onto a 9 ml cushion of 1 M sucrose (10 mM Tris · HCl, pH 8/1 mM EDTA) and centrifuged at $5000 \times g$ for 10 min at 4°C. Nuclei were frozen in a buffer containing 50 mM Tris · HCl, pH 8.0/75 mM NaCl/0.5 mM EDTA, pH 8.0/200 mM PMSF/5 mM DTT/50% glycerol and stored in liquid nitrogen. Nuclear nascent RNA was labeled and used as a probe on southern blots as described by Kooter et al. [16].

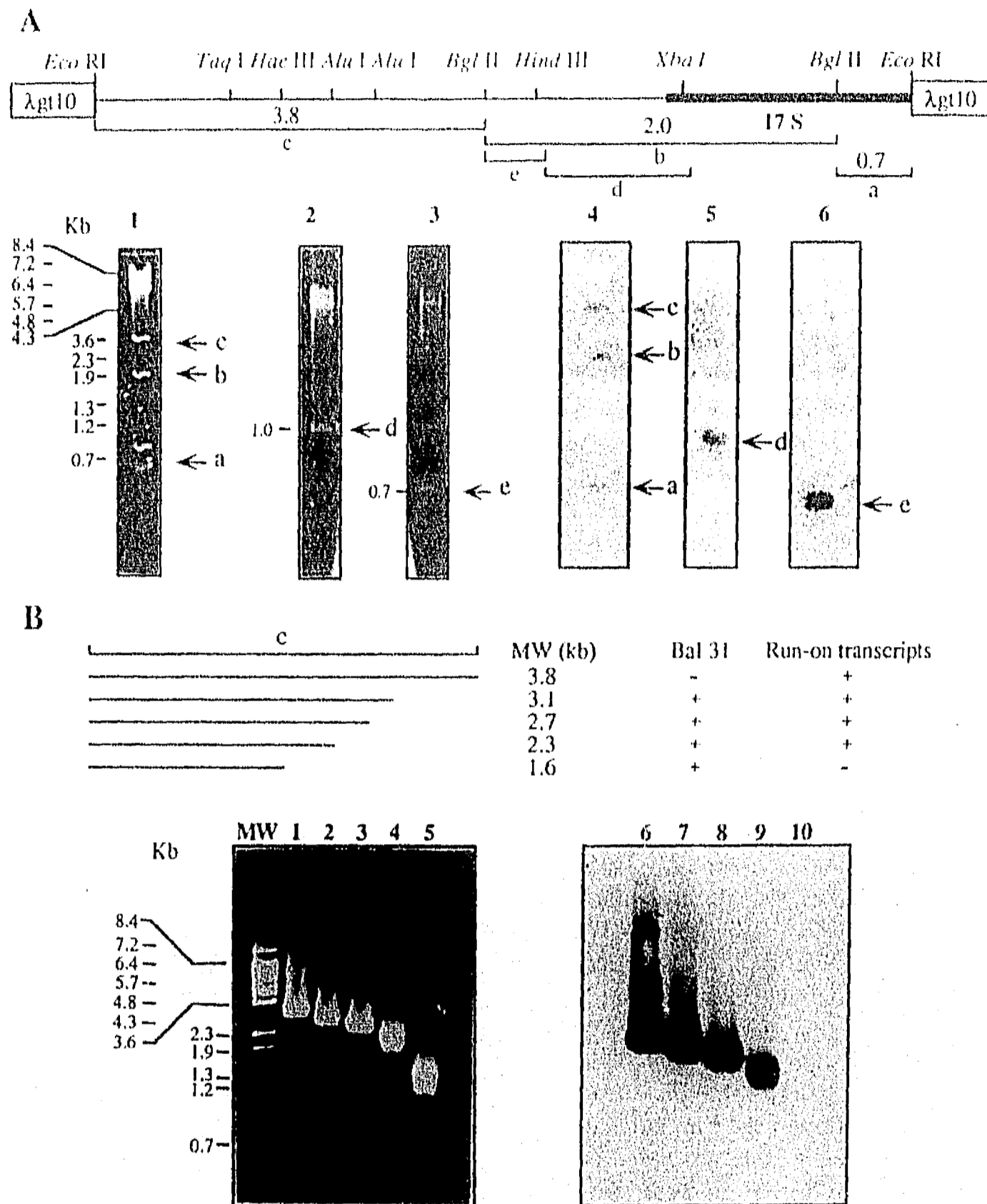


Fig. 1. Run-on mapping of the rDNA transcription unit. (A) Hybridization of labeled nascent transcripts to restriction fragments of the λ gt10-6.5 clone and to subclones d and e. Lanes 1-3 show ethidium-bromide-stained gels and lanes 4-6 show the Southern blot hybridizations of the same gels using labeled nascent transcripts as probes. Lanes 1 and 4, λ gt10-6.5 digested with *Eco*RI and *Bgl*II; lanes 2 and 5, subclone M13d, containing fragment d, digested with *Hind*III + *Xba*I; lanes 3 and 6, subclone M13e, containing fragment e, digested with *Eco*RI + *Hind*III. The restriction map of the λ gt10-6.5 clone is indicated at the top of the figure. Lambda/*Bst*EII fragments were used as molecular mass markers for the gel shown in lane 1. Numbers on the left of the gel (lanes 2 and 3) indicate the relative sizes of the restriction fragments. Black arrows indicate the localization of fragments a, b, c, d and e. (B) Hybridization of labeled nascent transcripts to the p3.8 clone treated with BAL 31. The p3.8 clone was linearized at the *Hind*III site of the polylinker and treated with 2 U of BAL 31 for different periods of time (0, 1, 2, 4 and 8 min). Lanes 1-5 show the ethidium bromide-stained gel and lanes 6-10 the Southern blot hybridization of the same gel with the nascent labeled transcripts. Lanes 1 and 6, no BAL 31 added, time 0; lanes 2 and 7, 1-min digest; lanes 3 and 8, 2-min digest; lanes 4 and 9, 4-min digest; lanes 5 and 10, 8-min digest. The diagram shows the relative sizes of the deletions obtained with the nuclease, and the column on the right shows the extent of hybridization (+, -) of the digested material with labeled nascent transcripts. Lambda/*Bst*EII fragments were used as molecular mass markers.

3. Results

3.1. Isolation of the upstream region of the rDNA transcription unit

In order to isolate a genomic fragment containing the upstream region of the 17S rDNA, a λ gt10 genomic library was screened using an oligonucleotide specific for *E. histolytica* 17S rRNA (Oligo 17S; 5'-GTGTTCTACAATTGTCGTGG-3' complementary to nt 251–273 of the 17S rDNA sequence of *E. histolytica*). From this screening, a 6.5-kb *Eco*RI fragment was isolated and characterized in detail (clone λ gt10–6.5). The restriction map of this DNA is identical to the map reported by Huber et al. [2]. The region contained between the *Taq*I and *Xba*I restriction sites of this clone was sequenced (restriction map at the top of Fig. 1). The 17S-specific oligonucleotide was used as a primer to begin sequencing within fragment a. Based on sequence analysis, the clone λ gt10–6.5 contains 1742 bp of the coding region of the 17S ribosomal gene [20].

3.2. Mapping the transcription unit with labeled nascent transcripts

As a first step to locate the approximate start site of the rDNA transcription unit, northern blots containing total cellular RNA were hybridized with the 17S oligonucleotide, as well as with fragments d, e and c. Only the 17S oligonucleotide and fragment d gave a hybridization signal (data not shown). We were unable to detect a signal with either fragment e or c. These results suggested the presence of a precursor starting within fragment d or, alternatively, the existence of a longer, rapidly processed precursor. In order to distinguish between these two alternatives, we performed assays using RNA labeled in isolated nuclei to obtain preferential labeling of nascent precursor molecules. [32 P]UTP-labeled nascent RNA was used to probe a Southern blot containing restriction fragments from the genomic clone. Results are presented in Fig. 1A. The ethidium bromide-stained gels show the fragments generated after digestion of the λ gt10–6.5 clone with *Eco*RI and *Bgl*II (lanes 1), as well as fragments d and e released by digestion of independent M13 subclones (lanes 2 and 3). The labeled nascent RNA

hybridized with several fragments that contained part of the coding region of the 17S structural gene: fragment a and fragment b (lanes 4), and with fragment d (lane 5). A strong hybridization signal was also observed with fragment e (lane 6), as well as

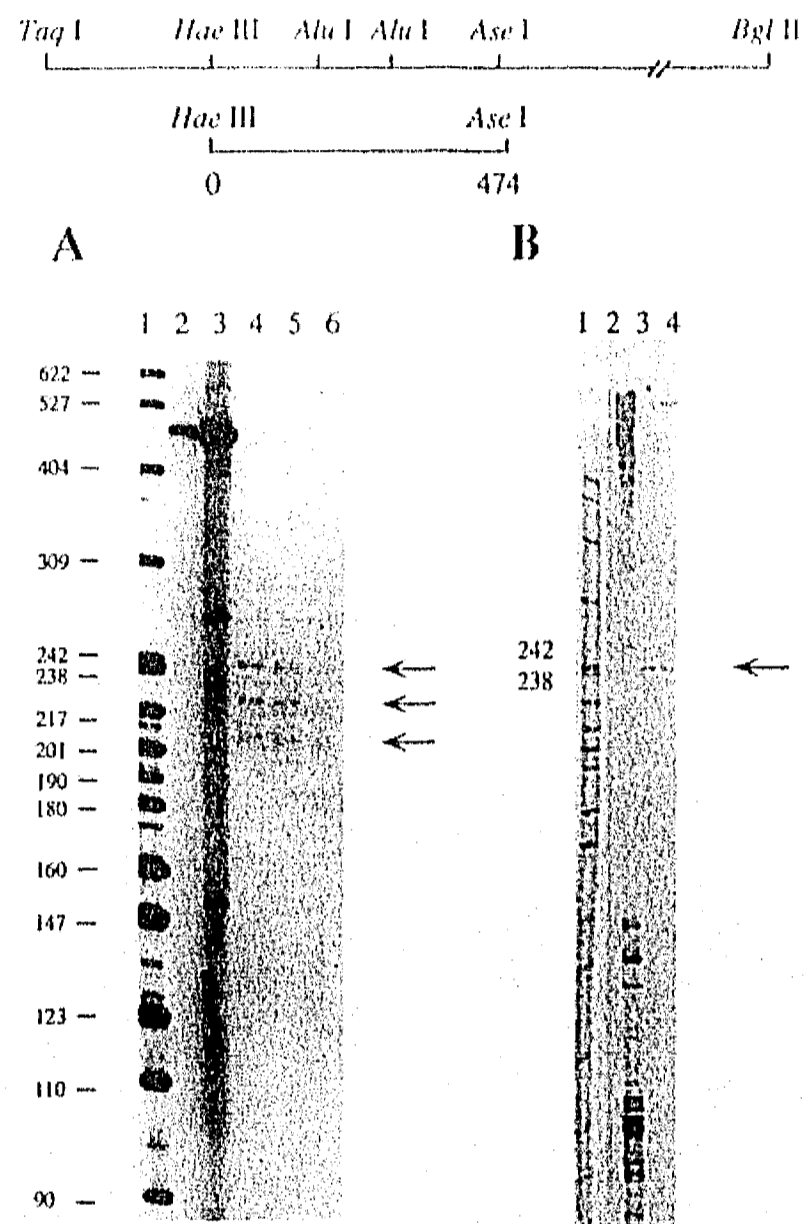


Fig. 2. (A) S1 protection assay. The *Hae*III-*Ase*I fragment from clone p3.8 was labeled with [γ - 32 P]ATP using polynucleotide kinase and hybridized with 10 μ g of total RNA in the presence of 80% formamide. After hybridization, the samples were treated either with 0, 20, 40, or 80 units of nuclease S1 (lanes 3, 4, 5 and 6, respectively). Lane 2 shows the labeled *Hae*III-*Ase*I fragment from clone p3.8. Size markers are end-labeled fragments of pBR322 digested with *Msp*I. (B) RNase protection assay of the in vitro capped rRNA precursor. A high-molecular-mass RNA fraction was obtained by gel filtration and labeled by in vitro capping (see Materials and methods). Labeled RNA was hybridized with the *Hae*III-*Ase*I fragment from clone p3.8 in the presence of 80% formamide. The heteroduplex was treated either with no RNase A (lane 2) or with 333 ng ml $^{-1}$ of RNase A (lanes 3 and 4). Size markers are end-labeled fragments of pBR322 digested with *Msp*I (lane 1).

with fragment c (lane 4). This result indicates that labeled nascent RNA hybridized far upstream from the 17S coding region.

In order to map the transcribed domain in the upstream region of the 17S structural gene, deletions of fragment c were probed with nascent RNA. p3.8 clone was generated (fragment c inserted in pBlue-script) and was cut with *Hind*III (on the vector) and partially digested with nuclease BAL 31. Aliquots were withdrawn at different digestion time intervals and cut with *Eco*RI. Partial digests of the fragment c ranged in size from 3.8 kb with no enzyme added, to 1.6 kb after 8 min exposure to the enzyme (Fig. 1B, lanes 1–5). Labeled nascent RNA hybridized strongly

to the non-digested clone (lane 6), as well as to the 3.1 kb (lane 7), the 2.7 kb (lane 8) and the 2.3 kb (lane 9) fragments; however, no hybridization signal was observed with the 1.6-kb fragment (lane 10). These data mapped the transcribed domain to the region localized between the 1.6-kb and the 2.3-kb BAL 31 digestion products (Fig. 1B).

3.3. Mapping the transcription start point by nuclease protection assays

A detailed restriction map of clone p3.8 is shown on top of Fig. 2. Based on results obtained with nuclear run-on assays (Fig. 1), a 474-bp *Hae*III-*Ase*I

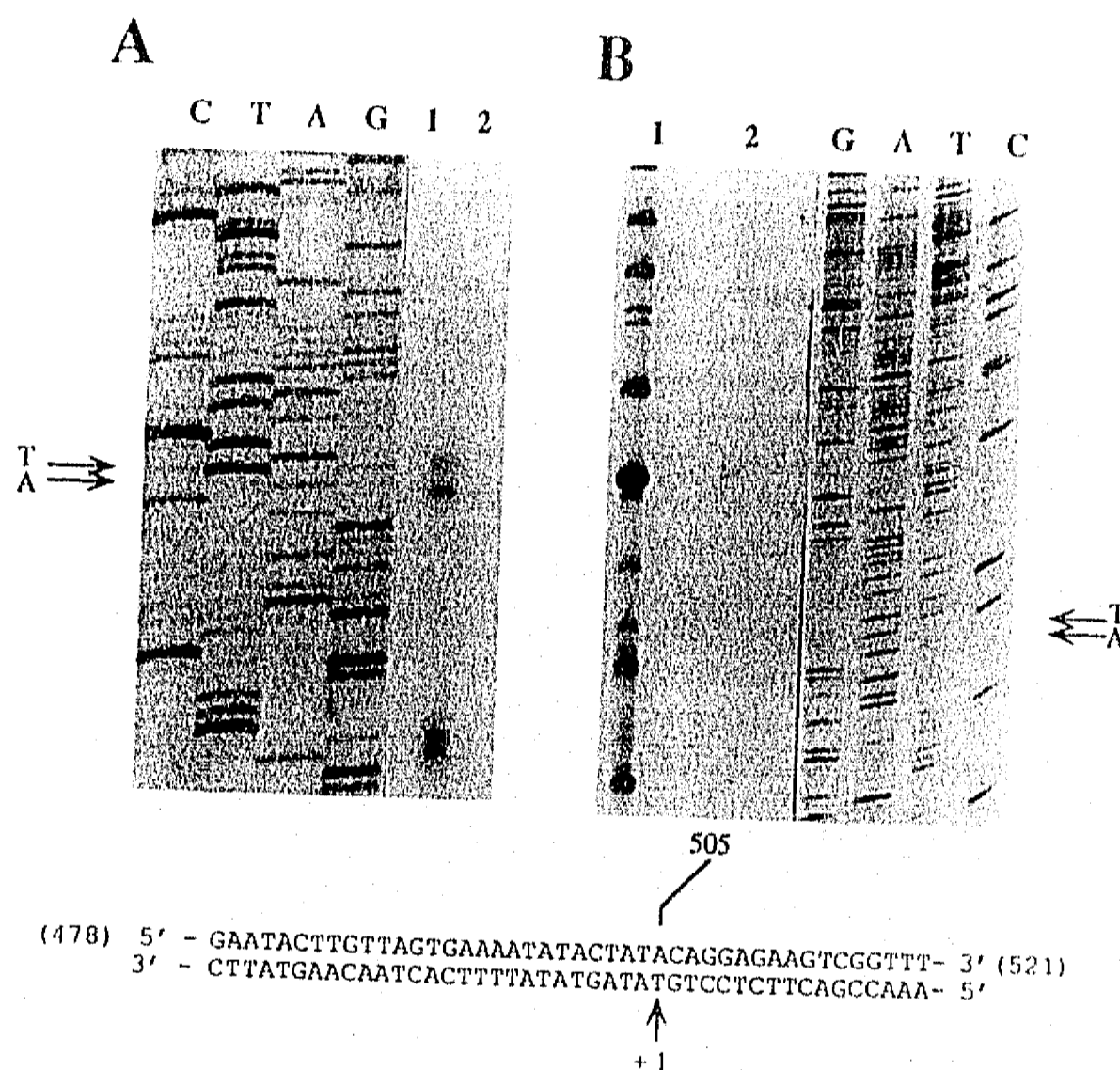


Fig. 3. High-resolution mapping of the 5' termini of the rRNA precursor. (A) Primer extension analysis. Total RNA from *E. histolytica* trophozoites was hybridized with ³²P-labeled oligonucleotide V (see Materials and methods) and incubated with reverse transcriptase to produce cDNA extensions. The products were separated on a 5% polyacrylamide sequencing gel using a sequence reaction of clone p3.8 primed with oligonucleotide V as molecular mass markers. Lane 1 shows the cDNA extension products and lane 2 shows a control sample treated with RNase A before reverse transcription. (B) TAP-RLPCR analysis of the rRNA precursor. The PCR products generated by ³²P-labeled oligonucleotide V and linker primer from total cellular RNA (lanes 2) (see Materials and methods and Results) were resolved in a 5% polyacrylamide sequencing gel using a sequencing reaction of clone p3.8 primed with oligonucleotide V as marker. Lane 1 shows end-labeled fragments of pBR322 digested with *Msp*I as molecular mass marker. The sequence between nt 478 and 521 is shown on the bottom of the figure to locate the mapped transcription start point (nt 505).

fragment was used as a probe for nuclease protection assays. S1 nuclease resistant fragments were analyzed on a sequencing gel. End-labeled fragments of pBR322 digested with *MspI* were used as molecular mass markers. The heteroduplex formed between the 5'-labeled probe (Fig. 2A, lane 2) and RNA yielded three protected fragments of 240, 225 and 215 bp when exposed to S1 nuclease (Fig. 2A, lanes 4–6). The ends of these protected fragments map to nt 509 ± 5 , 523 ± 5 and 533 ± 5 in the sequence shown in Fig. 4.

We tested the possibility that the RNA 5'-ends mapped by S1 nuclease protection might represent processing events rather than transcription initiation events. To discriminate between these two possibili-

ties, we performed RNase protection of in vitro capped RNA, using the same probe we used for the S1 nuclease protection assay. We used the enzyme guaninyl transferase, which catalyzes the attachment of a GMP moiety to a 5'-triphosphate or diphosphate end, but not to a 5'-monophosphate end. Thus, in vitro capping can be used to label specifically 5'-ends of primary transcripts. To obtain a rRNA precursor-enriched fraction, total cellular RNA from *E. histolytica* was fractionated by Sepharose 4B gel filtration. The labeling pattern of this RNA fraction yielded high molecular weight RNA (appearing as a smear), as well as some small RNA molecules that were specifically labeled by the enzyme (Fig. 2B, lane 2). The 474-bp *HaeIII*-*AseI* fragment (un-

```

1   CGACACGAGA AAAAGTCAA CAAAAAGTGC AAAAACCCTC CCCCTACTCA AACTACGAAT
61  TTTGAAAATG CCATTCTAC GAATTTTACC ATTTTTTTGT TGAATTGGA AGGAATGGAA
121 TAAAAAAAT TGAAATTTCA ATGAAAAGTG AACCTTTATT ACACAAAGTC TGGGAAAAAC
181 CATCAGATTA AACTAGACC AAGGATTTAA GTTGGTAAA GCTCTATTTG CCGAGAAAAG
241 ATGACGTGTG AACAACTTTT TTTCCGCATT TTCGGCCGAT ATATTGTAT GGATATTCTG
301 AATATATTTT TATTTTAGA AATATGGCGT CATCTTTTAA ATTTTCTTGG TTTTCTAAAA
361 ATAAATAACC GTTTATGAAT AAAAAAATGT TAATAGTAA ATTATAGTAA ATTAACATA
421 AAATTATAGT TTGATGTGGA TTTTCAAAA TTTTCGTTAG CTACAAATAG ACCAATAGAA
481 TACTTGTTAG TGAAAATATA CTATACAGGA GAAGTCGGTT TGAGGTCAGT TGAACAGTTT
541 TTAAGAAA GCAAAAAAC ATGTTCAATT GAAATCAGCT ATAAATAACG CCAACGAAAG
601 ATGGAATGAA TAGAATGAGG AAAACGAAA CGAAAGATGG TTGAGTTAGA ACGAAAGTGT
661 TTGAGTTAG AACGAAAGAT GGTGAGTTA GAACGAAAGA TGGTTGAGTT AACACGGAGG
721 CATTAGTAA TTGAGAATAA AAAATGATTA ATTTAAAGAT TAAGCATATT AGAGATTGAA
781 AAGAAAATGA AACAAAAAA GAATTTAAG ATTAAGCAGA TTAGAAATAG AAAAAAATA
841 TGAAACAAA AACAAATGA TATTAGTAA TGAAGAAATG AAATGATAA AAGATAGAAC
901 ATTGTTACG AGATTAAGT ATTGTGTTT TGGACCGATT CTAATGCAA GATTCAAGC
961 CTCAGAAAA TGATAGTAA TTTAGAATAT CTGAAATGA CAAATAATT AAGAAGMAAT
1021 TAATGGAGAA ATGAATTTA GAAATATAT AAGATTTTTC ACGAGATTAA AATATTGTGT
1081 TTCTGGACCG ATTCTAATGC AAGATTTTA AGCGTAGCA AATCAGATG GAATTTAGAA
1141 TAATGAAA TGACC AATA ATTAAGAAGA AATAATGGA GAAATGAAAT TAAGAAATTA
1201 PATAAGATTT TTCACGAGAT TAAATATTG TGTTCCTGGA CCGATTCTAA TGCAAAGATT
1261 CCAAGCGTAG CAAAATGAT AGTGATTTA GAATATCTGA AATGACCA ATAATTAAGA
1321 AGAAATTAAT GGAGAAATGA AATTAAGAAA TTATATAAGA TTTTACGA GATTAAAATA
1381 TTGTGTTTCT GGACCGATTC TAATGCAAG NTTTTAAGCG TAGCAAAAT GATATAGATT
1441 TTGGATGATC ATGGACAATG GATAAAAAA AGAAATCTA AAAATAAAA ATGAAAGGGA
1501 AATTAAGGAT GAGAATTGAC CAAAAAATGT CTCATTTTTC ACGAGATTAA AATATGTTAA
1561 TCTAGGACCA TCAGAGATGC AAGATTTTA ATTTCTGAA

```

Fig. 4. Sequence localization of the transcription start point. The repeated elements found with the dot matrix plot are indicated in the sequence. The sequence starts at the *TaqI* site shown on top of Fig. 1. Repeats are indicated with alternate white and black boxes for emphasis. A 21-bp element is repeated four times. An element of approx. 149 bp is localized downstream from the 21-bp repeats, and is also repeated four times.

labeled) was used to protect the radiolabeled-capped rRNA precursor. RNase A treatment of heteroduplexes formed between radiolabeled capped RNA and the 474-bp *HaeIII*-*AscI* fragment yielded a unique protected fragment of 240 nt (Fig. 2B, lanes 3 and 4). Thus, the 240-bp protected fragment observed in both nuclease protection assays (S1 nuclease and RNase A) represents the protection of a primary rRNA transcript which maps the transcription start point in the close vicinity of position 509 (509 ± 5) in the sequence shown in Fig. 4. The 225 and 215 bp protected fragments observed in the S1 mapping experiment probably represent S1 nuclease digestion artifacts or additional processing events.

3.4. High-resolution mapping of the transcription start point

High resolution primer extension experiments were carried out to define the exact site of transcription initiation of the rRNA precursor. End-labeled oligonucleotide V was used as primer. The extension reactions were performed using total cellular RNA, and the products were resolved on a denaturing polyacrylamide gel using the corresponding sequencing ladder as size marker. The two longest extension products, indicated by arrows, are shown in Fig. 3A. The sizes of the extended products are 124 and 123 nt (lane 1). A control sample treated with RNase A did not generate cDNA extensions. The 124-nt extension product is consistent with a transcription start point mapping at nt 505 in the sequence shown in Fig. 4.

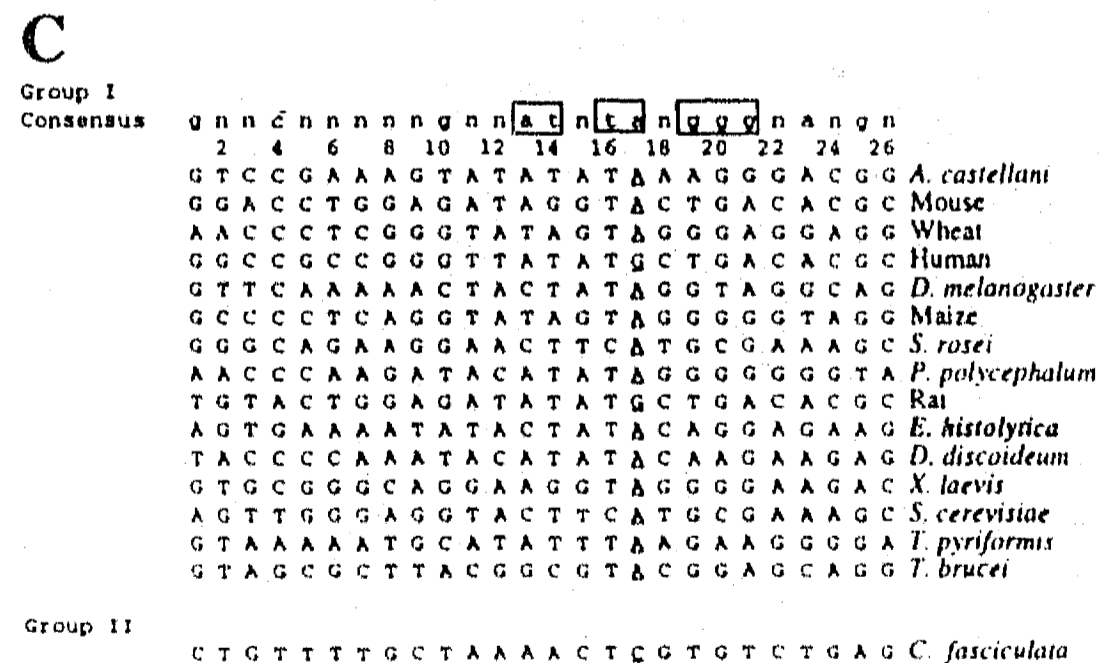
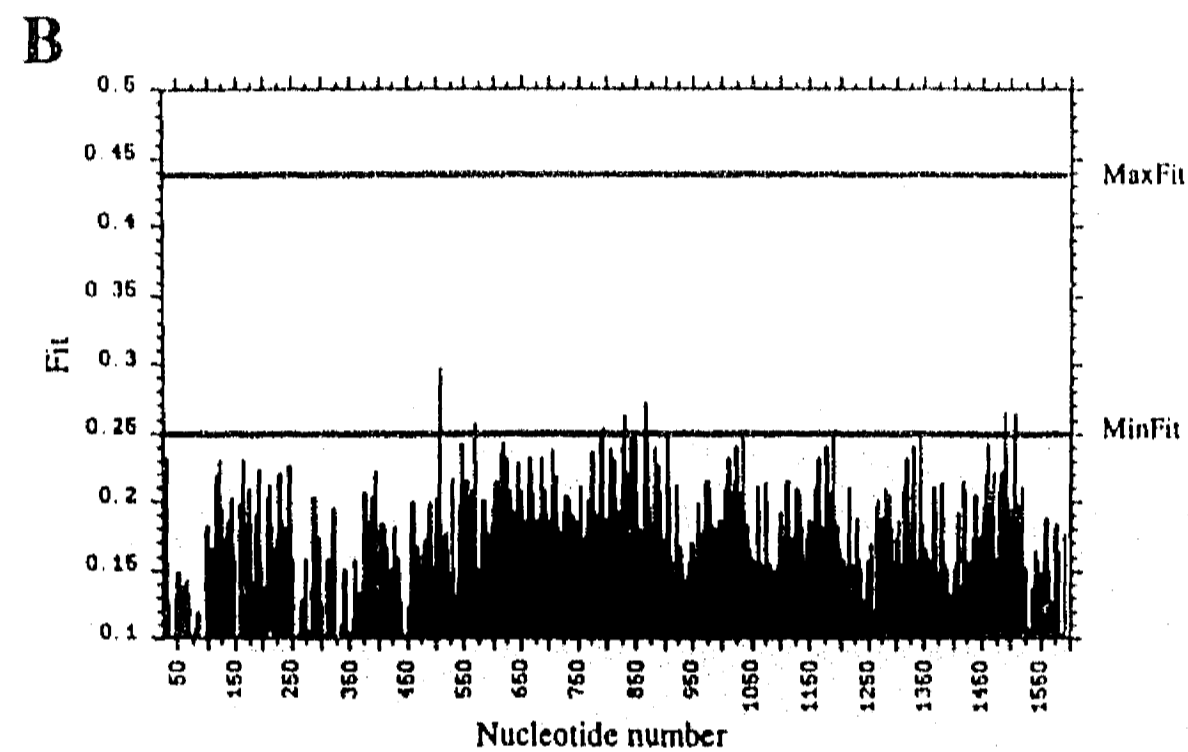
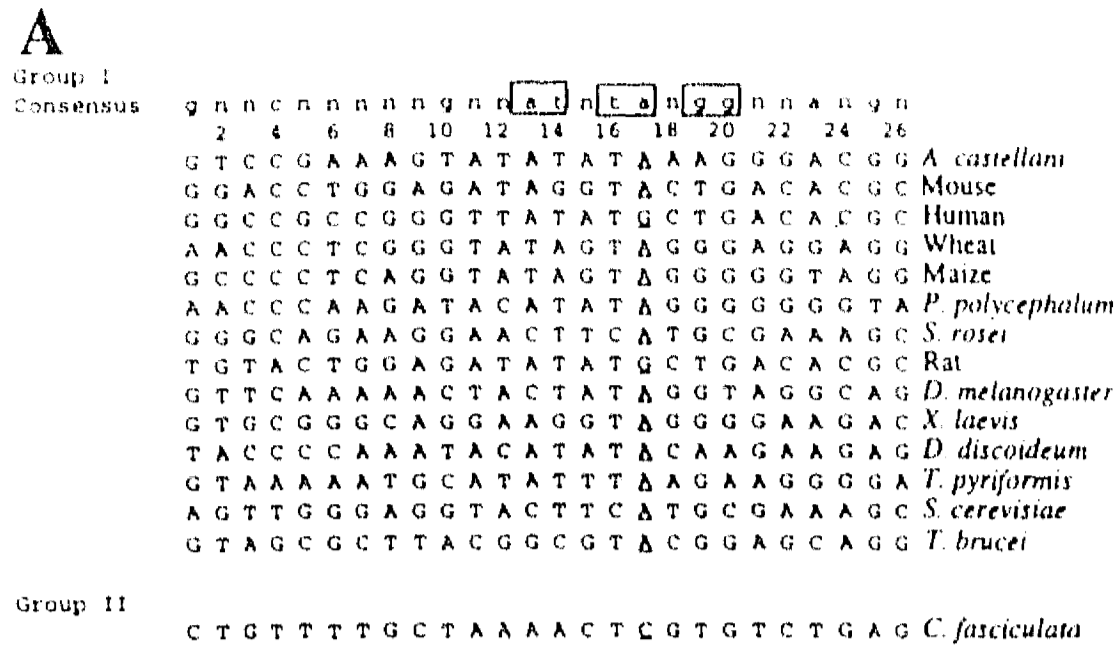
The common methods for mapping the 5' ends of transcripts, including nuclease protection and primer extension analysis, have been reported subject to various limitations and artifacts. In order to confirm the data we obtained from both nuclease protection assays and primer extension analysis, we performed an alternative procedure for mapping 5' termini of primary transcripts, referred to as TAP-RLPCR [19]. After removal of 5'-terminal pyrophosphate a specific oligonucleotide was ligated to the 5' ends of primary transcripts. cDNAs containing exclusively ribosomal RNA sequences were generated using a specific primer and reverse transcriptase. The cDNA was amplified using a nested 32 P-labeled specific primer (oligonucleotide V) and a DNA primer whose

sequence is identical to the RNA linker. The amplified products were resolved on a denaturing sequencing gel using the corresponding sequence ladder as marker. Fig. 3B, lane 2 shows the amplified product of total RNA. The size of the amplified product is 141 nt. Taking into account the length of the linker primer, plus two adenosines added by the template-independent terminal transferase activity of *Taq* polymerase, which results in the addition of a single nucleotide at each 3'-end of the fragment of the double-stranded fragment [21], the 5' end of the primary transcript maps at exactly the same position that the one obtained by primer extension, that is, the adenosine at nt 505 of the sequence shown in Fig. 4, within the sequence 5'-ATATACTATACAGG-3' (the start site is underlined). This site is located 2447 nt upstream from the 17S structural gene.

3.5. Sequence analysis of the promoter region of the rRNA precursor

Tandemly repeated DNA elements are commonly present in promoter regions of rRNA genes [22]. We looked for such elements in the sequence surrounding the transcription start point of *E. histolytica* using a dot matrix plot analysis (Gene Works, Intellegentics). Two regions with tandem repeats, both downstream the transcription start point, were observed on the dot matrix plot (data not shown). The smaller repetitive element is composed of four 21 bp repeats organized in tandem, located 123 nt downstream from the transcription start point (Fig. 4). The second type of repetitive element consists of four closely related repeats with lengths of 118, 149, 149 and 148 bp for the first, second, third and fourth repeat, respectively. These repeats are located 376 nt downstream from the transcription start point (Fig. 4).

We performed a computer analysis of the region encompassing roughly 3000 nt upstream from the 17S gene using the program PatternScan (InitScanAG option), which scans the entire sequence using a weight matrix similar to that reported by Staden [23], with the important difference that the sequences used to generate the weight matrix are first ranked and sorted for consensus coherence. The weight matrix was obtained using as a raw database the core promoter sequences from positions -16 to



+10 of the rRNA genes of 15 organisms. The PatternScan analysis, shown in Fig. 5A, built a preferred weight matrix using a subset of only 14 of the 15 initiator site sequences we provided as the consensus database. The sequence of the *C. fasciculata* ribosomal core promoter was rejected by the algorithm because it deviates strongly from the consensus base frequencies. The scanning plot generated by the program (5B) shows a consensus fit maximum at nt 505 of the first 1600 nt of the sequence starting at the *TaqI* site shown in Fig. 1, a site which corresponds exactly to the initiator base that we identified by primer extension. The weight matrix program found this maximum in the 1600-nt window shown in Fig. 5B using a database that did not contain the *E. histolytica* core promoter sequence. Sequences with a similar high value of consensus fit are infrequent in the 3200-nt domain that we examined. When the *E. histolytica* core sequence was included in the promoter database, the program ranked the *E. histolytica* core sequence in the 10th position among the ordered consensus fit scores (Fig. 5C).

4. Discussion

It is widely documented that whereas the sequences coding for the 18S, 5.8S and 28S rRNA

have been strongly conserved throughout evolution, both the size and the sequence of the transcribed spacer segments are highly diverged (for review, see Refs. 22,24). Our results show the presence of a 2447-nt transcribed spacer. Preliminary Northern blot analysis (not shown) had suggested that the ribosomal precursor molecules are rapidly processed, as has been observed in several other biological systems [22]. In various species the first processing event of pre-rRNA occurs so rapidly both in vivo and in vitro that the 5' end of the predominant pre-rRNA maps to this processing site. This has led, for example, to the erroneous identification of a transcription start point based on S1 mapping data in the mouse [25,26]. To identify the start site of the rDNA transcription unit in *E. histolytica*, we first mapped nascent transcripts, in order to locate transcriptionally active regions. We then employed a variety of complementary approaches to map the start point. S1 nuclease mapping, as well as RNase A protection assays of in vitro-capped RNA yielded a common 240-bp fragment. The capping/protection assay demonstrated the presence of either a diphosphate or a triphosphate terminus, by virtue of the biochemical specificity of the labeling reaction. Finally, the start point was identified at single-base resolution by primer extension analysis, and independent evidence for the presence of a triphosphate

Fig. 5. Weight matrix scanning of sequences upstream the rDNA genes of *E. histolytica* for segments matching a transcriptional initiation sequence consensus, using a -16 to +10 data window. (A) Sequence sort output of program PatternScan (InitScanAG option) showing a set of known ribosomal initiator sequences ordered according to the degree of fit to their own consensus, using a window of 26 nt. The initiator site base (#17 in this table) is underlined. The scoring matrix algorithm used is essentially the same as that described by Staden [23], but the program penalizes the fit score at the initiator position by 30% if the base at that position is not an A or a G. The sorting is based on an arbitrary fit cut value that is adjusted by the user of the program to increase or decrease consensus set homogeneity. Stringent (i.e., high) cut values may yield several different consensus sets from a single database. We used a low-stringency setting (0.45) because our database is small. At this setting the program sorted out the *C. fasciculata* on the basis of its large deviation from the consensus of the other 15 sequences, which comprises the majority set. (B) Consensus fit scores at each position in the first 1600 nt of the *E. histolytica* sequence (starting at the *TaqI* site shown in Fig. 1), as generated by the program PatternScan (InitScanAG option), using a weight matrix calculated from a subset that included 14 out of 15 sequences given as input to the program. The lines indicated by MaxFit and MinFit indicate the consensus fit scores of the 'best' sequence (*A. castellanii*) and the 'worst' sequence (*T. brucei*) in the set of 14 promoters used by the program to build a consensus weight matrix. The fit score of position 505 was 0.2970. When the sequence from nt 1600 to 3200 was scanned, three positions gave fit score values similar to that obtained in position 505 (0.2941, 0.2937 and 0.2980). (C) Results of running a new sorting phase of the PatternScan (InitScanAG option) program, this time including the *E. histolytica* transcriptional initiator sequence as the 16th item in the input database. The *C. fasciculata* sequence was again sorted out as a non member of the consensus set, and the weight matrix scores placed the *E. histolytica* sequence in the 10th position. Transcription initiation sequences are from *A. castellanii* [33,28], mouse [34], human [35], rat [36], *S. rosei* [37] *D. melanogaster* [38], *D. discoideum* [39], *P. polycephalum* [40], *X. laevis* [41], *T. pyriformis* [42] *S. cerevisiae* [43], and *T. brucei* [16], wheat [44], maize [45] and *C. fasciculata* [46]. The program PatternScan is available by anonymous ftp from ibt.unam.mx. It runs on Apple Macintosh computers with a math co-processor.

terminus was obtained by TAP-RLPCR analysis. The transcription start point was found to map at nt 505 of the sequence shown in Fig. 4.

Weak homologies have been reported between the initiator sequences that constitute the polymerase I core promoter [27]. A comparison of the transcriptional initiation sequences of rRNA genes has been reported recently using initiator sequences from 14 species and a consensus for the initiator was obtained [28]. Using the same rules, we derived a weak consensus for the 15 initiator sequences shown in Fig. 5C, from nt positions -16 to $+10$: $-g/r, n, n, C, n, n, n, r, g, n, n, a, T, r, T, \underline{A}, n, g/r, G, g/r, r, r, n, G, n$ (where n = any nucleotide; r = purine $\geq 75\%$; upper-case letters $\geq 66\%$ of the time; lower-case letters $\geq 50\%$ of the time; the initiator nucleotide is underlined). We chose not to include in our database sequences data from more than two closely related species to avoid introducing a bias in the consensus. Therefore, this consensus is not identical to that reported in [28]; only nucleotides from -4 to $+1$ and the guanosine at position $+3$ are conserved.

Some *E. histolytica* strains have been shown to contain additional transcripts encoded on the rRNA episome [29,30]; however, no such transcripts are found in the HK9 strain used in this study [30]. Our physical mapping data is supported by sequence motif analysis; computer analysis using Staden's weight matrix algorithm [23] identified a consensus fit maximum at nt 505 from a scanning of the first 1600 nt (starting at the *TaqI* site shown in Fig. 1) using a database made of initiator sequences of rRNA genes from 15 different organisms (not including the *E. histolytica* sequence). This site corresponds to the transcription start point that we identified experimentally.

A common structural organization for the rRNA promoter exists; it consists of a core domain around the transcription start point essential for initiation, and upstream control elements (UCE) required for efficient transcription [22,31]. UCE consist of repeated elements of variable length located upstream the transcription start point and it has been demonstrated that they influence rDNA transcriptional activity (for a review, see Ref. 22). Two copies of a 644-bp repeat of unknown function are present 669 nt upstream the transcription start point of *E. histolytica* [32]. Another non-transcribed repetitive ele-

ment (denominated p145 to indicate its size) has been identified in the rDNA episome of *E. histolytica* [2]. Our sequence data show the presence of two additional repetitive elements; the shorter one consists of four perfect copies of a 21-bp sequence and the longer one consists of four imperfect copies of a 149-bp sequence. Interestingly, both repetitive elements are part of the primary transcript. Functional assays have to be done to limit and to define a promoter region and to assess whether these repetitive elements could be involved in rRNA transcription regulation. Although the localization of these repetitive elements is atypical when compared to the common organization of regulatory elements in the rDNA of other eukaryotes, their vicinity to the transcription start point suggests a possible regulatory role. Since *E. histolytica* rDNA genes are located in a circular extrachromosomal DNA molecule, the possible regulatory role of the various types of repeats can only be assessed by functional assays. The future development of an in vitro transcription system and the use of transient transformation assays [7,8] for *E. histolytica* will define a functional rRNA promoter and may help to clarify the putative role of these various repetitive elements in rRNA transcriptional regulation.

Acknowledgements

We thank Dr. Patricia León, Dr. Felix Recillas, Dr. Alejandra Covarrubias, Dr. Luis Miguel Salgado for helpful suggestions, and Alejandro Olvera and Felipe Olvera for their expert assistance in the culture of *E. histolytica*. We also like to thank very specially Dr. Gloria Rudenko for her helpful advice regarding nuclear run-on analysis. This work was supported by grants from DGAPA/UNAM, the Rockefeller Foundation (RF 89088#79) and the Howard Hughes Medical Institute (75191-527103). P.M.L. is an International Research Scholar of the Howard Hughes Medical Institute.

References

- [1] Walsh, J.A. (1986) Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Rev. Infect. Dis.* 8, 228–238.

- [2] Huber, M., Koller, B., Gitter, C., Mirelman, D., Revel, M., Rozenblatt, S. and Garfinkel, L. (1989) *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules. *Mol. Biochem. Parasitol.* 32, 285–296.
- [3] Bhattacharya, S., Bhattacharya, A., Diamond, L.S. and Soldo, A.T. (1989) Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. *J. Protozool.* 36, 455–458.
- [4] Grodberg, J., Salazar, N., Oren, R. and Mirelman, D. (1990) Autonomous replication sequences in an extrachromosomal element of a pathogenic *Entamoeba histolytica*. *Nucleic Acids Res.* 18, 5515–5519.
- [5] Zurita, M., Alagon, A., Vargas, V.J. and Lizardi, P.M. (1991) The *Entamoeba histolytica* rDNA episome: nuclear localization, DNAase I sensitivity map, and specific DNA-protein interactions. *Mol. Microbiol.* 5, 1843–1851.
- [6] Albach, R.A., Booden, T., Boonlayangoor, P. and Downing, S. (1980) Concepts of function of peripheral non-chromatin and endosome in *Entamoeba histolytica*. *Arch. Invest. Med. (México)* 11, 63–74.
- [7] Purdy, J.E., Mann, B.J., Pho, L.T. and Petri, W.J. (1994) Transient transfection of the enteric parasite *Entamoeba histolytica* and expression of firefly luciferase. *Proc. Natl. Acad. Sci. USA* 91, 7099–7103.
- [8] Nickel, R. and Tannich, E. (1994) Transfection and transient expression of chloramphenicol acetyltransferase gene in the protozoan parasite *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. USA* 91, 7095–7098.
- [9] Diamond, L.S., Harlow, D.R. and Cunnick, C.C. (1978) A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans. R. Soc. Trop. Med. Hyg.* 72, 431–432.
- [10] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [12] Zurita, M., Bieber, D., Ringold, G. and Mansour, T.E. (1987) Cloning and characterization of a female genital complex cDNA from the liver fluke *Fasciola hepatica*. *Proc. Natl. Acad. Sci. USA* 84, 2340–2344.
- [13] Mulligan, R.M., Maloney, A.P. and Walbot, V. (1988) RNA processing and multiple transcription initiation sites result in transcript size heterogeneity in maize mitochondria. *Mol. Gen. Genet.* 211, 373–380.
- [14] Grosschedl, R., Weaver, D., Baltimore, D. and Costantini, F. (1984) Introduction of a mu immunoglobulin gene into the mouse germ line: specific expression in lymphoid cells and synthesis of functional antibody. *Cell* 38, 647–658.
- [15] Goldberg, D.A. (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* 77, 5794–5798.
- [16] Kooter, J.M., van der Spek, H., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J. and Borst, P. (1987) The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. *Cell* 51, 261–272.
- [17] Mulligan, R.M., Leon, P. and Walbot, V. (1991) Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. *Mol. Cell. Biol.* 11, 533–543.
- [18] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.R., Seidman, J.A. and Strahl, K. (1994) *Current Protocols in Molecular Biology*. Wiley, New York, NY.
- [19] Fromont, R.M., Bertrand, E., Pictet, R. and Grange, T. (1993) A highly sensitive method for mapping the 5' termini of mRNAs. *Nucleic Acids Res.* 21, 1683–1684.
- [20] Que, X. and Reed, S.J. (1991) Nucleotide sequence of a small subunit ribosomal RNA (16S-like rRNA) gene from *Entamoeba histolytica*: differentiation of pathogenic from nonpathogenic isolates. *Nucleic Acids Res.* 19, 5438.
- [21] Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Res.* 16, 9677–9686.
- [22] Sollner, W.B. and Mougey, E.B. (1991) News from the nucleolus: rRNA gene expression. *Trends Biochem. Sci.* 16, 58–62.
- [23] Staden, R. (1984) Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res.* 12, 505–519.
- [24] Palmer, J.M. and Fulk, W.R. (1990) Unraveling the complexities of transcription by RNA polymerase III. *Trends Biochem. Sci.* 15, 300–304.
- [25] Grummt, I. (1981) Mapping of a mouse ribosomal DNA promoter by in vitro transcription. *Nucleic Acids Res.* 9, 6093–6102.
- [26] Urano, Y., Kominami, R., Mishima, Y. and Muramatsu, M. (1980) The nucleotide sequence of the putative transcription initiation site of a cloned ribosomal RNA gene of the mouse. *Nucleic Acids Res.* 8, 6043–6058.
- [27] Sommerville, J. (1984) RNA polymerase I promoters and transcription factors [news]. *Nature* 310, 189–190.
- [28] Perna, P.J., Harris, G.H., Iida, C.T., Kownin, P., Bugren, S. and Paule, M.R. (1992) The start site of the *Acanthamoeba castellanii* ribosomal RNA transcription unit. *Gene Expr.* 2, 71–78.
- [29] Jansson, A., Gillin, F., Kagardt, U. and Hagblom, P. (1994) Coding of hemolysins within the ribosomal RNA repeat on a plasmid in *Entamoeba histolytica*. *Science* 263, 1440–1443.
- [30] Sehgal, D., Mittal, V., Ramachandran, S., Dhar, S.K., Bhattacharya, A. and Bhattacharya, S. (1994) Nucleotide sequence organisation and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 67, 205–214.
- [31] Sollner, W.B. and Tower, J. (1986) Transcription of cloned eukaryotic ribosomal RNA genes. *Annu. Rev. Biochem.* 55, 801–830.
- [32] Sehgal, D., Bhattacharya, A. and Bhattacharya, S. (1993) Analysis of a polymorphic locus present upstream from rDNA transcription units in the extrachromosomal circle of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 62, 129–130.
- [33] Kownin, P., Iida, C.T., Brown, S.S. and Paule, M.R. (1985) The ribosomal RNA promoter of *Acanthamoeba castellanii* determined by transcription in a cell-free system. *Nucleic Acids Res.* 13, 6237–6248.

- [34] Grunmt, I. (1981) Specific transcription of mouse ribosomal RNA in a cell free system that mimics control in vivo. Proc. Natl. Acad. Sci. USA 78, 727-731.
- [35] Financesek, I., Mizumoto, K., Mishima, Y. and Muramatsu, M. (1982) Human ribosomal RNA gene: nucleotide sequence of the transcription initiation region and comparison of three mammalian genes. Proc. Natl. Acad. Sci. USA 79, 3092-3096.
- [36] Financesek, I., Mizumoto, K. and Muramatsu, M. (1982) Nucleotide sequence of the transcription initiation region of a rat ribosomal RNA gene. Gene 18, 115-122.
- [37] Verbeet, M.P., Klootwijk, J., van Heerikhuizen, H., Fontijn, R., Vreugdenhil, E. and Planta, R.J. (1983) Molecular cloning of the rDNA of *Saccharomyces rosei* and comparison of its transcription initiation region with that of *Saccharomyces carlsbergensis*. Gene 23, 53-63.
- [38] Kohorn, B.D. and Rae, P.M. (1983) Localization of DNA sequences promoting RNA polymerase I activity in *Drosophila*. Proc. Natl. Acad. Sci. USA 80, 3265-3268.
- [39] Hoshikawa, Y., Iida, Y. and Iwabuchi, M. (1983) Nucleotide sequence of the transcriptional initiation region of *Dictyostelium discoideum* rRNA gene and comparison of the initiation regions of three lower eukaryotes' genes. Nucleic Acids Res. 11, 1725-1734.
- [40] Blum, B., Seebeck, T., Braun, R., Ferris, P. and Vogt, V. (1983) Localization and DNA sequence around the initiation site of ribosomal RNA transcription in *Physarum polycephalum*. Nucleic Acids Res. 11, 8519-8533.
- [41] Boseley, P., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. (1979) Sequence organization of the spacer DNA in a ribosomal gene unit of *Xenopus laevis*. Cell 17, 19-31.
- [42] Saiga, H., Mizumoto, K., Matsui, T. and Higashinakagawa, T. (1982) Determination of the transcription initiation site of *Tetrahymena pyriformis* rDNA using in vitro capping of 35S pre-rRNA. Nucleic Acids Res. 10, 4223-4236.
- [43] Kulkens, T., Riggs, D.L., Heck, J.D., Planta, R.J. and Nomura, M. (1991) The yeast RNA polymerase I promoter: ribosomal DNA sequences involved in transcription initiation and complex formation in vitro. Nucleic Acids Res. 19, 5363-5370.
- [44] Barker, R.F., Harberd, N.P., Jarvis, M.G. and Flavell, R.B. (1988) Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. J. Mol. Biol. 201, 1-17.
- [45] McMullen, M.D., Hunter, B., Phillips, R.L. and Rubenstein, I. (1986) The structure of the maize ribosomal DNA spacer region. Nucleic Acids Res. 14, 4953-4968.
- [46] Grondal, E.J., Evers, R. and Cornelissen, A.W. (1990) Identification and sequence analysis of the ribosomal DNA promoter region of *Crithidia fasciculata*. Nucleic Acids Res. 18, 1333-1338.