

**MAESTRÍA EN CIENCIAS GENÓMICAS**

---

**Expresión de genes *myb* SHAQKYF en *Entamoeba histolytica***

**T E S I S**

QUE PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIAS GENÓMICAS  
P R E S E N T A

**I.B.I. Helios Cárdenas Hernández**

**DIRECTORA DE TESIS**

DRA. Elisa Irene Azuara Liceaga

**Co-DIRECTORA DE TESIS**

DRA. Claudia Selene Zárate Guerra

## SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



## UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

### RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

#### DERECHOS RESERVADOS ©

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

# **COMITÉ TUTORIAL**

## **DIRECTOR**

Dra. Elisa Irene Azuara Liceaga

Profesor Titular. Posgrado en Ciencias Genómicas.  
Universidad Autónoma de la Ciudad de México.

## **CO-DIRECTOR**

Dra. Claudia Selene Zárate Guerra

Profesor Titular. Posgrado en Ciencias Genómicas.  
Universidad Autónoma de la Ciudad de México.

## **ASESORES**

Dra. Esther López Bayghen

Profesor Titular. Departamento de Genética y Biología Molecular.  
Cinvestav-IPN.

Dra. Minerva Camacho Nuez

Profesor Titular. Posgrado en Ciencias Genómicas.  
Universidad Autónoma de la Ciudad de México.

Dr. Mario César López Camarillo

Profesor Titular. Posgrado en Ciencias Genómicas.  
Universidad Autónoma de la Ciudad de México.

El presente proyecto de investigación fue realizado gracias al apoyo académico y económico que me otorgó la Universidad Autónoma de la Ciudad de México, a la cual agradezco.

El presente trabajo se realizó en el laboratorio 2 del Posgrado en Ciencias Genómicas de la Universidad Autónoma de la Ciudad de México bajo la dirección de la Dra. Elisa Irene Azuara Liceaga y la co-dirección de la Dra. Claudia Selene Zárate Guerra así como en el Laboratorio de la Dra. Esther Orozco Orozco de Patología Experimental del CINVESTAV-IPN.

Esta tesis se llevó a cabo gracias al financiamiento por parte de la Universidad Autónoma de la Ciudad de México (UACM), el Consejo Nacional de Ciencia y Tecnología (CONACyT) y al programa "Creación Joven" a través del Instituto de Ciencia y Tecnología del Distrito Federal (ICyT DF).

Para José Luis, Martha Lilia, Violeta , Dalia y Nancy.

Gracias por todo su apoyo.

# Índice

<b>1. Introducción</b>	<b>7</b>
1.1. Antecedentes generales del tema	7
1.1.1. Diversidad fenotípica	7
1.1.2. Expresión génica	7
1.2. Estado del arte	15
1.2.1. <i>Entamoeba histolytica</i>	15
1.2.2. Epidemiología de la amibiasis	15
1.2.3. Ciclo de vida de <i>Entamoeba histolytica</i>	15
1.2.4. Genoma de <i>Entamoeba histolytica</i>	16
1.3. Antecedentes particulares del proyecto	19
1.3.1. Proteínas MYB como factores de transcripción	19
1.3.2. Transcripción en <i>E.histolytica</i>	21
<b>2. Hipótesis y Objetivos</b>	<b>24</b>
2.1. Hipótesis	24
2.2. Objetivos	24
<b>3. Estrategia experimental</b>	<b>25</b>
3.1. Búsqueda de secuencias homólogas, alineamientos y reconstrucción filogenética.	25
3.2. Predicción de estructura y dominios funcionales de las proteínas EhMyb con el motivo SHAQKYF de <i>E.histolytica</i> .	26
3.3. Análisis de los niveles de expresión de los genes <i>ehmyb</i> SHAQKYF en <i>E.histolytica</i> .	26
3.3.1. Cultivo <i>in vitro</i> de trofozoitos de <i>E.histolytica</i>	26
3.3.2. Obtención de ADN genómico de <i>E.histolytica</i>	26
3.3.3. Diseño de los oligonucleótidos para PCR y clonación	27
3.3.4. RT-PCR en tiempo final en condiciones basales de cultivo	27
3.3.5. Cinéticas de amplificación para los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i>	31
3.3.6. RT-PCR en tiempo final en condiciones de estrés calórico.	33

3.3.7.	RT-PCR en tiempo final en condiciones de eritrofagocitosis.	33
3.4.	Clonación y Secuenciación de los genes <i>Ehmybs1</i> y <i>Ehmyb21</i> . . . . .	35
3.4.1.	Obtención de plásmidos y restricción enzimática. . . . .	35
3.4.2.	Reacción de secuenciación de los genes <i>Ehmybs1</i> y <i>Ehmyb21</i>	38
3.5.	Clonación, expresión y purificación de la proteína EhMybS3 de <i>E.histolytica</i> . . . . .	39
3.5.1.	Obtención del inserto para clonación del gen <i>Ehmybs3</i> . . . . .	39
3.5.2.	Obtención del vector pGEX-6P-1 . . . . .	41
3.5.3.	Ligación del gen <i>Ehmybs3</i> en el plásmido pGEX-6P-1 . . . . .	41
3.5.4.	Expresión de la proteína recombinante EhMybS3 . . . . .	42
3.5.5.	Purificación de la proteína EhMybS3r . . . . .	43
3.5.6.	Ensayo de inmunodetección de la proteína EhMybS3r . . . . .	43
3.6.	Análisis bioquímico mediante ensayos de retardamiento de la pro- teína EhMybS3r de <i>E.histolytica</i> . . . . .	44
3.6.1.	Hibridación de los oligonucleótidos . . . . .	44
3.6.2.	Marcaje de los oligonucleótidos . . . . .	45
3.6.3.	Ensayos de retardamiento en gel (EMSA) . . . . .	45
3.7.	Búsqueda de las posibles secuencias blanco de proteínas EhMyb con el motivo SHAQKYF en el genoma de <i>E.histolytica</i> . . . . .	46
<b>4.</b>	<b>Resultados</b>	<b>48</b>
4.1.	Reconstrucción filogenética . . . . .	48
4.1.1.	Búsqueda de las secuencias de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> . . . . .	48
4.1.2.	Alineamiento de las secuencias de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> . . . . .	50
4.1.3.	Selección del mejor método de sustitución. . . . .	50
4.1.4.	Reconstrucción del árbol filogenético . . . . .	50
4.2.	Reconstrucción filogenética de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> en relación con las de otros organismos. . . . .	52
4.2.1.	Búsqueda de las secuencias de proteínas Myb SHAQKYF de otros organismos. . . . .	52
4.2.2.	Selección del mejor método de sustitución. . . . .	54

4.2.3. Reconstrucción del árbol filogenético . . . . .	54
4.3. Predicción de estructura y dominios funcionales . . . . .	54
4.3.1. Predicción de estructura secundaria y terciaria. . . . .	54
4.3.2. Predicción de dominios funcionales. . . . .	57
4.3.3. Predicción de señales de localización nuclear. . . . .	60
4.4. Análisis de los niveles de expresión de los genes <i>ehmyb</i> SHAQKYF de <i>E.histolytica</i> en condiciones de cultivo. . . . .	60
4.4.1. Expresión de los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i> en condiciones basales de cultivo. . . . .	60
4.4.2. Cinéticas de ciclado para los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i> . . . . .	61
4.4.3. Corroboración de la identidad de los genes <i>Ehmybs1</i> y <i>Ehmybs2</i> por secuenciación. . . . .	68
4.5. Análisis de los niveles de expresión de los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i> de <i>E.histolytica</i> en condiciones de estrés calórico. . . .	68
4.6. Análisis de los niveles de expresión de los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i> de <i>E.histolytica</i> en condiciones de eritrofagocitosis. . .	71
4.7. Obtención de la proteína EhMybS3r de <i>E.histolytica</i> . . . . .	75
4.7.1. Expresión y Purificación . . . . .	75
4.8. Ensayos de retardamiento para la proteína EhMybS3 de <i>E.histolytica</i>	76
4.9. Búsqueda de posibles secuencias blanco de proteínas EhMyb SHAQKYF en el genoma de <i>E.histolytica</i> . . . . .	79
<b>5. Discusión</b>	<b>86</b>
<b>6. Conclusiones</b>	<b>90</b>
<b>7. Apendice A: Abreviaturas</b>	<b>91</b>
<b>8. Apéndice B: Soluciones</b>	<b>96</b>

## Índice de figuras

1. Elementos de un promotor eucarionte. . . . .	9
2. Inicio de la Transcripción en Eucariontes . . . . .	11
3. Funciones de un Activador Transcripcional. . . . .	13
4. Dominio de Unión a ADN Hélice-Vuelta-Hélice. . . . .	14
5. Ciclo de vida de <i>Entamoeba histolytica</i> . . . . .	17
6. Características del genoma de <i>Entamoeba histolytica</i> y comparación de éstas con las de otros organismos eucariontes . . . . .	18
7. Dominio de Unión a ADN de c-Myb. . . . .	22
8. Condiciones de amplificación. . . . .	31
9. Mapa del vector de expresión bacteriano pGEX-6P-1. . . . .	40
10. Segmento más conservado en el alineamiento de las proteínas EhMyb encontradas en el genoma de <i>E.histolytica</i> . . . . .	51
11. Árbol filogenético de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> . . . . .	53
12. Segmento más conservado en el alineamiento de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> y de otros organismos . . . . .	55
13. Árbol filogenético de las proteínas EhMyb SHAQKYF de <i>E.histolytica</i> en relación con las de otros organismos. . . . .	56
14. Estructura secundaria de las proteínas EhMybS1, EhMybS2 y EhMybS3. . . . .	58
15. Estructura terciaria de las proteínas EhMyb SHAQKYF . . . . .	59
16. Señal de localización nuclear de la proteína EhMybS3. . . . .	60
17. Expresión basal de los genes <i>ehmybs1</i> , <i>ehmybs2</i> y <i>ehmybs3</i> . . . . .	62
18. Cinética de ciclado para el gen <i>25s rARN</i> . . . . .	64
19. Cinética de ciclado para el gen <i>EhmybS1</i> . . . . .	65
20. Cinética de ciclado para el gen <i>EhmybS2</i> . . . . .	66
21. Cinética de ciclado para el gen <i>EhmybS3</i> . . . . .	67
22. Alineamiento de secuencias del gen <i>ehmybS1</i> . . . . .	69
23. Alineamiento de secuencias del gen <i>ehmybS1</i> . . . . .	70

24.	RT-PCR de los genes <i>ehmybs1</i> , <i>ehmybs2</i> y <i>ehmybs3</i> en respuesta al choque térmico . . . . .	72
25.	Cambios en la expresión de los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> , <i>Ehmybs3</i> y <i>peroxirredoxina</i> en condiciones de estrés calórico	73
26.	RT-PCR de los genes <i>Ehmybs1</i> , <i>Ehmybs2</i> y <i>Ehmybs3</i> durante la eritrofagocitosis. . . . .	74
27.	Expresión y purificación de la proteína EhMybS3r . . . . .	77
28.	Western Blot de la expresión y purificación de pEhMybS3r	78
29.	Ensayo de unión a ADN para la proteína EhMybS3r . . . . .	80
30.	Genes que contienen en su región promotora el elemento de reconocimiento a Myb T(C/T)AAC(G/T)G. . . . .	83
31.	Genes que contienen en su región promotora la secuencia AAC(A/T)GTT de <i>D. discoideum</i> . . . . .	84
32.	Genes que contienen en su región promotora la secuencia TTACCA de <i>O. sativa</i> . . . . .	85

## Índice de tablas

1.	Secuencias de los oligonucleótidos diseñados . . . . .	28
2.	Mezcla de reacción para el tratamiento de DNAsa del ARN.	30
3.	Mezcla para la reacción de retrotranscripción . . . . .	30
4.	Mezclas de reacción para PCR. . . . .	32
5.	Protocolo de PCR utilizado para la condición de estrés calórico . . . . .	34
6.	Protocolo de PCR utilizado para la condición de eritrofagocitosis . . . . .	36
7.	Mezcla de reacción para clonación en el vector TOPO TA 2.1 . . . . .	37
8.	Mezcla de reacción de secuenciación . . . . .	38
9.	Reacción de Secuenciación . . . . .	38
10.	Sondas utilizadas en los ensayos de EMSA . . . . .	46
11.	Proteínas EhMyb SHAQKYF de <i>E.histolytica</i> . . . . .	49

## Resumen

En este trabajo se inició el estudio de un grupo de proteínas que tienen el dominio de unión a ADN Myb de un solo repetido, el cual posee un motivo SHAQKYF. Como resultado de este trabajo identificamos que *E.histolytica* tiene 9 proteínas de tipo Myb SHAQKYF que llamamos EhMybS1-S9, encontramos que 3 de ellas (EhMybS1-S3) conforman un grupo monofilético y que tienen como ortólogos más cercanos a otras proteínas de plantas y del organismo *Dictiostellium discoideum*. Además observamos que estas proteínas conservan la estructura terciaria característica de los dominios de unión a ADN de tipo Myb. Posteriormente estudiamos si la expresión de estas tres proteínas podría ser modulada por diferentes estímulos, con lo cual encontramos que la proteína EhMybS3 cambia sus niveles de expresión bajo los estímulos de choque térmico y eritrofagocitosis. De esta manera los resultados de este trabajo soportan la idea de que la proteína EhMybS3 podría estar involucrada en la regulación de genes importantes para la respuesta a cualquiera de estos dos estímulos. Además en este trabajo mostramos que la proteína EhMybS3 es capaz de unirse a una secuencia de ADN que posee el elemento de reconocimiento a Myb. Finalmente, si la proteína EhMybS3 es un factor de transcripción, nos preguntamos que genes podrían ser regulados por ésta. Así, por medio del análisis de la secuencia de los promotores de los genes de *E.histolytica* encontramos alrededor de 88 genes candidatos a ser regulados por la proteína EhMybS3 ya que contienen en sus regiones promotoras las secuencias blanco de otras proteínas de tipo Myb. En conjunto los resultados de este trabajo son una primera aproximación al estudio de los posibles factores de transcripción Myb en *E.histolytica* y además soportan la idea de que la proteína EhMybS3 sea un factor de transcripción capaz de regular a otros genes involucrados en la respuesta al estrés o la virulencia en este parásito.

# 1. Introducción

## 1.1. Antecedentes generales del tema

### 1.1.1. Diversidad fenotípica

La cantidad de genes contenidos en los genomas de diferentes organismos varía notablemente, por ejemplo *Escherichia coli* tiene 4288 genes (6) mientras que el genoma humano contiene alrededor de 20 000 genes (13). En contraste, organismos con fenotipos muy diferentes como son el organismo multicelular *Arabidopsis thaliana* y el alga unicelular *Chlamydomonas reinhardtii* presentan genomas muy similares, en lo que a tamaño se refiere, de 125 y 120 Mb respectivamente (25, 36). Es claro que el contenido de los genomas varía notoriamente y no necesariamente en función de la complejidad del organismo, es decir, la gran variedad fenotípica, desde una bacteria hasta organismos eucariontes superiores, no es debida, en su totalidad, al aumento en el tamaño de sus genomas. Además, considerando a un organismo multicelular es clara la gran diversidad fenotípica que se observa entre las diferentes células que lo constituyen, por ejemplo, todas las células de la planta *Arabidopsis thaliana* contienen el mismo genoma, sin embargo aquellas que constituyen las raíces son muy diferentes a las de las hojas. Si bien todas las células de un mismo organismo contienen la misma información, cada tipo celular la utiliza de forma distinta, es decir expresa sus genes selectivamente. De esta manera cada célula utiliza la información contenida en su genoma de acuerdo a sus necesidades en un momento determinado. Así la expresión génica se presenta como un actor muy importante en esta gran diversidad fenotípica, donde no sólo hay que considerar que información codificada en el genoma si no como y cuando se expresa

### 1.1.2. Expresión génica

La expresión génica, como fue concebida en sus inicios, era el flujo unidireccional de la información contenida en los genes a través del ARN y su traducción a una molécula efectora (las proteínas), éste es el dogma central de la biología molecular propuesto por Francis Crick desde principios de los años setenta. Actualmente esta

visión ha cambiado debido a los avances en nuestro entendimiento de los procesos moleculares que tienen lugar en los organismos; en esencia, el flujo de información es muy similar con la sutil diferencia de que ya no es considerado unidireccional, es decir cierta información puede pasar del ARN al ADN por medio de la enzima transcriptasa reversa (mecanismo utilizado por los retrovirus), además el ARN también puede ser una molécula efectora (ribozimas, ARNi, microARN).

En términos generales, la expresión génica se lleva a cabo con dos procesos, conocidos como transcripción y traducción.

### **Transcripción en eucariontes**

La transcripción es el proceso mediante el cual la información contenida en el ADN es transcrita a una molécula de ARN, este proceso se lleva a cabo mediante la interacción de un complejo de proteínas con el ADN y se divide en tres fases: inicio, alargamiento y terminación (1). El inicio consiste en la unión de la enzima ARN polimerasa a una secuencia específica en el ADN llamada región promotora y en el comienzo de la transcripción de una molécula de ADN a ARN. Este paso a su vez puede ser dividido en 3 subpasos: la unión de la polimerasa al promotor (el conjunto de ADN y las proteínas unidas a éste se denomina complejo de iniciación), la separación de la doble hélice para formar la burbuja de transcripción (complejo abierto) y la síntesis de aproximadamente 30 ribonucleótidos sin que se disocie el complejo (complejo terciario estable) durante el cual ocurre el capping del extremo 5' del pre-mARN. Este último subpaso es la transición entre el proceso de iniciación y el alargamiento. Una vez comenzada la segunda fase, la ARN polimerasa no sólo continúa alargando, sino también desdobra el ADN frente a ella, disocia la cadena creciente de ARN del templado conforme avanza y además se corrige a sí misma. Finalmente, cuando la polimerasa ha transcrito todo el gen, ésta debe detenerse y liberar el producto de ARN, este es el paso denominado terminación. (52).

La transcripción en eucariontes es llevada a cabo por una enzima que tiene actividad de ARN polimerasa, estos organismos tienen 3 polimerasas nucleares (pol I, pol II y pol III). La transcripción de los genes que codifican para proteínas son transcritos por la **ARN polimerasa II**, de forma general todos estos genes comparten en su promotor 4 tipos de secuencias de ADN: el **Inr** (iniciador) alrededor del inicio de la transcripción, la caja **TATA** aproximadamente 25 nucleótidos

río arriba del Inr, el elemento de reconocimiento a TFIIB ( **BRE**) localizado río arriba de la caja TATA y el **DPE** (elemento río abajo del promotor) que, como su nombre lo indica, se encuentra río abajo del sitio de inicio de la transcripción. Un esquema de la estructura de este tipo de promotores se muestra en la figura 1.

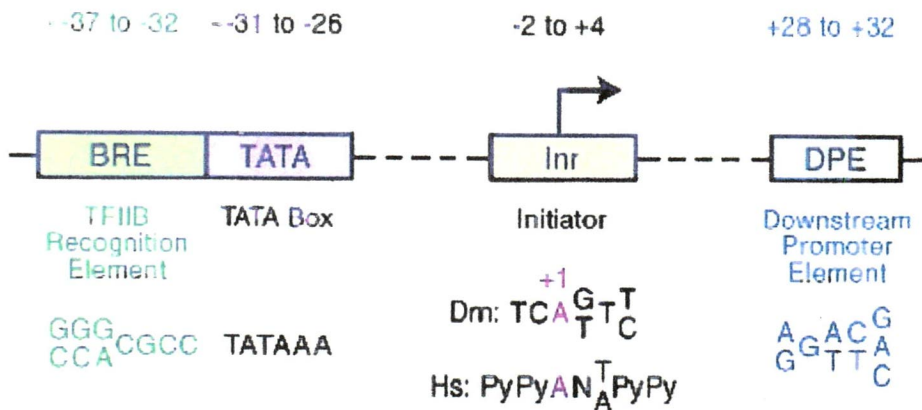


Fig. 1: Elementos de un promotor eucarionte.

Esquema en el que se muestran los diferentes elementos que se encuentran en el promotor mínimo de un gen donde se observa la localización y las diferentes secuencias relevantes para su actividad. Dm (*Drosophila*), Hs(humans). Tomado de Butler, 2002 (8).

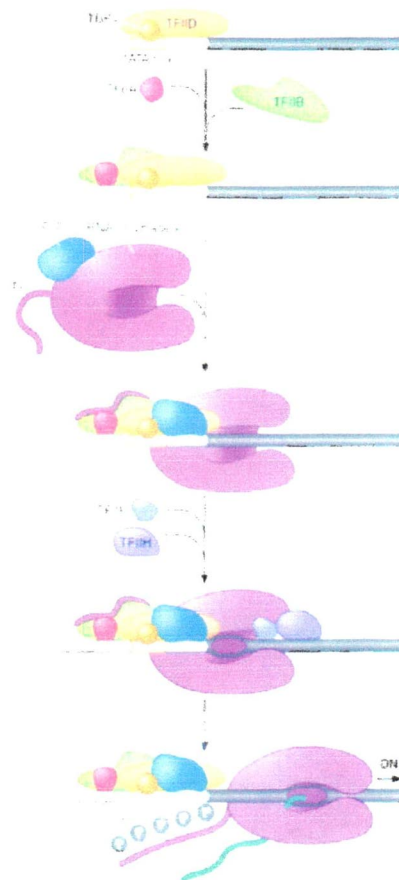
Este tipo de promotores a su vez se organizan en dos tipos, los que tienen la caja TATA y el Inr, los cuales pueden o no tener el elemento BRE y los que contienen el DPE y el Inr pero no caja TATA ni BRE (3). La ARN polimerasa II no es capaz de unirse por sí misma a sus promotores blanco y necesita de otras proteínas para hacerlo, conocidas como factores de transcripción generales (GTF). Éstos se denominan con las siglas TF (del inglés, Transcription Factor) seguidas del numeral romano que corresponde a la polimerasa con la que se asocian y una letra que los identifica (A, B, etc.). Los factores de transcripción generales actúan de forma secuencial en el ensamblaje de la maquinaria basal de la transcripción. Primero **TFIID** se une a la caja TATA o al DPE (dependiendo del tipo de promotor), ésta es una proteína con múltiples subunidades: entre las que se encuentran

a **TBP** (del inglés, TATA Binding Protein) y cerca de 13 factores asociados a la TBP denominados TAFs (del inglés, TBP Associated Factors) (8), esta proteína es la que media el reconocimiento y la unión de TFIID al promotor. Posteriormente **TFIIA** y **TFIIB** son reclutados, este nuevo complejo puede unir a la ARN polimerasa II, la cual a su vez previamente se ha asociado a **TFIIF**. Finalmente el complejo de iniciación se completa con el reclutamiento de **TFIIE** y **TFIIH**. Una vez ensamblado el complejo, tiene lugar la activación de la polimerasa por medio de la fosforilación del CTD mediada por **TFIIH** en una reacción acoplada a ATP y la apertura de la doble hélice, así la cadena templada entra al sitio activo de la polimerasa (complejo abierto), en este momento comienza la transcripción. En la figura 2 se muestra de forma general este proceso. Una vez que la polimerasa ha logrado sintetizar una cadena de aproximadamente 30 nucleótidos pierde contacto con el promotor y la maquinaria basal, y es el cuando comienza el paso de alargamiento.

En este punto múltiples factores, que favorecen al alargamiento, el procesamiento de la cadena de ARN creciente (capping, splicing, poliadenilación) transporte de ARN y la modificación de la cromatina, son reclutados. (22). La terminación de la transcripción esta íntimamente ligada a la poliadenilación de los ARN que se están sintetizando, una vez que la polimerasa llega al final del gen, ésta transcribe secuencias específicas que promueven el ensamblaje de la maquinaria de poliadenilación. En este punto el ARN es poliadenilado, para posteriormente ser transportado fuera del núcleo. Finalmente la polimerasa se disocia del templado. (52).

### **Regulación de la Transcripción**

Dado el alto costo energético y a que es el inicio del flujo de la información contenida en el genoma, la transcripción es uno de los niveles más importantes de la regulación de la expresión génica. Una de las principales características que definen a los organismos eucariontes es el alto grado de compactación de su ADN, lo cual representa un primer nivel de regulación debido a que el acceso al mismo está "restringido" y en consecuencia la maquinaria basal de la transcripción no puede ser ensamblada en los promotores de los genes. Además del acceso al ADN, la adecuada y específica formación del complejo de iniciación juega un papel muy



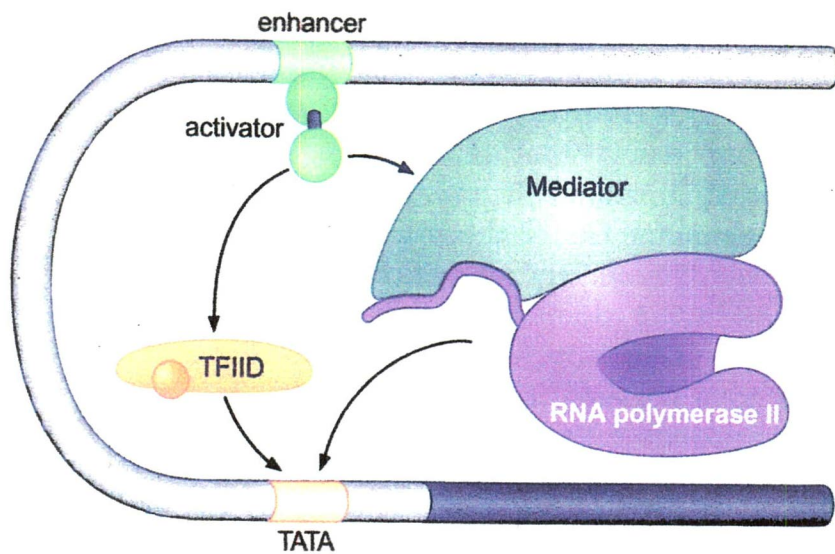
**Fig. 2: Inicio de la Transcripción en Eucariontes**

. Se muestran los diferentes factores generales que se unen durante esta fase. Primero TFIID se une al promotor, después TFIIA y TFIIB son reclutados. Posteriormente la ARN polimerasa II asociada a TFIIF, se une al promotor y finalmente el complejo de iniciación se completa con la asociación de los factores TFIIE y TFIIH. Tomado de Watson, 2004 (52).

importante en la regulación de la expresión génica. Estos mecanismos están mediados por elementos que actúan en *cis* (secuencias en el ADN que pueden influenciar la transcripción de un gen) y los elementos que actúan en *trans* que son proteínas (factores de transcripción) capaces de reconocer y unirse a secuencias específicas en el ADN. Existen varios elementos que actúan en *cis*, unos se localizan muy cerca del promotor (proximales) mientras que otros pueden estar hasta más de 1000 pb lejos río arriba del sitio de inicio de la transcripción (distales). Estas secuencias son reconocidas por los elementos que actúan en *trans* y se unen a ellas. Los elementos en *trans* ejercen su acción ya sea interaccionando directamente con la maquinaria basal o con otras proteínas que a su vez interactúan con dicha maquinaria (mediador), además algunas proteínas reguladoras pueden reclutar a otras capaces de remodelar la cromatina modulando así el acceso al ADN. Todos estos elementos a su vez pueden ser identificados como activadores transcripcionales (aumentan o promueven la transcripción) o represores transcripcionales (disminuyen o inhiben la transcripción). Uno de los mecanismos de activación es el reclutamiento de las proteínas involucradas en el ensamblaje de la maquinaria basal de la transcripción. Como se muestra en la figura 3, un factor de transcripción activador se une a un elemento en *cis* distal (potenciador) y recluta al mediador y a través del mismo a la ARN polimerasa y/o al factor TFIID (52). Este mecanismo tiene como finalidad posicionar en el lugar adecuado los elementos necesarios para ensamblar la maquinaria basal de la transcripción. Es necesario resaltar que en la realidad muchos activadores pueden actuar de forma conjunta y por diversos mecanismos (remodelación de la cromatina, reclutamiento, etc) para promover o reprimir la transcripción.

### **Factores de Transcripción**

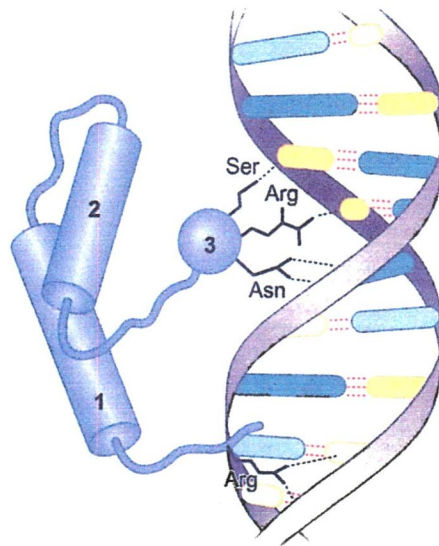
Los factores de transcripción son proteínas modulares que están compuestas de al menos dos dominios, uno es el dominio de unión a ADN (del inglés, DBD) y el otro es un dominio con el cual es capaz de interaccionar con otras proteínas, llamado dominio de regulación de la transcripción (también se le denomina dominio de activación, si se trata de una proteína que promueve la transcripción). En organismos eucariontes se han descrito varios dominios de unión a ADN que se designan de acuerdo a su estructura, uno de ellos es el homeodominio, el cual es



**Fig. 3: Funciones de un Activador Transcripcional.**

Se muestran los diferentes elementos que actúan de forma conjunta en la regulación de la transcripción, donde un activador transcripcional unido a un elemento en *cis* distal es capaz de interactuar con el complejo mediador, el cual a su vez recluta a la ARN polimerasa II. Además el activador puede reclutar a TFIID. Tomado de Watson, 2004 (52).

una estructura de la clase hélice-vuelta-hélice. Éste consiste en 3 alfa hélices como se muestra en la figura 4, una es la hélice de reconocimiento, la cuál es intercalada en el surco mayor del ADN y es la responsable de establecer interacciones del tipo puentes de hidrógeno con una secuencia específica en el ADN, las otras dos son necesarias para estabilizar la estructura (52). Los dominios de activación se clasifican en función del tipo de aminoácidos que los conforman y no de acuerdo a su estructura, esto debido a que pueden presentar diferentes estructuras dependiendo del factor de transcripción de que se trate. Existen de 3 diferentes tipos de dominios de activación: aquellas con regiones ricas en aa ácidos (aspartato y glutamato), en glutamina o en prolina. (3).



**Fig. 4: Dominio de Unión a ADN Hélice-Vuelta-Hélice.**

Se muestran las 3 hélices características del dominio, donde los residuos en la tercer hélice son los responsables del reconocimiento e interacción con secuencias específicas del ADN. Tomado de Watson, 2004 (52).

## 1.2. Estado del arte

### 1.2.1. *Entamoeba histolytica*

El parásito *Entamoeba histolytica* es el protozooario intestinal causante de la amibiasis humana. Esta enfermedad es principalmente entérica, aunque una parte de los individuos infectados presentan síntomas relacionados con la penetración de las amibas al epitelio intestinal, que se conoce como amibiasis invasora (17). En algunas ocasiones y cuando la enfermedad es mas severa, la diseminación del parásito puede hacerse a través el torrente sanguíneo permitiéndole invadir otros tejidos, en particular el hígado, resultando en abscesos hepáticos (41). Si bien este padecimiento es de control relativamente sencillo, no ha sido posible erradicarlo, evidenciando la importancia de dar continuidad a los estudios de la biología de este parásito, sin otro fin que el entendimiento, cada vez más profundo, de los mecanismos que le permiten sobrevivir como son la adhesión, la fagocitosis, la citólisis de células blanco y el enquistamiento.

### 1.2.2. Epidemiología de la amibiasis

*E. histolytica* es responsable de 100,000 muertes al año, lo que la convierte en la segunda causa de muerte provocada por parásitos protozoarios después de la malaria (54). En México, se han reportado a la semanas epidemiológicas 48 del 2007 y 13 del 2008, 539 813 y 114 333 nuevos casos respectivamente(10, 9), lo que la coloca como una enfermedad de alto impacto en la salud pública de nuestro país, dónde las condiciones de higiene y los programas de salud son insuficientes.

### 1.2.3. Ciclo de vida de *Entamoeba histolytica*

El ciclo de vida de *E. histolytica* comprende dos estadios bien definidos, el primero dentro del hospedero (humano), dónde se mantiene en forma de trofozoito, en el lumen del intestino delgado, y es capaz de dividirse para posteriormente invadir el intestino grueso, lugar dónde puede permanecer hasta 6 meses o más, aunque casi todos son eliminados en un periodo de 12 meses (19). Posteriormente debido a ciertas condiciones del micro-ambiente, el trofozoito cambia de forma a quiste, condición en la cual pueden ser expulsados del hospedero y sobrevivir

fuera de éste hasta que otro ser humano los ingiera a través de comida o agua contaminada con las heces de una persona infectada y así comenzar otra ronda de infección. Los mecanismos moleculares que regulan estos eventos no han sido caracterizados en su totalidad, aunque se han identificado algunos genes implicados en la patogénesis de las amibas (17).

#### 1.2.4. Genoma de *Entamoeba histolytica*

El primer borrador del genoma del parásito protozoario *E.histolytica* fue publicado en el 2005 (31), en un esfuerzo conjunto del Instituto Sanger y el Instituto de Investigación Genómica (TIGR) del Reino Unido y Estados Unidos de Norteamérica respectivamente. Actualmente el genoma de este parásito está siendo reanotado y puede accederse a él en el centro de recursos bioinformáticos del Instituto Nacional de Alergias y Enfermedades Infecciosas de Estados Unidos (NIAID) PATHEMA-ENTAMOEBA.

Recientemente una revisión de la estructura y contenido del genoma ha sido publicada (11), en esta revisión se reportan un genoma de 23.7 Mbp y un contenido total de 9938 genes en el genoma de este parásito, debido a la constante curación de la base de datos, el número actual de genes es de 8197. Un resumen de las características moleculares en comparación con otros organismos eucariontes se muestra en la figura 6. Un resultado del análisis del genoma ha sido la identificación del dominio Myb como el más común de los relacionados con la regulación transcripcional (11). Por otro lado es pertinente señalar las similitudes moleculares que comparte *E.histolytica* con el organismo amoeboide *Disctiostellium discoideum* por ejemplo el contenido de GC de 24.1 % y 22.5 % respectivamente. Por otro lado se ha descrito que estos dos organismos de tipo amoeboide comparten alrededor de 1, 510 familias de genes, representando 3,216 genes de *Disctiostellium discoideum* y 3, 833 genes de *E.histolytica*, en este contexto podemos pensar que la información generada respecto a la función de algunos genes en *Disctiostellium discoideum* podría ser una buena aproximación para comenzar el estudio de sus ortólogos en el parásito protozoario *E.histolytica* (49).

El genoma de *E.histolytica* es sin duda una fuente de información muy valiosa que representa un avance significativo en el entendimiento de los mecanismos que

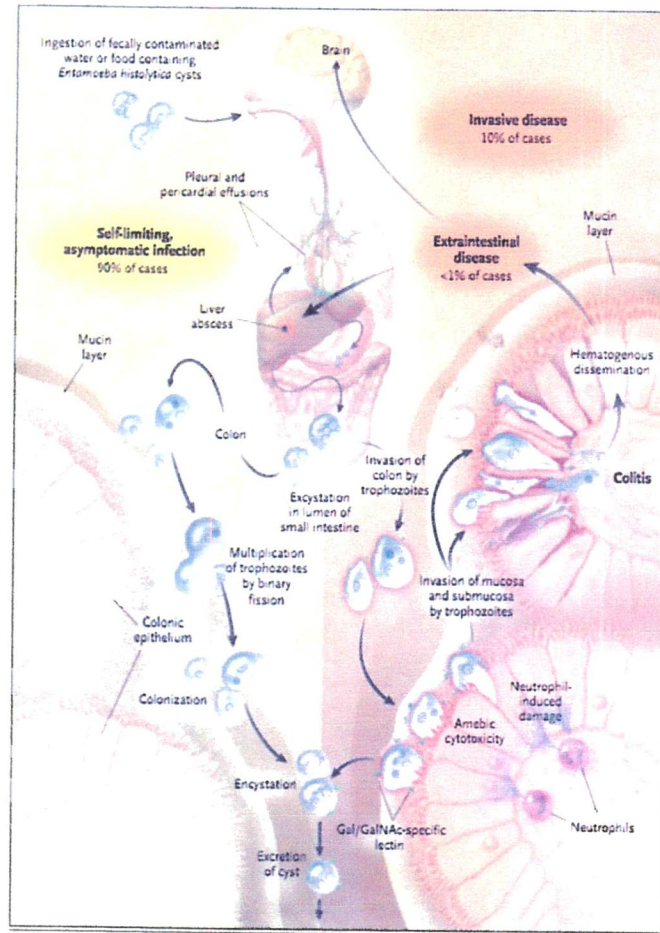


Fig. 5: Ciclo de vida de *Entamoeba histolytica*

. La infección comienza con la ingestión de agua o comida contaminada con quistes de *E. histolytica*. En el lumen del intestino se desenquista y cambia a la forma invasiva de trofozoito. En la mayoría de las infecciones los trofozoitos se agregan en la capa de mucina del intestino y forman nuevos quistes. Lo cual es una enfermedad autolimitante y asintomática. En algunos casos los trofozoitos inician la invasión del colón por medio de la adhesión y lisis de las células del epitelio. Posteriormente el parásito puede diseminarse al peritoneo, al hígado u otros sitios. Tomado de Haque, 2003 (23).

Statistic <sup>a</sup>	<i>Entamoeba histolytica</i>	<i>Plasmodium falciparum</i>	<i>Dictyostelium discoideum</i>	<i>Saccharomyces cerevisiae</i>	<i>Encephalitozoon cuniculi</i>
Genome Size (Mbp)	23.7	22.8	33.8	12.5	2.5
G + C content (%)	24.1	19.4	22.5	38	45.5
Gene number	9938	5268	12,500	5538	1997
Av. gene size (bp)	1167	2534	1756	1428	1077
% coding DNA	49.2	52.6	ND	70.5	ND
Av. protein size (aa)	389	761	518	475	359
Av. intergenic dist. (kb)	0.8	1.7	0.8	0.6	0.1
Gene density (kb per gene)	1.9	4.3	2.5	2.2 kb	1.1
% Genes with introns	25.2	54	69	5	<1
Av. intron size (bp)	102.1	179	146	ND	-
Av. number of introns/gene	1.5	2.6	1.9	1	1

Fig. 6: Características del genoma de *Entamoeba histolytica* y comparación de éstas con las de otros organismos eucariontes  
. Tomado de Clarck, 2007 (11).

le permiten a este organismo parasitar al hospedero, evadir la respuesta inmune, matar y fagocitar células blanco y el enquistamiento entre otros. De esta manera el análisis de la información contenida en el genoma se presenta como uno de los retos actuales más importantes en el estudio de la biología del parásito protozoario *E.histolytica*.

### 1.3. Antecedentes particulares del proyecto

#### 1.3.1. Proteínas MYB como factores de transcripción

Los genes *myb*, fueron primeramente identificados como la contraparte celular normal (*c-myb*) del oncogen *v-myb*, que es el gen transformador del virus de la mieloblastosis en aves, el cual codifica para una proteína nuclear que se une al ADN de forma secuencia específica y se ha propuesto funciona como un trans-activador transcripcional (30, 5). El gen prototipo *c-myb* es expresado predominantemente en células hematopoyéticas inmaduras y codifica para al menos 2 productos traduccionales de 75 y 89 kDa, respectivamente (14). Actualmente se han identificado un gran número de genes pertenecientes a esta familia que se encuentran conservados y distribuidos a lo largo de la espiral evolutiva. Se han encontrado homólogos en vertebrados, parásitos protozoarios, ciliados y particularmente en plantas (12, 39, 51, 45, 50, 46). Los productos protéicos de estos genes intervienen en múltiples procesos celulares, por ejemplo, en plantas las funciones de las proteínas Myb incluyen la regulación del metabolismo secundario, el control de la morfogénesis celular y la regulación de la formación de meristemos y el ciclo celular (29). Es pertinente señalar que si bien la mayoría de las proteínas de tipo Myb descritas actúan como factores de transcripción, existen algunas que no lo son, un ejemplo de éstas son las proteínas de unión a telomeros (TRFs) (38).

Las proteínas de tipo Myb se caracterizan por la presencia de un dominio de unión a ADN estructuralmente muy conservado, denominado dominio MYB, el cual esta conformado por 3 repetidos imperfectos de aproximadamente 52 aa en la región amino terminal de la proteína. Dentro de cada repetido se encuentran una serie de 3 triptofanos espaciados regularmente cada 18 o 19 aa (47). Cada uno de estos repetidos adopta una conformación hélice-vuelta-hélice, además se ha

propuesto que la tercera hélice de los repetidos 2 y 3 son las responsables de interaccionar directamente con el ADN, intercalándose en el surco mayor (27) como se muestra en la figura 7. Las proteínas Myb se clasifican de acuerdo con el número de repetidos que tengan en 3 categorías: las de 3 repetidos (R1R2R3), las de dos (R2R3) y las de un solo repetido (R2/R3). Las proteínas Myb canónicas (que se denominan A, B y C) de mamíferos poseen con 3 repetidos y reconocen la misma secuencia consenso en el ADN: [T/C]AAC[G/T]G denominado Elemento de Reconocimiento a MYB (del inglés, MRE), aunque el reconocimiento varía en afinidad para cada proteína (4). El dominio de activación se encuentra en la región carboxilo terminal adyacente al DBD, sin embargo estas regiones no están conservadas y varían de una proteína a otra, de ahí su multifuncionalidad. Por ejemplo, se ha descrito que *c-myb* es capaz de interaccionar con una gran variedad de factores de transcripción como son: p300/CBP, NF-M, C-EBP-beta, Ets2, PU.1-c/EBP-alfa y AML1 entre otros (38). Con respecto a los genes *myb* R2R3, éstos han sido descritos, en su mayoría en plantas, como parte de una familia multigénica en contraste con los genes canónicos en vertebrados (R1R2R3) con relativamente pocos miembros. Por ejemplo en *Arabidopsis thaliana* se han identificado 125 genes *myb*, en el maíz se han identificado más de 80, 30 en *Petunia hybrida* y aproximadamente 200 en el algodón. La función de la mayoría de éstos genes es desconocida pero el estudio de algunos de ellos ha permitido definir al menos 3 funciones importantes para estos factores de transcripción: control del metabolismo secundario, regulación de la morfogénesis celular y mediador en rutas de transducción de señales como puede ser en respuesta al estrés abiótico y al ataque por patógenos (28). Las proteínas con un solo dominio MYB conservan su capacidad de unión a ADN y han sido descritas en plantas, algas y en el organismo amoeboide *Dictyostelium discoideum* (46, 18, 32, 21). Dentro de las proteínas con un solo repetido, se encuentra una subfamilia que se caracteriza por la presencia de un motivo conservado en la tercer hélice del dominio y que es denominado SHAQKYF, donde la alanina sustituye al tercer triptofano característico de los dominios MYB canónicos. Estos genes han sido identificados en diferentes organismos, por ejemplo: el gen *ZmMybst1* relacionado con el desarrollo del endospermo en *Zea mays* (24). *Mybst1* posiblemente involucrado en la expresión de genes tejido-específicos en *Solanum tuberosum* (2),

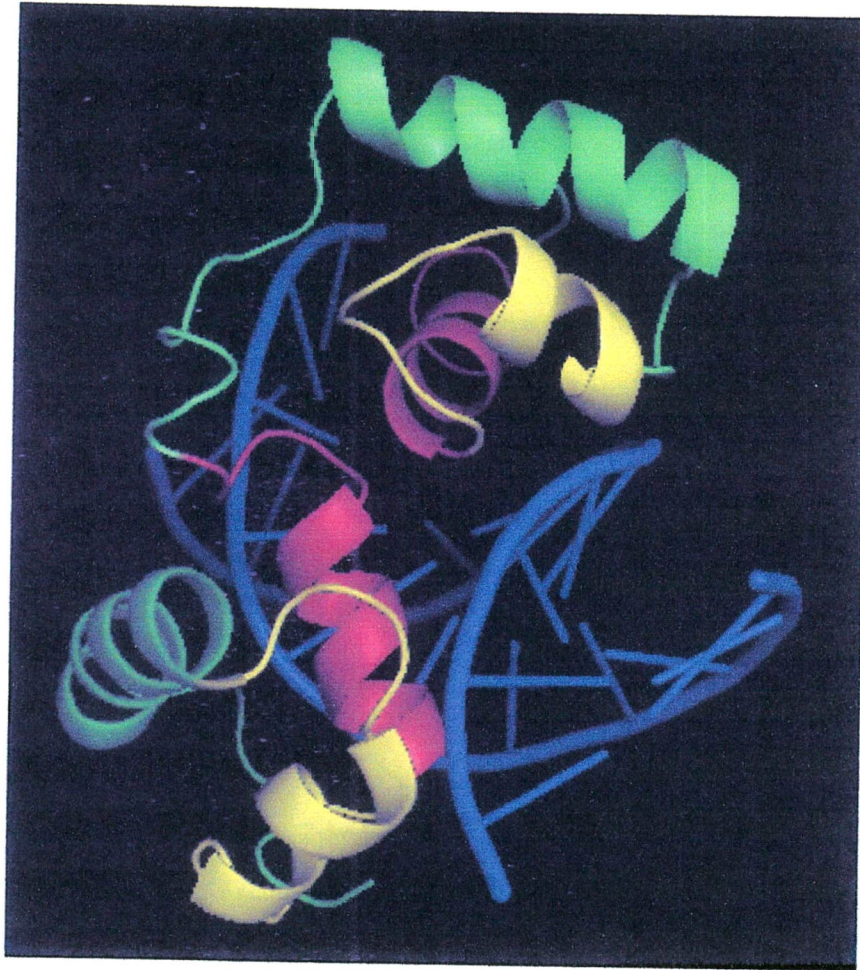
*OsMybS1*, *S2* y *S3* como reguladores de la expresión del gen de la *alfa-amilasa* en respuesta a hormonas y azúcar en *Oryza sativa* (32), *LeMybI* involucrado en la regulación de la expresión de los genes *RBCS1*, *S2* y *S3A* en *Lycopersicon esculentum* (45), *HvMybS3* como activador transcripcional durante el desarrollo del endospermo en *Hordeum vulgare* y *MybE* involucrado en la diferenciación celular en *Dictyostelium discoideum* (18). Además en algunos casos se ha determinado su capacidad de unión secuencia específica como en *Dictyostelium discoideum* donde MybE es capaz de unirse a la secuencia AAC[A/T]GTT (18) y en *Oryza sativa* las proteínas OsMYBS1,S2 y S3 a la secuencia TATCCA (32). Estos elementos varían respecto al descrito para las proteínas canónicas de mamíferos.

### 1.3.2. Transcripción en *E.histolytica*

En el genoma de *E.histolytica* se han identificado 10 homólogos para las subunidades de la ARN polimerasa II, faltando los de las subunidades 4 y 12. Además en lo que se refiere a los factores de transcripción generales se han encontrado homólogos sólo para TFII E, F, H y TFIID, para este último sólo se han identificado 6 de las 14 subunidades que lo conforman (11). Con respecto a la caracterización de los factores de transcripción en este parásito se han reportado: la proteína de unión a URE3 (URE3-BP) (20) involucrado en la regulación de los genes de ferredoxina y la subunidad pesada de la lectina Gal/GalNAc, EhEBP1 y EhEBP2 involucrados en la regulación del gen *hgI5* (48), además la proteína de unión a la caja TATA EhTBP (33) y Ehp53 (35) han sido caracterizados como los ortólogos de TBP y p53 de mamíferos, respectivamente.

En lo que se refiere a los elementos de control en *cis* de la transcripción en este parásito, se ha reportado una estructura tripartita para los promotores de *E.histolytica* que consiste en 3 elementos conservados: la caja TATA, el elemento GAAC y el INR (42). Es pertinente señalar que no todos los promotores de *E.histolytica* tienen esta estructura, por ejemplo la región promotora del gen *eh-rabB* carece de la caja TATA y el INR (44).

Es evidente lo poco que se conoce de la regulación transcripcional en este parásito, además considerando que las células eucarióticas contienen miles de genes que codifican para proteínas y su transcripción debe estar intrincadamente



**Fig. 7: Dominio de Unión a ADN de c-Myb.**

En azul se muestra el templado de ADN. Se muestra la parte correspondiente a los repetidos R2 y R3 del dominio de unión a ADN, responsables de la interacción con los ácidos nucleicos. En verde las hélices 1, en amarillo las segundas hélices (vuelta) y en rojo las terceras hélices, donde se encuentran los residuos que interaccionan directamente con el ADN, tanto para R2 como para R3. Tomado de Ogata, 1994 (37).

coordinada tanto en tiempo como en espacio, la necesidad de conocer más acerca de estos mecanismos y particularmente de los factores de transcripción involucrados en la regulación de genes relevantes para los procesos de patogénesis del parásito, se manifiesta por demás necesaria e interesante.

Por lo anteriormente expuesto, en este trabajo se inició el estudio de los genes *myb* con motivo SHAKQYF, tanto desde el punto de vista bioinformático, como una caracterización inicial de sus funciones biológicas.

## 2. Hipótesis y Objetivos

### 2.1. Hipótesis

Los genes *ehmyb* SHAQKYF predichos en el genoma de *E.histolytica* son proteínas de unión a ADN y su expresión puede ser modulada en respuesta a diferentes estímulos.

### 2.2. Objetivos

#### Objetivo General

Identificar el repertorio completo de las proteínas MYB con el motivo SHAQKYF en el genoma de *E.histolytica*, analizar su expresión y caracterizar la capacidad de unión de la proteína EhMybS3r.

#### Objetivos particulares

1. Obtener la filogenia de las proteínas EhMyb con el motivo SHAQKYF de *E.histolytica*.
2. Predecir la estructura secundaria y terciaria así como los dominios funcionales de las proteínas *Ehmybs1*, *Ehmybs2* y *Ehmybs3* de *E.histolytica*.
3. Analizar los cambios en los niveles de expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* en *E.histolytica* en condiciones de estrés calórico.
4. Analizar los cambios en los niveles de expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* en *E.histolytica* en condiciones de eritrofagocitosis.
5. Determinar la capacidad de unión a ADN de la proteína EhMybS3r con el motivo SHAQKYF de *E.histolytica*.
6. Buscar las posibles secuencias blanco de proteínas EhMyb SHAQKYF en los promotores de genes en el genoma de *E.histolytica*.

### 3. Estrategia experimental

#### 3.1. Búsqueda de secuencias homólogas, alineamientos y reconstrucción filogenética.

Utilizando como sonda la secuencia de una proteína Myb SHAQKYF previamente identificada, con número de acceso EHI.092160 de la base de datos Pathema Entamoeba (<http://pathema.jcvi.org/tigr-scripts/Entamoeba/PathemaHomePage.cgi>) se realizó una búsqueda utilizando el algoritmo Blast-p en el mismo servidor contra el genoma de *E.histolytica*, para identificar posibles genes parálogos. Además, también se realizó una búsqueda utilizando la herramienta psi-Blast en el servidor NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) contra todos los organismos para buscar ortólogos de las proteínas de tipo Myb SHAQKYF, utilizando tantas iteraciones como fueran necesarias hasta no obtener más hits.

Posteriormente, las secuencias obtenidas fueron alineadas utilizando el programa Muscle (15) utilizando los parámetros por default a excepción del número de iteraciones y el valor del Gap Open Score, utilizando 8 y -5 respectivamente.

Una vez obtenidos los alineamientos, éstos fueron cortados con el programa GBlocks 0.91b, seleccionando de esta manera la región más conservada. Los parámetros de corte se establecieron como sigue:

1. Número mínimo de secuencias para una posición conservada: 6
2. Número mínimo de secuencias para una posición flanqueante : 6
3. Número máximo de posiciones contiguas no conservadas: 32000
4. Longitud mínima del bloque: 10
5. Permitir gaps: Todos

Posteriormente se seleccionó el mejor modelo de sustitución de aminoácidos para los alineamientos recortados, utilizando el programa ProtTest (<http://darwin.uvigo.es/software/protest.html>). Con este modelo se construyeron los árboles filogenéticos por el método bayesiano con ayuda del programa Mr.Bayes, finalmente se editó el árbol con FigTree 2.1.

### **3.2. Predicción de estructura y dominios funcionales de las proteínas EhMyb con el motivo SHAQKYF de *E.histolytica*.**

La predicción de estructura secundaria para cada una de las secuencias se realizó con el servidor Jpred (<http://www.compbio.dundee.ac.uk/~www-jpred/>). La predicción de estructura terciaria se realizó por homología utilizando para este fin el servidor Phyre (<http://www.sbg.bio.ic.ac.uk/~phyre/>). Para la visualización de las estructuras se utilizó el programa MacPyMOL 2006.

Con el servidor pfam (<http://pfam.sanger.ac.uk/>) se realizó la predicción de dominios funcionales, además se buscó posibles señales de localización nuclear utilizando el servidor predictNLS (<http://cubic.bioc.columbia.edu/predictNLS/>) para cada una de las secuencias bajo estudio.

### **3.3. Análisis de los niveles de expresión de los genes *ehmyb* SHAQKYF en *E.histolytica*.**

#### **3.3.1. Cultivo *in vitro* de trofozoitos de *E.histolytica***

Los trofozoitos de *E. histolytica* de la cepa HMI:IMSS, amablemente donados por la Dra. Esther Orozco Orozco del Departamento de Infectómica y Patogénesis Molecular del CINVESTAV-IPN, se cultivaron en medio TYI-S-33 completetando con 3% de vitaminas de Diamond (Sigma) y suplementado con 20% de suero bovino (Equitech) penicilina (100 U/ml) y estreptomocina (100  $\mu$  g/ml) en botellas para cultivo y éstas se incubaron a 37° C.

#### **3.3.2. Obtención de ADN genómico de *E.histolytica***

El ADN genómico se obtuvo de trofozoitos de la cepa HMI:IMSS mediante el uso del kit Wizard Genomic ADN Purification (Promega) según las indicaciones del fabricante. Este kit esta basado en una extracción de 4 pasos en los cuales primeramente las células y los núcleos se rompen utilizando una solución de lisis de núcleos, en seguida se realiza una digestión con RNasa para posteriormente

realizar una precipitación de las proteínas con la solución precipitante, finalmente el ADN es precipitado con isopropanol y resuspendido en solución de rehidratación de ADN. El ADN genómico obtenido así se cuantificó por espectrofotometría a 260 nm utilizando la ecuación 3.1 donde FD es el factor de dilución empleado. Se hicieron alícuotas y se almacenó a -20 C. El ADN genómico obtenido así se empleo como control positivo en todas las reacciones de RT-PCR.

$$ADN_{\mu g/ml} = (A_{260nm})(50)(FD) \quad (1)$$

### 3.3.3. Diseño de los oligonucleótidos para PCR y clonación

El diseño de los oligonucleótidos se realizó de acuerdo a la secuencia de los genes completos encontrados en la base de datos de Pathema Entamoeba. Los oligonucleótidos se diseñaron para utilizarlos como cebadores para la amplificación de los genes bajo estudio en las reacciones de PCR y además para su clonación en un vector de expresión. Es por ello que se les adicionaron los sitios de corte para las enzimas *Bam*HI y *Xho*I para los oligos en sentido y antisentido respectivamente. Así mismo fue necesario adicionar 6 bases extras para permitir el corte de las enzimas de restricción en el producto de ADN amplificado (ver Tabla 1). En todos los genes analizados se verificó que no tuvieran sitios de corte interno para estas enzimas de restricción. Finalmente se realizó un blastn contra el genoma de *E.histolytica* para asegurarnos que los oligonucleótidos fueran específicos para cada gen en estudio.

### 3.3.4. RT-PCR en tiempo final en condiciones basales de cultivo

El ARNtotal se obtuvo de trofozoitos de la cepa HMI:IMSS mediante el uso del kit SV Total RNA Extraction (Promega) de acuerdo a las indicaciones del fabricante. Para ello entre  $5 \times 10^3$ - $5 \times 10^6$  trofozoitos obtenidos de cultivos confluentes se cosecharon en tubos cónicos de 50 ml a 2500 rpm por 15 min Se realizó un lavado con amortiguador PBS y se contaron en cámara de Neubauer en microscopio con el objetivo de 40X de acuerdo con la ecuación 3.2, donde CT es el total de células, NCC es el Número de Células Contadas y VT es el Volúmen Total.

Tabla 1: Secuencias de los oligonucleótidos diseñados

Gen	Orientación	Secuencia
<i>Ehmybs1</i>	Sentido	5'- <i>GAGTCGGGATCC</i> ATGTCTAGTTTTGGTGCATAC- 3'
<i>Ehmybs1</i>	Antisentido	5'- <i>GAGTCGCTCGAGCTA</i> ATTGGTATAAATCGGATG-3'
<i>Ehmybs2</i>	Sentido	5' - <i>GAGTCGGGATCC</i> ATGTCCCTTCACTGTTGTCAA-3'
<i>Ehmybs2</i>	Antisentido	5'- <i>GAGTCGCTCGAGTTA</i> ATCAGTATAAATTGGATG-3'
<i>Ehmybs3</i>	Sentido	5' - <i>GAGTCGGGATCC</i> ATGTCAATTTACACCTCATGAT-3'
<i>Ehmybs3</i>	Antisentido	5'- <i>GAGTCGCTCGAGTTA</i> ATAACATAAATCCAAAGT-3'
<i>Peroxisferredoxina</i>	Sentido	5' -TTGTCCAACAGAAATGATTG-3'
<i>Peroxisferredoxina</i>	Antisentido	5'- ATGACAATAAACTGAATCTA-3'
<i>25srARN</i>	Sentido	5' GCGGACGGCTCATTATAA 3'
<i>25srARN</i>	Antisentido	5' ATTGTCGTGGCATCCTAAC 3'

Estos oligonucleótidos se diseñaron tanto para la reacción de RT-PCR así como para clonación en un vector de expresión. En itálicas se muestran las bases agregadas para que la enzima pueda cortar adecuadamente, en negritas se muestra los sitios de corte de las enzimas *Bam*HI y *Xho*I

$$[CT] = (NCC)(10000)(VT) \quad (2)$$

Las pastillas obtenidas se resuspendieron en una solución de lisis celular, posteriormente se adicionó buffer de dilución (Promega), se mezcló por inversión y se incubó a 70 °C por 3 min, para después centrifugarse a 14 000 x g durante 10 min. Al sobrenadante así recuperado se le agregó etanol al 95 % y se transfirió a una columna de separación la cual fue centrifugada a 14 000 x g por 1 min. Posteriormente, se agregó solución de lavado (Promega) y se centrifugó nuevamente a 14 000 x g por 1 min. Para eliminar el ADN contaminante se agregaron 50  $\mu$ l de solución de DNAsa y se incubó durante 15 minutos a temperatura ambiente. La actividad de la DNAsa se detuvo agregando 200  $\mu$ l de solución de paro de DNAsa y se centrifugó a 14,000 x g durante 1 min. Se realizaron dos lavados con 600  $\mu$ l de solución de lavado y se centrifugó a 14,000 x g durante 1 min. Por último, el ARN total se eluyó con 100  $\mu$ l de agua libre de nucleasas y se centrifugó a 14000 x g por 1 minuto. El ARN total así obtenido se cuantificó por espectrofotometría a 260 nm utilizando la ecuación 3.3 donde FD es el factor de dilución empleado. Este fue alicuotado y se almacenó a -70 C hasta su uso. Una alicuota se utilizó para cargar un gel de agarosa al 1 % con bromuro de etidio para verificar la integridad del ARN total tcon base en el patrón de corrimiento electroforético en amortiguador TAE 1X. El gel se visualizó con luz UV con un fotodocumentador Molecular Imager Gel Doc XR (Biorad).

$$ARN \mu g/ml = (A_{260nm})(44)(FD) \quad (3)$$

### Reacción de retrotranscripción

Se utilizó 1  $\mu$ g de ARN total extraído de trofozoitos en condiciones basales de cultivo. Éste se trató con DNAsa de acuerdo a la reacción mostrada en la tabla 2 por 12 min a 37 °C. Una vez finalizado el tratamiento se adicionó EDTA 25mM y se incubó a 70 ° C por 12 min para desactivar a la enzima. La reacción de RT-PCR se realizó con 1  $\mu$ g de RNA total, 0.5  $\mu$ g de oligo dT (generosamente donado por el Dr. Cesar López Camarillo del Posgrado en Ciencias Genómicas), 5 mM de DTT, 1mM de dNTPs, 5mM de MgCl<sub>2</sub>, y 1 U de enzima Improm-II (Promega) tal y

Tabla 2: Mezcla de reacción para el tratamiento de DNasa del ARN.

Reactivo	$\mu\text{L}$
ARN ( $1\mu\text{g}/\mu\text{L}$ )	1
RNAsin	0.5
Dnasa Buffer 10 X	0.5
Dnasa ( $1\text{U}/\mu\text{L}$ )	2
H2O	1

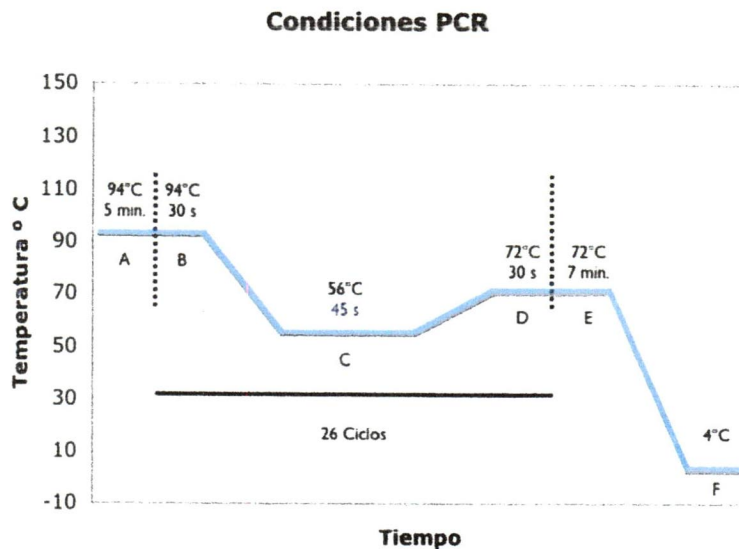
como se muestra en la tabla 3. Las condiciones de reacción fueron las siguientes: 5 minutos a temperatura ambiente (Alineamiento), una h a  $42\text{ }^{\circ}\text{C}$  (Extensión) y 15 minutos a  $70\text{ }^{\circ}\text{C}$  (Inactivación). El cDNA obtenido así se utilizó como templado para la amplificación de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* por PCR o se almacenó a  $-20\text{ }^{\circ}\text{C}$  hasta su uso.

Tabla 3: Mezcla para la reacción de retrotranscripción

Reactivo	Basal( $\mu\text{L}$ )	RT -( $\mu\text{L}$ )
ARN tratado ( $1\mu\text{g}$ total)	7	7
Oligo dT tratado	1	1
DTT 0.1 M	1	1
Buffer RT 5X	4	4
MgCl <sub>2</sub> 25 mM	4	4
dNTP 's 10 mM	2	2
RT Improm II	1	0
H2O	0	1
Volumen total	20	20

### Reacción en Cadena de la Polimerasa

Se utilizó  $1\ \mu\text{l}$  del cDNA obtenido como templado, para cada una de las reacciones de PCR. Para la mezcla de reacción se utilizaron las siguientes concentraciones finales de cada reactivo:  $0.5\ \text{mM}$  de cada oligonucleótido,  $0.2\ \text{mM}$  de dNTPs,  $2.5\ \text{mM}$  de MgCl<sub>2</sub> y  $1\ \text{U}$  de la enzima GoTaq DNA Polymerase (Promega). Para amplificar cada uno de los genes se utilizaron los oligonucleótidos específicos (Tabla 1)



**Fig. 8: Condiciones de amplificación.**

Las regiones marcadas como A y B indican el proceso de desnaturalización, la zona C es la de alineamiento, la D es la zona de polimerización, E la de extensión y F la de mantenimiento.

y como control de carga se usó el gen *25s rARN*. Así mismo se utilizó ADN genómico como templado para control positivo de cada uno de los genes. Las cantidades empleadas en cada protocolo se muestran en la tabla 4.

Las reacciones de PCR se realizaron de acuerdo al programa mostrado en la figura 8. Finalmente los productos de amplificación se corrieron en un gel de agarosa al 1% a 120V. El gel se visualizó con luz UV con un fotodocumentador Molecular Imager Gel Doc XR (Biorad).

### 3.3.5. Cinéticas de amplificación para los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3*

Para asegurarnos que los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* se amplificaran en la fase lineal de la curva de amplificación, se realizó una cinética de ciclado para cada uno de los genes bajo estudio. Para ello usamos el cDNA obtenido en

Tabla 4: Mezclas de reacción para PCR.

Reactivo	Gen( $\mu\text{L}$ )	RT -( $\mu\text{L}$ )	PCR(+)( $\mu\text{L}$ )	PCR(-)( $\mu\text{L}$ )
ADN genómico (100ng/ $\mu\text{L}$ )	0	0	1	0
cDNA (Basal)	1	0	0	0
cDNA (RT-)	0	1	0	0
PF (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5
PR (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5
Buffer Green (5X)	10	10	10	10
MgCl <sub>2</sub> (25 mM)	5	5	5	5
dNTP's (10 mM)	1	1	1	1
Go Taq (5 U/ $\mu\text{L}$ )	0.2	0.2	0.2	0.2
H <sub>2</sub> O	cbp 50	cbp 50	cbp 50	cbp 50
Volumen total	50	50	50	50

Se realizaron las siguientes 4 reacciones para cada uno de los genes en estudio en la cual se utilizó el cDNA de la condición basal y los primers son los específicos para cada gen. Las columnas 3,4 y 5 se refieren a los controles, en este caso utilizando los primers para el gen *25s rARN*.

las condiciones basales de cultivo y se realizó una cinética de 18, 20, 22, 24, 26, 28, 30, 32, 34 y 36 ciclos manteniendo las demás condiciones de la PCR como se describieron anteriormente.

### 3.3.6. RT-PCR en tiempo final en condiciones de estrés calórico.

Para analizar los cambios en la expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* en respuesta a estrés inducido por calor, los trofozoitos de *E. histolytica* de la cepa HMI:IMSS se cultivaron a 42 ° C por 4 hrs, en paralelo se incluyó un control el cual se mantuvo a 37 °C. Posteriormente, se extrajo el ARNtotal y se sintetizó el cDNA como se describió anteriormente. Las reacciones de PCR se realizaron como se muestra en la tabla 5. Finalmente los productos de amplificación se corrieron en un gel de agarosa al 1% a 120 V. El gel se visualizó con luz UV con un fotodocumentador Molecular Imager Gel Doc XR (Biorad).

### 3.3.7. RT-PCR en tiempo final en condiciones de eritrofagocitosis.

#### Obtención de eritrocitos humanos.

Se extrajeron 5 ml de sangre periférica de tipo O Rh -, se agregaron 10 ml de solución Alsever (ácido acético 0.055 %, citrato de sodio 0.8 %, NaCl 0.42 % y Dextrosa 2.05 %) y se centrifugó a 2500 rpm por 5 min Se desecho el sobrenadante y se repitieron 4 lavados con la misma solución centrifugando en las mismas condiciones para eliminar el suero y los leucocitos. Finalmente los eritrocitos se ajustaron a una concentración de  $1 \times 10^8$  células /ml.

#### Eritrofagocitosis

Se obtuvieron cultivos de trofozoitos de *E. histolytica* después de haber sido pasados por hígado de hámster dorados, *Cricetus auratus*, los cuales fueron generosamente donados por el Dr. Alfonso Olivos de la Unidad de Medicina Experimental del Hospital General de México.

Los trofozoitos se cosecharon y se resuspendieron en medio TYI sin suero ajustándolos a una concentración de  $1 \times 10^6$  células/ml. Se adicionaron eritrocitos en relación 1 trofozoito: 100 eritrocitos y se incubaron a 37 ° C por 5, 15 y 30 min. Al tubo correspondiente al tiempo 0 no se le adicionaron eritrocitos, pero

Tabla 5: Protocolo de PCR utilizado para la condición de estrés calórico

Reactivo	Basal( $\mu\text{L}$ )	EC ( $\mu\text{L}$ )	RT(-)( $\mu\text{L}$ )	PCR(+)( $\mu\text{L}$ )	PCR(-)( $\mu\text{L}$ )
ADN genómico (100ng/ $\mu\text{L}$ )	0	0	0	1	0
cDNA (Basal)	1	0	0	0	0
cDNA (EC)	0	1	0	0	0
cDNA (RT-)	0	0	1	0	0
PF (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5	0.5
PR (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5	0.5
Buffer Green (5X)	10	10	10	10	10
MgCl <sub>2</sub> (25 mM)	5	5	5	5	5
dNTP's (10 mM)	1	1	1	1	1
Go Taq (5 U/ $\mu\text{L}$ )	0.2	0.2	0.2	0.2	0.2
H <sub>2</sub> O	cbp 50	cbp 50	cbp 50	cbp 50	cbp 50
Volumen total	50	50	50	50	50

los trofozoitos se mantuvieron en las mismas condiciones que los demás. Una vez terminados los tiempos de interacción se adicionaron 40 ml de agua MilliQ estéril para lisar los eritrocitos no fagocitados durante 10 min Posteriormente se centrifugó a 2500 rpm por 15 min se desecho el sobrenadante y se lavó con 25 ml de PBS pH 7.0 Se centrifugó de nuevo a 2500 rpm por 15 min Finalmente, se extrajo el ARNtotal de acuerdo al protocolo del kit SV Total RNA Extraction (Promega) como se describió anteriormente.

Una vez extraído el ARNtotal éste se cuantificó por espectrofotometría a 260 nm y se verificó la integridad con base en el patrón de corrimiento electroforético en un gel de agarosa al 1 % el cual se corrió a 120 V y se visualizó con luz UV con un fotodocumentador Molecular Imager Gel Doc XR (Biorad).

Se realizaron las reacciones de PCR como se muestra en la tabla 6.

### **3.4. Clonación y Secuenciación de los genes *Ehmybs1* y *Ehmyb21***

Los productos de amplificación se clonaron en el vector TOPO TA 2.1 (Invitrogen). La reacción se llevó a cabo de acuerdo al protocolo del proveedor, en la tabla 7 se muestra los componentes de la reacción. La mezcla se incubó a temperatura ambiente por 5 min posteriormente se puso en hielo y se transformaron las células competentes con 2  $\mu$ l de la reacción de clonación por cada vial con células competentes One Shot *E.coli* DH5  $\alpha$  (Invitrogen, componente del kit) y se mezcló suavemente. Posteriormente se dio un choque térmico por 30 s a 42 °C y 15 min en hielo. Se recuperó a las células en medio SOC a 37 °C por 1 h a 200 rpm. Finalmente se espatularon 50  $\mu$ l de células en cajas petri con medio Luria con 50  $\mu$ g/ml de ampicilina y se incubaron a 37 °C toda la noche. Las colonias obtenidas se picaron y se crecieron en 5 ml de medio Luria Ampicilina líquido a 37 °C toda la noche.

#### **3.4.1. Obtención de plásmidos y restricción enzimática.**

Los cultivos correspondientes a las colonias obtenidas en la clonación se procesaron para la obtención de los plásmidos, primero se cosecharon a 14000 x g

Tabla 6: Protocolo de PCR utilizado para la condición de eritrofagocitosis

Reactivo	0' ( $\mu\text{L}$ )	5' ( $\mu\text{L}$ )	15' ( $\mu\text{L}$ )	30' ( $\mu\text{L}$ )	RT(-)( $\mu\text{L}$ )	PCR(+)( $\mu\text{L}$ )	PCR(-)( $\mu\text{L}$ )
ADN genómico (100ng/ $\mu\text{L}$ )	0	0	0	0	0	1	0
cDNA (Basal)	1	0	0	0	0	0	0
cDNA (Eritro)	0	1	1	1	0	0	0
cDNA (RT-)	0	0	0	0	1	0	0
PF (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5	0.5	0.5	0.5
PR (50 $\mu\text{M}$ )	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Buffer Green (5X)	10	10	10	10	10	10	10
MgCl <sub>2</sub> (25 mM)	5	5	5	5	5	5	5
dNTP 's (10 mM)	1	1	1	1	1	1	1
Go Taq (5 U/ $\mu\text{L}$ )	0.2	0.2	0.2	0.2	0.2	0.2	0.2
H <sub>2</sub> O	cbp 50	cbp 50	cbp 50	cbp 50	cbp 50	cbp 50	cbp 50
Volumen total	50	50	50	50	50	50	50

Tabla 7: Mezcla de reacción para clonación en el vector TOPO TA 2.1

Reactivo	( $\mu$ L)
Producto de PCR	0.5 a 4
Solución salina	1
H2O	cbp 5
Vector TOPO	1
Volumen final	6

durante 2 min y las pastillas se resuspendieron en solución TS, a las pastillas resuspendidas se les agregó solución ELT y se incubaron a temperatura ambiente por 5 min. Posteriormente los tubos se incubaron a 70 °C por 10 min. Una vez finalizada la incubación se realizó una centrifugación a 14 000 x g por 15 min Se recuperó el sobrenadante y se adicionó un volumen igual de PEG para su precipitación y se agitó suavemente. Se incubó por 20 min a temperatura ambiente y se centrifugó a 14,000 x g por 3 min La pastilla de ADN se secó y se resuspendió en agua MilliQ estéril.

El plásmido extraído se restringió con las enzimas *Bam*HI y *Xho*I (New England). Se cargo 1  $\mu$ l del producto de la restricción en un gel de agarosa al 1% para verificar la liberación del inserto correspondiente al tamaño completo de cada uno de los genes. Una vez verificada la liberación del inserto, se seleccionaron colonias y se procesaron para la obtención de ADN plasmídico de alta calidad y pureza para su secuenciación. En breve, las colonias seleccionadas se picaron y se crecieron en medio Luria Ampicilina 100  $\mu$ g/ml . Las células se procesaron para la obtención de plásmido con el kit Wizard plus SV minipreps DNA Purification System (Promega) de acuerdo a las instrucciones del fabricante. Se cosecharon las células por centrifugación y se adicionó solución de resuspensión celular (Promega), posteriormente se lisaron con solución de lisis y se incubó con solución de proteasa alcalina por 5 min a temperatura ambiente. Una vez terminada la incubación se agregó solución de neutralización (Promega) y se centrifugó a 14 000 x g por 10 min. El sobrenadante fue recuperado y se paso por una columna de purificación centrifugando a 16 000 x g por 1 min Se adicionó solución de lavado

Tabla 8: Mezcla de reacción de secuenciación

Reactivo	Cantidad
Big Dye	2 $\mu\text{L}$
Buffer	3 $\mu\text{L}$
Primer S o AS(10 picomolar)	1 $\mu\text{L}$
ADNplásmidico	400 ng
H2O	cbp 20 $\mu\text{L}$

Tabla 9: Reacción de Secuenciación

Temperatura ( $^{\circ}\text{C}$ )	Tiempo (min)	Ciclos
96	120	1
96	10	
50	5	40
60	240	
4	10	1

y se centrifugó a 16 000 x g por 2 min Finalmente se recuperó el ADN plasmídico de la columna con agua libre de nucleasas centrifugando a 16 000 x g por 1 min

### 3.4.2. Reacción de secuenciación de los genes *Ehmybs1* y *Ehmyb21*

Los plásmidos obtenidos se cuantificaron por espectrofotometría a 260/280 nm y se corrieron en un gel de agarosa al 1% teñido con bromuro de etidio. Para estimar la concentración de los plasmidos obtenidos se utilizó el marcador de talla molecular de 1 Kb(Invitrogen). Se cargaron 1  $\mu\text{l}$  de cada plásmido por pozo. Una vez cuantificados los plásmidos se realizó la reacción de secuenciación como se describe en la tabla 8. de acuerdo al kit ABI Prism BigDye v3.1 (Applied Biosystem) en un secuenciador automático 31300 Genetic Analyzer (Applied Biosystem).

La reacción de secuenciación se llevó a cabo en un termociclador Thermoline como se muestra en la tabla 9.

Las reacciones de PCR se purificaron utilizando el kit DyeEx 2.0 Spin Kit (QIAGEN). Para ello se preparó la columna por centrifugación a 750 x g por 3

min. La muestra se adicionó a la matriz de la columna y se centrifugó de nuevo para obtener los fragmentos amplificados libres de nucleótidos no incorporados. Posteriormente se obtuvo una pastilla al evaporar el agua en un concentrador (Eppendorf), a temperatura ambiente por 20 min y se resuspendió para finalmente inyectarla al equipo.

### 3.5. Clonación, expresión y purificación de la proteína EhMybS3 de *E.histolytica*.

Para la clonación y expresión del gen se eligió el vector pGEX-6P-1 (GE Healthcare), amablemente donado por el Dr. Luis Briebe de Langebio Cinvestav-IPN. Este vector utiliza el promotor *tac*, el cual se induce por el análogo de lactosa, isopropyl b-D thiogalactoside (IPTG), el mapa del vector se muestra en la figura 9. Además este vector tiene clonado el gen *lacIq*, el cual codifica para una proteína represora que en ausencia del IPTG previene la expresión de la proteína recombinante, lo cual permite el control de la expresión hasta que el inductor este presente. La proteína recombinante se genera como una proteína de fusión en el extremo amino terminal con la proteína Glutathione S-transferasa (GST) de *Schistosoma japonicum*, la cual tiene un peso molecular de 26 kDa. La obtención de la proteína recombinante fusionada con GST tiene como ventaja favorecer la solubilidad y el correcto plegado de proteínas heterólogas en *E. coli*. Las proteínas de fusión se purifican por medio de cromatografía de afinidad utilizando sepharosa-glutathion, posteriormente las impurezas se lavan y las proteínas de fusión se eluyen en condiciones no-desnaturalizantes usando glutatión reducido preservando la antigenicidad de la proteína y la función.

#### 3.5.1. Obtención del inserto para clonación del gen *Ehmybs3*

El producto de PCR del gen de interés se amplificó a partir de cDNA obtenido de condiciones basales de cultivo de *E. histolytica* y se purificó por columna utilizando el kit Wizard SV Gel Extraction (Promega). Para crear los sitios de clonación en el gen, se procedió a cortar con las enzimas correspondientes *Bam*HI y *Xho*I (New England Biolabs). La reacción se incubó a 37 °C por 2 h Una vez

**pGEX-6P-1 (27-4597-01)**  
 PreScission™ Protease  
 Leu Glu Val Leu Phe Gln Gly Pro Leu Gly Ser Pro Glu Phe Pro Gly Arg Leu Glu Arg Pro His  
 CTG GAA GTT CTG TTC CAG GGG CCC CTG GGA TCC CCG GAA TTC CCG GGT CGA CTC GAG CCG CCG CAT  
 BamHI EcoRI SmaI SalI XhoI NotI

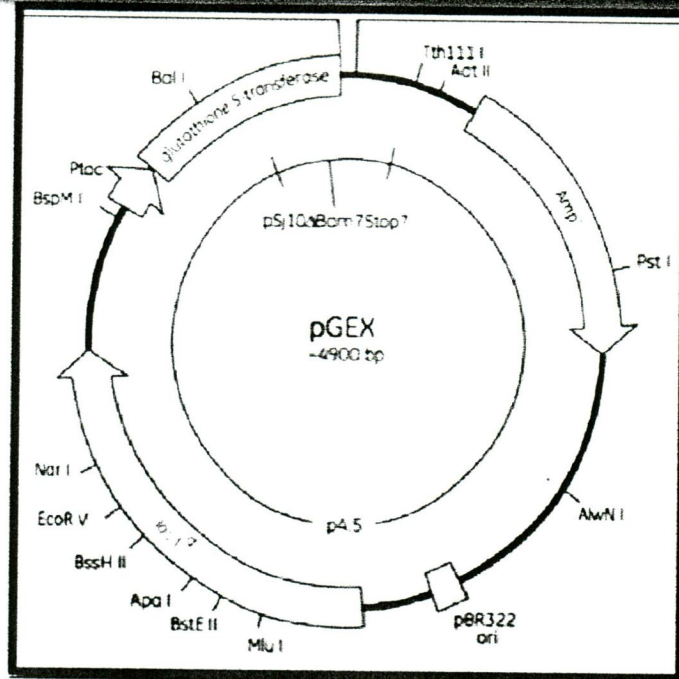


Fig. 9: Mapa del vector de expresión bacteriano pGEX-6P-1. En la figura se muestra el mapa del vector de expresión pGEX-6P-1, y su polilinker con los diferentes sitios de clonación.

terminado el tiempo de incubación los productos de la restricción se corrieron en un gel de agarosa al 1 % y se purificaron los fragmentos con el Kit Wizard SV Gel Extraction (Promega) el DNA fue eluido en 50  $\mu$ l de agua desionizada estéril.

### 3.5.2. Obtención del vector pGEX-6P-1

Para la obtención del vector pGex-6P-1 (GE Healthcare Life Science) éste fue transformado en células bacterianas DH5  $\alpha$  en cajas Petri con medio agar Luria Broth suplementado con ampicilina a una concentración de 100  $\mu$ g/ml. Del cultivo obtenido se seleccionó una colonia y se creció en 50 ml de medio Luria Broth suplementado con ampicilina (100  $\mu$ g/ml) por 16 h El cultivo bacteriano se centrifugó en una microcentrífuga Allegra 25R Refrigerated Benchtop Centrifuge (Beckman Coulter) y se procedió a la extracción del ADN plasmídico utilizando el Kit Midiprep (Qiagen) de acuerdo a las instrucciones del proveedor. La integridad del ADN se verificó mediante electroforesis. Se procedió a realizar una reacción de restricción utilizando las enzimas *Bam*HI y *Xho*I (New England Biolabs) como se describió anteriormente. El producto de corte se visualizó en un gel de agarosa al 1 % teñido con bromuro de etidio. Una vez que se verificó que las dos enzimas cortaron el vector, se procedió a preparar un gel preparativo de agarosa, sometiendo a electroforesis el volumen restante del producto de corte. Una banda de aproximadamente 4.9 kilobases (kb) que correspondió al vector fue cortada con una navaja y eluida del gel utilizando el Kit Wizard SV Gel Extraction (Promega) de acuerdo a las instrucciones del proveedor.

### 3.5.3. Ligación del gen *Ehmybs3* en el plásmido pGEX-6P-1

El vector y los insertos purificados se ligaron utilizando la enzima T4 ligasa (Fermentas) con el kit Rapid Ligation (Fermentas). Como control negativo de la ligación se preparó una reacción con el vector sin colocar el inserto y con la T4 ligasa. Si la reacción de restricción fue eficiente, ésta no deberá de ligarse dado que los sitios de restricción *Bam*HI y *Xho*I no son compatibles.

Posteriormente la reacción de ligación se utilizó para transformar células competentes y éstas se plaquearon en cajas de cultivo petri con Medio Luria ampicilina y se incubaron a 37 °C por toda la noche. Las colonias obtenidas se picaron y se

crecieron en 5 ml de medio Luria ampicilina líquido y se incubaron a 37 ° C toda la noche. Posteriormente los cultivos se cosecharon y se realizó la extracción del ADN plásmidico con Lisozima y se restringió utilizando las enzimas *Bam*HI y *Xho*I. Los productos se corrieron en un gel de agarosa al 1% y se verificó la liberación del inserto en la talla esperada (546 pb).

Se seleccionaron 4 colonias que liberaron inserto, se picaron de nuevo y se crecieron para las purificación del plásmido con el kit Wizard plus SV minipreps DNA Purification System (Promega) y para su secuenciación como se describió anteriormente para el vector TOPO TA 2.1.

#### 3.5.4. Expresión de la proteína recombinante EhMybS3

El plásmido purificado y secuenciado se utilizó para transformar células competentes *E.coli* de la cepa derivada de BL21 denominada Rossetta II II (generosamente donada por el Dr. Luis Brieba de Langebio Cinvestav), la cuál esta diseñada para la expresión de proteínas eucarioticas, ya que contiene 7 plásmidos extras con resistencia a cloranfenicol que codifican para tARNs que regularmente no se encuentran en *E.coli*. Las células transformadas se plaquearon en cajas de cultivo con medio Luria-Ampicilina (100 µg/ml)-cloranfenicol (34 µg/ml) (LAC) y se incubaron a 37 ° C toda la noche y con una agitación de 225 rpm. De las colonias obtenidas se picó una y se creció en 5 ml de medio líquido LAC a 37°C y con una agitación de 225 r.p.m toda la noche, éste cultivo celular se utilizó como inóculo para sembrar 500 ml de medio LAC, éste se creció hasta alcanzar una densidad óptica a 600 nm entre 0.4 y 0.6. Una vez que se alcanzó la DO se adicionó IPTG (Invitrogen) tal que la concentración final fuera de 0.5mM y se indujo por 3 h a 37°C y con una agitación de 225 rpm. En todos los cultivos se utilizó una concentración de 100 µg/ml de ampicilina y 34 µg/ml de cloranfenicol éste procedimiento también se realizó en paralelo para el plásmido vacío para expresar a la proteína GST que se utilizará como control en los ensayos de retardamiento. Se tomaron muestras de ambos cultivos antes de inducir e inducido. Las muestras tomadas se corrieron en un gel de poliacrilamida al 15% en condiciones desnaturizantes para verificar si la proteína se había inducido.

### 3.5.5. Purificación de la proteína EhMybS3r

Los cultivos ya inducidos se cosecharon en tubos cónicos de 50 ml por centrifugación a 2500 rpm por 15 min. Las pastillas se resuspendieron en amortiguador de unión 1X pH 7.3 y se procesaron en Prensa French (marca Thermo Electron) a 1500 psi de presión. El lisado bacteriano se centrifugó a 8000 rpm por 45 min. El sobrenadante se inyectó en el cromatógrafo FPLC Äktaprime plus y la proteína se purificó utilizando la columna GSTrap HP (GE Healthcare) y se recuperaron 10 fracciones.

Las fracciones se corrieron en un gel de poliacrilamida al 15% en condiciones desnaturalizantes para verificar la fracción enriquecida con la proteína de interés. Esta fracción se dializó en amortiguador PBS 1X pH 7.0 con EDTA 5 mM en una membrana de diálisis con un tamaño de poro tal que no permitiera el paso de proteínas mayores a 10 kDa durante toda la noche a 4°C. Finalmente la proteína se cuantificó por espectrofotometría a 280 nm para ello se determinó el coeficiente de extinción molar utilizando la herramienta de software para el cálculo de este parámetro de la compañía EnCor Biotechnology (<http://encorbio.com/protocols/Prot-MW-Abs.htm>). Para determinar la concentración de la proteína se utilizó la ecuación 3.4 donde CEM es el coeficiente de extinción molar y PM es el peso molecular. La proteína EhMybS3r se almacenó en alíquotas de 50 µl a -70 °C hasta su uso.

$$[mg/ml] = (A_{280nm}) / (CEM / PM) \quad (4)$$

### 3.5.6. Ensayo de inmunodetección de la proteína EhMybS3r

Con la finalidad de verificar la identidad de la proteína expresada y purificada se realizaron ensayos de inmunodetección utilizando un anticuerpo comercial anti-GST (GE Healthcare). Para ello las diferentes fracciones durante la cromatografía y los controles de inducción y no inducción se cargaron en un gel de poliacrilamida al 15% en condiciones desnaturalizantes y se corrieron a 120 V. El gel se transfirió a una membrana de nitrocelulosa de 0.45 µm utilizando buffer de transferencia al 20% de metanol en una cámara semihúmeda durante 30 min a 300 mAmp. La membrana se tiñó con rojo de Ponceau para verificar presencia e integridad de las

proteínas transferidas. Posteriormente la membrana se bloqueó con TBS-Tween 0.05 % al 5% de leche durante toda la noche. Una vez terminado el bloqueo la membrana se incubó con el anticuerpo primario anti-GST en dilución 1:10 000 durante 2 h a temperatura ambiente. Para eliminar el anticuerpo primario que no se unió específicamente se realizaron 3 lavados de 10 min con TBS-Tween 0.05 %. Finalmente la membrana se incubó con el anticuerpo secundario anti-chivo peroxidado (Jackson ImmunoResearch) en dilución 1: 35 000 durante 1 h y se realizaron 3 lavados con TBS-Tween al 0.05 %. La membrana se reveló por quimioluminiscencia utilizando el sustrato Inmobilon Western (MilliPore).

### **3.6. Análisis bioquímico mediante ensayos de retardamiento de la proteína EhMybS3r de *E. histolytica*.**

En este trabajo se utilizó un oligonucleótido que contenía la secuencia consenso del elemento de reconocimiento a Myb (ERM) del promotor *mim-1* (por sus siglas en inglés myb induced myeloid protein-1) de pollo. Este oligonucleótido se utiliza como secuencia consenso para caracterizar a factores tipo Myb en otros organismos como por ejemplo en el parásito *Plasmodium falciparum* (7). También se incluyó un oligonucleótido el cual contiene una versión mutada del ERM, llamado *mim-1M*. Las secuencias de los oligonucleótidos utilizados se muestran en la tabla 10. Los oligonucleótidos se sintetizaron en la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología de la UNAM. La concentración se determinó por espectrofotometría a 260 nm en un NanoDrop-1000 UV-Vis Spectrophotometer (ThermoScientific) en el Departamento de Genética del Cinvestav Unidad Irapuato.

#### **3.6.1. Hibridación de los oligonucleótidos**

Todos los oligonucleótidos se sintetizaron como cadena sencilla y posteriormente fueron utilizados en forma de doble cadena, para lo cual fueron hibridados bajo las siguientes condiciones: 53  $\mu$ M de cada uno de los oligonucleótidos de cadena sencilla se diluyeron en amortiguador Tm (TrisHCl 100 mM pH 8.5 y MgCl<sub>2</sub> 50 mM) en un volumen de 20  $\mu$ L. Estos se calentaron a 85 °C durante 5 minutos

y posteriormente fueron enfriados muy lentamente hasta alcanzar la temperatura ambiente. Los oligonucleótidos hibridados se almacenaron a -20 C hasta su uso.

### 3.6.2. Marcaje de los oligonucleótidos

Los oligonucleótidos empleados fueron diseñados con extremos cohesivos *Xba*I (5'-CTAG-3'), los cuales se rellenaron para marcaje terminal (5' y 3'). El marcaje de los oligonucleótidos de doble cadena se realizó empleando el fragmento Klenow de la ADN polimerasa I de *E. coli*. Se marcaron los oligonucleótidos mim-1 y mim-1M ambos a una concentración de 5.31  $\mu$ M, posteriormente se adicionó 1  $\mu$ l de cada nucleótido (10mM de los nucleótidos dGTP, dCTP y dTTP) y 10  $\mu$ Ci de dATP[ $\alpha$ P32], así mismo se adicionó 1  $\mu$ l de amortiguador Tm y 1U de la enzima Klenow (Amersham Biosciences) en un volumen final de 10  $\mu$ L. Se dejó incubar 30 minutos a temperatura ambiente. Para eliminar los nucleótidos no incorporados se utilizó el kit QIAquick Nucleotide Removal (Quiagen) según las indicaciones del fabricante. La reacción se diluyó en 100  $\mu$ l de solución PN, ésta se transfirió a una columna la cual fue centrifugada a 14 000 X g por 1 min Posteriormente se hicieron dos lavados con 500  $\mu$ l de solución PE centrifugándose cada vez a 14 000 x g por 1 min Finalmente, las sondas marcadas se eluyeron utilizando 150  $\mu$ l de solución EB de manera que quedaran a una concentración final de 0.0531  $\mu$ M (1 ng/ $\mu$ L).

### 3.6.3. Ensayos de retardamiento en gel (EMSA)

En estos experimentos se ponen de manifiesto las interacciones ADN-proteína ya que al haber reconocimiento de una secuencia dentro del ADN marcado por la proteína recombinante, la movilidad electroforética de la sonda será menor, observándose un retardo con respecto a la sonda libre.

Concentraciones de 20 $\mu$ M y 40 $\mu$ M de la proteína recombinante EhMybS3 se preincubarón por 10 min en hielo con amortiguador de unión (12 mM HEPES pH 7.9, 60 mM KCl, 1 mM EDTA pH 8, 4 mM Tris pH 7.9, 1 mM DTT , 5 mM MgCl<sub>2</sub> y Glycerol 10 %), solución de Sp-Mg (4 mM Spermidina y 4 mM MgCl<sub>2</sub>) y con o sin 100 ng de poly[d(I-C)] (Amersham Biosciences).

Tabla 10: Sondas utilizadas en los ensayos de EMSA

Sonda	Secuencia
mim-1	5' TCTAGAACACATTATA <b>ACGGTTTTTT</b> TAGC 3'
mim-1 mutado	5' TCTAGAACACATTAG <b>CCAGGGTTTTTT</b> TAGC 3'

En negritas se muestran las bases relevantes para la unión del factor de transcripción c-Myb y subrayadas las bases que se modificaron para generar la secuencia mutada.

Al término de los 10 min se agregó 0.0531  $\mu$ M de los oligonucleótidos de doble cadena marcados radioactivamente y se incubaron en hielo durante 10 min adicionales. Las muestras se depositaron en un gel nativo no desnaturante de poliacrilamida al 6 % y de baja fuerza iónica (0.5X TBE). La electroforesis se llevó a cabo a 150 V. El gel fue secado al vacío y la imagen fue revelada utilizando un detector Fosforimager (Biorad).

### 3.7. Búsqueda de las posibles secuencias blanco de proteínas EhMyb con el motivo SHAQKYF en el genoma de *E.histolytica*.

La información de secuencia y anotación del genoma de *E.histolytica* fue obtenida del servidor Pathema-Entamoeba. A partir del archivo con las anotaciones del genoma se obtuvieron 4 datos de cada gen predicho dentro del genoma: el nombre del gen, el número de acceso de la base de datos Pathema, la molécula de ADN en la que se encuentra el gen, y las posiciones de inicio y final dentro del contig. Con esta información se obtuvo la secuencia de cada uno de los genes de *E.histolytica*.

Una vez obtenidas las secuencias codificantes, se verificó que cada una correspondiera a un ORF, considerando que tuviera un ATG en el inicio, que fuera divisible entre 3 y que no tuviera codones de paro. Para todos aquellos genes que cumplieran con los criterios definidos se obtuvo la secuencia promotora en el contig correspondiente, tomando 1000 pb río arriba y 200 pb río abajo del ATG. Posteriormente, utilizando expresiones regulares, se localizaron las secuencias blanco consenso ya reportadas para las proteínas Myb canónicas y las proteínas Myb SHAQKYF. Además, se buscaron las expresiones regulares correspondientes a los

consensos de la caja TATA y el elemento GAAC ya reportados como sitios reguladores presentes en algunos promotores de genes de *E.histolytica*. Finalmente los genes que contenían en su promotor los posibles sitios de unión a proteínas Myb tanto canónicas como de tipo SHAQKYF se organizaron de acuerdo a su función putativa.

El manejo y obtención de la información a partir de los archivos obtenidos de la base de datos Pathema-Entamoeba se llevo a cabo mediante programas en el lenguaje de programación Python.

## 4. Resultados

### 4.1. Reconstrucción filogenética

#### 4.1.1. Búsqueda de las secuencias de las proteínas EhMyb SHAQKYF de *E.histolytica*.

Con el fin de determinar si existían más proteínas de la familia Myb con el motivo SHAQKYF, adicionales a las que ya habían sido previamente identificadas, en el genoma de *E.histolytica*, se realizó una búsqueda de blastp utilizando como sonda la secuencia de la proteína EhMybS1 (EHL\_092160) contra la base de datos de Pathema-Entamoeba. Nueve proteínas fueron identificadas las cuales se muestran en la tabla 11, es pertinente señalar que de todas las proteínas encontradas, 7 ya estaban anotadas como posible proteínas de unión a ADN de tipo Myb SHAQKYF y 2 de ellas (EhMybS7 y EhMybS8) aún no han sido anotadas y se presentan como hipotéticas. Éstas dos proteínas presentan sustituciones en el motivo SHAQKYF donde la serina es sustituida por una treonina (ver figura 10), además en el segundo tritofano característico de las proteínas Myb, también se encuentran sustituidos por lisina y alanina. Es interesante notar que la proteína EhMybS9 a pesar de tener sustituciones similares está anotada como una proteína Myb SHAQKYF.

Éstas proteínas tienen un tamaño pequeño que varía desde 132 a 189 aa de longitud con pesos moleculares variables desde 15.3 a 21.8 kDa. Lo cual contrasta con la proteína canónica c-Myb de humano la cual tiene una longitud de 640 aa y un peso molecular de 72.4 kDa o con la de *Dictiostellium discoideum* con 817 aa y 93 kDa respectivamente.

Tabla 11: Proteínas EhMyb SHAQKYF de *E.histolytica*

Nombre	Número de acceso	Pathema	Longitud (aa)	Descripción*
EhMybS1	XP_654568	EHI.092160	189	Familia Myb, SHAQKYF
EhMybS2	XP_649754	EHI.135150	165	Familia Myb, SHAQKYF
EhMybS3	XP_655261	EHI.136420	177	Familia Myb, SHAQKYF
EhMybS4	XP_648179	EHI.197980	188	Familia Myb, SHAQKYF
EhMybS5	XP_654390	EHI.008130	167	Familia Myb, SHAQKYF
EhMybS6	XP_655784	EHI.155580	165	Familia Myb, SHAQKYF
EhMybS7	XP_651291	EHI.051440	133	Proteína hipotética
EhMybS8	XP_656309	EHI.013340	132	Proteína hipotética
EhMybS9	XP_656862	EHI.038640	144	Familia Myb, SHAQKYF

Se muestran todas las proteínas del tipo EhMyb SHAQKYF encontradas en el genoma de *E.histolytica*, así como sus correspondientes números de acceso de la base de datos NCBI y Pathema. También se muestra la longitud en aminoácidos (aa). La asignación del nombre se hizo con base en: Eh (*E.histolytica*), Myb (proteína de tipo myb), S (SHAQKYF) y el número consecutivo para identificación. \* de acuerdo a la base de datos Pathema Entamoeba

#### 4.1.2. Alineamiento de las secuencias de las proteínas EhMyb SHAQKYF de *E.histolytica*.

Las secuencias tipo Myb SHAQYF encontradas en el genoma de *E.histolytica* fueron traducidas y alineadas con el programa Muscle. Posteriormente, utilizando la herramienta de software GBlocks se extrajo el segmento más conservado del alineamiento el cual se muestra en la figura 10. En éste se puede observar que los triptofanos característicos del repetido R3 de c-Myb están conservados o sustituidos por los aminoácidos tirosina, alanina, serina, lisina, cisteína o treonina, además se muestra el motivo característico de este tipo de proteínas denominado SHAQKYF. Como se puede observar en el alineamiento, algunas de las proteínas presentan variaciones del motivo como son la sustitución de la serina por treonina como es el caso de las proteínas EhMybS7, EhMybS8 y EhMybS9.

#### 4.1.3. Selección del mejor método de sustitución.

Para poder construir el árbol filogenético fue necesario seleccionar el modelo que explicara mejor los procesos de sustitución observados en el alineamiento. Para ello se utilizó el servidor ProtTest que determina el mejor modelo de sustitución de acuerdo con el valor del criterio de información de Akaike (AIC por sus siglas en inglés, Akaike's Information Criterion). El mejor modelo de acuerdo a AIC fue el modelo general reversible, con una distribución gamma para las frecuencias entre sitios (RtREV+G) de acuerdo a AIC.

#### 4.1.4. Reconstrucción del árbol filogenético

Finalmente, utilizando el alineamiento y el modelo de sustitución RtREV+G se reconstruyó un árbol bayesiano utilizando Mr. Bayes. Este árbol se editó utilizando la herramienta de software Fig Tree. El árbol obtenido se muestra en la figura 11. En este árbol se muestra que las proteínas conforman 3 grupos definidos, de esta manera tenemos a las proteínas EhMybS7, EhMybS8 y EhMybS9 en el clado azul, que conforman un grupo de proteínas que tienen como característica el motivo THAQKYF, por otro lado las proteínas EhMybS5 y EhMybS6 el clado rojo, se caracterizan por la sustitución del segundo triptofano por cisteína o

```

c-myb_WWNHWNSTM  WTDEEDRIIYQAHKRIGNRWAE-----IAKLLPGRTDNAIKNHWNSTMRR
EhMybS1_WWSHAQKYF WTDEEHALFLEGLSLYHRDWKR-----IEQHVKTKTVVQIRSHAQKYFLK
EhMybS2_WWSHAQKYF WTDEEHALFVEGLSLYHKDWKR-----IEGHVKTKTVVQIRSHAQKYFLK
EhMybS3_WWSHAQKYF WTDAEHAKEFVEGLALFHKDWKK-----IKEYIGTKTVVQIRSHAQKYFLK
EhMybS4_WYSHAQKFF WTDEERSKFEAEAFNKYGRDYAK-----ISAFIGSKTIYQVRSHAQKFFKK
EhMybS5_WCSHAQKYF WTKEECLSEKIKFLNINCGQALDFRLISMVVKTRTPQVRSHAQKYFLK
EhMybS6_WTSHAQKYF WTDVECIAEKKAIKYFGYTTARGINTFLISAFVGTTRTPQVRSHAQKYFMK
EhMybS7_WKTHAQKYF WHDREHIREFLVCIKFKKATKCSGLPVKQIAMYIYTRNGIQVRTHAQKYFKN
EhMybS8_WATHAQKYF WSNDEHTRFMVCLCFINATA CKGLPVLKISKYVKTRTSVQVRTHAQKYFSD
EhMybS9_WSTHAQKFF WSDEEHTKEIKAITWLGCTSTRRLPVKLI SKFVGTTRTPVQVRTHAQKFFDA
consensus          * . . . . * . . . . * . . . . * . . . . * . . . . * . . . . *

```

Fig. 10: Segmento más conservado en el alineamiento de las proteínas EhMyb encontradas en el genoma de *E.histolytica*.

Se muestran los triptofanos o sus sustituciones en relación al dominio R3 de la proteína de tipo Myb canónica c-Myb de humano.

treonina y mantienen el motivo SHAQKYF intacto. Por último en el clado verde las proteínas EhMybS1, EhMybS2, EhMybS3 y EhMybS4 conservan los dos primeros triptofanos y el motivo SHAQKYF, es pertinente señalar que la proteína EhMybS4 a pesar de formar parte de este grupo presentan una sustitución del segundo triptofano por una tirosina y una fenilalanina por una tirosina en el motivo SHAQKYF. Con base en este análisis decidimos continuar el estudio con aquellas proteínas que conservaran en mayor grado la estructura primaria con respecto a otras proteínas de tipo Myb SHAQKYF ya descritas en la literatura, por lo que se seleccionaron las proteínas EhMybS1, EhMybS2 y EhMybS3.

## **4.2. Reconstrucción filogenética de las proteínas EhMyb SHAQKYF de *E.histolytica* en relación con las de otros organismos.**

### **4.2.1. Búsqueda de las secuencias de proteínas Myb SHAQKYF de otros organismos.**

Después de haber encontrado nueve proteínas de tipo myb que presentaban alguna variación del motivo SHAQKYF en *E.histolytica*, nos interesaba saber que tan distribuidos se encuentran a lo largo de la espiral evolutiva. Para ello se llevó a cabo una búsqueda con el algoritmo PSI-BLAST en la base de datos del NCBI, utilizando como sonda la secuencia de la proteína EhMybS3. Esta búsqueda arrojó proteínas homólogas que contenían el motivo SHAQKYF en otros organismos incluyendo algas, plantas y al organismo amoeboide *Disctiostellium discoideum*. Adicionalmente por medio de esta búsqueda encontramos otro grupo de proteínas similares que contenían dos triptofanos y la variante SYARQYF que se encuentran en mamíferos. Todas las secuencias obtenidas en la búsqueda fueron alineadas con las proteínas EhMybS1, EhMybS2 y EhMybS3, utilizando el programa Muscle, y el alineamiento resultante fue cortado con la herramienta de software Gblocks, este alineamiento se muestra en la figura 12, en el se muestran los dos primeros triptofanos conservados y el motivo SHAQKYF o sus variantes, en la mayoría de ellos el tercer triptofano se sustituye por alanina. Además se pueden

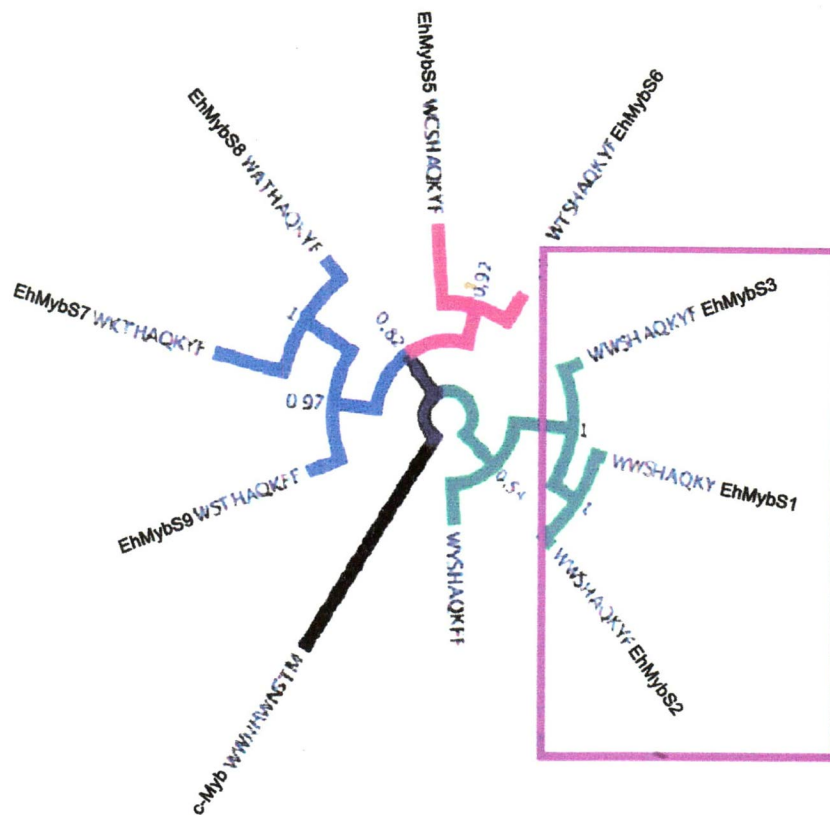


Fig. 11: Árbol filogenético de las proteínas EhMyb SHAQKYF de *E. histolytica*.

En azul se muestran las proteínas EhMyb con la variación THAQK[Y/F]F. En rojo dos proteínas con sustituciones en el segundo triptofano. En verde 3 proteínas que conservan los dos primeros triptofanos y el motivo SHAQKYF sin variaciones y una que presenta sustitución en el segundo triptofano. En negro se muestra la proteína canónica de tipo Myb de humano (c-Myb). En el recuadro se señalan las 3 proteínas que mantienen tanto los dos primeros triptofanos como el motivo SHAQKYF intactos. En cada nodo se muestran los valores de soporte (probabilidad posterior)

observan los aminoácidos conservados glutamina, isoleucina y arginina adyacentes al motivo SHAQKYF que se han descrito como importantes en unión al ADN.

#### 4.2.2. Selección del mejor método de sustitución.

Para poder construir el árbol filogenético de las proteínas alineadas fue necesario también seleccionar el mejor método de sustitución que explicara el alineamiento de las proteínas EhMyb SHAQKYF y las de otros organismos, para esto se utilizó el servidor ProtTest. El mejor modelo de acuerdo a AIC fue: **WAG+G** de acuerdo a AIC.

#### 4.2.3. Reconstrucción del árbol filogenético

Finalmente con el alineamiento y el modelo de sustitución adecuado se construyó el árbol filogenético con Mr. Bayes y se editó utilizando la herramienta de software Fig Tree. El árbol obtenido se muestra en la figura 13. Podemos observar 3 grupos bien definidos dónde en uno de ellos identificamos diferentes organismos animales incluyendo mamíferos, equinodermos y artrópodos dentro de éstos podemos observar los motivos SYARQYF y SRVQKYF. En el mismo clado encontramos algunas proteínas de plantas y a la proteína canónica c-myb. De manera interesante las proteínas EhMybS1, EhMybS2 y EhMybS3 se agrupan mayoritariamente con plantas, algas y el organismo amoeboide *Disctiostellium discoideum* donde observamos la presencia conservada del motivo SHAQKYF y los aminoácidos adyacentes glutamina, isoleucina y arginina. Resulta interesante que los genes derivados de diferentes especies de plantas no tengan un origen monofilético, y que algunos de estos genes se agrupen con genes derivados de animales superiores y otros con genes de protozoarios. Para dilucidar el mecanismo evolutivo que dio origen a esta familia de genes es necesario llevar a cabo un análisis más detallado.

### 4.3. Predicción de estructura y dominios funcionales

#### 4.3.1. Predicción de estructura secundaria y terciaria.

Como parte de una caracterización inicial de las proteínas EhMyb, se llevó a cabo la predicción de su estructura secundaria. Ésto con el fin de determinar si

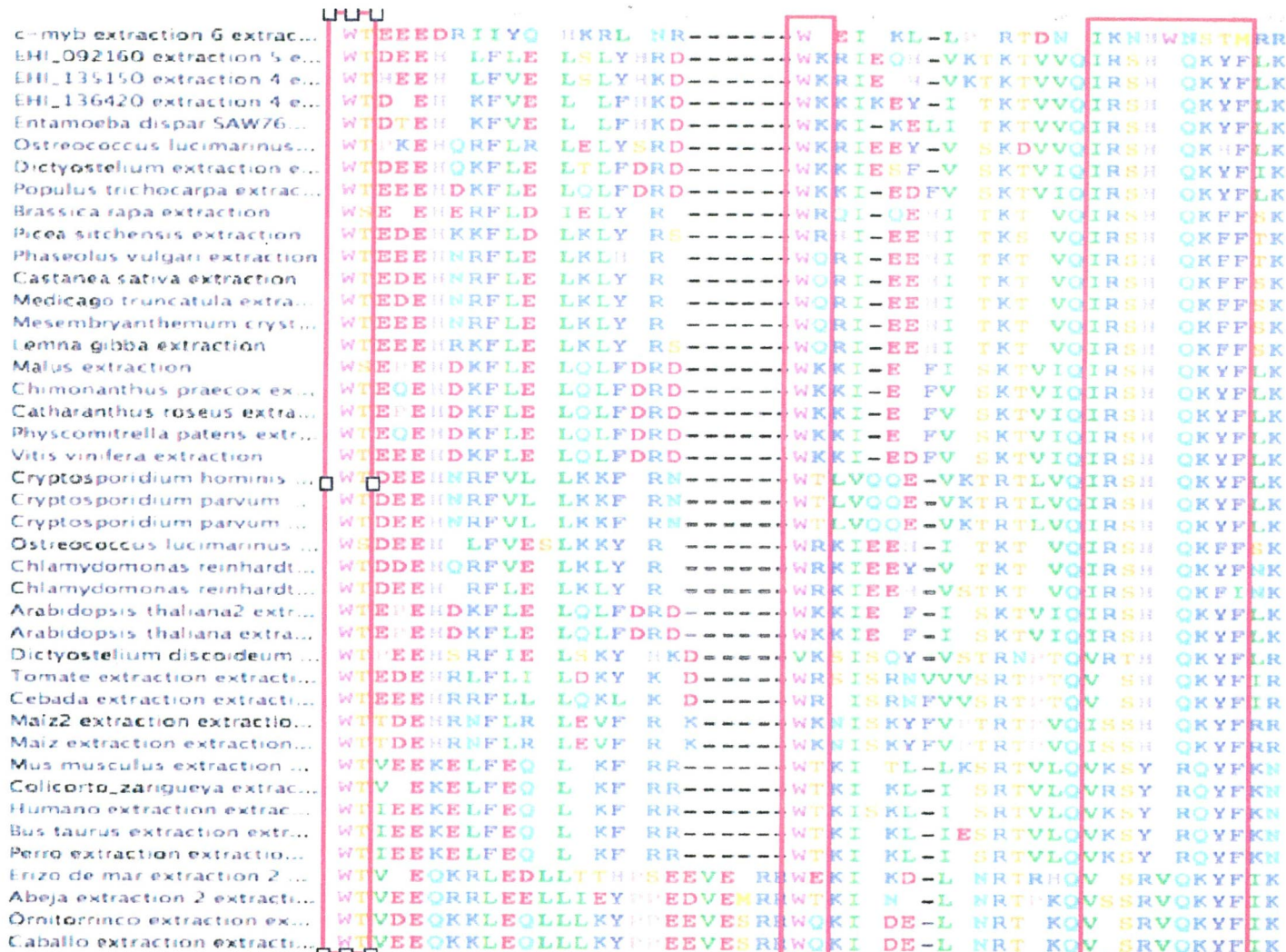


Fig. 12: Segmento más conservado en el alineamiento de las proteínas EhMyb SHAQKYF de *E.histolytica* y de otros organismos

. Se muestran los triptofanos o sus sustituciones en relación al dominio R3 de la proteína de tipo Myb canónica c-Myb de humano.



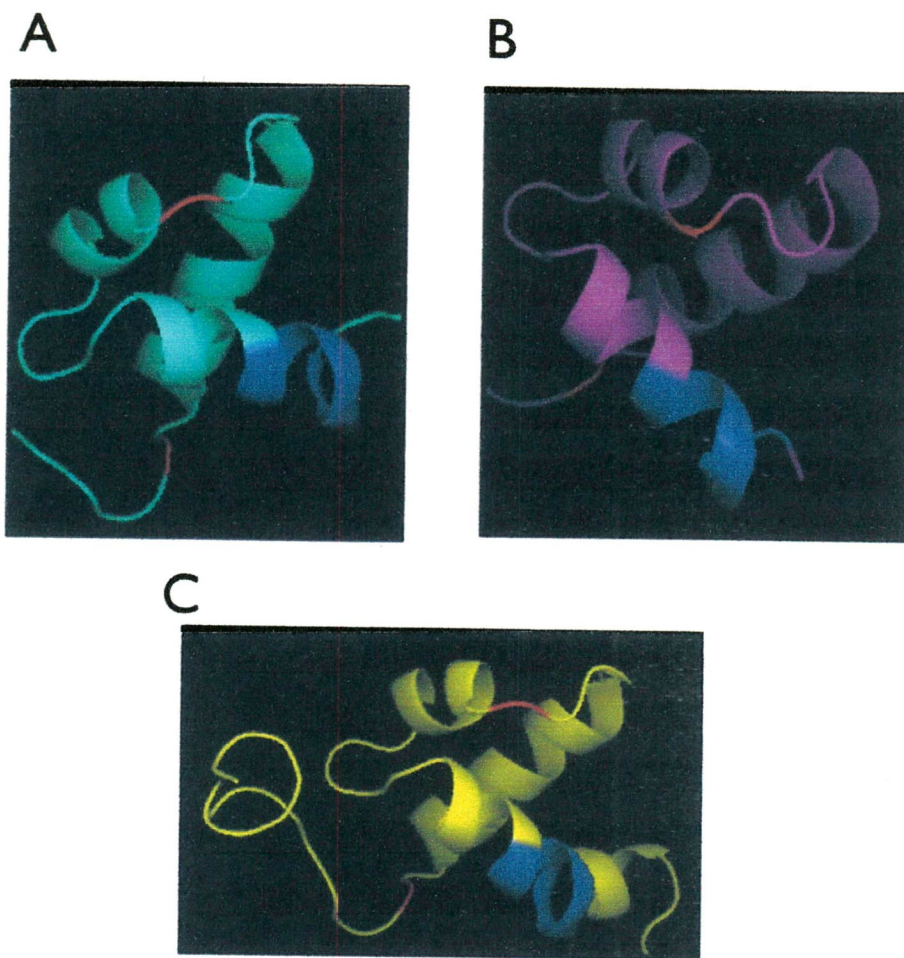
dichas secuencias presentaban la estructura hélice vuelta hélice característica de las proteínas Myb. Estos análisis se realizaron solo con las proteínas EhMybS1, EhMybS2 y EhMybS3 que son las únicas que preservan los triptofanos y el motivo SHAQKYF intactos y que son las que continuamos analizando en este trabajo a nivel de estructura y dominios funcionales. Utilizando el servidor Jpred se obtuvieron las predicciones de estructura secundaria de las tres proteínas. Éstas se muestran en la figura 14. Como se puede observar en la figura, las tres proteínas presentan el motivo hélice vuelta hélice esperado, el cuál se encuentra ubicado hacia el extremo amino terminal de la proteína de manera similar a la ubicación del dominio de unión a ADN en c-Myb.

Posteriormente se llevó a cabo la predicción de estructura terciaria, la cual se realizó por homología utilizando el servidor Phyre, el archivo pdb se visualizó con el programa PyMol. Las estructuras se muestran en las figuras 15. En todos los casos se observa que la estructura del dominio de unión a ADN MYB se conserva.

#### **4.3.2. Predicción de dominios funcionales.**

Finalmente, para concluir la caracterización inicial de las proteínas empleando estrategias bioinformáticas, se buscaron dominios funcionales utilizando el servidor Pfam. En las 3 proteínas bajo estudio se descubrió la existencia de un dominio de unión a ADN de tipo MYB. Para el caso de la proteína EhMybS1 se localizó el posible dominio de unión al ADN (DBD), el cuál se extiende desde el residuo 50 hasta el 95, para la proteína EhMybS2 se localizó el posible DBD del aminoácido 39 hasta el aminoácido 84, y para la proteína EhMybS3 el dominio de unión a ADN comienza en el aminoácido 49 y finaliza en el aminoácido 94.





**Fig. 15: Estructura terciaria de las proteínas EhMyb SHAQKYF**  
 . Se muestran las estructuras correspondientes a las proteínas EhMybS1, EhMybS2 y EhMybS3 en los paneles A,B y C respectivamente. En azul se resaltan el motivo SHAQKYF de cada una de ellas y en rojo los triptofanos espaciados.

### 4.3.3. Predicción de señales de localización nuclear.

Para determinar si las proteínas EhMybS1, EhMybS2 y EhMybS3 bajo estudio tenían alguna de las señales de localización nuclear conocidas, las secuencias se analizaron en el servidor PredictNLS. En el caso de las proteínas EhMybS1 y EhMybS2 no se detectó ninguna señal de localización nuclear, sin embargo, para la proteína EhMybS3 se identificó una señal de localización nuclear la cual se extiende del aminoácido 36 hasta el 76, la cual es una región rica en lisinas y argininas aminoácidos característicos de este tipo de regiones como se muestra en la figura 16. Además se predijo con un 94% de probabilidad como una proteína nuclear.

Input Sequence (NLS's in Red)	MSISPHDSFVPOYPLSPGSSQSTSSSPLPSDRLVYTKKQRKQYTTTKKREVWTDAAEAKF VEGLALPHKDWKKIKKEYIGTKTVVOIRSHAQKYFLKLNKTAPPQPFLLTFLKNFSVQOSI TKNSCOPPSPQFHDHIGDSTNCTSSAFSPVRDYADYIQIDGLLGLLPCYLITLIDLCY
Sequence Length	177
NLS's found. <i>No gives position of Motif</i>	• KKQRKQYTTTKKREVWTDAAEAKFVEGLALPHKDWKKIK 36

Fig. 16: Señal de localización nuclear de la proteína EhMybS3. Se muestra la salida del servidor ProtTest. En rojo la posible señal de localización nuclear encontrada por este servidor.

## 4.4. Análisis de los niveles de expresión de los genes *ehmyb* SHAQKYF de *E.histolytica* en condiciones de cultivo.

### 4.4.1. Expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* en condiciones basales de cultivo.

En el genoma de *E. histolytica* se encuentran codificados 9 genes que contienen el dominio de unión a ADN MYB con el motivo SHAKYF, de estos genes se han

reportado cambios en su expresión, en algunos de ellos en respuesta a estímulos como la invasión intestinal o bien se ha determinado su expresión como específica del estadio de trofozoito. Con la finalidad de conocer la abundancia relativa de los transcritos de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* en condiciones basales de cultivo, decidimos utilizar cDNA sintetizado a partir de ARN total de trofozoitos de *E. histolytica*. Mediante ensayos de RT-PCR en tiempo final exploramos su abundancia relativa utilizando oligonucleótidos específicos que amplificaron a cada uno de los genes completos. De esta manera se obtuvieron los productos de amplificación de los genes *ehmybs1*, *ehmybs2* y *ehmybs3* de 582, 510 y 546 pb respectivamente, como se muestra en la figura 17. En esta figura además se muestra que existe una expresión diferencial para cada uno de ellos siendo el gen *ehmybs2* el que se expresa de manera más abundante y el gen *ehmybs3* el que se expresa en menor cantidad en condiciones basales de cultivo relativo a la expresión del gen control *25srARN*. Además el gen *ehmybs1* presenta una expresión similar al control. Estos resultados nos indican una abundancia distinta de los mensajeros de cada uno de estos genes.

Se ha descrito que la expresión del gen *c-myb* cambia sus niveles en respuesta a diferentes estímulos (43), además otras proteínas de tipo Myb estudiadas en otros organismos como son *Oriza sativa* y *Chlamydomonas reinhardtii* cambian su expresión en condiciones de estrés (32, 46). En este contexto nos planteamos estudiar si la expresión de los genes *ehmybs1*, *ehmybs2* y *ehmybs3* podría ser modificada en respuesta a algún tipo de estrés y dado que en *E. histolytica* el modelo de estrés más estudiado es el inducido por choque térmico, decidimos evaluar los niveles de expresión bajo este estímulo.

#### 4.4.2. Cinéticas de ciclado para los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3*.

Dadas las diferencias en los niveles de expresión basales de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3*, previo al análisis de cambios de expresión en respuesta a diferentes estímulos, se decidió realizar cinéticas de ciclado para asegurar que la PCR se realice en la parte lineal de la curva de amplificación. Para esto se utilizó cDNA de condiciones basales de cultivo. De igual forma se realizó la cinética

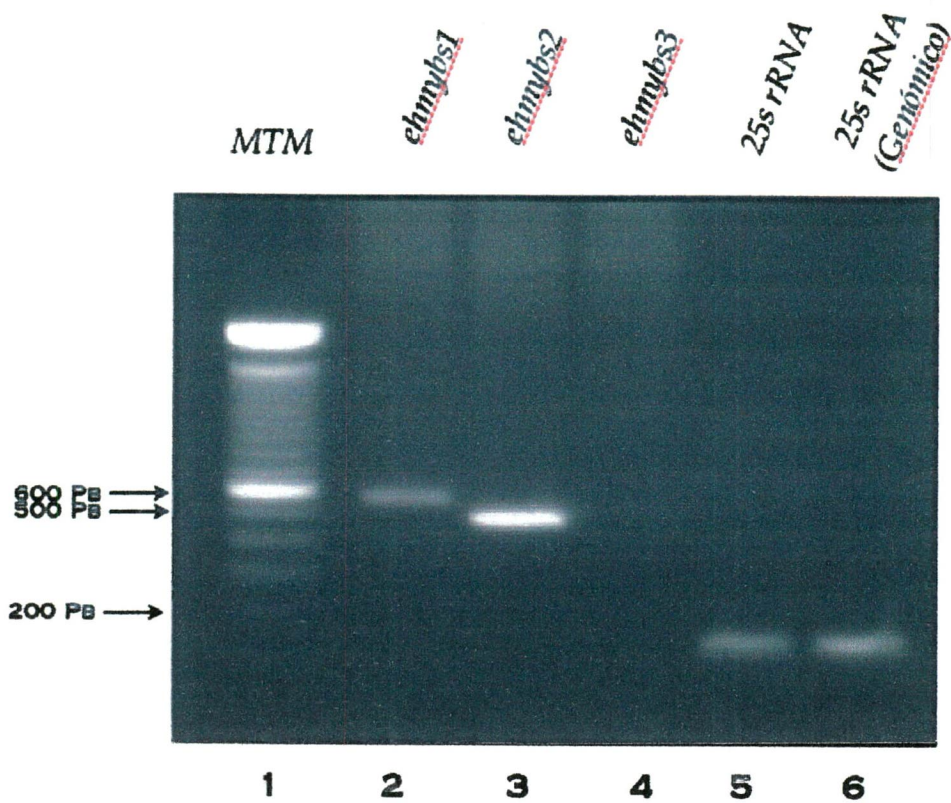
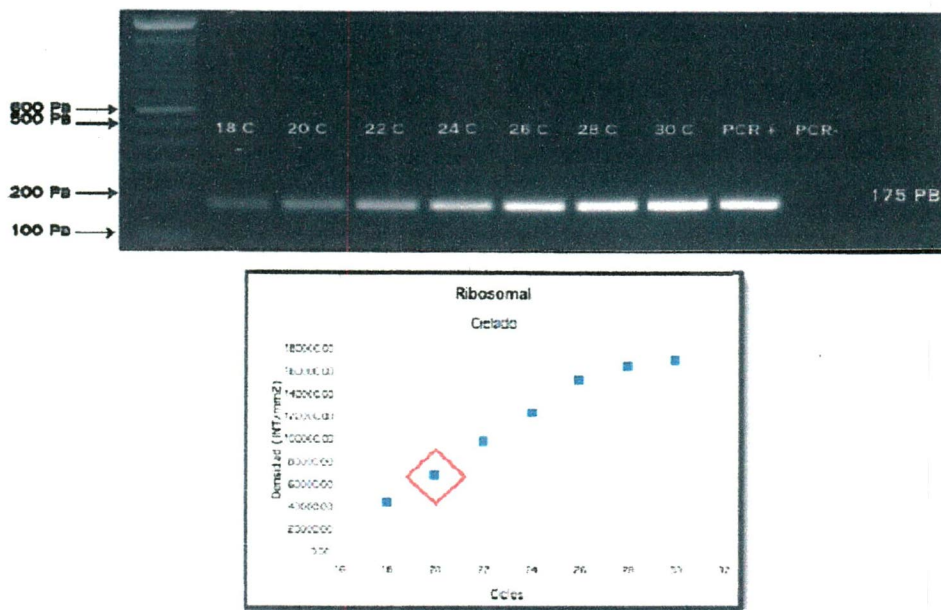


Fig. 17: Expresión basal de los genes *ehmybs1*, *ehmybs2* y *ehmybs3*. Se muestran los amplificados obtenidos de los genes bajo estudio utilizando como controles el amplificado del gen 25srARN a partir del mismo cDNA y a partir de ADN genómico ambos de 175pb. MTM.Marcador de Talla Molecular.

para el gen *25s rARN* el cual fue utilizado como control en todas las reacciones de RT-PCR. Los resultados de este experimento se muestran en las figuras 18, 19, 20 y 21. Mediante estos experimentos pudimos determinar que para el gen control *25s rARN* el ciclo óptimo para su amplificación es 20, en el caso del gen *ehmybs1*, el cual tiene niveles de expresión altos, el número óptimo de ciclos lo determinamos en 26. Al igual que éste, para el gen *ehmybs2* se seleccionaron 26 ciclos. Por último para el gen *ehmybs3*, el cual es el gen cuya expresión basal es menor, se seleccionaron 32 ciclos.



**Fig. 18: Cinética de ciclado para el gen *25s rARN***

. Se muestran los amplificadores obtenidos a los ciclos 18, 20, 22, 24, 26, 28 y 30. En el panel superior se muestra el gel de agarosa al 1.5% con los productos de amplificación para el gen bajo estudio a partir de cDNA de condiciones basales de cultivo y los controles tanto positivo como negativo de la PCR, ambos a partir de ADN genómico. En el panel inferior se muestra la gráfica de la densitometría en unidades de Intensidad/ $mm^2$  vs. Ciclos. En el recuadro rojo dentro de la gráfica se señala el número de ciclos seleccionado.

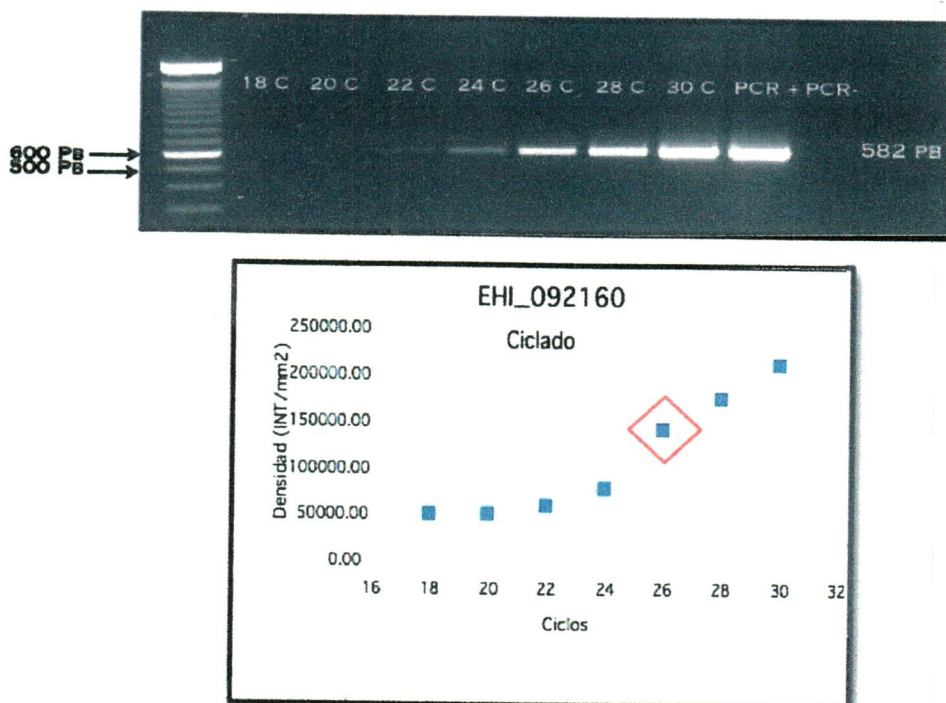
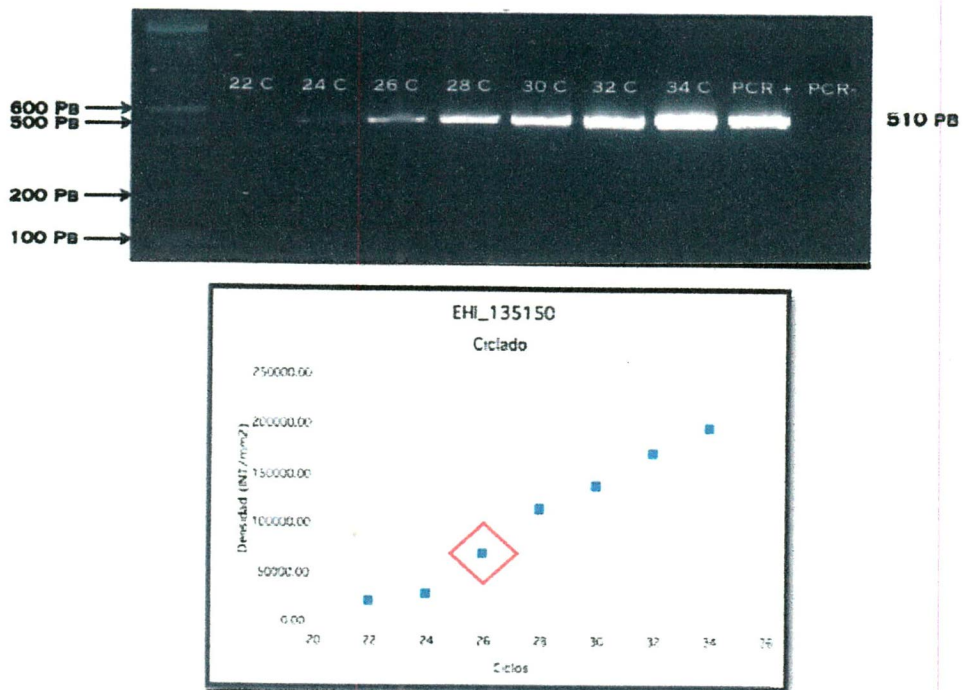
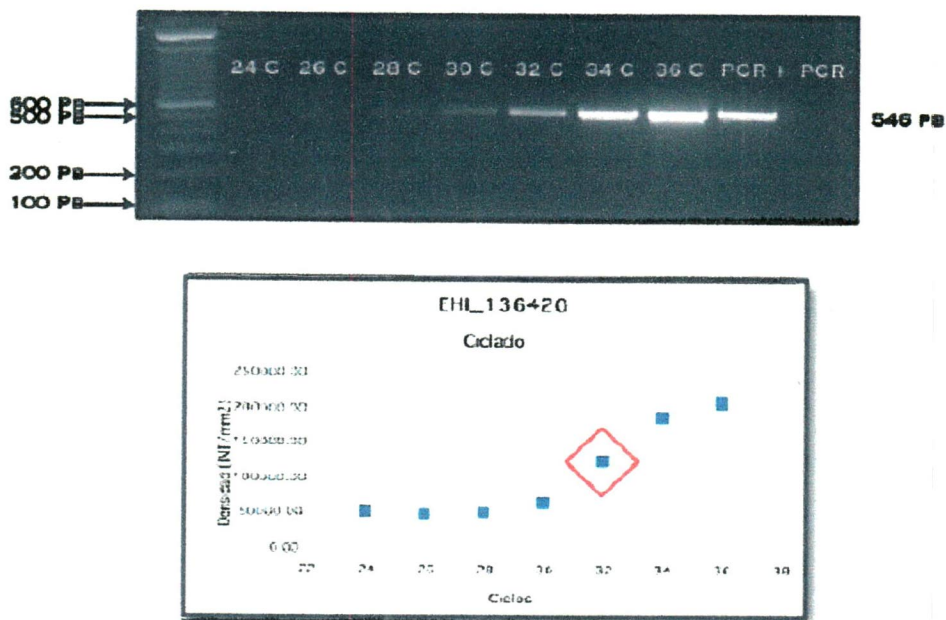


Fig. 19: Cinética de ciclado para el gen *EhmgbS1*. Se muestran los amplificadores obtenidos a los ciclos 18, 20, 22, 24, 26, 28, y 30. En el panel superior se muestra el gel de agarosa al 1.5% con los productos de amplificación para el gen bajo estudio a partir de cDNA de condiciones basales de cultivo y los controles tanto positivo como negativo de la PCR, ambos a partir de ADN genómico. En el panel inferior se muestra la gráfica de la densitometría en unidades de Intensidad/ $mm^2$  vs. Ciclos. En el recuadro rojo dentro de la gráfica se señala el número de ciclos seleccionado.



**Fig. 20: Cinética de ciclado para el gen *EhmybS2***

. Se muestran los amplificadores obtenidos a los ciclos 22, 24, 26, 28, 30, 32 y 34. En el panel superior se muestra el gel de agarosa al 1.5% con los productos de amplificación para el gen bajo estudio a partir de cDNA de condiciones basales de cultivo y los controles tanto positivo como negativo de la PCR, ambos a partir de ADN genómico. En el panel inferior se muestra la gráfica de la densitometría en unidades de Intensidad/ $mm^2$  vs. Ciclos. En el recuadro rojo dentro de la gráfica se señala el número de ciclos seleccionado.



**Fig. 21: Cinética de ciclado para el gen *EhmybS3***

. Se muestran los amplificadores obtenidos a los ciclos 24, 26, 28, 30, 32, 34 y 36. En el panel superior se muestra el gel de agarosa al 1.5% con los productos de amplificación para el gen bajo estudio a partir de cDNA de condiciones basales de cultivo y los controles tanto positivo como negativo de la PCR, ambos a partir de ADN genómico. En el panel inferior se muestra la gráfica de la densitometría en unidades de Intensidad/ $mm^2$  vs. Ciclos. En el recuadro rojo dentro de la gráfica se señala el número de ciclos seleccionado.

#### 4.4.3. Corroboración de la identidad de los genes *Ehmybs1* y *Ehmybs2* por secuenciación.

Se corroboró la identidad de los mensajeros de los genes *Ehmybs1* y *Ehmybs2* por secuenciación. Para esto los productos de PCR se clonaron en el vector TOPO TA 2.1 ya que es un vector que permite la clonación rápida a partir de productos de PCR. Se purificaron y secuenciaron 9 colonias positivas de cada uno de estos genes, las secuencias alineadas se muestran en las figuras 22 y 23. Estas secuencias se alinearon utilizando el programa ClustalW donde se verificó que de las nueve colonias secuenciadas para los genes *Ehmybs1* y *Ehmybs2* todas correspondieron a una sola población de ARN mensajeros.

#### 4.5. Análisis de los niveles de expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* de *E.histolytica* en condiciones de estrés calórico.

Utilizando cDNA sintetizado a partir de ARN extraído de trofozoitos de *E.histolytica* sometidos a estrés calórico por 4 h a 42 ° C, se realizó la reacción de amplificación de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3*. Como control de cambio se usó el gen de la *peroxirredoxina* con número de acceso EHL122310 en la base de datos Pathema, el cual en ensayos de microarreglos se ha reportado que disminuye hasta 16 veces en sus niveles de expresión en respuesta al choque térmico (53). Como control de carga se utilizó el gen *25s rARN*. Los resultados de esta RT-PCR en tiempo final se muestran en la figura 24.

Para obtener los datos semi-cuantitativos se realizó la densitometría de los geles obtenidos, con estos resultados se construyeron las gráficas de densitometría. Todos los valores obtenidos se normalizaron con los del gen *25s rARN*. Los cambios en los niveles de expresión se reportan como porcentaje en correlación con la expresión basal (100%). Estos resultados se muestran en la figura 25 en la que podemos observar que solo en el gen *ehmybs3* encontramos cambios en su expresión en respuesta al estrés inducido por choque térmico. Como control incluimos al gen de la *peroxirredoxina* el cual se observa que durante este estímulo se apaga su expresión tal y como fue reportado previamente. Todos nuestros datos se





normalizaron con base base a los datos obtenidos con el gen control *25s rARN*.

#### **4.6. Análisis de los niveles de expresión de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* de *E.histolytica* en condiciones de eritrofagocitosis.**

El estudio de la regulación génica es fundamental para elucidar el funcionamiento de cualquier sistema biológico en respuesta a un estímulo en particular. En este proyecto planteamos analizar condiciones que modifiquen la expresión de genes *ehmyb* con estímulos relacionados con la virulencia del parásito. Diferentes ensayos *in vitro* o *in vivo* como la destrucción de monocapas celulares, ensayos de eritrofagocitosis o la inducción de absesos hepáticos en gerbos se han empleado para investigar la importancia de diferentes moléculas en la patogenicidad de *E. histolytica*. En este proyecto estamos interesados en estudiar cambios en la expresión de los genes en respuesta a la interacción con eritrocitos, este modelo ha sido ampliamente utilizado por diferentes grupos para evaluar la virulencia de este parásito. Utilizando cDNA sintetizado a partir de ARN extraído de trofozoitos de *E.histolytica* sometidos a eritrofagocitosis por 0, 5, 15 y 30 min de interacción, se realizó la reacción de amplificación de los genes *ehmybs1*, *ehmybs2* y *ehmybs3*. Como control de carga se utilizó el gen *25s rARN*. Los resultados de esta RT-PCR en tiempo final se muestran en la figura 26.

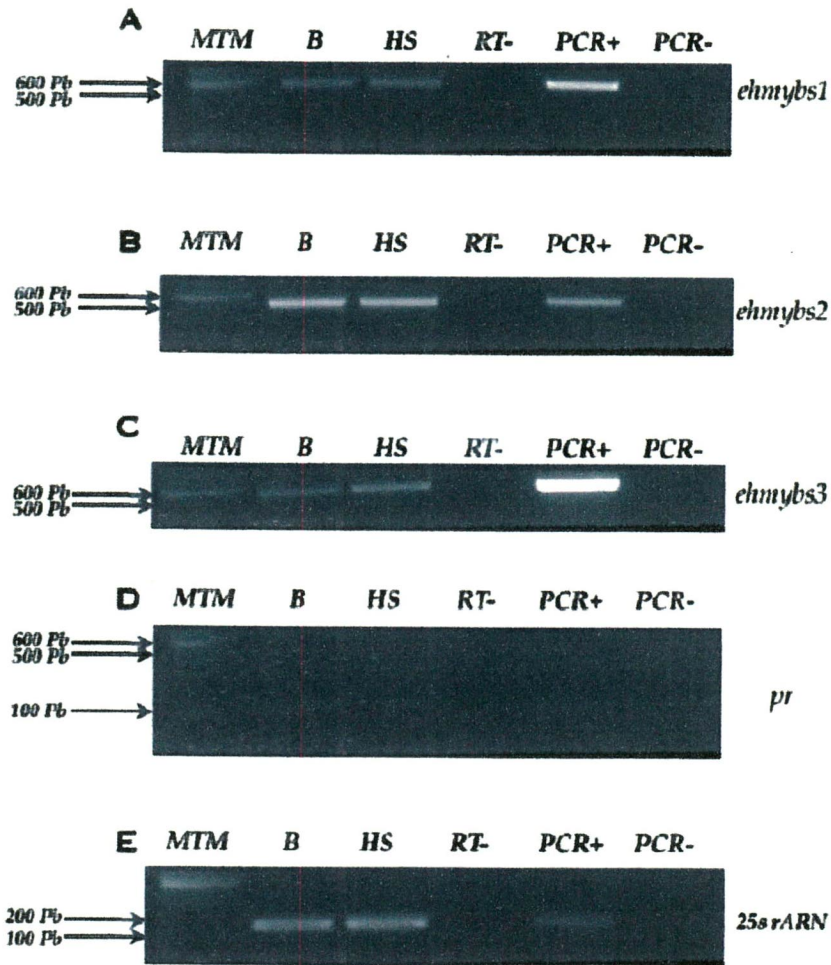


Fig. 24: RT-PCR de los genes *ehmybs1*, *ehmybs2* y *ehmybs3* en respuesta al choque térmico

. Se muestra en el panel A al gen *Ehmybs1*, en el B al gen *Ehmybs2*, en el C *ehmybs3*, en el D al gen *peroxirredoxina(pr)*, y en el E al gen *25s rARN*. Para cada panel se indican: MTM. Marcador de Talla Molecular, B. Condiciones basales de cultivo, EC. Condiciones de Estrés Calórico a 42 ° C por 4 hrs., RT-. Control de no contaminación por ADN genómico. PCR +. Control de la reacción de amplificación utilizando como templado ADN genómico. PCR -. Control de la reacción de amplificación sin templado.

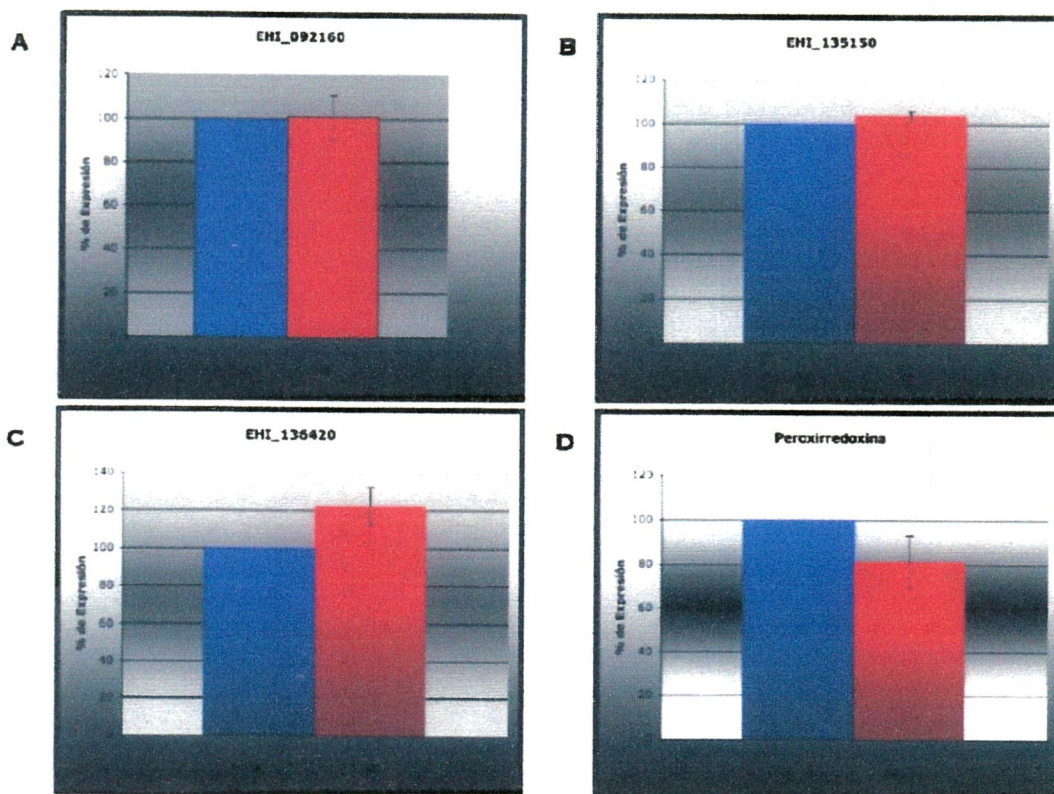


Fig. 25: Cambios en la expresión de los genes *Ehmybs1*, *Ehmybs2*, *Ehmybs3* y *peroxirredoxina* en condiciones de estrés calórico. Se muestran los paneles A, B, C, y D correspondientes a *Ehmybs1*, *Ehmybs2*, *Ehmybs3*, *peroxirredoxina(pr)* respectivamente. Para cada panel se indican en azul la condición de cultivo basal, y en rojo la condición de estrés inducido por calor. Se muestran las barras de la desviación estándar. Los datos representados están normalizados con los valores de densidad del gen *25s ARN*.

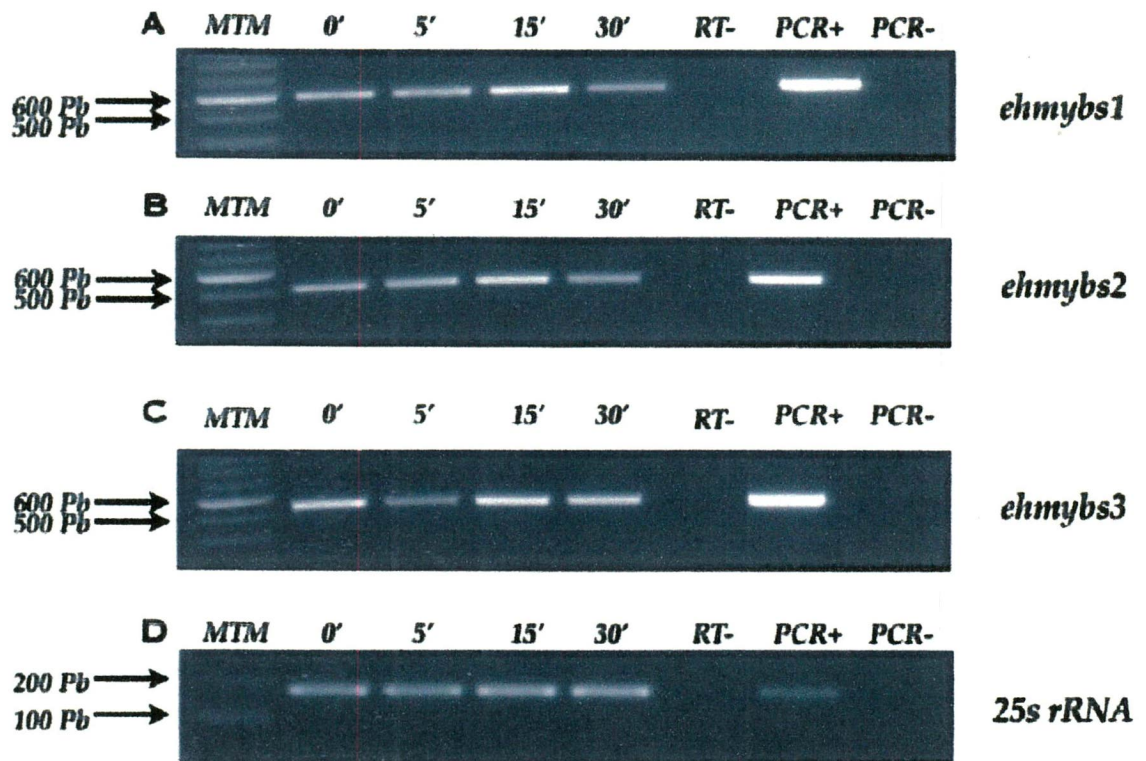


Fig. 26: RT-PCR de los genes *Ehmybs1*, *Ehmybs2* y *Ehmybs3* durante la eritrofagocitosis.

Se muestra en el panel A al gen *Ehmybs1*, en el B al gen *Ehmybs2*, en el C *ehmybs3*, en el D al gen *25s rARN*. Para cada panel se indica: MTM. Marcador de Talla Molecular, la expresión a diferentes tiempos de interacción con eritrocitos 0', 5', 15' y 30'. RT-. Control negativo de la transcripción reversa. PCR +. Control de la reacción de amplificación utilizando como templado ADN genómico. PCR -. Control de la reacción de amplificación sin templado.

## 4.7. Obtención de la proteína EhMybS3r de *E.histolytica*.

Para caracterizar a la proteína EhMybS3 necesitábamos obtenerla de forma recombinante y pura, por lo que el gen *ehmybs3* de *E. histolytica* se clonó y expresó de manera recombinante en bacterias. Con la proteína recombinante se realizaron ensayos de interacción ADN-proteína para analizar su capacidad de unión a una secuencia específica en el ADN.

### 4.7.1. Expresión y Purificación

La expresión de la proteína recombinante se realizó en la bacteria *E.coli* en la cepa derivada de BL21 denominada Rossetta II ya que esta contiene plásmidos con tARN que son raros en bacterias, de esta forma esta cepa está diseñada para favorecer la expresión de proteínas eucariontes. Al analizar el extracto de proteínas de nuestro cultivo por SDS PAGE, observamos que en el cultivo de la cepa BL 21 Rosetta II inducido con IPTG aparecía una banda de aproximadamente 46 kDa, correspondiente al peso teórico de la proteína de fusión recombinante (derivado de la suma del peso molecular de EhMybS3es que es de 20 kDa y GST, cuyo peso es de 26 kDa). En la figura 27 se muestra el experimento de la inducción de la expresión y se observa la banda en el peso esperado correspondiente a la proteína EhMybS3r de *Entamoeba histolytica*.

La purificación se realizó utilizando la prensa French y por cromatografía. De igual forma se expresó la GSTr para utilizarse como control en los ensayos de retardamiento de la movilidad electroforética. Las muestras tomadas en todos los puntos de la expresión y purificación se cargaron en un gel de poliacrilamida al 10%. Para visualizar las proteínas, el gel se tiñó con azul brillante de comassie. Este resultado se muestra en la figura 27.

Posteriormente se realizó un ensayo de inmunodetección utilizando un anticuerpo anti-GST, para verificar la expresión y el grado de purificación de la proteína recombinante EhMybS3 fusionada a GST, como se muestra en la figura 28

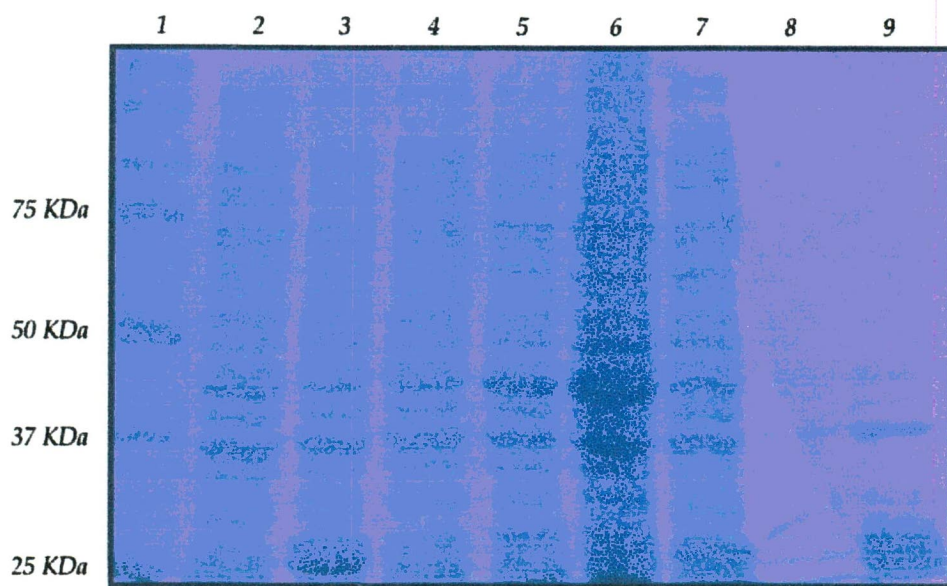
Con el protocolo utilizado fuimos capaces de purificar a la proteína EhMybS3r, sin embargo como se observa en el gel teñido con Coomassie, además de la banda en el peso esperado, se observan bandas con pesos moleculares menores, que pensamos son productos de degradación. A pesar de la degradación de la proteína,

consideramos que la cantidad de proteína íntegra es suficiente para la realización de los ensayos de retardamiento, dado que la cantidad requerida de proteína es muy pequeña. En purificaciones posteriores se deberán utilizar otros inhibidores de proteasas durante la purificación para prevenir en lo posible la degradación.

#### 4.8. Ensayos de retardamiento para la proteína EhMybS3 de *E.histolytica*

El dominio de unión a ADN Myb se encuentra presente en diferentes miembros de familias de factores de transcripción, de los cuales la proteína c-Myb es la proteína canónica. c-Myb reconoce de forma específica la secuencia [T/C]AAC[G/T]G, la cual se denomina elemento de reconocimiento a Myb (ERM) (4). En este contexto y considerando que las predicciones de estructura secundaria y terciaria indican que la proteína EhMybS3 tiene un dominio de unión a ADN similar en estructura al repetido R3 de c-Myb, nos preguntamos si la proteína EhMybS3r tenía la capacidad de unirse al ADN y si podría reconocer al ERM ya que se ha descrito que proteínas con el motivo SHAQKYF reconocen diferentes variaciones de esta secuencia.

Es por esto que realizamos un ensayo de unión a ADN (EMSA) y para ello utilizamos el oligonucleótido mim-1, el cual se obtuvo de una secuencia que se encuentra en el promotor del gen *mim-1* el cual contiene al ERM TAACGG. El oligonucleótido mim-1 se ha utilizado para caracterizar diferentes proteínas de tipo Myb entre las que se encuentran la proteína PfMyb1 de *Plasmodium falciparum* (7). De esta forma realizamos ensayos de retardamiento utilizando el oligonucleótido mim-1 y dos concentraciones molares de la proteína EhMybS3r. Como se muestra en la figura 29 este ensayo nos permitió identificar la presencia de 5 complejos ADN-proteína. De los cuales, los complejos I, IV y V fueron competidos en diferente grado con 100 ng de poli[dI-dC]. Los complejos II y III no fueron competidos por el poli[dI-dC] por lo que pensamos que son complejos producto de la unión específica de la proteína EhMybS3r con el ERM contenido en el oligonucleótido mim-1. La formación de dos complejos con la proteína EhMybS3r puede ser debido a la presencia de diferentes poblaciones de la proteína recombinante



**Fig. 27: Expresión y purificación de la proteína EhMybS3r**  
 1. MTM: Marcador de talla molecular. 2: Cultivo de GSTr sin inducir. 3: Cultivo de GSTr Inducido. 4: Cultivo de pEhMybS3r sin inducir. 5: Cultivo de pEhMybS3r inducido. 6: Fracción insoluble de la purificación de pEhMybS3r después de la prensa French . 7: Fracción soluble de la purificación de pEhMybS3r después de la prensa French 8: Fracción 1 de la purificación de pEhMybS3r por FPLC. 9. Fracción 2 de la purificación de pEhMybS3r por FPLC.

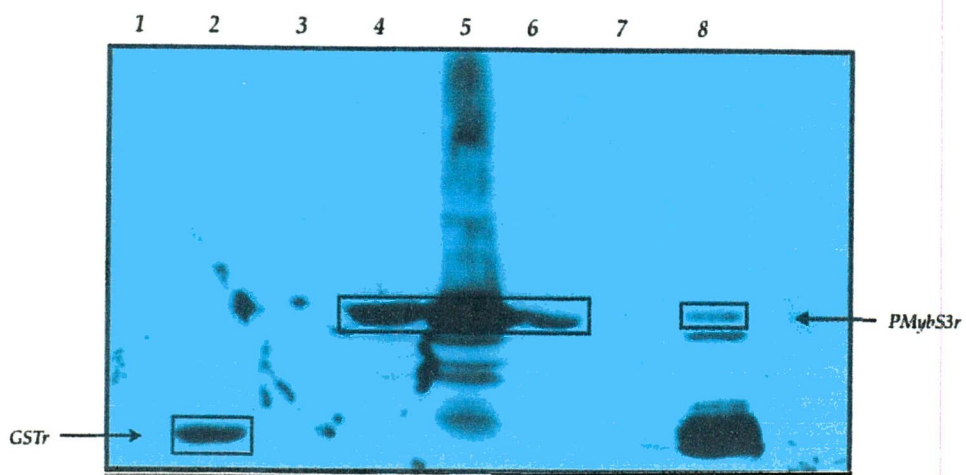


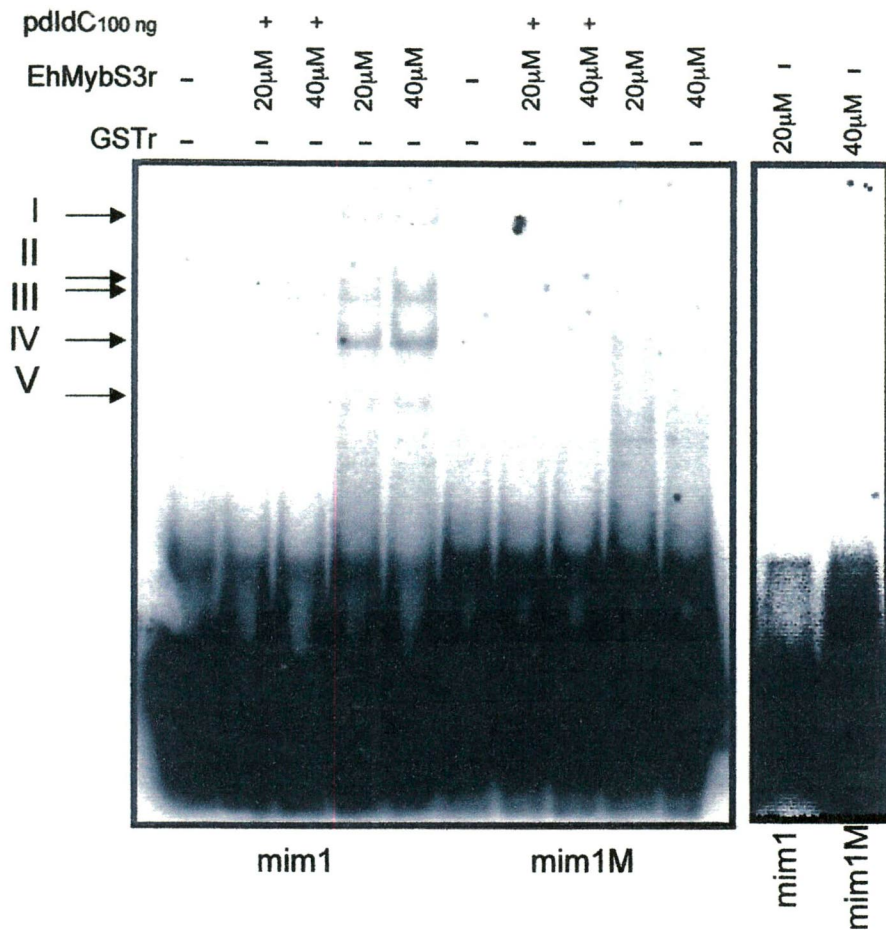
Fig. 28: **Western Blot de la expresión y purificación de pEhMybS3r**  
 1. MTM: Marcador de talla molecular. 2: Cultivo de GSTr sin inducir. 3: Cultivo de GSTr Inducido. 4: Cultivo de pEhMybS3r sin inducir. 5: Cultivo de pEhMybS3r inducido. 6: Fracción insoluble de la purificación de pEhMybS3r después de la prensa French . 7: Fracción soluble de la purificación de pEhMybS3r después de la prensa French. 8: Fracción 1 de la purificación de pEhMybS3r por FPLC. 9. Fracción 2 de la purificación de pEhMybS3r por FPLC.

producto de la degradación que observamos durante el proceso de purificación. Pensamos que estas diferentes poblaciones mantienen su capacidad de unirse al ADN debido a que conservan el dominio de unión al ADN formando así complejos EhMybS3r-ADN con diferente movilidad electroforética. Además, para conocer si la secuencia **TAACGG** es necesaria para la unión de la proteína EhMybS3r se utilizó una sonda la cual tenía mutada las bases **TAAC** por **GCCA**, las cuales son relevantes para la unión del factor de transcripción. Estos experimentos también se realizaron en presencia o ausencia del competidor poli[dI-dC]. Observamos que no se formaron complejos ADN-proteína con el oligonucleótido mim-1M; a excepción de uno en ausencia de poli[dI-dC] el cual tiene una movilidad electroforética similar al complejo IV previamente identificado como inespecífico. Por otro lado se incluyó como control a la proteína recombinante GST dado que la proteína EhMybS3r está fusionada a GST, con la finalidad de descartar la posibilidad de que los complejos observados con nuestra proteína recombinante fueran debidos a la GSTr o a otras que se co-purifiquen con la misma. Estos resultados sugieren que la proteína EhMybS3r tiene capacidad de unión a ADN y que requiere de la secuencia **TAACGG** contenida en el oligonucleótido mim-1.

#### 4.9. Búsqueda de posibles secuencias blanco de proteínas EhMyb SHAQKYF en el genoma de *E.histolytica*.

Ya que la proteína EhMybS3r tiene la capacidad de unirse a un elemento de reconocimiento a Myb y ya que las secuencias que reconocen las proteínas Myb SHAQKYF en los promotores de los genes del parásito *E. histolytica* no han sido descritas, decidimos, para finalizar este trabajo, buscar posibles genes blanco en el genoma de *E.histolytica* que contuvieran estas secuencias en sus promotores. Para ello se escribieron una serie de programas utilizando el lenguaje Python. Con dichos programas se obtuvieron todos los promotores correspondientes a 5835 genes del parásito representando el 65% de todos los genes contenidos en su genoma. Posteriormente se buscaron, utilizando expresiones regulares, las secuencias correspondientes a los elementos de reconocimiento a Myb, la caja TATA y el elemento GAAC, esto con la finalidad de que los resultados estuvieran en un contexto

# BIBLIOTECA UACM



**Fig. 29: Ensayo de unión a ADN para la proteína EhMybS3r**  
 . En el panel izquierdo se muestra el ensayo de la proteína EhMybS3r y se señalan los complejos formados. En el panel derecho se muestran los ensayos de unión a ADN de la GSTr con las sondas mim-1 y mim-1M.

biológico. Las secuencias Myb que se buscaron en el genoma de *E. histolytica* fueron el ERM canónico **T(C/T)AAC(G/T)G** y dos secuencias que son reconocidas por proteínas que contienen el motivo SHAQKYF: la secuencia de reconocimiento **AAC[A/T]GTT** identificada en el promotor del gen *pstO* de *Dictiostellium discoideum* (18), la cual es reconocida por la proteína Myb SHAQKYF MybE y la secuencia de reconocimiento **TATCCA** del gen de la  $\alpha$ -amilasa la cual es reconocida por las proteínas Myb SHAQKYF OsMYBS1, S3 y S3 (32). A partir de esta búsqueda pudimos obtener una serie de genes candidatos a ser regulados por proteínas de tipo Myb SHAQKYF en *E. histolytica*. De manera interesante encontramos que la secuencia canónica para las proteínas Myb se encuentra presente en alrededor de 40 promotores y su ubicación varía desde la posición -57 (respecto al ATG) como es el caso del promotor del gen *pgp-6*, hasta la posición -997 para un gen putativo que codifica para una proteína con dedos de zinc. También es importante destacar que encontramos promotores de genes relacionados con diversas funciones desde genes que participan en la respuesta al estrés por choque térmico hasta genes involucrados con la virulencia en este parásito e incluso genes relacionados con el proceso de fagocitosis como es el caso del gen putativo que codifica para una proteína RAB GTPasa. Los posibles genes blanco que contienen el ERM se muestran en la figura 30.

Dado que las proteínas con el motivo SHAQKYF se agrupan filogenéticamente con las de algunas plantas y con las del organismo amoeboide *Dictiostellium discoideum* proponemos que las proteínas que contienen el motivo SHAQKYF en *E. histolytica* podrían reconocer secuencias similares a las ya descritas para estos organismos. Es por esto que decidimos buscar ambas secuencias en los promotores de los genes del parásito, además dado que *Dictiostellium discoideum* y *E. histolytica* tienen un contenido de GC similar (alrededor del 22%) las secuencias de sus promotores podrían tener secuencias de reconocimiento similares. En esta búsqueda encontramos 18 genes candidatos y los elementos encontrados varían en su ubicación desde posiciones cercanas al ATG como es el caso del gen putativo *pgp-2* en la posición -202, hasta posiciones más alejadas como es el caso del gen putativo que codifica para la proteína Histona H2A donde se encuentra a -985. Por otro lado, con esta búsqueda encontramos 2 genes putativos relacionados con

la respuesta a estrés, en los cuales identificamos anteriormente que contienen el ERM, como es el caso del que codifica para la proteína activadora de la ubiquitina, donde se identificó el ERM en la posición -265 y el elemento de reconocimiento descrito en *Dictiostellium discoideum* en la posición -202. Esto nos sugiere que las proteínas de tipo Myb en el parásito pueden actuar de forma sinérgica en la regulación de la transcripción. Además es interesante mencionar que con este análisis se identificaron a los genes que codifican para las proteínas histonas H2B y H2A como posibles blancos ya que se sabe que otras proteínas de este tipo son reguladas por proteínas Myb. Por ejemplo el gen de la histona H2A.Z en humano y los genes de las histonas H3 y H2A en *Plasmodium falciparum*. Los resultados de esta búsqueda se muestran en la figura 31.

Finalmente incluimos en esta análisis la secuencia reconocida por las proteínas OsMybS1, S2 y S3 de *Oryza sativa*, esto con la finalidad de ampliar nuestra búsqueda de posibles genes blanco, además no podemos descartar la posibilidad de que reconozcan secuencias parecidas a las que se encuentran en los promotores de los genes de algunas plantas, esto debido a la relación filogenética que guardan las proteínas de tipo Myb SHAQKYF con las de otras pertenecientes a las plantas. Con esta búsqueda encontramos 20 posibles genes blanco, donde la posición de la secuencia de reconocimiento varía desde la localización -264 como es el caso del gen putativo que codifica para la proteína Pgp-2 hasta la localización -985 en el gen que codifica para una GTPasa de la familia Rho. Además dentro de los genes candidatos identificados, encontramos a algunos relacionados con la respuesta a estrés inducido por calor como es el gen putativo que codifica para una proteína de choque térmico y genes relacionados con virulencia como es el caso del gen putativo que codifica para una proteína de la familia AIG1. Los resultados de esta búsqueda se muestran en la figura 32.

Por otro lado es pertinente señalar que para todas los elementos de reconocimiento, también se buscaron las posiciones correspondientes a la secuencia consenso de la caja TATA y el elemento GAAC, esto en concordancia con la estructura de los promotores de *E. histolytica* que ha sido descrita en la literatura (42).

Número de acceso	Nombre	T(C/T)AAC(G/T)G	ATTTAAA	GAAC
<i>Traducción</i>				
>EHI_038620_1200_1000	40S ribosomal protein S16, putative	-807	-63	[-39, -22]
>EHI_148470_1200_1000	aspartate-ammonia ligase, putative	-377	-48	[-41]
<i>Membrana</i>				
>EHI_101230_1200_1000	P-glyco protein 6, putative	-57	-36	[-26]
>EHI_097990_1200_1000	158792 GrfA protein, putative	-82	-66	[-35]
<i>Transporte Vesicular</i>				
>EHI_052740_1200_1000	dynammin-like protein	-287	-44	[-31]
>EHI_015260_1200_1000	Rab GTPase activating protein, putative	-254	-89	[-52]
>EHI_160900_1200_1000	Sec1 family protein	-258	-43	[-23]
<i>Enquistamiento</i>				
>EHI_118670_1200_1000	Skp1 protein, putative	-391	-70	[-25]
>EHI_142040_1200_1000	exocyst complex component exo70, putative	-451	-85	[-59, -26]
<i>Señalización</i>				
>EHI_039850_1200_1000	Rho guanine nucleotide exchange factor, putative	-287	-93	[-14]
>EHI_148890_1200_1000	calmodulin, putative	-936	-71	[-26]
>EHI_128070_1200_1000	protein kinase, putative	-381	-89	[-52]
<i>Regulación ADN/ARN</i>				
>EHI_192080_1200_1000	zinc finger protein, putative	-997	-71	[-16]
>EHI_079290_1200_1000	EF-hand calcium-binding domain containing protein	-211	-69	[-26]
>EHI_069570_1200_1000	methyltransferase-like protein 2, putative	-762	-85	[-46, -24]
>EHI_167080_1200_1000	acetyltransferase, putative	-121	-43	[-32]
>EHI_065690_1200_1000	ankyrin repeat protein, putative	-374	-72	[-44, -38]
>EHI_023890_1200_1000	nuclear movement protein, putative	-807	-43	[-32]
>EHI_064680_1200_1000	alpha-soluble NSF attachment protein, putative	-679	-47	[-9]
>EHI_082500_1200_1000	structure specific recognition protein, putative	-748	-42	[-33]
<i>Proteólisis / respuesta a choque térmico</i>				
>EHI_035540_1200_1000	ubiquitin-activating enzyme, putative	-265	-51	[-19]
>EHI_072140_1200_1000	heat shock protein, putative	-815	-44	[-36]
>EHI_170060_1200_1000	ubiquitin-like protein	-703	-70	[-19, -10]
>EHI_009640_1200_1000	WD repeat protein	-783	-47	[-18]
>EHI_056440_1200_1000	WD domain containing protein	-747	-71	[-55]
>EHI_001960_1200_1000	WD domain containing protein	-525	-91	[-69]
<i>Virulencia</i>				
>EHI_010000_1200_1000	BspA-like leucine rich repeat protein, putative	-865	-83	[-37]
>EHI_066620_1200_1000	BspA-like leucine rich repeat protein, putative	-616	-35	[-9]
>EHI_051290_1200_1000	leucine rich repeat protein 1	-637	-35	[-9, -4]
<i>Metabolismo</i>				
>EHI_165070_1200_1000	short chain dehydrogenase family protein	-405	-37	[-20]
>EHI_178670_1200_1000	aldose reductase, putative	-856	-46	[-39]
>EHI_112890_1200_1000	apyrase, putative	-667	-55	[-26]
>EHI_026480_1200_1000	(2r)-phospho-3-sulfolactate synthase, putative	-733	-38	[-28, -22]
>EHI_179820_1200_1000	3'(2'),5'-bisphosphate nucleotidase, putative	-636	-84	[-25]
>EHI_111810_1200_1000	tryptophanase, putative	-579	-47	[-27]
>EHI_025710_1200_1000	iron-sulfur flavoprotein, putative	-340	-42	[-24]

Fig. 30: Genes que contienen en su región promotora el elemento de reconocimiento a Myb T(C/T)AAC(G/T)G.

En la tabla se indica el nombre de cada gen y la función con la que está relacionado, su número de acceso en la base de datos de Pathema, y la localización del ERM, así como la localización de la caja TATA (ATTTAAA) y la secuencia GAAC.

Número de acceso	Nombre	AAC(A/T)GTT	ATTTAAA	GAAC
<b>Respuesta a estímulos extracelulares</b>				
>EHI_067390_1200_1000	Rho family GTPase	-728	-40	[-24]
>EHI_148580_1200_1000	choline/ethanolamine kinase, putative	-741	-44	[-17]
<b>Traducción</b>				
>EHI_148470_1200_1000	aspartate-ammonia ligase, putative	-467	-48	[-41]
<b>Proteólisis/ Respuesta a Choque Térmico</b>				
>EHI_008380_1200_1000	aminopeptidase, putative	-677	-39	[-18]
>EHI_035540_1200_1000	ubiquitin-activating enzyme, putative	-202	-51	[-19]
>EHI_009640_1200_1000	WD repeat protein	-782	-47	[-18]
<b>Transporte vesicular</b>				
>EHI_052740_1200_1000	dynamín-like protein	-615	-44	[-31]
>EHI_162550_1200_1000	vesicular integral membrane protein VIP36, putative	-770	-62	[-34]
>EHI_160900_1200_1000	Sec1 family protein	-257	-43	[-23]
>EHI_082500_1200_1000	structure specific recognition protein, putative	-579	-42	[-33]
<b>Regulación de ADN/ARN</b>				
>EHI_188730_1200_1000	histone H2B	-935	-51	[-34]
>EHI_126210_1200_1000	histone H2A, putative	-985	-47	[-30]
>EHI_065690_1200_1000	ankyrin repeat protein, putative	-197	-72	[-44, -38]
>EHI_167080_1200_1000	acetyltransferase, putative	-635	-43	[-32]
<b>Metabolismo</b>				
>EHI_165070_1200_1000	short chain dehydrogenase family protein	-482	-37	[-20]
>EHI_025710_1200_1000	iron-sulfur flavoprotein, putative	-339	-42	[-24]
<b>Membrana</b>				
>EHI_186600_1200_1000	P-glycoprotein-2, putative	-202	-86	[-39]
>EHI_075410_1200_1000	P-glycoprotein 5, putative	-981	-81	[-46]

Fig. 31: Genes que contienen en su región promotora la secuencia AAC(A/T)GTT de *D. discoideum*.

En la tabla se indica el nombre de cada gen y la función con la que está relacionado, su número de acceso en la base de datos de Pathema, y la localización de la secuencia AAC(A/T)GTT, así como la localización de la caja TATA (ATTTAAA) y la secuencia GAAC.

Número de acceso	Nombre	TATCCA <sup>1</sup>	ATTTAAA <sup>2</sup>	GAAC <sup>3</sup>
<i>Traducción</i>				
>EHI_098840_1200_1000	40S ribosomal protein S4, putative	-624	-91	[-5]
<i>Membrana</i>				
>EHI_101230_1200_1000	P-glyco protein 6, putative	-835	-36	[-26]
>EHI_186600_1200_1000	P-glycoprotein-2, putative	-264	-86	[-39]
<i>Tráfico Vesicular</i>				
>EHI_068240_1200_1000	Rho family GTPase	-985	-47	[-39, -21]
>EHI_180900_1200_1000	Sec1 family protein	-697	-43	[-23]
<i>Proteólisis / respuesta a choque térmico</i>				
>EHI_072140_1200_1000	heat shock protein, putative	-539	-44	[-36]
>EHI_086080_1200_1000	proteasome alpha subunit, putative	-387	-67	[-22]
<i>Virulencia</i>				
>EHI_024230_1200_1000	cysteine synthase 2, putative	-376	-45	[-27]
>EHI_180930_1200_1000	cysteine synthase 2, putative	-376	-45	[-27]
>EHI_199470_856_656	AIG1 family protein, putative	-362	-31	[-17]
>EHI_051290_1200_1000	leucine rich repeat protein 1	-980	-35	[-9, -4]
>EHI_008380_1200_1000	aminopeptidase, putative	-721	-39	[-18]
<i>Señalización</i>				
>EHI_152350_1200_1000	casein kinase, putative	-681	-57	[-50]
>EHI_007340_1200_1000	protein phosphatase, putative	-322	-35	[-20]
<i>Ciclo celular</i>				
>EHI_189910_1200_1000	meiotic checkpoint regulator, putative	-699	-72	[-35]
<i>Metabolismo</i>				
>EHI_098560_1200_1000	hexokinase	-591	-42	[-33, -7]
>EHI_099700_1200_1000	NAD(FAD)-dependent dehydrogenase, putative	-454	-46	[-27]
>EHI_128020_1200_1000	glucosidase, putative	-371	-43	[-24]
>EHI_178670_1200_1000	aldose reductase, putative	-743	-46	[-39]
>EHI_112990_1200_1000	2,3-bisphosphoglycerate-independent phosphoglycerate mutase, putative	-374	-46	[-33]

Fig. 32: Genes que contienen en su región promotora la secuencia TTACCA de *O. sativa*. En la tabla se indica el nombre de cada gen y la función con la que está relacionado, su número de acceso en la base de datos de Pathema, y la localización la secuencia TTACCA, así como la localización de la caja TATA (ATTTAAA) y la secuencia GAAC.

## 5. Discusión

La regulación de la transcripción es uno de los puntos clave que permite la diversidad fenotípica entre diferentes organismos. En el caso de *E.histolytica* se conoce poco acerca de factores que regulan la transcripción. En este trabajo nos enfocamos en el estudio de factores de transcripción Myb con el motivo SHAKQYF. Resulta interesante que el análisis de la secuencia del genoma completo de *E.histolytica* revela la presencia de múltiples copias de genes cuyos productos protéicos tienen el motivo de unión a ADN myb. Una de las ideas que se ha consolidado en la biología evolutiva, es que la generación de novedades evolutivas depende en gran medida de que un número relativamente pequeño de elementos de regulación génica adquieran nuevos papeles e intervengan en nuevas rutas de regulación. . Es probable que la duplicación y la divergencia sean condiciones para que se dé la evolución de funciones nuevas, especialmente en genes involucrados en regulación de la transcripción. La idea detrás de este concepto es que que éstos genes se encuentran fuertemente restringidos selectivamente, ya que son necesarios para el funcionamiento óptimo de un organismo, y la exploración de nuevas funciones les está excluida. En cambio, la duplicación de estos genes permitiría que las copias parálogas quedaran en libertad de explorar el espacio funcional sin poner en riesgo la viabilidad del organismo. Aunque no está claro el papel que las diferentes copias de genes tipo myb juegan en el ciclo de vida de *E.histolytica*, el hecho de que se conservaron es indicativo de una diversificación o especialización de sus funciones.

Los eventos de duplicación de genes pueden ocurrir en distintos puntos a lo largo de la historia evolutiva, sin embargo resulta importante destacar que los eventos que dieron origen a las tres proteínas myb con motivo SHAKQYF en las que se centró este estudio parecen haber ocurrido en un período relativamente reciente, siendo las más cercanas a éstas las proteínas derivadas de organismos amoeboides como es *Entamoeba dispar* y *Dictiostellium discoideum*. Se sabe que *E.histolytica* comparte muchas características moleculares con el organismo *Dictiostellium discoideum*, por ejemplo con base en la comparación del genoma de ambos organismos se ha identificado que comparten 1 510 familias de genes (49), por lo cual no es sorprendente que algunas proteínas de ambos organismos estén

relacionadas filogenéticamente, como es el caso de las proteínas EhMybS1, S2 y S3 de *E. histolytica* y MybE de *Dictiostellium discoideum*.

El cambio de especificidad de los factores de transcripción puede ocurrir en la alteración ya sea de su dominio de unión a DNA, que potencialmente modificaría la secuencia blanco en el ADN, o en el dominio transactivador que le permite al factor de transcripción interactuar con otras proteínas. En este caso nos enfocamos a analizar los cambios en el dominio de unión a ADN. Se ha reportado que las sustituciones de los triptofanos por otros aminoácidos aromáticos o por alanina o leucina no afectan la capacidad de unión a ADN en la proteína canónica c-Myb (47), y éste es precisamente el tipo de mutaciones presentes en las proteínas analizadas en este trabajo, lo que sugiere que las secuencias identificadas en nuestro análisis preliminar podrían conservar su capacidad de unir DNA, ésta idea fue apoyada por los datos de predicción de estructura secundaria y terciaria de dichas proteínas. Cuando se evaluó la capacidad de unión a ADN de la proteína recombinante EhMybS3r a una sonda que contiene un elemento de reconocimiento a Myb canónico denominado mim-1, se observó la formación de complejos ADN-proteína, lo que sugiere que a pesar de la mutaciones que presenta la proteína con respecto a un dominio myb canónico, conserva la capacidad de unirse al ADN y que además posee cierta afinidad por el elemento de reconocimiento a Myb ya que cuando se cambian las bases relevantes para la unión (mim-1 mutada) no se formaron complejos. Es pertinente señalar que no podemos definir si esta capacidad de unión es secuencia específica ya que no se realizaron ensayos con competidores específicos. Sin embargo, los resultados obtenidos en este trabajo nos permiten plantear la idea de que los cambios observados en la secuencia de aminoácidos del dominio myb cambie la especificidad de unión o la fuerza de la interacción con el ADN, lo que a su vez se reflejaría en la especialización o diversificación de las funciones de ésta familia de factores transcripcionales.

Algunas proteínas Myb han sido descritas como reguladores transcripcionales involucradas en el control de la expresión de genes relacionados con respuesta a diferentes tipos de estrés, por ejemplo OsMYBS1,S2 y S3 en la regulación del gen  $\alpha$  - amilasa en respuesta a la privación de azúcar en *Oryza sativa*, LCR1 involucrado en la respuesta a la concentración de CO<sub>2</sub> en *Chlamydomonas reinhardtii* y

AtPHR1 relacionado con la respuesta a privación de fosfato en *Arabidopsis thaliana*. Un modelo de estrés estudiado ampliamente en *E.histolytica* es la respuesta al estrés calórico, en este contexto en ensayos con microarreglos se detectaron 3.6 veces de cambio en la expresión del gen EHL\_063550 que codifica para una proteína MYB de tipo R2R3 (34). Resulta interesante que la expresión del gen *ehmybs3* haya aumentado en condiciones de estrés calórico. Dado que otras proteínas de tipo Myb han sido relacionadas con la regulación de genes que participan en la respuesta a determinado estrés, no es sorprendente que el gen *ehmybs3* cambie sus niveles de expresión bajo la condición estudiada. Si bien es necesario continuar con los estudios para poder establecer si este gen está involucrado en la respuesta al choque térmico, este resultado representa una primera aproximación a las funciones que podrían desempeñar las proteínas Myb SHAQKYF en el parásito *Entamoeba histolytica*.

Otra característica de las proteínas de tipo Myb es que éstas han sido relacionadas con el desarrollo o diferenciación celular en una gran variedad de organismos como son mamíferos, plantas e incluso algunos parásitos protozoarios (38), (40), (51). Por ejemplo el gen *zmmyst1* que codifica para una proteína myb de tipo SHAQKYF involucrado en el desarrollo y diferenciación celular en *zea mays* (24), otra proteína de tipo Myb SHAQKYF, HvMYBS3 ha sido relacionada con el desarrollo del endospermo en *Hordeum vulgare*, LeMybI involucrado en la regulación de la expresión de los genes tejido específicos RBCS1,S2 y S3A en *Lycopersicon esculentum* (45) y MybE involucrado en la diferenciación celular en *Dictyostelium discoideum* (18) entre otros. *E.histolytica* es un organismo que durante su ciclo de vida presenta dos estadios celulares: quiste y trofozoito. En este contexto se han identificado genes regulados por el desarrollo que se expresan de forma específica en uno u otro estadio, entre estos los genes myb EHL\_063550 y *ehmybs2* (16) se han reportado como genes específicos del trofozoito. Es bien conocido que los trofozoitos de este parásito llevan acabo diferentes mecanismos que le permiten invadir al hospedero como son la adhesión, la citólisis y la fagocitosis de células blanco. Considerando a los genes ehmyb SHAQKYF como posibles factores de transcripción pensamos que estos podrían estar involucrados en la regulación de otros genes relacionados con cualquiera de los procesos específicos del trofozoito. En este trabajo

con la finalidad de estudiar a los genes *ehmybs1*, *ehmybs2* y *ehmybs3* decidimos evaluar sus niveles de expresión en un modelo de laboratorio que representara un proceso específico del trofozoito: la eritrofagocitosis. Es importante destacar que en los tres genes se observó una disminución en su expresión que sugiere su participación en eventos relacionados con la eritrofagocitosis, lo cual implicaría una relación directa con la virulencia de este parásito.

Si las proteínas EhMyb SHAQKYF son factores de transcripción de tipo Myb, entonces deberían tener la capacidad de unirse al ADN y probablemente reconocer secuencias específicas como es el caso de las proteínas canónicas que se unen al consenso: [T/C]AAC[G/T]G denominado Elemento de Reconocimiento MYB (MRE) (4) o las proteínas Myb SHAQKYF como en *Dictyostelium discoideum* donde MybE es capaz de unirse a la secuencia AAC[A/T]GTT (18) o en *Oryza sativa* donde las proteínas OsMYBS1,S2 y S3 son capaces de unirse a la secuencia TATCCA (32). En este trabajo se encontraron posibles elementos de regulación myb en genes involucrados en diversas funciones, si bien sólo un número limitado de promotores parece contener estos elementos. Es pertinente señalar que como criterios de búsqueda además de las secuencias antes mencionadas también se consideró que los hits encontrados estuvieran en un contexto biológico, es decir que los promotores tuvieran una caja TATA y que el elemento se encontrara río arriba de la misma. Además se buscó si contenían el elemento GAAC, esto en correspondencia con la estructura tripartita de los promotores de *E.histolytica* sugerida por Purdy en 1996 (42). El INR no se consideró como un criterio de exclusión ya que se conocen promotores que no lo contienen como es el caso del gen *ehrabB* (44). Cabe mencionar que este análisis es preliminar y que la secuencia específica que reconocen estas proteínas no se conoce, por lo que será necesario llevar a cabo un análisis más profundo conforme se vaya ampliando la información disponible acerca de este tipo de factores de transcripción.

En conjunto la evidencia presentada en este trabajo apoya la idea de que las proteínas myb con motivo SHAKQYF son factores de transcripción, aunque es necesario dilucidar los procesos en los que intervienen dentro del ciclo de vida de *E.histolytica*.

## 6. Conclusiones

1. *E.histolytica* contiene en su genoma 9 genes que codifican para proteínas del tipo EhMyb SHAQKYF.
2. Las proteínas EhMyBS1, EhMyb S2 y EhMybS3 no presentan sustituciones en los dos primeros triptofanos característicos del dominio de unión a ADN Myb y el motivo SHAQKYF se mantiene sin variaciones.
3. Las proteínas EhMyBS1, EhMyb S2 y EhMybS3 conforman un grupo monofilético.
4. Las proteínas EhMyBS1, EhMyb S2 y EhMybS3 tienen como parálogos más cercanos a las proteínas del organismo amoeboide *Dictiostellium discoideum* y el alga *Ostreococcus lucimarinus*.
5. La proteína EhMybS3 puede estar involucrada en la regulación de genes relacionados con la respuesta al choque térmico y la eritrofagocitosis.
6. La proteína EhMybS3r tiene capacidad de unión a ADN y reconoce a la secuencia de reconocimiento a Myb del oligonucleótido mim-1.
7. Existen genes en *E.histolytica* que contienen en sus promotores posibles secuencias blanco para las proteínas de tipo Myb SHAQKYF entre los que destacan genes de respuesta a estrés por choque térmico, genes involucrados con virulencia y genes relacionados con señalización.

## 7. Apendice A: Abreviaturas

**EhMyb** Proteína de *Entamoeba histolytica* con dominio Myb

**ADN** Ácido desoxirribonucleico

**PCR** Reacción en cadena de la polimerasa

**EHMybS1-S9** Proteína de tipo Myb SHAQKYF de *Entamoeba histolytica*

**RT-PCR** Retrotranscripción-reacción en cadena de la polimerasa

**EhMybS3r** Proteína de tipo Myb SHAQKYF de *Entamoeba histolytica* 3 recombinante

**EMSA** Ensayo del cambio de la movilidad electroforética

**Mb** Mega bases

**ARN** Ácido ribonucleico

**ARNi** Ácido ribonucleico de interferencia

**pre-m-ARN** ARN mensajero precursor

**pol I,II III** polimerasa I, II y III

**Inr** Elemento iniciador

**BRE** Elemento reconocible por TFIIB

**DPE** Elemento de reconocimiento a río abajo

**GTF** Factor de transcripción General

**TF** Factor de transcripción

**TBP** Proteína de unión a la caja TATA

**TAFS** Factores asociados a la TBP

**CTD** Dominio carboxilo terminal

**ATP** Adenosin trifosfato

**pb** Pares de bases

**DBD** Dominio de unión a ADN.

**aa** aminoácidos

**TIGR** Instituto de Investigación Genómica

**Mpb** Mega pares de bases

**c-mb** Contraparte celular del gen del virus de la mieloblastosis.

**v-myb** Virus de la mieloblastosis.

**kDa** Kilo Daltones

**TRF** Factor de unión a los repetidos teloméricos

**R1, R2 y R3** Repetidos 1,2 y 3.

**MRE** Elemento de reconocimiento a Myb.

**URE** Elemento regulador río arriba.

**NCBI** Centro Nacional de Información Biotecnológica

**min** Minutos

**U** Unidades enzimáticas

**mL** Mililitro

**µg** Microgramos

**°C** Grados centrigados

**hrs** Horas

**FD** Factor de dilución

**nm** Nanometros

**rpm** Revoluciones por minuto

**g** Gravedades

**UV** Ultra violeta

**mM** Mili Molar

**DTT** Ditioneitol

**dNTPs** Desoxinucleótidos trifosfatados

**cDNA** ADN complementario

**rARN** ARN ribosomal

**V** Volts

**PS** Primer sentido

**PA** Primer Antisentido

**PBS** Solución Amortiguadora de fosfatos

**s** Segundos

**dT** Desoxitimidina

**tARN** ARN de transferencia

**LAC** Medio Luria Ampicilina Cloranfenicol

**FPLC** Cromatografía Líquida Rápida de Proteínas

**pH** Potencial de hidrógeno

**PM** Peso molecular

**CEM** Coeficiente de Extinción Molar

**mA** Mili amperios.

**TBS** Solución amortiguadora de Tris salina

**$\mu$ M** Micromolar

**$\mu$**  Microlitros

**$\mu$ Ci** Micro Curies

**dATP** 2'-deoxyadenosina 5'-trifosfato

**dGTP** 2'-deoxyguanosina 5'-trifosfato

**dCTP** 2'-deoxycitosina 5'-trifosfato

**dTTP** 2'-deoxytimidina 5'-trifosfato

**ng** Nanogramos

**AIC** Criterio de Información de acuerdo a Akaike

**MTM** Marcador de Talla Molecular

**mm<sup>2</sup>** milímetros cuadrados

**INT** Intensidad

**RT-** Control de retrotranscripción negativo.

**PCR+** Control positivo de la reacción en cadena de la polimerasa.

**PCR-** Control negativo de la reacción en cadena de la polimerasa.

**DO** Densidad Óptica

**SDS PAGE** Electroforésis en gel de poliacrilamida desnaturizante.

**GSTr** Proteína Glutation S-transferasa recombinante

**poly(dI-dC)** Homopolimero de desoxiinosina y desoxicitosina

**IPTG** Inductor Isopropil  $\beta$ -D-1-tiogalactopiranosida

**mm-1** Gen que codifica para la proteína mieloides 1 inducida por Myb

**mim-1M** sonda mim-1 mutada

**ELT** Solución EDTA-Lisozima-Triton

**TS** Solución Tris HCL-Sacarosa

**PEG** Polietilenglicol-NaCl

**RtREV+G** Modelo de sustitución para inferencia de la filogenía de retrovirus y retrotranscriptasa.

**WAG+G** Modelo de sustitución de Whelang y Goldman

## 8. Apéndice B: Soluciones

### Medios de cultivos

LB: 10 g de triptona, 5 g de extracto de levadura y 10 g de NaCl. Agua c.b.p. 1L.  
Se ajusta el pH a 7.5 con NaOH. Se esteriliza por autoclave.

### TYI-S-33

Biosate 300 g, Glucosa anhidra 91 g, NaCl 20 g,  $\text{KH}_2\text{PO}_4$  6 g,  $\text{K}_2\text{HPO}_4$  10 g, L(+)  
Ac. ascórbico 2.5 g, L-Cisteína 12.5 g, Agua c.b.p. 8.7 L.

### Solución Alsever (500mL)

Ácido cítrico 0.275 g, Citrato de Sodio 4.0 g, NaCl 2.1 g, Dextrosa 10.25 g, agua  
c.b.p. 500 mL.

### Amortiguador de carga para proteínas (2X)

Tris-HCl 100mM, 4% de SDS (p/v), 20% de glicerol, 2% de 2-mercaptoetanol  
y 0.25% de azul de bromofenol (p/v). Se ajusta el pH a 6.8.

### Amortiguador de corrida para electroforesis de proteínas

Tris-base 0.025M, glicina 200mM y SDS 0.1%.

### TAE (50X). 1L

242 g de Tris-base, 5.7% de ácido acético glacial (v/v), 18.6 g de  $\text{Na}_2\text{EDTA}$  2  
 $\text{H}_2\text{O}$ . Agua c.b.p. 1 L. Se ajusta el pH a 8.3.

### TBE(10X)

Tris-base 44.5 mM, ácido bórico 44.5 mM y EDTA 0.9mM.

### Amortiguador de transferencia(1X)

Tris-HCl 25mM, glicina 129 mM y metano al 20%.

### Amortiguadores para purificación de proteína por FPLC

#### Amortiguador de unión

20 mM fosfato de sodio, 0.15 M NaCl. Se ajusta el pH a 7.3

#### Amortiguador de elución

50 mM Tris-HCl, 10 mM glutatión reducido. Se ajusta el pH a 8.0.

### PBS pH 7.4

NaCl 137 mM,  $\text{KH}_2\text{PO}_4 \cdot 12\text{H}_2\text{O}$  10 mM y KCl 2.7 mM. Se ajusta el pH a 7.4.

### Solución de azul de coomassie(1L)

Coomassie R-250 al 0.25%, 500 mL de metanol y 70 mL de ácido acético. Agua  
c.b.p. 1L.

**Solución desteñidora**

Etanol 30%, ácido acético 10% y agua bidestilada 60%.

**Amortiguador de unión 5X**

HEPES pH 7.9 12 mM, kCl 60 mM, DTT 1 mM, MgCl<sub>2</sub> 4 mM, espermidina 4 mM y glicerol 10%.

## Referencias

- [1] **Avison, M. B.** 2007. *Measuring Gene Expression. The basics.* Taylor and Francis Group.
- [2] **Baranowskij, N., C. Froberg, S. Prat, and L. Willmitzer.** 1994. A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. *EMBO J.* **13**:5383–5392.
- [3] **Becker, W. M., L. J. Kleinsmith, and J. Hardin.** 2003. *The World of the Cell.* Benjamin Cummings, 5th ed.
- [4] **Bergholtz, S., T. O. Andersen, K. B. Andersson, J. Borrebaek, B. Lüscher, and O. S. Gabrielsen.** 2001. The highly conserved DNA-binding domains of A-B- and c-Myb differ with respect to DNA-binding, phosphorylation and redox properties. *Nucleic Acids Res.* **29**:3546–3556.
- [5] **Biedenkapp, H., U. Borgmeyer, A. E. Sippel, and K. H. Klempner.** 1988. Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**:835–7.
- [6] **Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–74.
- [7] **Boschet, C., M. Gissot, S. Briquet, Z. Hamid, C. Claudel-Renard, and C. Vaquero.** 2004. Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 *Plasmodium falciparum* clones. *Mol. Biochem. Parasitol.* **138**:159–63.
- [8] **Butler, J. E., and J. T. Kadonaga.** 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16**:2583–2592.

- [9] **CENAVECE**. 2007. Boletín epidemiología 2007. Tech. rep.
- [10] **CENAVECE**. 2008. Boletín epidemiología 2008. Tech. rep.
- [11] **Clark, C. G., U. C. M. Alsmark, M. Tazreiter, Y. Saito-Nakano, V. Ali, S. Marion, C. Weber, C. Mukherjee, I. Bruchhaus, E. Tannich, M. Leippe, T. Sicheritz-Ponten, P. G. Foster, J. Samuelson, C. J. Noël, R. P. Hirt, T. M. Embley, C. A. Gilchrist, B. J. Mann, U. Singh, J. P. Ackers, S. Bhattacharya, A. Bhattacharya, A. Lohia, N. Guillén, M. Duchêne, T. Nozaki, and N. Hall**. 2007. Structure and content of the *Entamoeba histolytica* genome. *Adv. Parasitol.* **65**:51–190. PMID: 18063096.
- [12] **Colin J. Davidson, L. A. H., Rabindra Tirouvanziam, and J. S. Lipsick**. 2005. Functional evolution of the vertebrate myb gene family: b-myb, but neither a-myb nor c-myb, complements drosophila myb in hemocytes. *Genetics* **169**:215–229.
- [13] **ConsortiumInternational, H. G. S.** 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**:931–945.
- [14] **Dudek, H., and E. P. Reddy**. 1989. Identification of two translational products for c-myb. *Oncogene* **4**:1061–6. PMID: 2674850.
- [15] **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797. PMC390337.
- [16] **Ehrenkauf, G. M., R. Haque, J. A. Hackney, D. J. Eichinger, and U. Singh**. 2007. Identification of developmentally regulated genes in *Entamoeba histolytica*. *Cell. Microbiol.* **9**:1426–1444.
- [17] **Espinosa-Cantellano, M., and A. Martínez-Palomo**. 2000. Recent developments in amoebiasis research. *Curr. Opin. Infect. Dis.* **13**:451–456.
- [18] **Fukuzawa, M., N. V. Zhukovskaya, Y. Yamada, T. Araki, and J. G. Williams**. 2006. Regulation of *Dictyostelium* prestalk-specific gene expres-

- sion by a SHAQKY family MYB transcription factor. *Development* **133**:1715–24.
- [19] **Gathiram, V., and T. F. Jackson.** 1987. A longitudinal study of asymptomatic carriers of pathogenic zymodemes of *Entamoeba histolytica*. *S. Afr. Med. J.* **72**:669–72.
- [20] **Gilchrist, C. A., C. F. Holm, M. A. Hughes, J. M. Schaeffer, B. J. Mann, and W. A. Petri.** 2001. Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J. Biol. Chem.* **276**:11838–11843.
- [21] **Gomez, E., J. Royo, Y. Guo, R. Thompson, and G. Hueros.** 2002. Establishment of cereal endosperm expression domains: identification and properties of a maize transfer cell-specific transcription factor, ZmMRP-1. *Plant Cell* **14**.
- [22] **Hahn, S.** 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**:394–402.
- [23] **Haque, R., C. D. Huston, M. Hughes, E. Houpt, and W. A. Petri.** 2003. Amebiasis. *N. Engl. J. Med.* **348**:1565–1573.
- [24] **Inderjit Singh, S. E. N. N., Robert B. Meeley, and O.-A. Olsen.** 2003. *Zea mays* ZmMybst1 cDNA, encodes a single Myb-repeat protein with the VASHAQKYF motif. *J. Exp. Bot.* **54**:1117–1119.
- [25] **International Arabidopsis Genome Initiative, T. A.** 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796–815.
- [26] **Iyer, L. M., V. Anantharaman, M. Y. Wolf, and L. Aravind.** 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* **38**:1–31.
- [27] **J. Frampton, T. J. Gibson, S. N. N., G. Döderlein, and T. Graf.** 1991. Proposed structure for the DNA-binding domain of the Myb oncoprotein based on model building and mutational analysis. *Protein Eng.* **4**:891–901.

- [28] **Jiang, C., J. Gu, S. Chopra, X. Gu, and T. Peterson.** 2004. Ordered origin of the typical two- and tree-repeat Myb genes. *Gene* **326**:13–22.
- [29] **Jin, H., and C. Martin.** 1999. Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* **41**:577–85.
- [30] **Klempnauer, K. H., H. Arnold, and H. Biedenkapp.** 1989. Activation of transcription by v-myb: evidence for two different mechanisms. *Genes Dev.* **3**:1582–9.
- [31] **Loftus, B., I. Anderson, R. Davies, U. C. M. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Ńozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabbinowitsch, H. Ńorbertczak, C. Price, Z. Wang, N. Guillén, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser, and N. Hall.** 2005. The genome of the protist parasite *entamoeba histolytica*. *Nature* **433**:865–8.
- [32] **Lu, C.-A., T. hua David Ho, S.-L. Ho, and S.-M. Yu.** 2002. Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of alpha-amylase gene expression. *Plant Cell* **14**:1963–80.
- [33] **Luna-Arias, J. P., R. Hernandez-Rivas, G. de Dios-Bravo, J. Garcia, L. Mendoza, and E. Orozco.** 1999. The TATA-box binding protein of *Entamoeba histolytica*: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy. *Microbiology* **145 ( Pt 1)**:33–40.
- [34] **MacFarlane, R. C., P. H. Shah, and U. Singh.** 2005. Transcriptional profiling of *Entamoeba histolytica* trophozoites. *Int. J. Parasitol.* **35**:533–42.

- [35] Mendoza, L., E. Orozco, M. A. Rodríguez, G. García-Rivera, T. Sánchez, E. García, and P. Gariglio. 2003. Ehp53, an *Entamoeba histolytica* protein, ancestor of the mammalian tumour suppressor p53. *Microbiology* 149:885–93.
- [36] Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Marechal-Drouard, W. F. Marshall, L.-H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.-L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernandez, H. Fukuzawa, D. Gonzalez-Ballester, D. Gonzalez-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J. V. Moroney, J. Moseley, C.Ñapoli, A. M. Nedelcu, K.Ñiyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J.-P. Ral, D. M. Riano-Pachon, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.-J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V.Ñ. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. Luo, D. Martinez, W. C. A. Ngau, B. Otilar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I. V. Grigoriev, D. S. Rokhsar, and A. R. Grossman. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318:245–250.
- [37] Ogata, K., S. Morikawa, H.Ñakamura, A. Sekikawa, T. Inoue,

- H. Kanai, A. Sarai, S. Ishii, and Y. Nishimura.** 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79**:639–48.
- [38] **Oh, I. H., and E. P. Reddy.** 1999. The myb gene family in cell growth, differentiation and apoptosis. *Oncogene* **18**:3017–3033.
- [39] **Ong, S.-J., H.-M. Hsu, H.-W. Liu, C.-H. Che, and J.-H. Tai.** 2007. Activation of multifarious transcription of an adhesion protein ap65-1 gene by a novel myb2 protein in the protozoan parasite *Trichomonas vaginalis*. *J. Biol. Chem.* **282**:6716–6725.
- [40] **Oppenheimer, D. G., P. L. Herman, S. Sivakumaran, J. Esch, and M. D. Marks.** 1991. A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* **67**:483–93.
- [41] **Palomo, A. M., and M. Espinosa-Cantellano.** 1998. Amoebiasis: New understanding and new goals. *Parasitol. Today (Regul. Ed.)* **14**:1–3.
- [42] **Purdy, J. E., L. T. Pho, B. J. Mann, and W. A. Petri.** 1996. Upstream regulatory elements controlling expression of the *entamoeba histolytica* lectin. *Mol. Biochem. Parasitol.* **78**:91–103.
- [43] **Ramsay, R. G., and T. J. Gonda.** 2008. Myb function in normal and cancer cells. *Nat. Rev. Cancer* **8**:523–34.
- [44] **Romero-Díaz, M., C. Gómez, I. López-Reyes, M. B. Martínez, E. Orozco, and M. A. Rodríguez.** 2007. Structural and functional analysis of the *entamoeba histolytica* EhrabB gene promoter. *BMC Mol. Biol.* **8**:82.
- [45] **Rose, A., I. Meier, and U. Wienand.** 1999. The tomato i-box binding factor LeMYBI is a member of a novel class of myb-like proteins. *Plant J.* **20**:641–52.
- [46] **Rubio, V., F. Linhares, R. Solano, A. C. Martín, J. Iglesias, A. Leyva, and J. Paz-Ares.** 2001. A conserved myb transcription factor involved

in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev.* **15**:2122–33.

- [47] **Saikumar, P., R. Murali, and E. P. Reddy.** 1990. Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. *Proc. Natl. Acad. Sci. U.S.A.* **87**:8452–6.
- [48] **Schaenman, J. M., C. A. Gilchrist, B. J. Mann, and W. A. Petri.** 2001. Identification of two *entamoeba histolytica* sequence-specific URE4 enhancer-binding proteins with homology to the rna-binding motif rrm. *J. Biol. Chem.* **276**:1602–9.
- [49] **Song, J., Q. Xu, R. Olsen, W. F. Loomis, G. Shaulsky, A. Kuspa, and R. Suggang.** 2005. Comparing the dictyostelium and *entamoeba* genomes reveals an ancient split in the conosa lineage. *PLoS Comput. Biol.* **1**:e71.
- [50] **Stracke, R., M. Werber, and B. Weisshaar.** 2001. The R2R3-MYB gene family in *arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**:447–56.
- [51] **Tie Yang, R. P., and A. Baroin-Tourancheau.** 2003. Myb genes in ciliates: A common origin with the myb protooncogene? *Protist* **154**:229–238.
- [52] **Watson, J. D.** 2004. *Molecular Biology of the Gen.* Jim Smith, 5th ed.
- [53] **Weber, C., G. Guigon, C. Bouchier, L. Frangeul, S. Moreira, O. Sismeiro, C. Gouyette, D. Mirelman, J. Y. Coppee, and N. Guillén.** 2006. Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *entamoeba histolytica*. *Eukaryotic Cell* **5**:871–875.
- [54] **World Health Organization.** 1997. WHO/PAHO/UNESCO report, a consultation with experts on amoebiasis. *Epidemiological Bulletin* **18**.

El jurado designado por la Universidad Autónoma de la Ciudad de México aprobó esta tesis el día 29 de Agosto del 2008, en la Ciudad de México D.F. para optar al Grado de Maestro en Ciencias Genómicas, a el I.B.I. Helios Cárdenas Hernández

Dra. Elisa Irene Azuara Liceaga

---

Dra. Claudia Selene Zárate Guerra

---

Dra. Esther López Bayghen

---

Dra. Minerva Camacho Nuez

---

Dr. M. César López Camarillo

---

Coordinación de Certificación y Registro

**UACM**  
Universidad Autónoma  
de la Ciudad de México

---

*Nada Humano me es ajeno*

Pedro Ahumada



## **TESIS URGENTES**

REP. DE CUBA No. 99 DESP. 23-A  
CENTRO HISTORICO MEXICO, D.F.

**TEL. 5512-8469**  
**PART. 2619-0742**

PRESUPUESTOS DE 9:30 A 7:00 P.M.  
SABADOS DE 9:30 A.M. A 3:00 P.M.