UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN

"Caracterización estructural del factor de transcripción EhYY1 en *Entamoeba histolytica*"

T E S I S

QUE PARA OBTENER EL TITULO DE:

LICENCIADO EN BIOQUÍMICA DIAGNÓSTICA

P R E S E N T A

ALAN YEUDIEL HERNANDEZ HERNANDEZ

ASESORA: DRA. MARÍA DEL CONSUELO GÓMEZ GARCÍA

COASESORA: DRA. MARITERE DOMÍNGUEZ ROJAS

CUAUTITLÁN IZCALLI, ESTADO DE MÉXICO, 2024

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN

SECRETARIA GENERAL

DEPARTAMENTO DE TITULACIÓN

VNIVERSIDAD NACIONAL AVFNºMA DE MEXICO

UNAM CUAUTITLÁN

PROTESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado: "Caracterización estructural del factor de transcripción EhYY1 en Entamoeba histolytica "

_____, que presenté para obtener el título/grado de: Licenciado en Bioquímica Diagnóstica, es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Entidad Académica, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de titulación/graduación.

Atentamente

Alan Yeudiel Hernandez Hernandez N°418067457

Nombre y número de cuenta del egresado(a)

VNIVERSIDAD NACIONAL AVTONOMA DE MEXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN

UNAM CUAUTITLÁN DEPARTAMENTO DE TITULACIÓN

ASUNTO: VOTO APROBATORIO

DR. DAVID QUINTANAR GUERRERO
DIRECTOR DE LA FES CUAUTITLAN
PRESENTE

ATN: DRA. MARÍA DEL CARMEN VALDERRAMA BRAVO
Jefa del Departamento de Titulación
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de tesis y examen profesional**

Caracterización estructural del factor de transcripción EhYY1 en Entamoeba histolytica.

Que presenta el pasante: **Alan Yeudiel Hernández Hernández**
Con número de cuenta: **418067457** para obtener el título de: **Licenciado en Bioquímica Diagnóstica**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO**.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 18 de Junio de 2024.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	Dra. Sandra Díaz Barriga Arceo	
VOCAL	Dra. Azucena Lee Mendoza	
SECRETARIO	Dra. Maritere Domínguez Rojas	
1er. SUPLENTE	Dr. Jorge Luis de la Rosa Arana	
2do. SUPLENTE	Dra. Sandra Socorro Martínez Robles	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional

MCVB/cga*

DEDICATORIA

A un hombre lo forman sus vivencias y las personas que lo acompañan durante su vida, por eso le dedico este trabajo, a toda mi familia: hermanas, tíos y abuelos, que aun en la distancia siempre estuvieron cerca de mí.

A mi padre Miguel Ángel Hernández Canseco y a mi madre Silvia Julia Hernández Santiago quienes siempre me han apoyado e incentivado a enfrentar retos, les agradezco profundamente su apoyo por estos años que estuve fuera de casa, sin ustedes no habría podido llegar hasta aquí. Gracias por creer en mí, por sacrificarse y por ser unos padres maravillosos. Ustedes me han dado la fuerza para superar todo lo que se ha cruzado en mi camino y me han moldeado para ser una mejor persona.

A mis amigos Rafael, José Luis, Mario y Sabina que siempre estuvieron conmigo cuando tenía ganas de renunciar a todo, gracias por sus palabras de aliento en los momentos más difíciles, durante estos años compartieron mis lágrimas para luego arrancarme sonrisas.

A mi abuelo Alfonso Marcial Hernández Villavicencio que tristemente falleció antes de poder verme graduado, sé que a él le hubiera hecho muy feliz el poder verme con un título en la mano.

"Si he logrado ver más lejos ha sido porque he subido a hombros de gigantes"

- *Isaac Newton*

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México, en especial a la Facultad de Estudios Superiores Cuautitlán, mi agradecimiento por brindarme una educación de calidad y de manera gratuita. Durante cuatro años, sus instalaciones fueron testigos de etapas significativas de mi vida, donde desarrollé mi pensamiento crítico y adquirí los conocimientos necesarios para mi formación como profesional.

A mi asesora, la Dra. María del Consuelo Gómez García, mi más sincero agradecimiento por compartir conmigo sus conocimientos. Gracias por brindarme una nueva perspectiva hacia el mundo científico, por creer en mí y por permitirme participar en congresos y formar parte del Laboratorio de Biología Molecular I del Instituto Politécnico Nacional en la Escuela Nacional de Medicina y Homeopatía.

A mi coasesora, la Dra. Maritere Domínguez Rojas, mi gratitud por el tiempo dedicado a este trabajo. Agradezco enormemente su apoyo con la documentación, sus revisiones y valiosas sugerencias, las cuales enriquecieron significativamente este proyecto de investigación.

A la Dra. Olivia, agradezco su amistad y generosidad al compartir conmigo su tiempo y conocimientos. Su paciencia y apoyo han sido muy importantes, sus enseñanzas serán reflejadas en futuras continuaciones de este trabajo.

A mi amigo Diego Barranco, por su invaluable colaboración en la elaboración de imágenes, su ayuda ha sido fundamental en la presentación visual de este proyecto.

ÍNDICE

CONTENIDO

LISTA DE ABREVIATURAS ..... *I*

LISTA DE FIGURAS ..... *II*

LISTA DE TABLAS ..... *III*

RESUMEN *1*

ABSTRACT *2*

I INTRODUCCIÓN *3*

1.1 Generalidades 3

Reservorio 3

Hospedadores 3

Dosis infectiva mínima (DIM) 3

Morfología 3

Supervivencia ambiental 4

Mecanismo de propagación y transmisión 4

1.2 Taxonomía 4

1.3 Ciclo de Vida 5

1.4 Epidemiología 6

1.5 Patogenia y manifestaciones clínicas 7

1.6 Tratamiento 8

II ANTECEDENTES *9*

2.1 Genoma de *Entamoeba histolytica* 11

2.2 Transcripción 13

2.3 Factores de transcripción 14

2.4 Promotores 16

2.5 Promotor del gen *EhPgp5* 17

2.6 Secuencias consenso 19

2.7 Factor YY1 20

2.7.1 Función 20

2.7.2 Estructura 22

III JUSTIFICACIÓN *25*

IV OBJETIVOS *26*
4.1 Objetivo general 26
4.2 Objetivos particulares 26
V MATERIALES Y MÉTODOS *27*
5.1 Obtención de la secuencia EhYY1 27
5.2 Comparación de EhYY1 con otros organismos 27
5.3 Comparación de EhYY1 con otras especies de *Entamoeba* 28
5.4 Análisis filogenético 28
5.5 Modelado de la proteína 28
5.6 Análisis estructural de EhYY1 29
5.7 Obtención de secuencia consenso de unión a DNA 29
5.8 Docking molecular Proteína-DNA 29
5.9 Mutaciones a la secuencia de unión a DNA. 30
VI RESULTADOS *31*
6.1 Obtención de la secuencia EhYY1 31
6.2 Comparación de EhYY1 con otros organismos 35
6.2.1 Posiciones y número de dominios dedos de zinc de diferentes especies 40
6.3 Comparación de EhYY1 con otras especies de *Entamoeba* 43
6.4 Análisis filogenético 48
6.5 Modelado de la proteína EhYY1 50
6.6 Análisis estructural de EhYY1 55
6.7 Obtención de secuencia consenso de unión a DNA 63
6.8 Docking molecular Proteína-DNA 64
6.9 Mutaciones a la secuencia de unión a DNA 72
VII DISCUSIÓN *77*
VIII CONCLUSIONES *84*
IX PERSPECTIVAS *85*
X APÉNDICE *86*
XI BIBLIOGRAFÍA *89*

LISTA DE ABREVIATURAS

aa	Aminoácido
DNA	Ácido desoxirribonucleico
EhPgp5	Gen de la P-glicoproteína 5 de *Entamoeba histolytica*
EhPGP5	Proteína P-glicoproteína 5 de *Entamoeba histolytica*
Ehyy1	Gen Yin-Yang 1 de *Entamoeba histolytica*
EhYY1	Proteína Yin-Yang 1 de *Entamoeba histolytica*
***EhPgp5*-YY1S**	Secuencia YY1 del promotor *EhPgp5* sentido
***EhPgp5*-YY1As**	Secuencia YY1 del promotor *EhPgp5* anti-sentido
HSE	Elemento de respuesta a choque térmico
HSTF	Factores de transcripción de choque térmico
HYY1	Yin-Yang 1 de Humano
HYY1-CS	YY1 de Humano secuencia consenso sentido
HYY1-CAs	YY1 de Humano secuencia consenso anti-sentido
kDa	Kilodaltons
MDR	Resistencia a múltiples fármacos
pb	Pares de bases
RNA	Ácido ribonucleico
YY1	Yin-Yang 1
ZFP	Proteínas de dedos de zinc

LISTA DE FIGURAS

Figura 1. Ciclo biológico de *E. histolytica* 5
Figura 2. Casos nuevos de amibiasis intestinal a nivel nacional del año 2012-2023 6
Figura 3. Descripción general del genoma de *E. histolytica* HM-1:IMSS Clona 6 12
Figura 4. Esquema del promotor *E. histolytica* 16
Figura 5. Mecanismo de resistencia a fármacos mediante *EhPgp5* 17
Figura 6. Región promotora del gen *EhPgp5* 17
Figura 7. Actividad de las secuencias consenso HSE y YY1 en el promtor del gen *EhPgp5* 18
Figura 8. Lista de secuencias que fueron reconocidas por YY1. 19
Figura 9. Representación esquemática de la proteína YY1 de humano. 22
Figura 10. Dedo de zinc C2H2 23
Figura 11. Diagrama metodológico general. 27
Figura 12.Alineamiento múltiple de secuencias YY1 de diferentes organismos 32
Figura 13. Alineamiento de secuencias proteica EhYY1 33
Figura 14. Alineamiento múltiple de secuencias YY1 de diferentes organismos contra EhYY1 36
Figura 15. Alineamiento de las secuencias de aminoácidos correspondientes a los genes YY1 de diferentes especies de Entamoeba con la secuencia de EhYY1. 44
Figura 16. Porcentaje de abundancia de aminoácidos 47
Figura 17. Árbol filogenético de secuencias YY1 y proteínas con dedos de zinc (ZFP) 48
Figura 18. Modelo 3D de proteínas EhYY1 50
Figura 19. Prueba lDTT para EhYY1 51
Figura 20. Dominio de unión al DNA de EhYY1. 52
Figura 21. Diagrama de Ramachandran EhYY1. 53
Figura 22. Z-Score de EhYY1 54
Figura 23. Estructura secundaria EhYY1 55
Figura 24. Ubicación de procesos postraduccionales en EhYY1 62
Figura 25. Alineamientos secuencia consenso de unión a DNA para YY1. 63
Figura 26. Logotipo de secuencia consenso YY1. 63
Figura 27. Modelo de la secuencia de unión a DNA 64
Figura 28. Docking molecular EhYY1- *EhPgp5* 65
Figura 29. Docking molecular EhYY1- DNA humano. 67
Figura 30. Docking molecular HYY1- Secuencia del promotor *EhPgp5* 69
Figura 31.Comparación estructural EhYY1 y HYY1 70
Figura 32. Superposición HYY1 y su DNA con EhYY1 71
Figura 33. Nucleótidos cruciales en la secuencia consenso de unión a DNA de EhYY1. 72
Figura 34. Acoplamiento molecular entre mutaciones a la secuencia de unión a DNA y EhYY1.. . 73
Figura 35. Acoplamiento molecular entre mutaciones a la secuencia de unión a DNA y HYY1. 75

LISTA DE TABLAS

Tabla 1.Factores de transcripción reportados para E. histolytica. 14
Tabla 2. Promotores reportados para E. histolytica 16
Tabla 3.Características YY1. 24
Tabla 4. Porcentaje de identidad y homología YY1. 37
Tabla 5.Características fisicoquímicas YY1. 38
Tabla 6. Posición de dedos de zinc ubicados en el extremo carboxilo terminal. 40
Tabla 7.Secuencias YY1 encontradas en otras especies de Entamoebas. 43
Tabla 8.Porcentaje de identidad y homología EhYY1. 45
Tabla 9.Características fisicoquímicas EhYY1. 46
Tabla 10. Resumen de la estructura secundaria EhYY1.. 57
Tabla 11. Dedos de zinc EhYY1.. 58
Tabla 12. Sitios de fosforilación.. 59
Tabla 13. Sitios de SUMOilación EhYY1. 61
Tabla 14.Sitios de acetilación para EhYY1. 62
Tabla 15. Interacciones de EhYY1 con el promotor de EhPgp5.. 66
Tabla 16. Interacciones de EhYY1 con el promotor de EhPgp5.. 68
Tabla 17. Interacciones de HYY1- Secuencia del promotor EhPgp5.. 70

RESUMEN

Entamoeba histolytica es el agente causal de la disentería amebiana, a nivel mundial cada año se reportan 50 millones de personas infectadas ocasionando una mortalidad que oscila entre 40,000 y 110,000 casos por año, además se han identificado fenotipos de resistencia a múltiples fármacos (MDR) como emetina y metronidazol representando un problema de salud pública.

Uno de los genes implicados en el fenotipo MDR es el *EhPgp5* cuya sobreexpresión resulta en un aumento de la P-glicoproteína *EhPGP5*, las drogas que ingresan al interior de la célula son expulsadas a través del poro que forma esta proteína. En el promotor de este gen (- 170 hasta - 136 pb) se identificó un elemento de respuesta a choque térmico (HSE), el cual es el principal responsable de la expresión del gen *EhPgp5* cuando los trofozoítos son expuestos a emetina. Además, está presente un motivo CdxA y una posible secuencia de unión para el factor YY1, un factor asociado con la activación transcripcional del gen *EhPgp5*.

Sin embargo, en otros organismos como el humano, el factor YY1 es una proteína fundamental en el control transcripcional de genes como *hdac2*, *FKBP3* y *Myc* relacionados con en la progresión del ciclo celular y en procesos de desarrollo entre otros. Los mecanismos por los cuales se lleva a cabo la transcripción de la amiba están poco estudiados. El factor YY1, se ha caracterizado por actuar como un activador o un represor, de ahí su nombre de Yin-Yang. En *E. histolytica* se detectó la sobreexpresión del gen Ehyy1 en trofozoítos expuestos a estrés térmico y estrés por emetina sugiriendo su posible papel en este tipo de respuesta. Por tal motivo, en el presente trabajo se llevó a cabo mediante técnicas *In silico* la identificación del gen homólogo YY1 en el genoma de *E. histolytica*, generamos el modelo 3D del factor EhYY1 y de la secuencia consenso de unión al DNA para este factor localizada en el promotor del gen *EhPgp5*, realizamos ensayos de Docking molecular para caracterizar la interacción proteína-DNA y mediante mutaciones en la secuencia consenso para YY1 se identificaron los nucleótidos relevantes en la interacción DNA-proteína.

ABSTRACT

Entamoeba histolytica is the causal agent of amoebic dysentery, worldwide each year 500 million people are reported infected, causing a mortality that ranges between 40,000 and 110,000 cases per year, in addition, multidrug resistance (MDR) phenotypes have been identified to emetine and metronidazole, representing a public health problem.

One of the genes involved in the MDR phenotype is *EhPgp5*, whose overexpression results in an increase in the P-glycoprotein *EhPGP5*; drugs that enter the cell are expelled through the pore formed by this protein. In the promoter of this gene (- 170 to - 136 bp) a heat shock response element (HSE) was identified, which is mainly responsible for the expression of the *EhPgp5* gene when trophozoites are exposed to emetine. Additionally, a CdxA motif and a possible binding sequence for the YY1 factor, a factor associated with transcriptional activation of the *EhPgp5* gene, are present.

However, in other organisms such as humans, the YY factor is a fundamental protein in the transcriptional control of genes such as *hdac2*, *FKBP3* and *Myc* genes related to cell cycle progression and development processes, among others. The mechanisms by which amoeba transcription is carried out are poorly studied; however, 6 transcription factors are known to be involved in stress response gene expression. The YY1 factor has been characterized by acting as an activator or a repressor, hence its name Yin-Yang. In *E. histolytica*, overexpression of the Ehyy1 gene was detected in trophozoites exposed to heat stress and emetine stress, suggesting its possible role in this type of response. For this reason, in the present work the identification of the homologous gene YY1 in the genome of *E. histolytica* was carried out using *In silico* techniques, we generate the 3D model of the *EhYY1* factor and the consensus DNA binding sequence for this factor located in the promoter of the gene *EhPgp5*, we performed molecular docking assays to characterize the protein-DNA interaction and through mutations in the consensus sequence for YY1, the relevant nucleotides in the DNA-protein interaction were identified.

I INTRODUCCIÓN

1.1 Generalidades

Entamoeba histolytica es un protozoo que pertenece al filo Sarcomastigophora es de distribución mundial y parasita al 10% de la humanidad. Esta ameba es el agente causal de la amebiasis, amebosis o disentería amebiana en el hombre y otros mamíferos. Los estudios realizados en la última década sobre la biología y comportamiento de *E. histolytica* permiten concluir que existen varias cepas de esta especie, que, si bien son morfológicamente indistinguibles entre sí, difieren en su patogenicidad (Fundación iO, 2019).

Reservorio: Humano (intestino), suelo húmedo, agua (aguas residuales), alimentos y fómites.

Hospedadores: Humanos y otros mamíferos (primates, cánidos, felinos, porcinos, roedores, bovinos).

Dosis infectiva mínima (DIM): Se cree que la ingestión de un solo quiste puede causar la infección.

Morfología: Su ciclo de vida comprende dos estados, la forma invasiva vegetativa ameboide (trofozoíto) y la forma de resistencia e infectante (quiste).

El trofozoíto es anaerobio facultativo, con forma irregular ameboide alargada y puede medir de 10 a 60 micras (µm) de diámetro, aunque el tamaño más habitual es de 12 a 15 µm. En el citoplasma tiene un único núcleo con un cariosoma central, cromatina periférica fina distribuida regularmente y vacuolas que pueden contener los eritrocitos fagocitados del huésped. A partir del citoplasma se forman prolongaciones o pseudópodos con los que se desplaza (INSST, 2022).

Fase de prequiste. En esta fase las amibas son células incoloras redondas u ovales, más pequeñas que los trofozoítos, pero mayores que los quistes e inmóviles, aun cuando a veces emiten pseudópodos muy cortos y anchos. En su citoplasma no se encuentran ya restos de eritrocitos ni vacuolas.

Fase de quiste. Los quistes son esféricos ligeramente asimétricos, hialinos con una membrana lisa aparente, aunque un poco delgada y que no se tiñe; su diámetro varía entre 5 y 20 µm Cuando se trata de un quiste inmaduro sólo presenta un núcleo como sucede en el trofozoíto y cuando es un quiste maduro, por lo general tiene cuatro núcleos con características morfológicas iguales a las de las formas activas. En raras ocasiones se pueden encontrar hasta 8 núcleos. En el citoplasma tanto de los quistes inmaduros como en los maduros generalmente hay dos tipos de inclusiones alimenticias una masa de glucógeno de bordes borrosos y las barras cromatoidales que se presentan como varillas refringentes con bordes redondeados.

Fase de metaquiste. En esta fase la amiba tiene forma semejante a la de los trofozoítos, aunque con un tamaño mucho menor (Martínez, 2019).

Supervivencia ambiental: Los quistes pueden sobrevivir en el agua, en el suelo húmedo y en las heces un tiempo variable en función de la temperatura ambiental. Sin embargo, los quistes son sensibles a la desecación por lo que no sobreviven mucho tiempo en suelos o superficies secas. Los trofozoítos apenas sobreviven en el ambiente exterior.

Mecanismo de propagación y transmisión: La transmisión se produce principalmente por la ingesta accidental de los quistes maduros presentes en el agua y los alimentos contaminados (ingesta de verduras o frutas crudas y mal lavadas). Puede transmitirse de persona a persona por vía fecal-oral, a través del contacto con manos, objetos (monedas, billetes, juguetes, etc.) o superficies contaminadas con heces infectadas, pues los quistes salen con las heces de la persona infectada y pueden permanecer viables durante unos 10 minutos en la piel seca y unos 45 minutos debajo de las uñas. También puede transmitirse por contacto sexual, sexo oral-anal (INSST, 2022) (OPS, 2003).

1.2 Taxonomía

Cavalier-Smith et al. en 2004 describió una de las clasificaciones taxonómicas más aceptadas para *E. histolytica*, que es la siguiente:

- Reino: Protozoa
- Subreino: Neozoa
- Superclase: Sarcomastigota
- Phylum: Ameobozoa
- Subphylum: Conosa
- Género: *Entamoeba*
- Especie: *histolytica*

1.3 Ciclo de Vida

Su ciclo de vida comprende dos estadios: la forma invasiva vegetativa ameboide (trofozoíto) y la forma de resistencia e infectante (quiste), como se explica en la figura 1.
Su ciclo de vida es directo (un solo hospedador). Cuando los quistes maduros son ingeridos por un hospedador, estos se desenquistan en el intestino delgado dando lugar a los trofozoítos. Los trofozoítos se multiplican por fisión binaria y se desplazan hacia el intestino grueso; a medida que avanzan hacia el exterior dejan de alimentarse y se rodean de una pared resistente transformándose así en quistes. Tanto los quistes como los trofozoítos son eliminados en las heces del hospedador. Una vez en el exterior, los trofozoítos apenas sobreviven, y, aunque sean rápidamente ingeridos por un hospedador no son capaces de sobrevivir a la acción de los jugos gástricos. Sin embargo, los quistes sobreviven en el exterior desde horas hasta meses en función de las condiciones ambientales, pero solo los quistes maduros son infecciosos (INSST, 2022) (CDC, 2020).

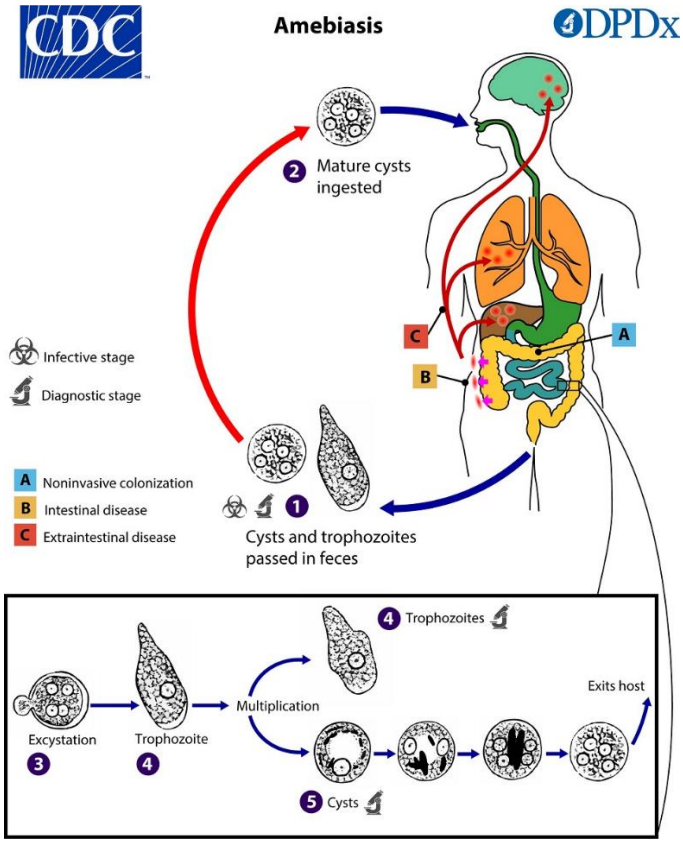


Figura 1. Ciclo biológico de E. histolytica. (1) La infección por Entamoeba histolytica (y E. dispar) se produce a través de la ingestión de quistes maduros (2) de alimentos, agua o manos contaminados con heces. También puede ocurrir exposición a quistes infecciosos y trofozoítos en la materia fecal durante el contacto sexual. El desenquistamiento (3) se produce en el intestino delgado y los trofozoítos (4) se liberan, que migran al intestino grueso. Los trofozoítos pueden permanecer confinados a la luz intestinal (A: infección no invasiva) y los individuos continúan eliminando quistes en las heces (portadores asintomáticos). Los trofozoítos pueden invadir la mucosa intestinal (B: enfermedad intestinal) o los vasos sanguíneos, alcanzando sitios extraintestinales como el hígado, el cerebro y los pulmones (C: enfermedad extraintestinal). Los trofozoítos se multiplican por fisión binaria y producen quistes (5). (Center for Disease Control, 2019)

1.4 Epidemiología

La importancia mundial de la amebiasis es generalizada, y la mayor carga de amebiasis recae en quienes residen en los países en desarrollo, en particular en los trópicos y subtrópicos, donde la higiene y el acceso al saneamiento son inadecuados. Millones de personas están infectadas con *E. histolytica*, lo que convierte a la colitis amebiana en una de las principales causas de diarrea, que se estima que mata a más de 55 000 personas cada año (Shirley, 2019).

En México a pesar de que las enfermedades diarreicas han ido a la baja y que esta patología ha presentado una disminución en el número de casos e incidencia desde 2020, continúa manteniéndose entre las primeras veinte causas de morbilidad.

La Secretaría de Salud registró una marcada disminución en la tasa de incidencia de esta enfermedad a nivel nacional entre los años 2012 y 2023 (Figura 2). Se pasó de 374,055 casos a 118,357 nuevos casos, lo que representa una reducción a aproximadamente un tercio de los casos.

De acuerdo con la distribución geográfica de la amebiasis en el país, los estados del sur y sureste son los que presentan la mayor incidencia. El estado de Tabasco presentó la incidencia más elevada seguida de Oaxaca, Campeche y Guerrero. Estos cuatro estados en conjunto constituyen el 22.12% de la totalidad de los casos presentados durante este periodo. A pesar de que Veracruz, Puebla y Colima se encuentran agrupados dentro de los estados con menor número de casos estos presentaron una incidencia por arriba del promedio nacional y los estados con menor incidencia fueron: Nuevo León, Sonora y Baja California (Secretaria de Salud, 2023)

Figura 2. Casos nuevos de amibiasis intestinal a nivel nacional del año 2012-2023. Horizontalmente se muestran los años 2012-2020, verticalmente el número de nuevos casos de amibiasis. (Secretaria de Salud, 2023)

1.5 Patogenia y manifestaciones clínicas

Amebiasis, o disentería amebiana, es un término que se utiliza para describir una infección causada por el protozoo *E. histolytica*. La mayoría de las infecciones son asintomáticas, pero puede producirse una enfermedad intestinal invasiva que se manifiesta con varias semanas de calambres, dolor abdominal, diarrea acuosa o con sangre y pérdida de peso. Se ha descrito una enfermedad diseminada extraintestinal como absceso hepático, neumonía, pericarditis purulenta e incluso amebiasis cerebral.

Aproximadamente el noventa por ciento de las infecciones por *Entamoeba* son asintomáticas. Los factores de riesgo que se asocian con una mayor gravedad y mortalidad de la enfermedad incluyen la edad joven, el embarazo, la malignidad, la desnutrición, el alcoholismo y el uso de corticosteroides (WHO, 2003).

Se han postulado varias hipótesis para explicar por qué ciertos pacientes tienen enfermedad asintomática, mientras que otros progresan a enfermedad invasiva y actualmente se están probando. Se ha descubierto que la virulencia de la cepa, el entorno y la susceptibilidad genética del hospedador, el estado inmunitario, la edad y el sexo predicen la gravedad de la enfermedad (Manual MSD, 2020).

La presencia de ciertas enzimas se ha relacionado con un mayor riesgo de enfermedad invasiva. Por ejemplo, se necesitan glicosidasas como la sialidasa, N-acetilgalactosamidasa y N-acetilglucosaminidasa para eliminar los polisacáridos ramificados de las células de mucina. Esto permite que los trofozoítos degraden la barrera protectora de la mucosa y posteriormente penetren en el epitelio colónico aumentando el riesgo de metástasis a sitios distantes. Además, se ha demostrado la presencia de la enzima glucósido hidrolasa B-amilasa al analizar el transcriptoma *E. histolytica* virulenta. Las especies que carecen de esta enzima no pudieron atravesar la barrera del moco y provocar una enfermedad invasiva. Otros mecanismos implicados en la destrucción de las células epiteliales y las células inflamatorias son, la secreción de proteínasas, la lisis de las células diana dependientes del contacto, la apoptosis y la formación de ameboporos que provocan la citólisis de las células infectadas (Kantor, 2018).

1.6 Tratamiento

El diagnóstico y el tratamiento específicos están justificados en todas las infecciones causadas por *E. histolytica*, incluso en portadores asintomáticos, no solo por el potencial de desarrollar una enfermedad invasiva, sino también para disminuir la propagación de la enfermedad. La amebiasis se controla principalmente mediante el tratamiento con derivados del metronidazol y emetina, aunque los nitroimidazoles con vidas medias más prolongadas, como tinidazol, secnidazol y ornidazol, permiten períodos de tratamiento más cortos y tienden a tolerarse mejor (Kantor, 2018).

Para los síntomas gastrointestinales y la amebiasis extraintestinal, se usa uno de los siguientes medicamentos:

Metronidazol por vía oral 500 a 750 mg 3 veces al día (entre 12 y 17 mg/kg 3 veces al día en los niños) durante 7 a 10 días.

Emetina en adultos se usa 1 mg por kg de peso corporal al día, administrada durante 5 a 10 días. Se usa cuando otros tratamientos no están disponibles o son ineficaces. Es crucial ajustar la dosis según la respuesta del paciente y la presencia de efectos secundarios, puede causar arritmias, insuficiencia cardíaca congestiva y miocarditis (Fonte, 2001).

Tinidazol 2 g por vía oral 1 vez al día (50 mg/kg [máximo 2 g] por vía oral 1 vez al día en los niños > 3 años) durante 3 días para los síntomas gastrointestinales leves a moderados, 5 días para los síntomas gastrointestinales graves y 3 a 5 días para el absceso amebiano hepático (Manual MSD, 2020).

Han aumentado los informes de casos de tratamientos farmacológicos fallidos, lo que sugiere que la resistencia a los fármacos puede ocurrir en parásitos dentro del hospedador humano, resistencia a los medicamentos era poco común en la amebiasis. Sin embargo, en el laboratorio se han generado cepas resistentes a los medicamentos metronidazol y emetina tras la exposición continua a concentraciones del fármaco en aumento de concentraciones (Samarawickrema, 1997).

II ANTECEDENTES

Los orígenes de la infección causada por *E. histolytica* probablemente se remontan a la antigüedad cuando la especie humana se separó de sus ancestros inmediatos en el curso de la evolución. La literatura médica antigua destaca la fiebre y la disentería como dos enfermedades prominentes. Documentos antiguos hacen referencia a los síntomas clásicos de la disentería amebiana (Espinosa-Cantellano, M; Palomo, A. Marinez, 1987). A partir del siglo V a. C., diversos compendios hipocráticos permiten identificar el término "disentería" para describir esta afección, así como las primeras descripciones de enfermedades intestinales ulcerosas (460-337 a. C.).

Antes de la llegada de los españoles a México en 1519, las tribus indígenas trataban la amebiasis utilizando ipecacuana (*Carapichea ipecacuanha*), una planta que contiene el alcaloide emetina (Arellano, 2017).

En 1755, un científico llamado Rösel Van Rosenhof realizó una investigación de muestras que contenían células que mostraban una asombrosa habilidad para cambiar de forma mientras se movían. Estas células fueron inicialmente llamadas "pequeños Proteos" por Van Rosenhof, pero más tarde recibieron el nombre de "amibas", derivado del término griego "amoibe", que significa "cambio" (Cabello, 2008).

En 1870 Lewis y en 1871 Cunningham realizaron el descubrimiento y aislamiento de amebas a partir de muestras fecales de enfermos de cólera, estableciendo un vínculo entre ellas. la presencia de este microorganismo y la enfermedad.

En 1883, Koch identificó cinco casos de disentería en Egipto, de los cuales dos desarrollaron abscesos hepáticos amebianos con múltiples úlceras infectadas que contenían amebas tanto en las lesiones intestinales como en los abscesos hepáticos (Espinosa-Cantellano, 1987).

En 1893, Quincke y Ros propusieron la existencia de diferentes especies de amebas localizadas en el tubo digestivo humano y dilucidaron el ciclo biológico de *E. histolytica* al descubrir la forma quística, sugiriendo su implicación en la propagación de la enfermedad y, por tanto, como la forma causal de infección (Espinosa-Cantellano, 1987).

En 1925, Emilie Brumpt propuso la existencia de dos especies distintas de amebas: una capaz de causar disentería y enfermedades invasivas, y otra que no causa ninguna patología en humanos. Esta segunda especie fue nombrada *Entamoeba dispar*.

En México el Dr. Bernardo Sepúlveda comenzó a promover el estudio de la amibiasis en 1936, coordinando la publicación de un número especial de la revista del Centro de Asistencia Médica para Enfermos Pobres dedicado a la amibiasis e impulsando numerosos seminarios internacionales sobre la enfermedad (Arellano, 2017).

En 1967, se aisló la primera ameba en forma axénica del intestino de un paciente hospitalizado en el Centro Médico Nacional IMSS, denominándola HM-1:IMSS. Actualmente, esta cepa es ampliamente utilizada en laboratorios de investigación en todo el mundo (Olivos-García, 2011).

En 1997, un comité de expertos de la Organización Mundial de la Salud (OMS) aceptó esta hipótesis. Estos descubrimientos han impulsado en la última década el desarrollo de pruebas para la detección de antígenos de *E. histolytica* mediante anticuerpos monoclonales dirigidos a las lectinas Gal/GalNAc, así como la técnica de reacción en cadena de la polimerasa (PCR), que ayudan a diferenciar entre ambas especies (Velázquez, 2008).

En 2023, el Sistema Nacional de Vigilancia Epidemiológica (SINAVE) de la Secretaría de Salud reportó más de medio 100,000 de casos de amibiasis intestinal en México, reflejando la persistencia de la enfermedad en el país.

2.1 Genoma de *Entamoeba histolytica*

La falta de condensación de los cromosomas en *E. histolytica* ha presentado desafíos para determinar con precisión el número de cromosomas. Sin embargo, se ha informado la presencia de 14 grupos cromosómicos relacionados con una ploidía de cuatro, cuyos tamaños oscilan entre 0.3 y 2.2 Mb. En 2005, Loftus y su equipo completaron el análisis completo del genoma de *E. histolytica*. Sin embargo, en 2010, Lorenzi y colaboradores realizaron un nuevo análisis del genoma de este parásito, revelando que el genoma es rico en adenina y timina, con una composición del 77,6%. El genoma contiene 20.799.072 pares de bases (pb) y se estima que hay 8.201 genes distribuidos en 818 andamios, con una longitud media de 1.260,9 pb por gen. Las regiones intergénicas varían en tamaño de 400 a 2300 pares de bases (Lorenzi, 2010).

Se encontró que los genes predichos contienen 10.754 exones, en su mayoría exones únicos, aunque aproximadamente el 6% de los genes pueden contener 2 o más intrones. Aproximadamente el 10% del genoma está compuesto por genes que codifican RNAt organizados en matrices en tándem, que varían en longitud de 490 a 1775 pares de bases. Además, el genoma de este parásito contiene una serie de moléculas circulares de DNA conocidas como plásmidos, uno de los cuales tiene un tamaño aproximado de 26 kb y contiene los genes que codifican los RNA mensajeros. Las proteínas previstas son de corta longitud, con una media de 389 aminoácidos (Loftus, 2005).

Sin embargo, este primer ensamblaje estaba altamente fragmentado debido a la naturaleza repetitiva del genoma, que incluye elementos nucleares intercalados cortos (SINE), elementos nucleares intercalados largos (LINE) y matrices de ARNt que contienen repeticiones cortas en tándem. En 2020, se realizó un nuevo ensamblaje del genoma de *Entamoeba histolytica HM-1 Clona 6* (Figura 3) con resolución a nivel cercano al cromosoma, utilizando las técnicas PacBio y Hi-C. El genoma resultante está compuesto de 38 andamios con una longitud total de 26,879,087 pb y contiene un total de 8,734 genes codificadores de proteínas según la predicción informática hecha por Companion (Kawano-Sugaya, 2020).

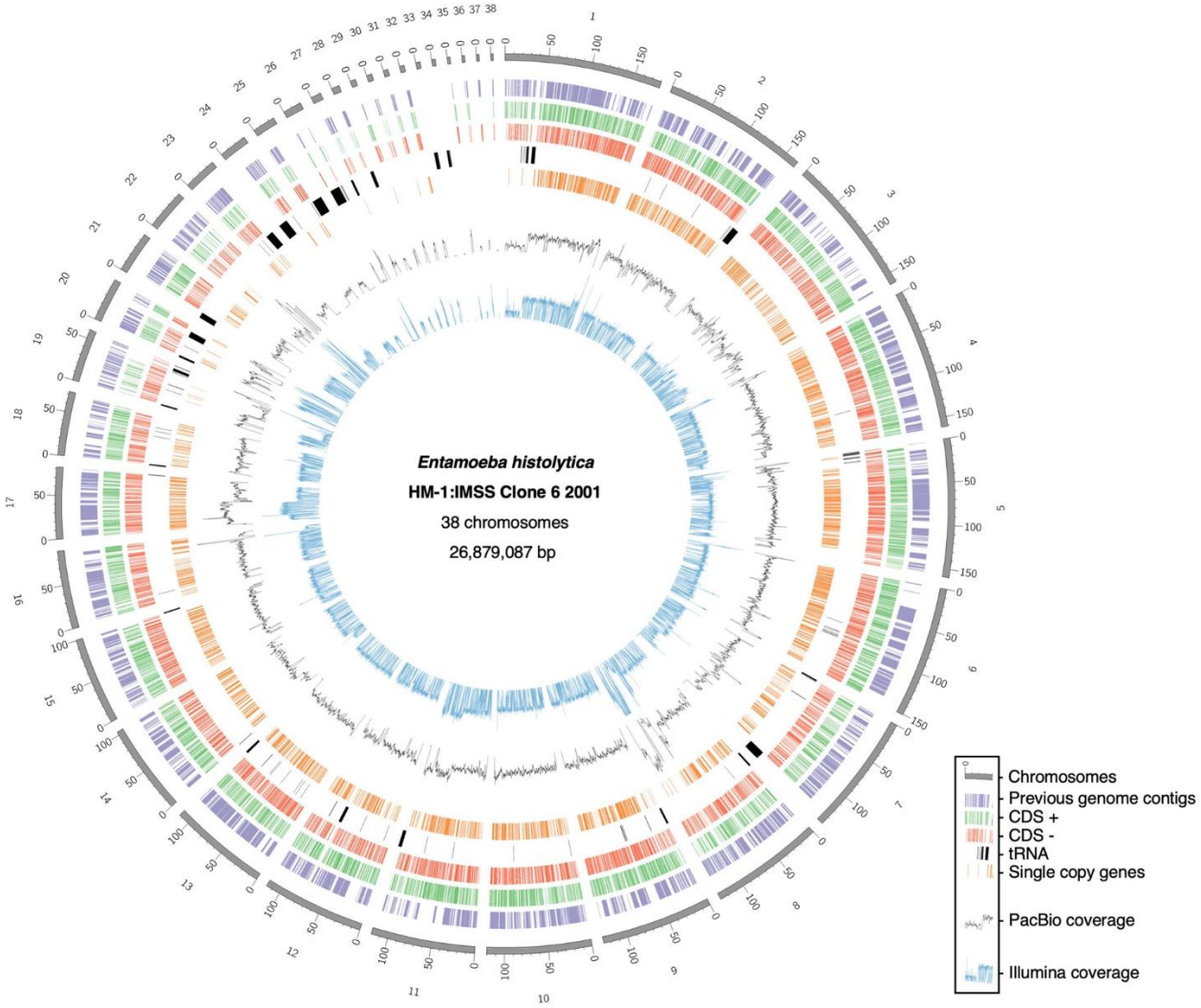


Figura 3. Descripción general del genoma de E. histolytica HM-1:IMSS Clona 6. Este diagrama de círculos muestra las estructuras generales con longitud, anotaciones y gráficos de coberturaDe afuera hacia adentro, se muestran los andamios recién ensamblados y su longitud (bloques grises; con una escala de 1:10 kb), los contigs del genoma anterior (códigos de barras violetas), secuencias codificantes + hebra (códigos de barras verdes), secuencias codificantes - hebra (códigos de barras naranjas), genes de ARNt (códigos de barras negros), genes de copia única (códigos de barras naranjas), cobertura de lectura de PacBio (una línea negra) y cobertura de lectura de Illumina (una línea azul). (Tomado de Kawano-Sugaya, 2020)

2.2 Transcripción

La transcripción es un proceso mediante el cual se produce una copia de ARN a partir de la secuencia de un gen. Esta copia, conocida como RNA mensajero (RNAm), sale del núcleo de la célula y entra en el citoplasma, donde dirige la síntesis de la proteína para la cual codifica (Hogart, 2011).

En el caso de *E. histolytica*, la transcripción de genes que codifican proteínas sigue un patrón monocistrónico, lo que significa que cada secuencia de DNA codificante contiene suficiente información para generar una única transcripción. Además, cada gen tiene su propio promotor y puede transcribirse y regularse de forma independiente (Gómez, 2010).

Se ha observado que la transcripción ocurre a través de una RNA polimerasa II con propiedades distintas a las de los eucariontes superiores (organismos pluricelulares), ya que es resistente a α-amanitina, un inhibidor de la RNA polimerasa II. Esta enzima es responsable de agregar nucleótidos a la cadena de RNA que es complementaria a las cadenas molde de DNA. Al comienzo de un gen, se encuentra una región no codificante llamada "promotor", la cual contiene secuencias específicas que permiten que la RNA polimerasa se una al DNA y dé inicio al proceso de transcripción. Es importante destacar que las RNA polimerasas de los eucariontes no pueden reconocer los promotores por sí solas, por lo que requieren la ayuda de proteínas conocidas como "factores de transcripción" (Krap, 2005).

El ensamblaje completo de los factores de transcripción y la RNA polimerasa se unen al promotor, formando un complejo de pre-iniciación de la transcripción (PIC). Este complejo incluye la RNA polimerasa II, así como los factores generales TFIIA, TFIIB, TFIID, TFIIE, TFIIF y TFIIH, junto con un mediador. En el PIC se establecen interacciones entre el DNA y factores generales, permitiendo que la RNA polimerasa II se una al promotor de doble cadena. La proteína de unión a TATA (TBP), que forma parte del complejo TFIID, se une a la caja TATA ubicada río arriba del sitio de inicio de la transcripción (TSS). Posteriormente, TBP dobla el DNA, que se estabiliza por TFIIA, TFIIB y TFIIF. Este último complejo facilita la localización del TSS por la RNA polimerasa II y estabiliza la interacción entre la RNA polimerasa II y el mediador. TFIIE interactúa con TFIIF y ayuda a estabilizar el complejo de preiniciación. Por último, TFIIE recluta TFIIH, el cual mediante su actividad ATPasa y helicasa genera tensión en el DNA, desenrollándolo y formando la burbuja de transcripción. TFIIF mantiene la burbuja abierta para que la ARN polimerasa II inicie la síntesis de RNA (Schier, 2020).

2.3 Factores de transcripción

Los factores de transcripción son proteínas que se unen a secuencias específicas del DNA llamadas "sitios de unión" o "elementos de respuesta". Son los encargados de controlar la expresión génica ya que pueden activar o reprimir la expresión de un gen al influir en la RNA Polimerasa para unirse y transcribir el DNA. Los factores de transcripción descritos hasta el momento para *E. histolytica* se muestran en la tabla 1.

Tabla 1.Factores de transcripción reportados para E. histolytica.

Factor	Función	Características
EhURE3-BP	Es un determinante en la virulencia ya que se descubrió que su expresión cambia la morfología del trofozoíto y promueve la invasión del colon y el hígado (Gilchrist, 2001).	*Proteína de 22.6 kDa *Dos motivos de EF-Hand *Actividad regulada por calcio
EhEBP 1 y 2	Contienen dos secuencias homólogas al motivo RRM (Motivo de reconocimiento de RNA) también estan presentes en proteínas de unión a DNA (Schaenman, 2001).	*Polipéptidos de 28 y 18 kDa *Reconocen la secuencia URE4
EhC/EBP	La proteína de unión a CCAAT/potenciador (EhC/EBP) regula la transcripción del gen EhPgp1, que está implicado en el fenotipo de resistencia a múltiples fármacos (MDR) (Marchat, 2003).	*Proteína de 25 kDa * Presenta un dominio bZIP altamente conservado
TBP	La proteína es capaz de interaccionar con la secuencia consenso de la caja TATA (Luna-Arias, 1999).	* Proteína de 26 kDa *Su dominio funcional muestra un 55% de identidad con TBP de *Homo sapiens*
Ehp53	Es similar a p53 supresora de tumores, se acumula en células irradiadas con luz UV (Mendoza, 2003).	*Proteína de 53 kDa *Localizada en el núcleo de la célula
EhHMGB1	Modulan la expresión de genes involucrados en la invasión y colonización del parásito (Abhyankar, 2008).	*Proteína de 13.4 kDa * Es capaz de plegar el DNA *in vitro* de forma no específica
STAT	La familia STAT son transductores de señales y activadores de transcripción (Cruz-Vera, 2003).	*Proteínas citoplasmáticas *Forman homo/hetero dímeros que son trasladados al núcleo

EhCudA	No se ha caracterizado, por lo que se desconoce qué genes o vías podría regular en *E. histolytica* (Yamada, 2008).	*Proteína de 34 kDa
EhMyb10	Se une específicamente a un elemento de reconocimiento Myb (MRE) de consenso, que se encuentra en varios promotores de *E. histolytica* de genes implicados en la transducción de señales, el transporte vesicular y la respuesta al choque térmico (Gilchrist, 2006).	*Proteína de 17.9 kDa *Presente en el promotor del gen *Ehhsp70*
EhPC4	Regula la expresión de genes implicados en el metabolismo de carbohidratos y ácidos nucleicos, la segregación cromosómica y la citocinesis (Hernández de la Cruz, 2016).	*Proteína de 20 kDa *Regula la expresión de EhABP16
EhHRM-BP	Activa la expresión de genes involucrados en la protección al estrés nitrosativo y oxidativo (Pearson, 2013).	*Proteína de 50 kDa *Su sobreexpresión incrementa la virulencia de *E. histolytica*
HSTF	Factores de Transcripción de Choque Térmico activan y se unen al elemento de choque térmico (HSE) presente en los promotores de genes de estrés en respuesta a diferentes condiciones de estrés (Rojas, 2017) (Bello et al, 2021) (Pérez-Mora, 2024).	* Una familia de siete factores de transcripción * El análisis del promotor central del gen MDR *EhPgp5* demostró la presencia de una secuencia HSE

**Se muestran tres columnas con el factor, su función y características propias del factor. (C. Gómez, datos por publicar).*

Los factores de transcripción en *E. histolytica* muestran una amplia diversidad en sus funciones, desde la regulación de la virulencia y la respuesta al estrés hasta la transducción de señales y la regulación específica de la transcripción genética. Esto refleja la necesidad del parásito de adaptar su expresión génica a una variedad de condiciones ambientales y desafíos del hospedador. La mayoría de estos factores se unen a secuencias específicas de DNA para regular la transcripción de genes. Algunos, como EhEBP 1 y 2, también tienen la capacidad de interactuar con RNA, lo que puede indicar roles adicionales en la regulación post-transcripcional. Factores como EhC/EBP y EhHRM-BP están directamente implicados en la adaptación del parásito a condiciones adversas, como la presencia de fármacos o estrés oxidativo, lo que les confiere una ventaja en la supervivencia y patogenicidad. Las proteínas tienen tamaños que oscilan entre 13.4 kDa y 53 kDa, y poseen dominios específicos que determinan su capacidad de unión y regulación del DNA. Estos dominios incluyen motivos como EF-Hand, bZIP, y secuencias RRM, que son cruciales para su función reguladora.

2.4 Promotores

Un promotor es una secuencia de DNA a la que se unen las proteínas encargadas de iniciar la transcripción, se ubica río arriba del gen que va a ser transcrito.

En los promotores de *E. histolytica* se identificaron tres elementos reguladores conservados, que conforman el "núcleo promotor" (Figura 4) el cual contiene una caja TATA atípica (TATTTAAA), un segundo elemento conocido como elemento iniciador o Inr (AAAAATTCAA) que se superpone con el sitio de inicio de la transcripción, y finalmente un tercer elemento denominado GAAC con una posición variable entre la caja TATA y la secuencia Inr y con la habilidad de determinar el sitio de inicio de la transcripción independientemente de estos dos elementos (Vanacova, 2003).

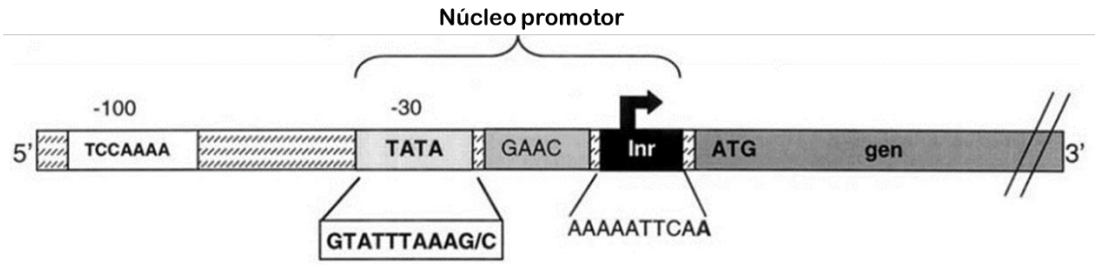


Figura 4.Esquema del promotor E. histolytica. Los números indican la posición de cada secuencia, por debajo la secuencia completa de los motivos. El Núcleo del promotor muestra a sus tres elementos conservados "TATA", "GAAC" e "Inr" con una flecha se señala el inicio de la transcripción (Modificado de Vanacova, 2003).

Son pocos los promotores que han sido descritos en *E. histolytica* en seguida se presentan en la tabla 2.

Tabla 2. Promotores reportados para E. histolytica

Promotor de genes	Función
Ehcp112 y *Ehadh112*	Codifican para el complejo EhCPADH, involucrado en mecanismos de patogenicidad.
Ehgl2 y *Ehgl5*	Codifican para la subunidad pesada de la lectina Gal/GalNac.
EhrabB	Sistema de transporte de vesículas.
Ehpgp1 y *Ehpgp5*	Codifican para proteínas de resistencia a múltiples fármacos.

**Se muestra el promotor y la función del gen. (Ramírez, 2010)*

2.5 Promotor del gen *EhPgp5*

E. histolytica presenta el fenotipo de resistencia a múltiples fármacos (MDR) el cual está dado por la sobreexpresión de los genes MDR (*EhPgp1*, *EhPgp5 y EhPgp6*) y la síntesis de las proteínas transportadoras PGP o P-glicopróteínas codificadas por éstos, las cuales transportan entre otras cosas los fármacos que ingresan a la amiba y los expulsan al exterior, evitando su efecto dañino en la amiba (Pérez, 1998). Uno de los genes implicados en el fenotipo MDR es el *EhPgp5* cuya sobreexpresión inducida por la presencia del fármaco antiamibiano emetina resulta en un aumento significativo de la P-glicoproteína volviendo resistente a este parásito, los fármacos que ingresan a la célula son expulsadas a través del poro que forma esta molécula.

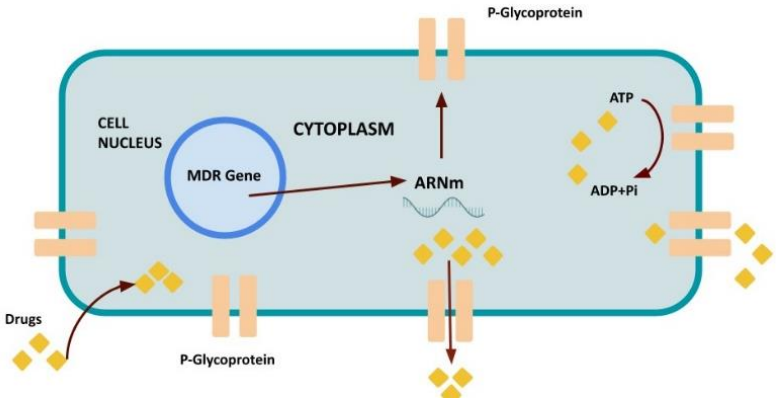


Figura 5. Mecanismo de resistencia a fármacos mediante EhPgp5. Los fármacos ingresan a E. histolytica y son expulsados por los poros formados por las P-glicopróteínas.

La región del promotor contiene una caja TATA (TATTTAAA) en la posición –31 pb y un sitio único de inicio de la transcripción (ATTCG), localizado tres nucleótidos río arriba del ATG (Pérez, 1998).

En la región del promotor de - 170 hasta - 136 pb del gen *EhPgp5* se identificó un elemento de respuesta a choque térmico (HSE) (- 151 a - 136 pb), además se reveló la presencia de probables secuencias consenso para los factores CdxA (- 170 a - 164 pb) y YY1 (-163 a -153 pb) (Figura 6) (Nieto, 2017).

Figura 6. Región promotora del gen EhPgp5. Se muestra en color verde la secuencia de unión a CdxA, rojo la posible secuencia de unión del Factor de Transcripción EhYY1 y en azul la secuencia HSE. (Modificado de Nieto, 2017)

En 2017 se analizó la funcionalidad de esta secuencia y se encontró que el elemento HSE es el principal responsable de la inducción de la expresión del gen *EhPgp5* cuando los trofozoítos son expuestos a emetina. Sin embargo, se observaron cambios en la actividad transcripcional del promotor del gen *EhPgp5* cuando se eliminó y mutó la secuencia YY1. Estos cambios se reflejaron en una disminución en la expresión del gen reportero CAT. Estos resultados sugieren que la secuencia YY1 probablemente esté involucrada en la activación del gen *EhPgp5*, aunque su contribución puede no ser tan fuerte como la aportada por la secuencia HSE. Estos hallazgos respaldan la idea de que múltiples elementos regulatorios actúan de manera coordinada para controlar la expresión del gen *EhPgp5* en *E. histolytica*, lo que contribuye a su resistencia a fármacos.

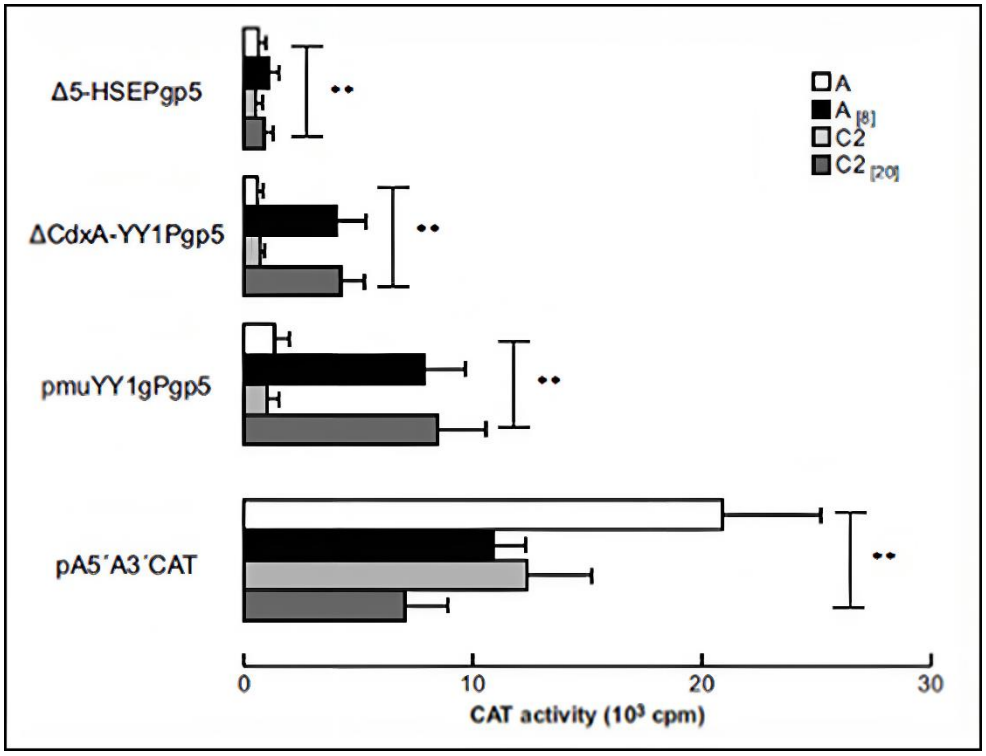


Figura 7. Actividad de las secuencias consenso HSE y YY1 en el promtor del gen EhPgp5. Las barras muestran el promedio de actividad CAT (cpm) para los plásmidos transfectados con HSE eliminado, CdxA-YY1 eliminado, YY1 mutado. La eficiencia de los ensayos de transfección fue monitoreada por las actividades dadas por el pA5 ' A3 ' Plásmido CAT. En la esquina superior derecha se muestra la simbología Clona A, Clona A 8 µM emetina, Clona C2, Clona C2 20 µM emetina. (Tomado de Nieto, 2017)

2.6 Secuencias consenso

Una secuencia consenso, representa una versión promedio de los nucleótidos más conservados, esta secuencia es el resultado del alineamiento y comparación de un grupo de múltiples secuencias que tienen la misma homología. Es decir, cada uno de los nucleótidos presentes en TATTTAAA (se sugiere que es la caja TATA de los genes amebianos) está presente en la mayoría de los promotores y casi todas las secuencias promotoras difieren de la secuencia consenso en tan solo unos nucleótidos (Berg, 2008) (Liljas, 2013).

Las secuencias consenso se localizan en las regiones promotoras de los genes principalmente. Sin embargo, algunas no solo son conservadas en secuencia, sino también en posición como lo es también la caja TATA que se ubica entre las posiciones de -25 a -35 pb (Shi, 2006)

Un ejemplo de secuencia consenso se obtuvo en 1995, cuando Hyde-DeRuyscher y colaboradores seleccionaron un conjunto de secuencias aleatorias de nucleótidos y las hicieron interactuar con la proteína YY1 humana. Recuperaron las secuencias con las cuales YY1 pudo acoplarse y generaron una lista (Figura 8). Posteriormente, realizaron un alineamiento de dichas secuencias y, con ellas, identificaron los nucleótidos más conservados para elaborar una secuencia consenso.

En 2017 Nieto y col. hallaron un secuencia de unión a DNA para el factor de transcripción YY1 en la region promotora del gen *EhPgp5*.

Figura 8. Lista de secuencias que fueron reconocidas por YY1. Las letras mayúsculas señalan la secuencia consenso conservada para YY1. Tomado de (Hyde-DeRuyscher, 1995)

2.7 Factor YY1

2.7.1 Función

YY1 también llamado DELTA, INO80 complex subunit S, INO80S, NF-E1, UCRBP, Yin and Yang 1 protein, YIN-YANG-1 (Orphanet, 2023) es un factor de transcripción de expresión ubicua que contiene dedos de zinc y que actúa tanto como represor como activador de la transcripción. Esta proteína interacciona con proteínas reguladoras clave tales como la Proteína de Unión a TATA (TBP), el TFIIB, Proteínas E1a de Adenovirus y RNA polimerasa II entre otras. YY1 tiene funciones fundamentales en la embriogénesis, diferenciación, replicación y proliferación celular. Ejerce sus efectos sobre los genes involucrados en estos procesos a través de su capacidad para iniciar, activar o reprimir la transcripción según el contexto en el que se encuentre, través del reclutamiento de cofactores, la activación o represión por interrupción de sitios de unión o cambios conformacionales del DNA (Omim, 2023).

YY1 fue identificado en 1991 por Shi como una proteína con actividad represora del promotor P5 en ausencia de la oncoproteína E1A, mientras que, en su presencia, YY1 era capaz de activar dicho promotor. Esta dualidad funcional llevó a su denominación como "Yin Yang 1", en referencia al concepto chino que describe las dos fuerzas opuestas fundamentales, "Yin" para la represión y "Yang" para la activación.
Se han propuesto numerosos mecanismos de acción para YY1 y su acción activadora o represora:

- Procesos post-traduccionales: La acetilación del dominio C-terminal de YY1 reduce su capacidad de unión al ADN in vitro, mientras que la acetilación del dominio central es necesaria para que YY1 actúe como un represor completo. YY1 acetilado puede unirse a un potenciador esencial 5 kb aguas arriba del gen homeobox Otx2 para provocar su expresión (Gordon, 2006).

- Interacción con cofactores: YY1 puede actuar como un factor de transcripción de unión al ADN tradicional, interactuando con una lista extensa de cofactores que median la activación o represión de la transcripción. Une los cofactores directamente a los promotores, como el factor 1 asociado a YY1 (YY1AP), que estimula la transcripción, y también recluta correpresores del ADN, como los miembros de la familia Smad que reprimen la señalización del TGF-β y la diferenciación celular. La superposición de los sitios de unión al ADN de YY1 y los activadores transcripcionales proporciona otra explicación para la represión mediada por YY1, ya que la unión de YY1 puede bloquear el factor activador (Thijs, 2020).

- Competencia por el sitio de unión: La superposición de los sitios de unión al ADN de YY1 y los activadores transcripcionales proporciona otra explicación para la represión mediada por YY1. Esto se debe a que la unión al ADN de YY1 y la unión de un activador en un locus determinado pueden ser mutuamente excluyentes. En este sentido, la unión de YY1 puede bloquear la acción del factor activador al competir por el mismo

sitio de unión en el ADN. Esta competencia por la unión al ADN puede resultar en la inhibición de la actividad del activador transcripcional (Atchison, 2003).

- Interacción con complejos de remodelación de la cromatina: YY1 puede interactuar con complejos de remodelación de la cromatina que regulan la transcripción, regulando la accesibilidad de la cromatina. Interactúa con histonas desacetilasas (HDAC) asociadas con el silenciamiento de genes, así como con histonas acetiltransferasas (HAT) como p300 y la proteína de unión a CREB (CBP), activando la transcripción (Yao, 2001).

- Regulación de la transcripción mediante la formación de bucles potenciadores-promotores: YY1 regula la transcripción mediante la formación de bucles potenciadores-promotores. Al controlar el bucle de cromatina, YY1 puede habilitar y estabilizar las interacciones potenciador-promotor para activar la expresión (Thijs, 2020).

El estudio de Oei y Shi en 2001 destacó una interacción física entre YY1 y la enzima PARP, responsable de la síntesis de polímeros de ADP-ribosa. La sobreexpresión de YY1 en células HeLa condujo a una acumulación de poli (ADP-ribosa) y aceleró la reparación del DNA después del daño genotóxico. Esto sugiere una relación funcional y física entre estas proteínas en procesos de reparación y transcripción del DNA.

Por otro lado, Sui, G. y col., 2004 encontraron que la eliminación de YY1 resultó en la acumulación de p53 debido a una reducción de la ubiquitinación de p53 *in vivo*. Por el contrario, la sobreexpresión de YY1 estimuló la ubiquitinación y degradación de p53.

Más aún, Gordon S. y col., 2006 presentaron una revisión del factor de transcripción YY1 con la conclusión de que, YY1 puede tener el potencial de actuar como iniciador de la tumorigénesis y, por lo tanto, puede servir como marcador tumoral de diagnóstico y pronóstico; además, puede proporcionar un objetivo eficaz para la quimioterapia y/o inmunoterapia antitumoral.

Se ha demostrado que YY1 es un factor de transcripción que participa en el control de múltiples mecanismos Santiago y Col., encontraron en 2007 que la sobreexpresión de YY1 inhibía la formación de neoíntima en vasos sanguíneos de humanos, conejos y ratas. YY1 desestabilizó p53 al inducir su ubiquitinación y degradación proteosomal. Se concluyó que YY1 suprime el crecimiento de las células del músculo liso y la reparación de heridas arteriales a través de p53, p21 y SP1.

Además, Lee M. y col., descubrieron en 2012 que la expresión de YY1 y BRCA1 se correlacionó positivamente en el examen histológico de tejido mamario tumoral y humano normal, con una expresión generalmente más baja de ambas proteínas en los cánceres de mama. La sobreexpresión de YY1 provocó la detención del ciclo celular en células de cáncer de mama transfectadas e inhibió la formación de tumores después de la inyección en ratones.

2.7.2 Estructura

El factor YY1 presente en mamíferos es una proteína altamente conservada compuesta por 414 aminoácidos, con un peso molecular de 44 kDa. En el caso de *Xenopus*, aunque su secuencia de aminoácidos es muy similar a la de mamíferos, es ligeramente más pequeña con 373 aminoácidos. Sin embargo, mantiene los cuatro dominios de zinc, característicos de ese factor, lo que indica una conservación en su estructura. (Chavez, 2008)

A lo largo de la secuencia proteica del factor YY1 de mamífero, se han identificado diversos sitios funcionales (Figura 9). La región N-terminal de EhYY1 contiene una zona rica en histidinas, flanqueada por dos regiones de aminoácidos de carácter ácido que funcionan como dominio de activación transcripcional. La primera abarca desde el residuo 1 al 54, y la segunda del 81 al 153. (Austen, 1996)
Se ha observado que la eliminación en el clúster de histidinas no afecta la actividad de transactivación. En el centro de la molécula se encuentra una región rica en glicinas y alaninas, comprendida entre los aminoácidos 154 y 198, la cual corresponde al dominio de interacción con la histona desacetilasa (HDAC). Junto al extremo C-terminal, esta región forma el dominio de represión transcripcional. Las proteínas con eliminaciones en cualquiera de las dos regiones acídicas o en la región espaciadora entre la región rica en glicina y alanina, y el dominio de unión al ADN, muestran una reducción del 50% de la actividad transcripcional. Además, se identifica una región central adicional llamada PHR (aminoácidos 205-226), la cual es responsable del reclutamiento de proteínas homeobox (Luke, 2006).

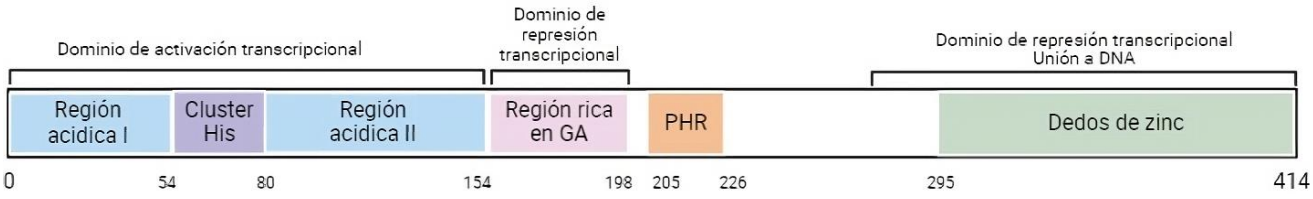


Figura 9. Representación esquemática de la proteína YY1 de humano.

Todos los factores YY1 caracterizados hasta el momento presentan cuatro dedos de zinc del tipo C2H2 localizados hacia el extremo carboxilo terminal los cuales corresponden a la secuencia específica del dominio de unión al DNA (Figura 10) (Arranz, 2015). Los dominios de dedos de zinc forman estructuras únicas muy sólidas, sostenidas en la mayoría de los casos por residuos aminoacídicos de cisteínas (C) e histidinas (H). La posición y distribución del dominio C2H2 es importante para que se dé un plegamiento correcto y estable del motivo de la proteína (Chavez, 2008).

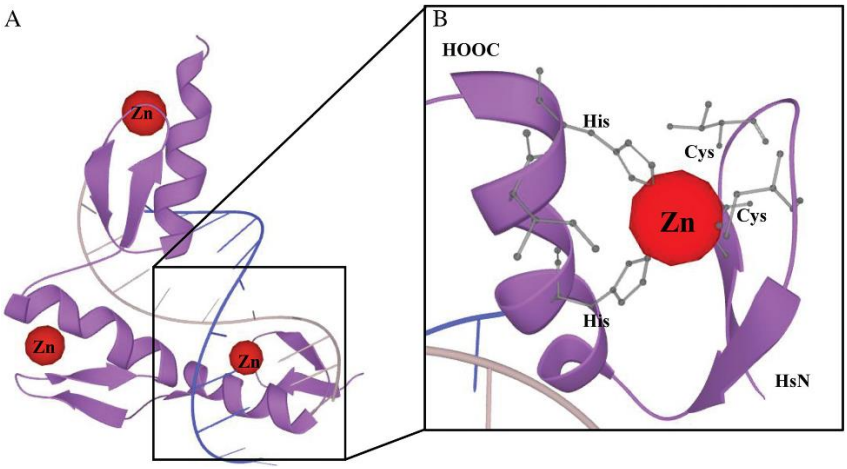


Figura 10. Dedo de zinc C2H2. Se muestra un ejemplo de proteína con tres motivos C2H2 unida a DNA (B) ampliación del dominio C2CH2. (Ke Wang et al., 2018)

En *E. histolytica*, Chávez en el 2008 identificó el factor YY1, cuyas características se muestran en la tabla 3 y se comparan con el factor YY1 humano. En la amiba, el factor es mucho más pequeño en el gen y, por consiguiente, en la proteína. Sin embargo, el dominio de unión al DNA está altamente conservado, ya que también está formado por cuatro dedos de Zinc localizados en el extremo carboxilo terminal (Tabla 6). En cuanto a su localización, estudios de microscopía confocal muestran que HYY1 responde a los eventos de los puntos de control de la síntesis de ADN. Al inicio de la síntesis de ADN, cuando las células entran en la fase S, el patrón de localización de YY1 cambia del citoplasma al núcleo. Más tarde, pasado el punto medio de la fase S, YY1 vuelve a ser principalmente citoplasmático (Beyrouthy, 2007). Aún no se han realizado estudios para determinar la localización de EhYY1, pero al tratarse de un factor de transcripción, se esperaría ubicarlo en el núcleo y en el citoplasma, igual que su homólogo humano.

Tabla 3.Características YY1.

Características YY1	*Homo sapiens*	*Entamoeba histolytica*
Ubicación	14q32.2 Durante la mitosis muestra que se distribuye al citoplasma durante la profase y permanece excluida del DNA hasta la telofase temprana	Desconocido
Gen (pb)	1242	528
Proteína (aa)	414	175
Dedos de zinc	4	4
Aminoácidos del sitio de unión a DNA	114	113

**Comparación de Homo sapiens y Entamoeba histolytica. (Gilchrist, 2001)*

III JUSTIFICACIÓN

E. histolytica es el agente causal de la disentería amebiana, que se manifiesta con varias semanas de calambres, dolor abdominal, diarrea acuosa o sanguinolenta y pérdida de peso además se ha descrito como enfermedad extraintestinal diseminada causando absceso hepático, neumonía, pericarditis purulenta e incluso amebiasis cerebral. Su distribución geográfica es mundial, se estima que en todo el mundo hasta 50 millones de personas están infectadas por *E. histolytica* y es responsable de más de 100,000 muertes al año principalmente en países en vías de desarrollo. México es un país endémico, ya que tiene una alta tasa de mortalidad que va de 40 a 110 mil muertes al año, lo que representa un problema de salud pública.

E. histolytica presenta el fenotipo de resistencia a múltiples fármacos (MDR) debido a la sobre expresión del gen *Ehpgp5*. En el promotor de este gen se ha identificado una posible secuencia consenso de unión para el factor de transcripción YY1 un factor asociado con la activación transcripcional del gen *Ehpgp5*. YY1 regula una amplia gama de genes que desempeñan funciones críticas en el desarrollo, la diferenciación y el ciclo celular además de poseer un carácter dual del que deriva su nombre Yin-Yang 1 (YY1) ya que puede tener una participación como activador o represor de genes.

Dado que en condiciones de estrés con emetina *Ehyy1* presenta una sobre expresión en los trofozoítos de *E. histolytica* es probable que este factor tenga un papel importante control transcripcional y en el funcionamiento de la amiba. Particularmente por su posible asociación con genes que le confieren resistencia a fármacos.

El mecanismo general de transcripción en *E. histolytica* no ha sido descrito con detalle y dadas sus características es particularmente importante estudiar al factor YY1, así como el papel que tiene en el control transcripcional en este parásito.

IV OBJETIVOS

4.1 Objetivo general

Caracterizar la estructura del factor de transcripción EhYY1 en *Entamoeba histolytica* mediante la identificación de dominios conservados en su secuencia proteica y la generación de un modelo tridimensional que permita el análisis de la interacción del factor de transcripción con su secuencia consenso.

4.2 Objetivos particulares

1) Identificar la secuencia de aminoácidos del factor EhYY1.
2) Localizar dominios conservados en la proteína EhYY1.
3) Generar un modelo tridimensional de la proteína EhYY1.
4) Realizar un análisis filogenético del factor EhYY1.
5) Evaluar la interacción EhYY1-DNA a través de un ensayo de Docking.

V MATERIALES Y MÉTODOS

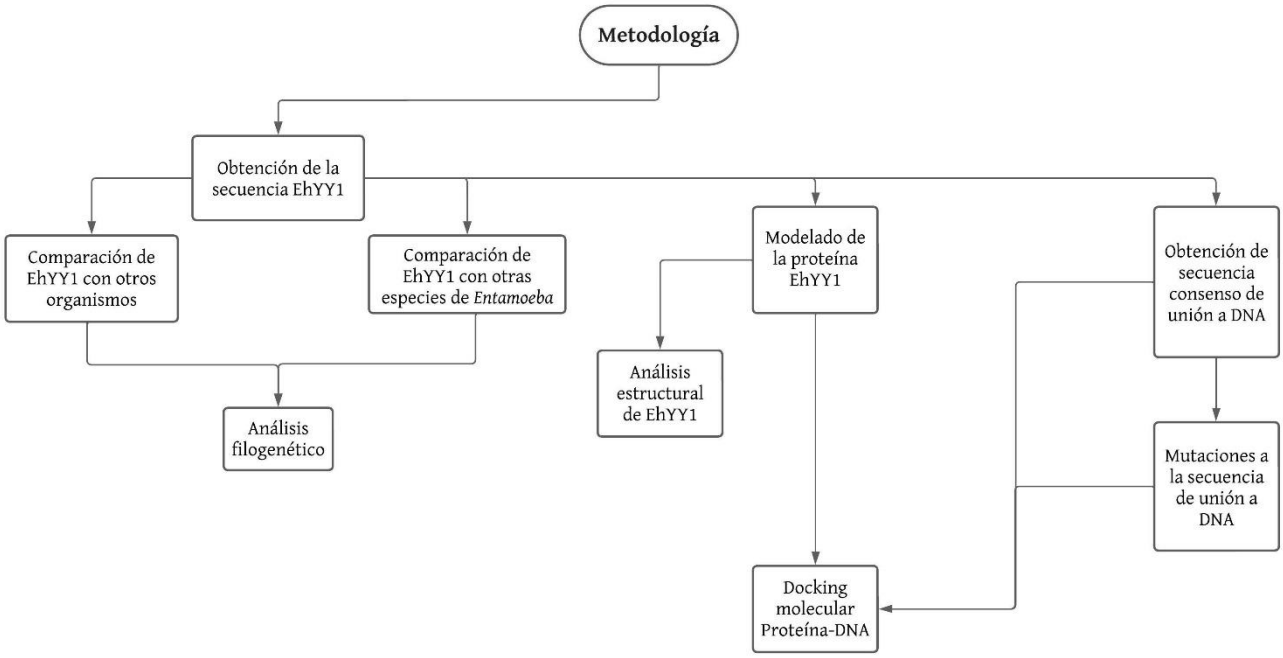


Figura 11. Diagrama metodológico general.

5.1 Obtención de la secuencia EhYY1

Mediante alineamientos múltiples se creó una secuencia consenso YY1 de 414 aa a partir de 9 secuencias de organismos ya caracterizadas para este factor de transcripción; *Homo sapiens* (P25490), *Mus musculus* (Q00899), *Xenopus laevis* (Q6DDI1), *Danio rerio* (Q7T304), *Mus caroli* (XP_021034542), *Callithrix jacchus* (XP_002754331), *Theropithecus gelada* (XP_025248879), *Sapajus apella* (XP_032143307) y *Pan troglodytes* (XP_009426747). Con la secuencia obtenida se realizó una búsqueda en AmoebaDB (VEuPathDB, 2019) en donde se obtuvo la secuencia "*Entamoeba histolytica* HM-1:IMSS hypothetical protein" (con número de acceso XM_645371.1 de 175 aa).

5.2 Comparación de EhYY1 con otros organismos

La secuencia EhYY1 obtenida se comparó mediante un alineamiento de secuencias múltiples con las 9 proteínas YY1 mencionadas en la plataforma Clustal Omega (EBI, 2022). Se identificaron las posiciones de los aminoácidos más conservados y se elaboró una tabla de identidad y homología. Además, se obtuvieron sus propiedades fisicoquímicas a través de la plataforma ProtParam (ProtParam tool, 2023) y se recopilaron en una tabla. Por último, se

llevó a cabo una búsqueda en la plataforma PROSITE (Swiss Institute of Bioinformatics, 2023) para identificar los dominios proteicos de las distintas YY1.

5.3 Comparación de EhYY1 con otras especies de *Entamoeba*

Se realizó la búsqueda de secuencias semejantes a EhYY1 a través del servidor AmoebaDB, en los genomas de varias cepas y especies de *Entamoeba*: *E. histolytica* KU27 (EHI5A_224080), *E. histolytica* HM-1: IMSS (EHI_163860), *E. histolytica* HM-3: IMSS (KM1_264880), *Entamoeba dispar* SAW760 (EDI_155470), Laredo (EMO_053350) y *Entamoeba invadens* IP1 (EIN_359510). Se identificaron las posiciones de los aminoácidos más conservados y se elaboró una tabla de identidad y homología. Además, se obtuvieron sus propiedades fisicoquímicas a través de la plataforma ProtParam y se recopilaron en una tabla. Finalmente, se generó un gráfico que muestra el porcentaje de abundancia de aminoácidos en las distintas secuencias YY1 de la amiba.

5.4 Análisis filogenético

Con la finalidad de conocer las relaciones evolutivas entre EhYY1 y otros organismos, se realizó un árbol filogenético a través del programa MEGA-X, con los 9 factores de transcripción YY1 de las especies anteriormente mencionadas, las secuencias YY1 de amibas y 5 proteínas de dedos de zinc correspondientes a los mejores puntajes de Blast para EhYY1 las cuales fueron *Pseudomonas stutzeri, Araneus ventricosus, Xiphophorus couchianus, Carassius auratus y Phytophthora cinnamomi*.

5.5 Modelado de la proteína

Se realizó el modelado de EhYY1 con dos métodos distintos para la totalidad de la proteína se hizo uso de Inteligencia Artificial (IA) en la plataforma AlphaFold. Subiendo la secuencia proteica de EhYY1 a su plataforma y descargando los resultados. (DeepMind y EMBL-EBI, 2023).

Posteriormente se realizó un modelado por homología para el dominio de unión a DNA (113 aa) a través de los servidores SWISS-MODEL del servidor EXPASY (SWISS-MODEL, 2019) usando como template la proteína humana YY1 (PDB ID: 1UBD) de 114 aa. Además, la calidad del modelo obtenido EhYY1 de 113aa se evaluó a través de PDBSUM (PDBsum, 2019) mediante un diagrama de Ramachandran y Prosa web (Wiederstein & Sippl, 2007) con Z-score.

Se realizó una superposición de las estructuras 3D entre la proteína modelada (EhYY1) y la proteína humana (HYY1) a través de la plataforma online Raptor X (University of Chicago, 2019) con el fin de conocer el porcentaje de homología estructural.

5.6 Análisis estructural de EhYY1

Se realizó un análisis de la secuencia proteica de EhYY1 (113 aa) mediante la plataforma PDBSum (EBI, 2023). Este análisis proporcionó información sobre la estructura secundaria de la proteína y la ubicación de sus dominios de dedos de zinc. Adicionalmente se escaneo la secuencia de aminoácidos para la búsqueda de potenciales sitios de fosforilación en NetPhos 3.1 (NetPhos - 3.1, 2023), sitios de SUMOilación en GPS-SUMO 2.0 (GPS-SUMO 2.0, 2022) y GPS-PAIL 2.0 para los sitios de acetilación (GPS-PAIL 2.0, 2014).

5.7 Obtención de secuencia consenso de unión a DNA

Debido a que existe más de una secuencia YY1 de unión a DNA reportada, se decidió crear una secuencia consenso, ser realizo la búsqueda de estas secuencias en los servidores Transfact y Pubmed, a partir de ellas se realizaron alineamientos múltiples para determinar los nucleótidos más conservados y posteriormente se usó el servidor WebLogo (WebLogo, 2019) para crear un logo de secuencias.

5.8 Docking molecular Proteína-DNA

Para analizar la interacción proteína-DNA primero se modelaron las estructuras 3D de las secuencias consenso para YY1 de humano de doble cadena, en sentido HYY1-CS [5’-AGGGTCTCCATTTTGAAGCG-3’] y la cadena anti-sentido HYY1-CAs [3’-CGCTTCAAAATGGAGACCCT -5’] ambas de 20 pb de la misma manera para la amiba se modelaron sus secuencias YY1 localizadas en de promotor *EhPgp5*, en sentido *EhPgp5*-YY1S [5’- ACCATTTTGGG-3’] y anti-sentido *EhPgp5*-YY1As [3’-CCCAAAATGGT-5’] de 11 pb, con el software UCSF Chimera 1.15.

Se realizaron múltiples ensayos Docking dirigidos, utilizando los servidores PatchDock (PatchDock, 2019) y HDOCK SERVER, (Huazhong University of Science and Technology, 2019) en el primer ensayo se evaluó la unión que presenta la proteína EhYY1 modelada contra la secuencia de consenso de humano (HYY1-CS y HYY1-CAs). Posteriormente se evaluó la unión de HYY1 con la secuencia YY1 del promotor *EhPgp5* en amiba (*EhPgp5*-YY1S y *EhPgp5*-YY1As). Finalmente se realizó un Docking entre EhYY1 y su secuencia YY1 de *EhPgp5* (*EhPgp5*-YY1S y *EhPgp5*-YY1As).

Nota: Para los ensayos de acoplamiento molecular se usaron solo los aminoácidos correspondientes a sitio de unión a DNA; Humano (144 aa) y *E. histolytica* (113 aa).

5.9 Mutaciones a la secuencia de unión a DNA.

A través de UCSF Chimera 1.15 (Pettersen, 2019) se realizaron tres tipos de mutaciones al DNA; scramble, inversion e intercambio, en la secuencia consenso de unión al DNA humano [5′-T**CCATTT**TGAA-3′] de 11 pb y en la secuencia YY1 del promotor *EhPgp5* [5'-A**CCATTT**TGGG-3'] de 11 pb, específicamente en 5 de sus pares de bases correspondientes a la secuencia [5′-CATTT- 3′] identificada como critica para la interacción proteína-DNA.

Generándose las siguientes secuencias para Entamoeba; Eh-scramble: [5′-AC**aatct**TGGG-3′], Eh-inversión: [5′-AC**tttac**TGGG-3′] e Eh-intercambio: [5′-AC**t**AT**c**TTGGG-3′], posteriormente se realizó un Docking con EhYY1.

De igual forma para el DNA humano; H-scramble: [5′-TC**aatct**TGAA-3′], H-inversión: [5′-TC**tttac**TGAA-3′] e H-intercambio: [5′-TC**t**AT**c**TTGAA- 3′], posteriormente se realizó un Docking con HYY1.

VI RESULTADOS

6.1 Obtención de la secuencia EhYY1

Con la finalidad de identificar si *E. histolytica* presenta un gen YY1 en su genoma, primeramente, obtuvimos las secuencias YY1 de *H. sapiens* (P25490), *M. musculus* (Q00899), *X. laevis* (Q7T304), *D. rerio* (Q6DDI1), *M. caroli* (XP_021034542), *C. jacchus* (XP_002754331), *T. gelada* (XP_025248879), *S. apella* (XP_032143307) y *P. troglodytes* (XP_009426747) y realizamos un alineamiento de secuencias múltiples en BLAST (Figura 12), el resultado mostró que comparten entre ellas de un 68 a un 100% de identidad (Tabla 4).

```
B.rerio           MASGETLYIEAEGAAVPAELVELQEIEVET-----IETTVIGGEDEDED---------- 44
X.laevis          MASGDTLYIASDGSEMPAEIVELHEIEVECIPVETIETTVVGDDEDDDDDQ--------- 51
M.musculus        MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDEDGGGGDHG 60
M.caroli          MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEEDDDEDGGGGDHG 60
C.jacchus         MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEDDDDDEDGGGGDHG 60
T.gelada          MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEAEEDEDDEDGGGGDHG 60
S.apella          MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG 60
H.sapiens         MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG 60
P.troglodytes     MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEDDDDEDGGGGDHG 60
                  ****:**** ::*: :***:***:*****      *****:* : ::::

B.rerio           ------------------EEEPAMIALQPLHTGDG-----EPGVILLQTREEVVGEDAE 80
X.laevis          ------------------GHHQQPMIALQPLDSDDLVH-PHHQEVILVQTREEVVGGDDS 92
M.musculus        GGGGGHGHAGHHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 120
M.caroli          GGGGGHGHAGHHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 120
C.jacchus         GGG-GHGHAVHH--HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 117
T.gelada          GGG-GHGHAGHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
S.apella          GGG-GHGHAVHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
H.sapiens         GGG-GHGHAGHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
P.troglodytes     GGG-GHGHAGHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
                             ...  ******* :.*      .  ***:******** * .

B.rerio           ---LRGDAFEEQILIPVPTPG-ADAQYIGQTLLTVAGRSSG----------GRAGRRAAA 126
X.laevis          -DLRADDGYEDQILIPVPVPAG-EDEYIEQTLVTVAGKSSSG---------GRMKKGGGG 141
M.musculus        DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGA--SSGGGRVKKG--- 175
M.caroli          DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGA--SSGGGRVKKG--- 175
C.jacchus         DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 174
T.gelada          DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
S.apella          DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
H.sapiens         DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
P.troglodytes     DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
                        *.:*:*******.*.  : :** ***:***. ...          **  :

B.rerio           GGKKTGKKSYLSA---------AESSGRKWEQKQVQIKTLEGEFSVTMWASDDKKELEHE 177
X.laevis          SGKKSSKKSYLSG---------TEPSGRKWEQKQVQIKTLEGEFSVTMWASDDKKDIDHE 192
M.musculus        GGKKSGKKSYLGGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
M.caroli          GGKKSGKKSYLGGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
C.jacchus         GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 234
T.gelada          GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
S.apella          GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
H.sapiens         GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
P.troglodytes     GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
                  .***:.*****..         :: ..:*********************:**:**:::**
```

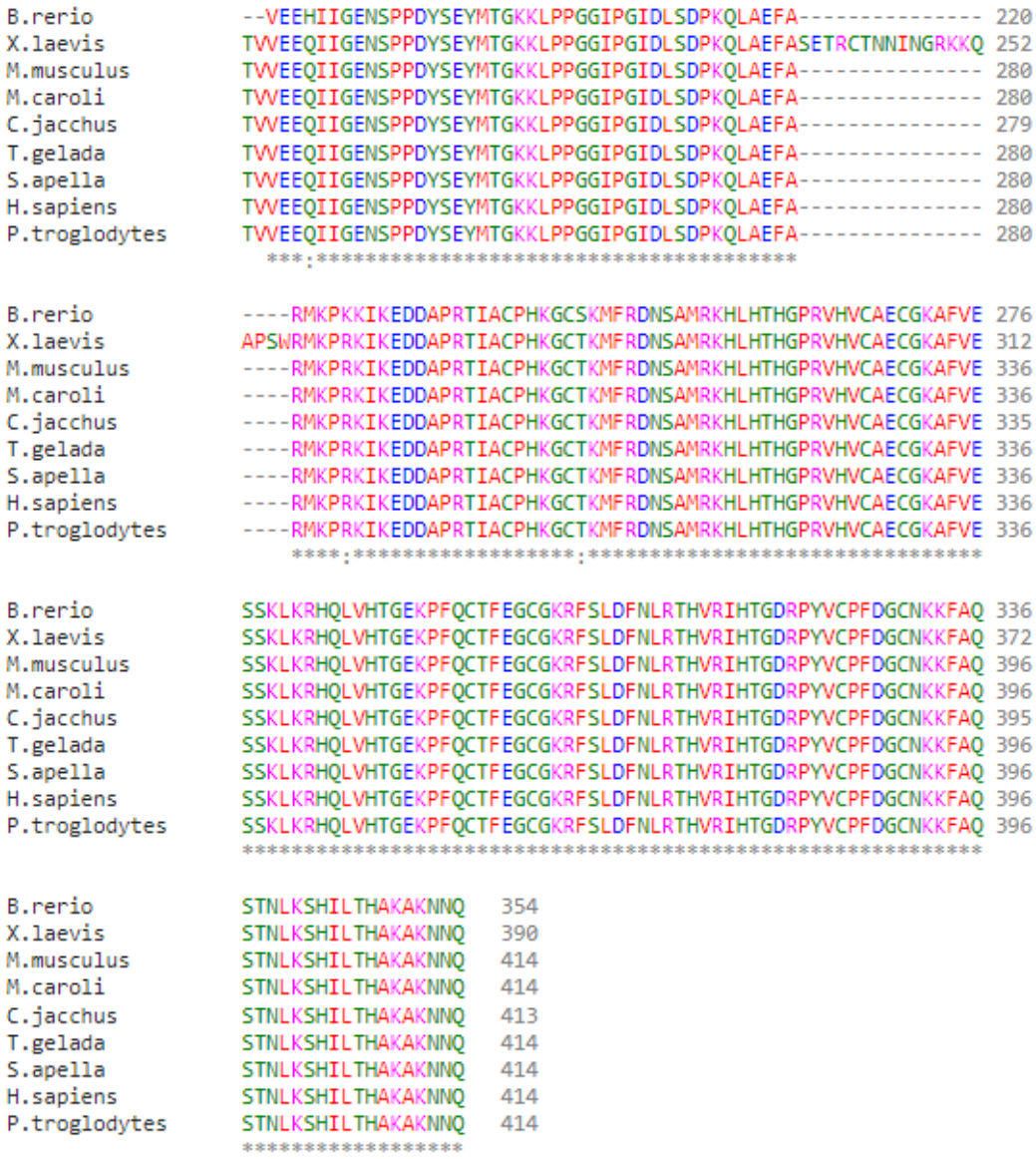


Figura 12. Alineamiento múltiple de secuencias YY1 de diferentes organismos. El asterisco señala los aminoácidos idénticos, los dos puntos señalan los aminoácidos similares y un punto señala los aminoácidos que comparten propiedades fisicoquímicas

Posteriormente utilizando el alineamiento múltiple obtenido y la herramienta EMBOSS Cons (Institute European Bioinformatics, 2023) se creó una Secuencia Consenso Predictiva YY1:

MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEDDDDDEDGGGG
DHGGGGGHGHAGHHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREE
VVGGDDSDGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSS
GGGRVKKGGGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFS
VTMWSSDEKKDIDHETVVEEQIIGENSPPDYSEYMTGKKLPPGGIPGIDLSDPKQLA
EFARMKPRKIKEDDAPRTIACPHKGCTKMFRDNSAMRKHLHTHGPRVHVCAECG
KAFVESSKLKRHQLVHTGEKPFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPF
DGCNKKFAQSTNLKSHILTHAKAKNNQ

Secuencia 1. Secuencia Consenso Predictiva de aminoácidos codificantes para EhYY1, en negrillas la primera metionina

Con dicha secuencia realizamos un escrutinio en el genoma de *E. histolytica* (Bioinformatics Resource Center , 2023). El resultado mostró la presencia de una secuencia de 528 pb (ID XM_645371.1), que codifica para una proteína de 175 aa con un peso molecular esperado de 20.5 kDa. La posible proteína YY1 de *E. histolytica* (a la que ahora llamaremos EhYY1) mostró de 58 % de identidad y 37% de homología con nuestra Secuencia Consenso Predictiva, se encuentra descrita bajo el nombre de "*Entamoeba histolytica* HM-1:IMSS hypothetical protein".

```
EHI_163860A-p1
 | transcript=EHI_163860A | gene=EHI_163860 | organism=Entamoeba_histolytica_HM-1:IMSS
| gene_product=hypothetical
protein | transcript_product=hypothetical protein |
location=DS571965:6-533(-) | protein_length=175 | sequence_SO=supercontig
| SO=protein_coding_gene | is_pseudo=false
Length=175

 Score = 89.4 bits (220),  Expect = 1e-21, Method: Compositional matrix adjust.
 Identities = 42/113 (37%), Positives = 66/113 (58%), Gaps = 3/113 (3%)

Query  298  CPHKGCTKMFRDNSAMRKHLHTHGP-RVHVCAECGKAFVESSKLKRHQLVHTGEKPFQCT  356
            CPH+GC K +    ++ +H+  H   + HVC  CGKAF+  S+ + H  +HTG KP+ C
Sbjct  28   CPHEGCGKCYDSEESLHRHMSRHSTVKEHVCEFCGKAFLRKSECEIHVRIHTGVKPYCCN  87

Query  357  FEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAK  409
               C KRF+   +L+ H+  H+ ++P+ CPF GC  +F +  + K H+  H K
Sbjct  88   I--CQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFKRKNDAKKHLRIHVK  138


 Score = 62.4 bits (150),  Expect = 4e-12, Method: Compositional matrix adjust.
 Identities = 36/85 (42%), Positives = 46/85 (54%), Gaps = 4/85 (5%)

Query  330  CGKAFVESSKLKRHQLVHTGEKPFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDG  389
            CGK +     L RH   H+  K   C F  CGK F        HVRIHTG +PY C  +
Sbjct  33   CGKCYDSEESLHRHMSRHSTVKEHVCEF--CGKAFLRKSECEIHVRIHTGVKPYCC--NI  88

Query  390  CNKKFAQSTNLKSHILTHAKAKNNQ  414
            C K+FA++T+LK H+L H+  K  Q
Sbjct  89   CQKRFARATDLKIHMLYHSDEKPFQ  113
```

Figura 13. Alineamiento de secuencias proteica EhYY1. Secuencia consenso predictiva (Querty) y "EhYY1" (Sbjct)

La secuencia de nucleótidos de *Ehyy1* es la siguiente:

atggaaagccaaatagaacgtagtgtgaatatcatttatgataaatttattagtaagaaacatacaactggaaagtatatatgtccacat
gaaggttgtgggaagtgttatgatagtgaagagtctcttcatagacatatgagtcgacattctactgttaaagaacatgtctgtgagtttt
gtggtaaagctttcttaagaaaatcagagtgtgaaattcatgttagaattcatacaggagttaagccatattgttgtaatatttgtcagaaa
cgttttgctcgtgctactgatcttaagattcatatgctttaccattctgatgaaaagccatttcaatgtccattcccaggttgtactttacgttt
taaacgtaaaaatgatgctaaaaaacacttaagaatacatgtcaaaaaaattcatggaaatgttacagaactaatgtctccccatcattc
agctttcaaagctgttcctaaacaattaattcgtcttgaatcttatggaataatgtatcattgtaat**taa**

Secuencia 2. Secuencia de bases del gen Ehyy1. En negrillas están señalados los codones de inicio y paro de la traducción respectivamente. Secuencia con numero de acceso XM_645371.1

La secuencia correspondiente a la traducción del gen *Ehyy1* es la siguiente:

MESQIERSVNIIYDKFISKKHTTGKYICPHEGCGKCYDSEESLHRHMSRHSTVKEHV
CEFCGKAFLRKSECEIHVRIHTGVKPYCCNICQKRFARATDLKIHMLYHSDEKPFQ
CPFPGCTLRFKRKNDAKKHLRIHVKKIHGNVTELMSPHHSAFKAVPKQLIRLESYGI
MYHCN

Secuencia 3. Secuencia de aminoácidos codificantes para Ehyy1, en negritas la primera metionina.

6.2 Comparación de EhYY1 con otros organismos

Con la finalidad de conocer la similitud del factor EhYY1 con respecto a los factores YY1 de otros organismos realizamos un alineamiento de las secuencias YY1 de nueve organismos distantes evolutivamente, así como con la secuencia EhYY1. Los resultados mostraron regiones en las secuencias altamente conservadas en el EhYY1 como lo son los dedos de zinc, lo que denota el grado de conservación de este factor y de los dominios implicados en su función como factor de transcripción.

```
Entamoeba_histolytica  ------------------------------------------------------------ 0
B.rerio                MASGETLYIEAEGAAVPAELVELQEIEVET-----IETTVIGGEDEDED---------- 44
X.laevis               MASGDTLYIASDGSEMPAEIVELHEIEVECIPVETIETTVVGDDEDDDDDQ--------- 51
M.musculus             MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG 60
M.caroli               MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEEDDDEDGGGGDHG 60
C.jacchus              MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEDDDDDEDGGGGDHG 60
T.gelada               MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEAEEDEDDEDGGGGDHG 60
S.apella               MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG 60
H.sapiens              MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG 60
P.troglodytes          MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEDDDDDEDGGGGDHG 60

Entamoeba_histolytica  ------------------------------------------------------------ 0
B.rerio                -------------------EEEPAMIALQPLHTGDG-----EPGVILLQTREEVVGEDAE 80
X.laevis               ------------------GHHQQPMIALQPLDSDDLVH-PHHQEVILVQTREEVVGGDDS 92
M.musculus             GGGGGHGHAGHHHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 120
M.caroli               GGGGGHGHAGHHHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 120
C.jacchus              GGG-GHGHAVHH--HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 117
T.gelada               GGG-GHGHAGHHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
S.apella               GGG-GHGHAVHHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
H.sapiens              GGG-GHGHAGHHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118
P.troglodytes          GGG-GHGHAGHHH-HHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDDS 118

Entamoeba_histolytica  ------------------------------------------------------------ 0
B.rerio                ---LRGDAFEEQILIPVPTPG-ADAQYIGQTLLTVAGRSSG----------GRAGRRAAA 126
X.laevis               -DLRADDGYEDQILIPVPVPAG-EDEYIEQTLVTVAGKSSSG---------GRMKKGGGG 141
M.musculus             DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGA--SSGGGRVKKG--- 175
M.caroli               DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGA--SSGGGRVKKG--- 175
C.jacchus              DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 174
T.gelada               DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
S.apella               DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
H.sapiens              DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175
P.troglodytes          DGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKG--- 175

Entamoeba_histolytica  ------------------------------------------------------------ 0
B.rerio                GGKKTGKKSYLSA---------AESSGRKWEQKQVQIKTLEGEFSVTMWASDDKKELEHE 177
X.laevis               SGKKSSKKSYLSG---------TEPSGRKWEQKQVQIKTLEGEFSVTMWASDDKKDIDHE 192
M.musculus             GGKKSGKKSYLGGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
M.caroli               GGKKSGKKSYLGGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
C.jacchus              GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 234
T.gelada               GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
S.apella               GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
H.sapiens              GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
P.troglodytes          GGKKSGKKSYLSGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHE 235
```

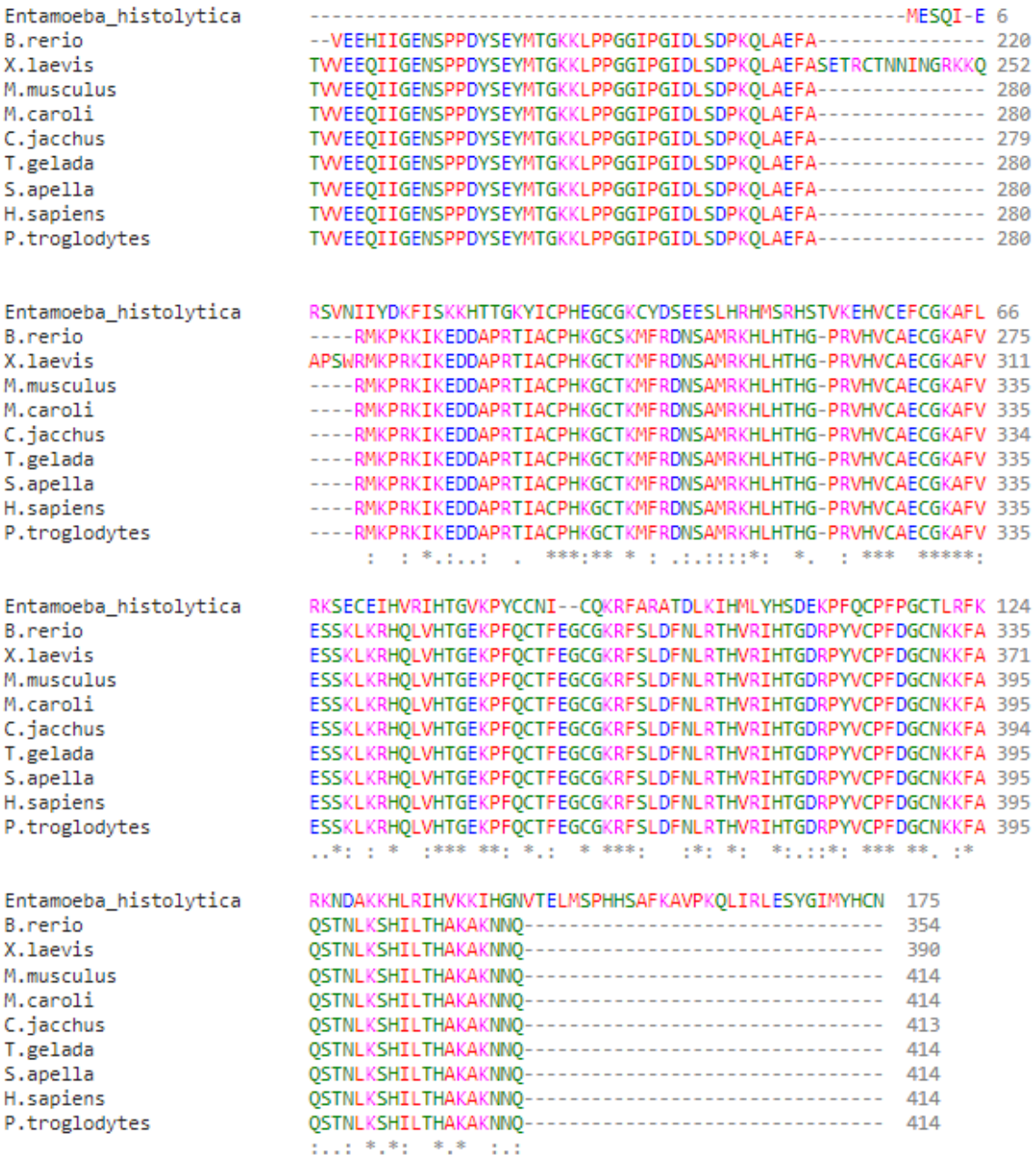


*Figura 14. Alineamiento múltiple de secuencias YY1 de diferentes organismos contra EhYY1. Realizado en Clustal Omega (https://www.ebi.ac.uk/Tools/msa/). * se muestran los residuos idénticos; con: se muestran los residuos semejantes y . se muestran los residuos que comparten las mismas propiedades fisicoquímicas.*

Los datos de la comparación de EhYY1 con otros organismos muestran una identidad del 37.17% (Tabla 4), lo que significa que en más de un tercio de sus posiciones se tiene el mismo aminoácido en comparación con las secuencias de otros organismos. Además, presenta un alto grado de homología del 66%, indicando que, dos tercios de las posiciones contienen aminoácidos similares que le dan las mismas propiedades bioquímicas a las secuencias. Estos valores ofrecen una perspectiva valiosa sobre la conservación y las similitudes en la estructura y función de YY1 en diversos organismos.

Asimismo, se obtuvieron las características fisicoquímicas de YY1, las cuales se muestran en la Tabla 5.

Tabla 4. Porcentaje de identidad y homología YY1.

YY1	%	E. histolytica	M. musculus	X. laevis	B. rerio	H. sapiens	M. caroli	C. jacchus	T. gelada	S. apella	P. troglodytes
E. histolytica	Identity	100									
	Homology	100									
M. musculus	Identity	37.17	100								
	Homology	66	100								
X. laevis	Identity	37.17	75.58	100							
	Homology	66	100	100							
B. rerio	Identity	37.17	68.60	72.56	100						
	Homology	66	100	100	100						
H. sapiens	Identity	37.17	98.56	74.42	70.29	100					
	Homology	66	100	100	100	100					
M. caroli	Identity	37.17	99.76	75.35	68.84	98.56	100				
	Homology	66	100	100	100	100	100				
C. jacchus	Identity	37.17	97.84	76.21	69.49	99.28	97.60	100			
	Homology	66	100	100	100	100	100	100			
T. gelada	Identity	37.17	98.08	75.35	70.29	99.52	97.84	98.79	100		
	Homology	66	100	100	100	100	100	100	100		
S. apella	Identity	37.17	98.32	75.81	70.53	99.76	98.08	99.52	99.28	100	
	Homology	66	100	100	100	100	100	100	100	100	
P. troglodytes	Identity	37.17	98.32	76.04	69.32	99.76	98.08	99.52	99.28	99.52	100
	Homology	66	100	100	100	100	100	100	100	100	100

**Se muestra una tabla comparativa entre diferentes especies.*

Tabla 5.Características fisicoquímicas YY1.

YY1	Gen ID	Genomic sequence (bp)	Exons	Introns	aa	Mw (kDa)	pI	(GRAVY)	Instability index	Aliphatic index
H. sapiens	P25490	1242	1	1	414	44712.57	5.80	-0.764	48.56	61.23
M. musculus	Q00899	1242	-	-	414	44716.61	5.83	-0.763	47.72	61.47
X. laevis	Q7T304	1170	-	-	390	43293.47	5.71	-0.782	49.19	64.72
B. rerio	Q6DDI1	1050	-	-	350	38440.36	5.67	-0.610	49.91	70.00
M. caroli	XP_021034542	1242	-	-	414	44730.63	5.83	-0.763	48.51	61.47
C. jacchus	XP_002754331	1239	-	-	413	44603.48	5.77	-0.747	47.86	62.08
T. gelada	XP_025248879	1242	-	-	414	44668.56	5.85	-0.752	47.45	61.47
S. apella	XP_032143307	1242	-	-	414	44754.65	5.80	-0.753	48.56	61.93
P. troglodytes	XP_009426747	1242	-	-	414	44698.54	5.80	-0.764	47.77	61.23
E. histolytica HM-1: IMSS	EHI_163860	528	1	-	175	20.5	9.33	-0.621	58.81	65.71

**Se muestra el ID, secuencia genómica, exones, intrones, aminoácidos, punto isoeléctrico, Gran Promedio de Hidropaticidad (GRAVY), índice de inestabilidad e índice alifático*

Con el objetivo de profundizar en la comprensión de las similitudes entre las proteínas YY1, se creó una tabla la cual presenta varios parámetros fisicoquímicos.

El pI (punto isoeléctrico): es el pH para el cual una proteína presenta carga neta compensada (igual número de cargas positivas que negativas) y en el mismo, la solubilidad de la proteína es mínima, el precipitado presentara mayor contenido de proteína. (Peinado, 2023) El pI de YY1 de los distintos organismos resulta ser una medida altamente conservada mostrando la característica acida de la proteína, podemos observar que se mantiene en un rango de 5.77 a 5.85 para los mamíferos y ligeramente menor para los organismos anfibios *(X. leavis*) y peces (*B. rerio*).

GRAVY (Grand Average Of Hydropathicity): el valor GRAVY para un péptido o proteína se calcula como la suma de los valores de hidropatía de todos los aminoácidos, dividida por el número de residuos en la secuencia. Donde valores positivos indican proteínas hidrofóbicas y valores negativos hidrofílicas. (ProtParam tool, 2023) El GRAVY de los YY1

ya caracterizados se muestra en un rango de -0.610 a -0.782 lo que nos indica la naturaleza hidrofílica de la proteína.

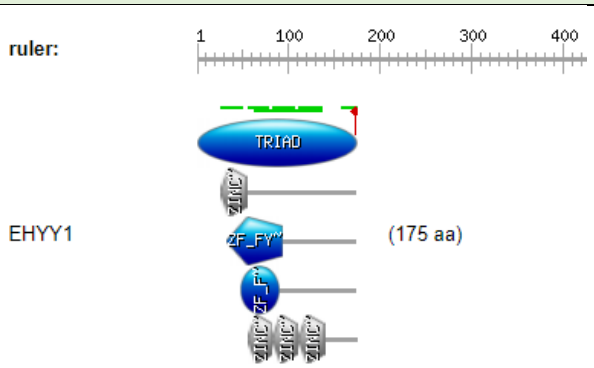
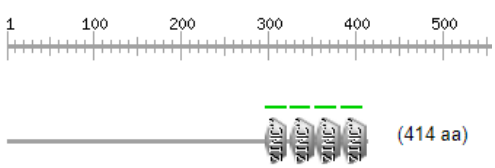
Instability Index (Índice de Inestabilidad): proporciona una estimación de la estabilidad de su proteína en un tubo de ensayo. Una proteína cuyo índice de inestabilidad es menor que 40 se predice como estable, un valor superior a 40 predice que la proteína puede ser inestable. El YY1 de los distintos organismos se encuentra en un rango de 47.45 a 49.91 es decir que la proteína no es estable fuera del ambiente celular.

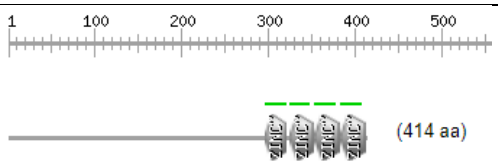
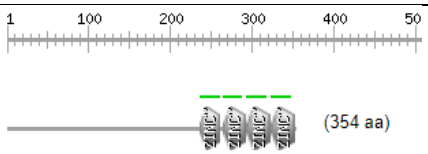
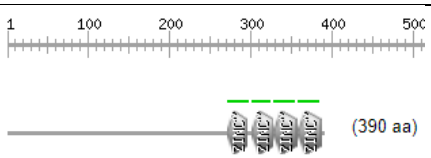
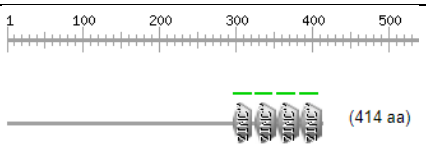
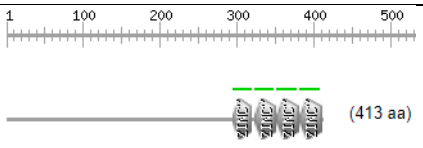
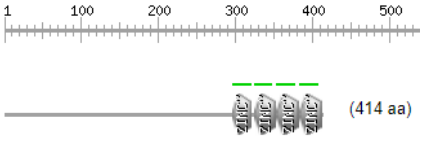
Aliphatic Index (Índice Alifático): El índice alifático de una proteína se define como el volumen relativo ocupado por las cadenas laterales alifáticas (alanina, valina, isoleucina y leucina). Con un alto índice alifático las proteínas son más estables térmicamente. Los aminoácidos alifáticos también son de naturaleza hidrófoba. Un ejemplo puede ser las citotoxinas en el rango de 66,5 a 84,33 indica que estas proteínas son térmicamente estables (Panda, 2012). YY1 muestra valores de 61.23 a 70 lo que nos habla de la buena estabilidad térmica de la proteína.

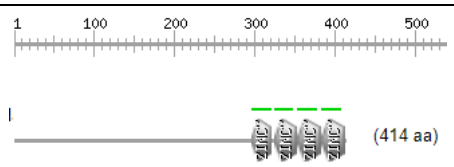
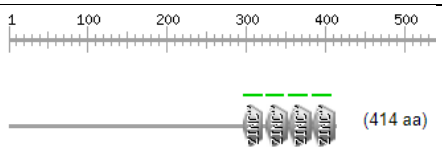
6.2.1 Posiciones y número de dominios dedos de zinc de diferentes especies

El escaneo de dominios de proteína en PROSITE (Swiss Institute of Bioinformatics, 2023) arrojo que al comparar la secuencia de aminoácidos EhYY1 con las previamente reportadas para otros organismos, encontramos que las secuencias de mamíferos son las más grandes de todas con 414 aa, mientras que la más corta corresponde a la YY1 de *E. histolytica* con 175 aa (Tabla 6) (Wai, 2016). Un dominio característico de los factores YY1 son los dedos de Zinc, los cuales son fundamentales para la interacción con el DNA. Interesantemente encontramos que el EhYY1 presenta cuatro dedos de zinc ubicados en las posiciones 26-55, 56-83, 84-111 y 112-141 aa de la secuencia (Tabla 6). Y al compararlo con los factores YY1 de otros organismos con los que realizamos el alineamiento se observa que éstos también presentan este mismo número y tipo de dedos de Zinc, resaltando la conservación de estos y probablemente la relevancia funcional de este dominio de unión al DNA.

Tabla 6. Posición de dedos de zinc ubicados en el extremo carboxilo terminal.

Especie	Dominio	Posiciones (aa) Dominio dedo de zinc tipo C2H2
E. histolytica (Secuencia EhYY1)		26-55 56-83 84-111 112-141 *1 – 175: TRIAD perfil de supradominio. *33-94: dedo de zinc FYVE *49-91: dedo de zinc FCS
H. sapiens		296-320 325-352 353-382

		383-412
M. musculus		296-320 325-352 353-382 383-412
X. laevis		236-260 265-292 293-322 323-352
B. rerio		272-296 301-328 329-358 359-388
P. troglodytes		296-320 325-352 353-382 383-412
C. jacchus		395-319 324-351 352-381 382-411
M. caroli		296-320 325-352 353-382 383-412

T. gelad		296-320 325-352 353-382 383-412
S. apella		296-320 325-352 353-382 383-412

**Con asterisco (*) se muestran dominios con bajo nivel de confianza.*

6.3 Comparación de EhYY1 con otras especies de *Entamoeba*

Debido al grado de conservación de este factor YY1 en diferentes especies incluída *E. histolytica*, decidimos investigar si otras especies del mismo género como *Entamoeba dispar SAW760*, *Entamoeba moshkovskii* Laredo y *Entamoeba invadens* IP1, así como clonas del mismo género como *E. histolytica* KU27, *E. histolytica HM-1:IMSS* y *E. histolytica HM-3:IMSS* presentan en su genoma la secuencia que codifica para este factor. Para lo cual realizamos un Blast en Uniprot y AmoebaDB usando como templado la secuencia EhYY1. Interesantemente identificamos en todos los genomas analizados la presencia de una potencial secuencia YY1 como se muestra en la Tabla 7.

Tabla 7.Secuencias YY1 encontradas en otras especies de Entamoebas.

Especie	Secuencia	ID
E. histolytica KU27	MESQIERSVNIIYDKFISKKHTTGKYICPHEGCGKCYDSEES LHRHMSRHSTVKEHVCEFCGKAFLRKSECEIHVRIHTGVKP YCCNICQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFK RKNDAKKHLRIHVKKIHGNVTELMSPHHSAFKAVPKQLIR LESYGIMYHCN	EHI5A_224080
E. histolytica HM-1: IMSS	MESQIERSVNIIYDKFISKKHTTGKYICPHEGCGKCYDSEES LHRHMSRHSTVKEHVCEFCGKAFLRKSECEIHVRIHTGVKP YCCNICQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFK RKNDAKKHLRIHVKKIHGNVTELMSPHHSAFKAVPKQLIR LESYGIMYHCN	EHI_163860
E. histolytica HM-3: IMSS	MESQIERSVNIIYDKFISKKHTTGKYICPHEGCGKCYDSEES LHRHMSRHSTVKEHVCEFCGKAFLRKSECEIHVRIHTGVKP YCCNICQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFK RKNDAKKHLRIHVKKIHGNVTELMSPHHSAFKAVPKQLIR LESYGIMYHCN	KM1_264880
E. dispar SAW760	MESQIEHSVNIIYDKFISKKNTTGKYMCPHEGCGKCYDSEE SLHRHMNRHSTIKEHICEFCGKAFLRKSECEIHVRIHTGVKP YCCSICQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFK RKSDAKKHLRIHVKKIHGNVTELISPHCSAFKVVPKQLISIE PYGVMYYCN	EDI_155470
Entamoeba moshkovskii Laredo	MQNNIEQSVNAIYEKYISKKKTNGKYVCPHKGCGKRYETE ESLQRHLNRHSTVKEHVCDYCGKSFLRKSECEIHKRIHTGV KPFNCSVCGKSFSRATDLKIHMVYHSDEKPFHCPFPGCNLR FKRKSDAKKHLRIHVKKSNGVVNDLIAPRCSAFKVVPKKFI HTEVDGICYSCN	EMO_053350
E. invadens IP1	MQAGLEQSVNIIYDKFISKKDVKGKYVCPYMGCGKTYLSE ESLTRHLNRHSTVKEHVCDVCGKAFLRKSECEIHMRIHTG VKPFSCSLCNKKFARATDLKIHMVYHSDEKPFKCPFPGCTL SFKRKSDAKKHLRIHVKKTQCNFQSLITQVVPSAFSPIPKRF IKVEKTELVYDCN	EIN_359510

**Se muestra su correspondiente número de acceso de la plataforma AmoebaDB.*

Posteriormente y para conocer el grado de similitud entre las secuencias identificadas se realizó un alineamiento con la secuencia de aminoácidos del EhYY1 y las seis secuencias identificadas. Los alineamientos mostraron que las cepas *HM-1: IMSS*, *HM-3: IMSS*, *KU27* y las especies *SAW760*, *Laredo* y *IP1* presentan una secuencia YY1 altamente conservada como se observa en la Figura 15. El tamaño de las secuencias de aminoácidos es prácticamente la misma (175-176 aa) y presentan porcentajes de identidad y homología de 69 - 100 %. (Tabla 8). Asimismo, la composición y abundancia de aminoácidos de las secuencias YY1 en estas cepas y especies de Entamoeba se encuentra conservada (Figura 16).

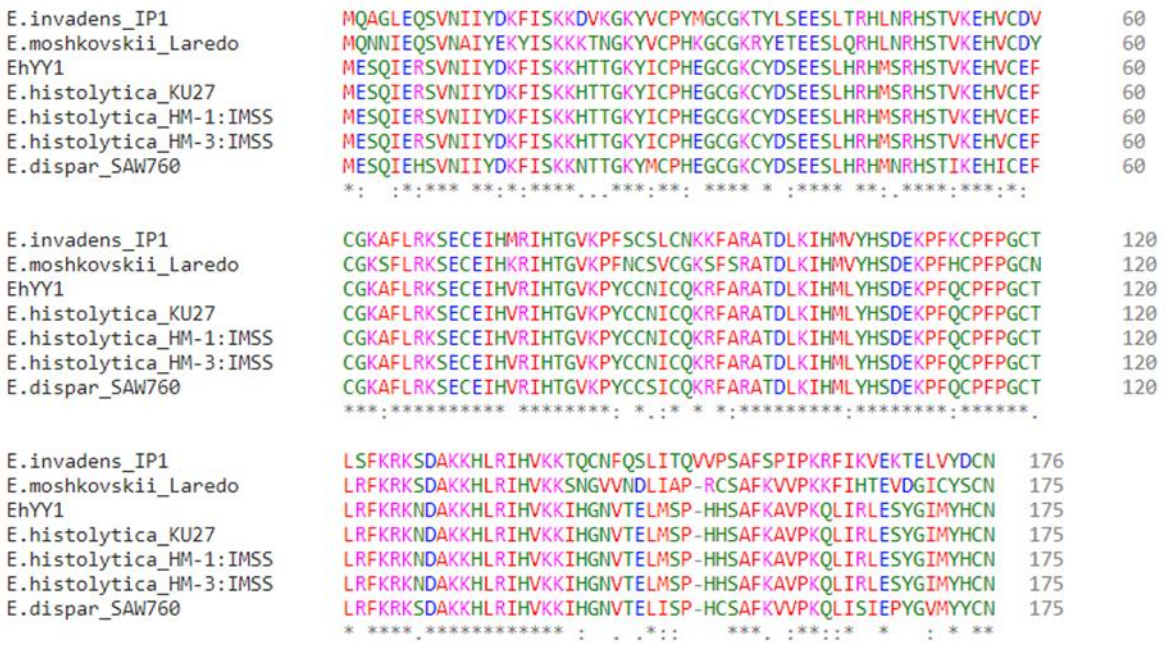


*Figura 15. Alineamiento de las secuencias de aminoácidos correspondientes a los genes YY1 de diferentes especies de Entamoeba con la secuencia de EhYY1. * se muestran los residuos idénticos, : se muestran los residuos semejantes y . se muestran los residuos que comparten las mismas propiedades fisicoquímicas*

En general la secuencia de aminoácidos se mantiene altamente conservada. Las especies *E. histolytica KU27*, *E. histolytica HM-1: IMSS*, *E. histolytica HM-3: IMSS y* tienen una similitud del cien 100% con EhYY1. Mientras que *E. moshkovskii Laredo*, *E. dispar SAW760* y *E. invadens* solo difieren en algunos aminoácidos, siendo este último quien presenta un aminoácido más que el resto de las secuencias (176 aa) (Figura 15, Tabla 8).

Tabla 8.Porcentaje de identidad y homología EhYY1.

	%	YY1	HM-1: IMSS	HM-3: IMSS	SAW760	Laredo	IP1	KU27	Humana
YY1	Identity	100							
	Homology	100							
HM-1: IMSS	Identity	100	100						
	Homology	100	100						
HM-3: IMSS	Identity	100	100	100					
	Homology	100	100	100					
SAW760	Identity	90.86	90.86	90.86	100				
	Homology	100	100	100	100				
Laredo	Identity	71.43	71.43	71.43	73.14	100			
	Homology	100	100	100	100	100			
IP1	Identity	69.32	69.32	69.32	70.45	71.59	100		
	Homology	100	100	100	100	100	100		
KU27	Identity	100	100	100	90.86	71.43	69.32	100	
	Homology	100	100	100	100	100	100	100	
Humana	Identity	37.17	37.17	37.17	33.87	40.17	35.48	37.17	100
	Homology	66	66	66	71	66	69	66	100

**Se muestra una tabla comparativa entre las diferentes especies de Entamoeba.*

Con el objetivo de profundizar en la comprensión de las similitudes entre las proteínas YY1 de diferentes cepas y especies de *Entamoeba*, se creó una tabla la cual presenta varios parámetros fisicoquímicos.

Tabla 9.Características fisicoquímicas EhYY1.

YY1	Gen ID	Secuencia Genómica (bp)	Exones	Intrones	aa	Mw (kDa)	pI	(GRAVY)	Instability index	Aliphatic index
E. histolytica HM-1: IMSS (EhYY1)	EHI_163860	528	1	-	175	20.5	9.33	-0.621	58.81	65.71
E. histolytica KU27	EHI5A_224080	528	1	-	175	20.5	9.33	-0.621	58.81	65.71
E. histolytica HM-3: IMSS	KM1_264880	528	1	-	175	20.5	9.33	-0.621	58.81	65.71
E. dispar SAW760	EDI_155470	528	1	-	175	20.4	9.14	-0.521	54.40	67.37
Laredo	EMO_053350	528	1	-	175	20.2	9.52	-0.722	45.81	62.29
E. invadens IP1	EIN_359510	531	1	-	176	20.2	9.47	-0.426	49.17	71.36

**Se muestra el ID, secuencia genómica, exones, intrones, aminoácidos, punto isoeléctrico, Gran Promedio de Hidropaticidad (GRAVY), índice de inestabilidad e índice alifático*

El pI de YY1 de las distintos cepas y especies de *Entamoeba* resulta ser un parametro altamente conservado mostrando una característica básica en la proteína, podemos observar que se mantiene en un rango de 9.14 a 9.52. Por otra parte el GRAVY de las secuencias YY1 se muestra en un rango de -0.426 a -0.722 lo que nos indica la naturaleza hidrofílica de la proteína.En cuanto al Índice de Inestabilidad las secuencias se encuentran en un rango de 45.81 a 58.81 es decir que la proteína no es estable fuera del ambiente celular. Finalmente, el Índice Alifático de estas secuencias muestra valores de entre 62.29 a 71.36 lo que nos habla de la buena estabilidad térmica de la proteína.

En este gráfico de abundancia, se puede apreciar que las distintas cepas y especies de *Entamoeba* exhiben cantidades de aminoácidos bastante similares entre sí. La diferencia más destacada reside en los porcentajes de histidinas e isoleucinas, que son menos abundantes en especies como *E. invadens* y *Laredo* en comparación con las cepas *E. histolytica SAW760*, *E. histolytica HM-3 IMSS* y *E. histolytica HM-1 IMSS* (EhYY1).

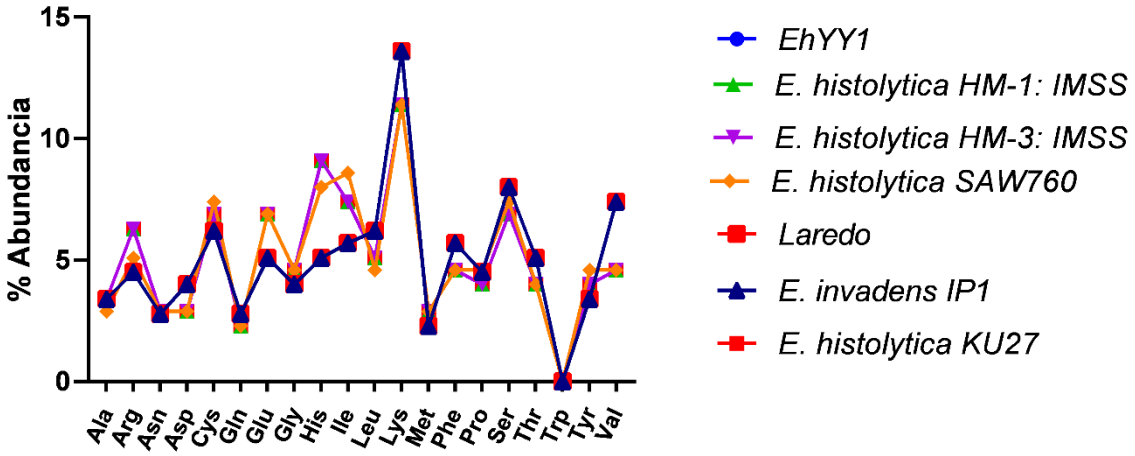


Figura 16.Porcentaje de abundancia de aminoácidos. Se muestra un gráfico comparativo de las diferentes secuencias de YY1 encontradas en Entamoeba y su abundancia relativa de aminoácidos.

6.4 Análisis filogenético

Debido a las similitudes observadas entre EhYY1 y las ZFP, así como con los factores YY1, se llevó a cabo un estudio evolutivo para esclarecer su origen y su ubicación genética. Construimos el árbol filogenético basado en las secuencias de aminoácidos de nueve factores de transcripción YY1 de otros organismos como el humano, primates, ratón, rata, anfibios, peces y la proteínas YY1 de amibas, además cinco proteínas de dedos de zinc (ZFP) con un mayor porcentaje de homología con EhYY1 provenientes de bacterias, arácnidos, peces y oomicetos (Figura 17).

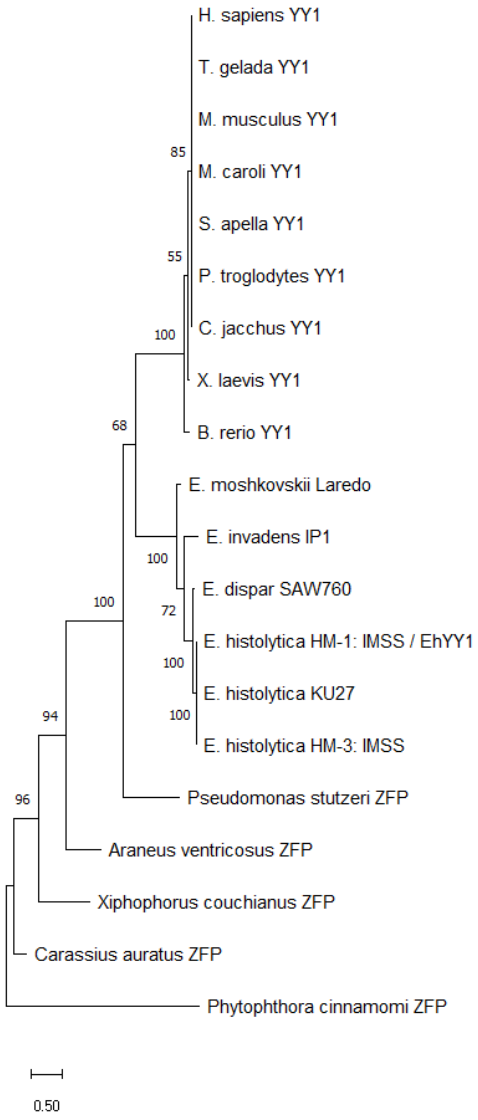


Figura 17. Árbol filogenético de secuencias YY1 y proteínas con dedos de zinc (ZFP)

El método empleado para la creación del árbol filogenético fue “máximum likelihood” el cual busca el árbol máximo verosímil, es decir, el árbol más probable que se generó con los datos de secuencias genéticas dados, utilizando un modelo de evolución y calculando la probabilidad de que los datos se ajusten a cada árbol propuesto. Luego, selecciona el árbol con la probabilidad más alta como la mejor estimación de la relación filogenética.

A su vez para evaluar la robustes del árbol filogenético creado se usó el método “Bootstrap” que es una técnica estadística que implica la creación de múltiples árboles filogenéticos a partir de muestras repetidas de los datos originales. Para este árbol se generaron 100 repeticiones, el número que se encuentra en cada nodo del árbol representa el "valor de bootstrap". Un valor de alto, por ejemplo, del 95%, indica una fuerte evidencia de que la rama en cuestión refleja una relación filogenética real y es, por lo tanto, más confiable.

El resultado muestra la formación de dos grandes clados que comparten un ancestro común (la ZFP de *Phytophthora cinnamomi*), en el primero se agrupan todos los factores de transcripción YY1 y las proteínas YY1 de Entamoeba. El segundo grupo, las proteínas de dedos de zinc (ZFP) se encuentran en las ramas inferiores lo que indica una separación evolutiva entre los factores de transcripción y las proteínas de dedos de zinc (ZFP).

6.5 Modelado de la proteína EhYY1

Se optó por utilizar inteligencia artificial para el modelado de EhYY1 debido a la limitación de no contar con una plantilla adecuada para el modelado por homología (el cristal proteico de HYY1 existente solo contiene la region de unión a DNA). Este enfoque tiene la ventaja de no depender de una estructura de proteína existente como plantilla, ya que se basa en técnicas de aprendizaje profundo y utiliza extensos conjuntos de datos que contienen secuencias y estructuras de proteínas para predecir con alta precisión las estructuras tridimensionales de las proteínas a partir de su secuencia de aminoácidos.

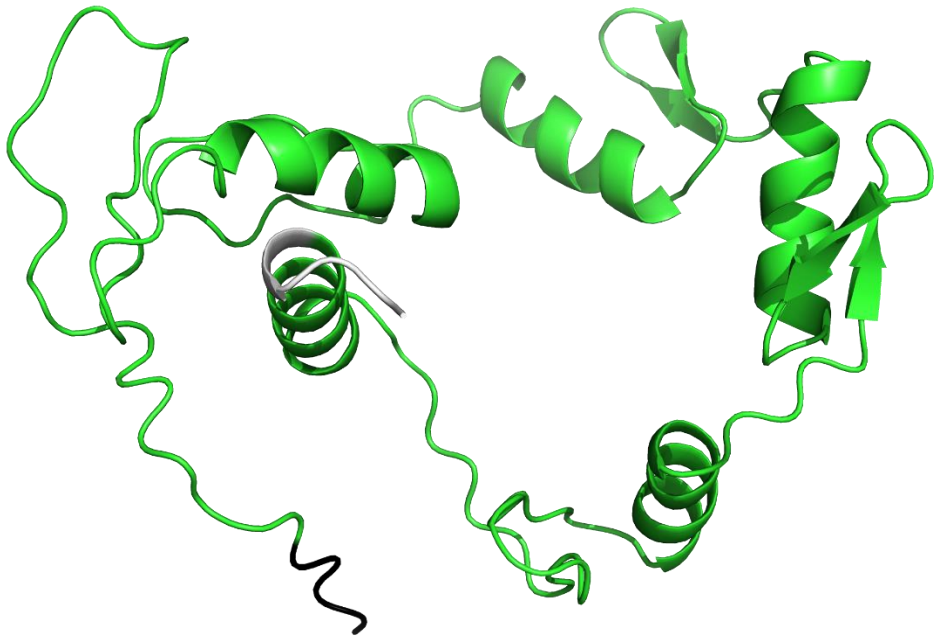


Figura 18. Modelo 3D de la proteína EhYY1. Se observa la proteína completa de EhYY1 (175 aa), extremo amino terminal en blanco y extremo carboxilo terminal en negro.

El resultado obtenido es una proteína de 175 aa con una alta presencia de" Loops" en su estructura, estas son difíciles de obtener experimentalmente por cristalografía pues al ser estructuras flexibles pueden adoptar múltiples conformaciones.

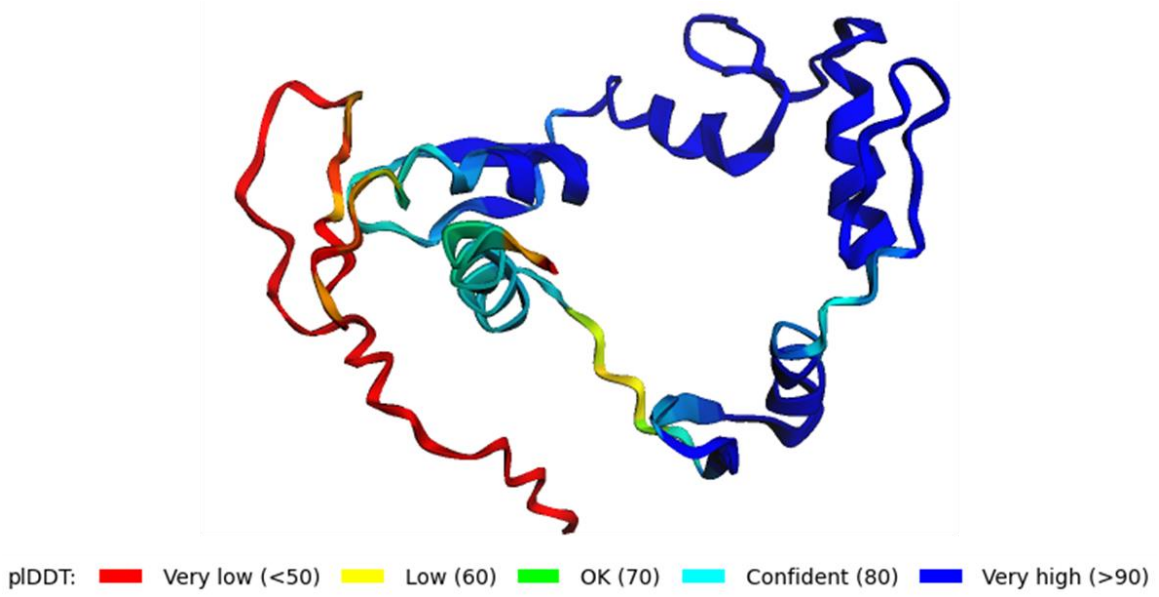


Figura 19. Prueba lDTT para EhYY1. En la parte inferior se muestran con diferentes colores los valores de confianza para cada zona de EhYY1.

Para evaluar la calidad del modelo generado de EhYY1, se utilizó la prueba lDDT (Local Distance Difference Test), que se utiliza para comparar la similitud entre la estructura predicha de una proteína (EhYY1) y la estructura real de proteínas homologas e incluye la evaluación de la plausibilidad estereoquímica (Mariani, 2013). Un mayor porcentaje de lDDT indica una mayor similitud y, por ende, una mejor calidad del modelo. En el caso de EhYY1, se observa que el extremo carboxilo terminal muestra un porcentaje "muy bajo". Esto se debe a que, como se mencionó, las estructuras de Loop son flexibles y pueden adoptar diversas conformaciones. Por otro lado, la región de dedos de zinc se modeló con una confianza "muy alta", indicando una representación precisa de esta parte crucial de la estructura de la proteína.

Por la fiabilidad del modelado por homología se decidió utilizar este método para el modelado del sitio de unión a DNA. Se obtuvieron 7 modelos, la proteína fue modelada a partir del dominio de unión a DNA del cristal de HYY1 que presenta 114 aminoácidos, Swiss Model utilizó 113 aa (25-137 pb) para el modelado, ya que estos corresponden con las características fisicoquímicas de la proteína usada como plantilla:

KYICPHEGCGKCYDSEESLHRHMSRHSTVKEHVCEFCGKAFLRKSECEIHVRIHTG
VKPYCCNICQKRFARATDLKIHMLYHSDEKPFQCPFPGCTLRFKRKNDAKKHLRIH
V

Secuencia 4. Dominio de unión a DNA de EhYY1. Secuencia seleccionada para modelar a EhYY1 (25-137 pb)

No obstante, seleccionamos el modelo número 7 debido a que demostró la mejor calidad entre los modelos generados (Figura 20). Esta elección se respalda con una serie de pruebas diseñadas para evaluar la calidad del modelo. Entre estas pruebas se incluyen el Diagrama de Ramachandran y el Z-score, que proporcionaron indicadores claros de la calidad y validez de la estructura tridimensional predicha. Estos análisis confirmaron que el modelo 7 era el más confiable y representaba de manera precisa la estructura de EhYY1 en términos de conformación y estabilidad.

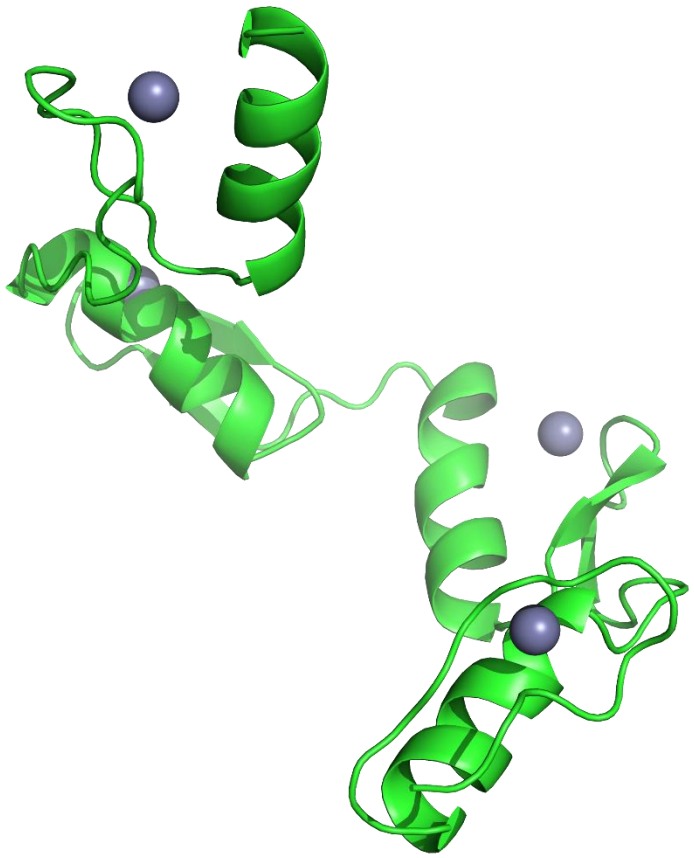


Figura 20. Dominio de unión al DNA de EhYY1. Se muestra en verde la estructura proteica y en gris el Zn al que se encuentra enlazado.

El diagrama de Ramachandran proporciona información crucial sobre la estructura secundaria probable de una proteína. En este diagrama, los ángulos Phi están representados en el eje x, mientras que los ángulos Psi están en el eje y. Los puntos azules en el diagrama representan aminoácidos y su ubicación en los cuadrantes indica las posibles estructuras secundarias. La hélice α izquierda se observa en el primer cuadrante, las hoja β en el segundo cuadrante, la hélice α derecha y los giros o bucles en el tercer cuadrante. El cuarto cuadrante representa una zona desfavorecida, es decir, combinaciones estructuralmente incompatibles (Ramachandran, 1963). Este análisis ayuda a prever las estructuras secundarias probables del péptido en función de los ángulos Phi y Psi.

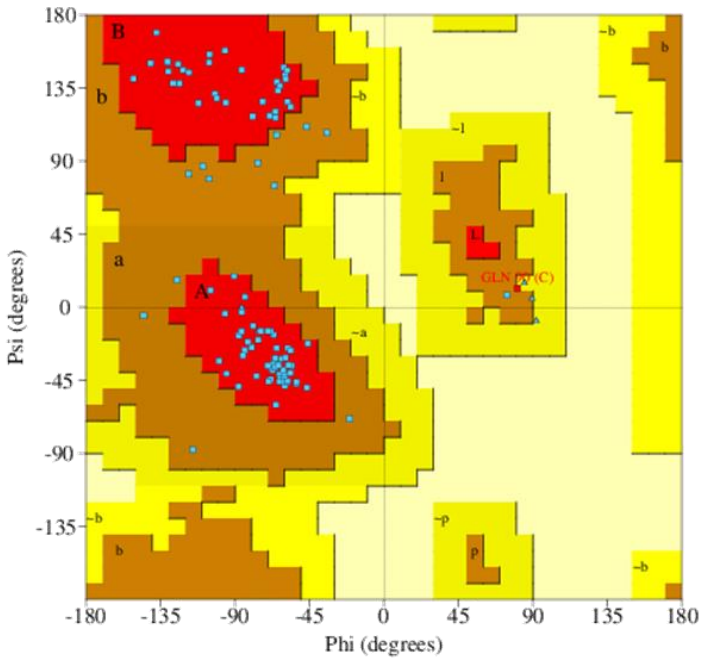


Figura 21. Diagrama de Ramachandran EhYY1. RFM (rojo), RAP (Café), RGP (amarillo) y RN (amarillo claro)

Al calcular los porcentajes en distintas regiones del gráfico, se encontró que el 87.1% de las conformaciones carecen de impedimentos estéricos, clasificándolas como Regiones más Favorecidas (RFM). Las Regiones Adicionales Permitidas (RAP) representan el 11.9%, y las Regiones Generosamente Permitidas (RGP) solo el 1%. No se identificaron conformaciones en Regiones No Permitidas (RN), indicando ausencia de impedimentos estéricos significativos en los aminoácidos. Además, se incorporó el factor G (valor fuera del gráfico), que proporciona una medida de rareza de una propiedad. Un valor inferior a -0.5 se considera inusual; sin embargo, el valor obtenido fue -0.15, indicando que las propiedades de la proteína son comunes y no atípicas.

Adicionalmente, se realizó un análisis PROSA (Protein Structure Analysis), una herramienta crucial para validar y evaluar modelos tridimensionales de proteínas derivados de predicciones o modelado molecular. Este análisis compara estructuras y calcula un puntaje Z (Z-score) que indica la diferencia entre la energía potencial de nuestro modelo y la de estructuras experimentales conocidas obtenidas mediante rayos X o resonancia magnética nuclear (Wiederstein, 2007).

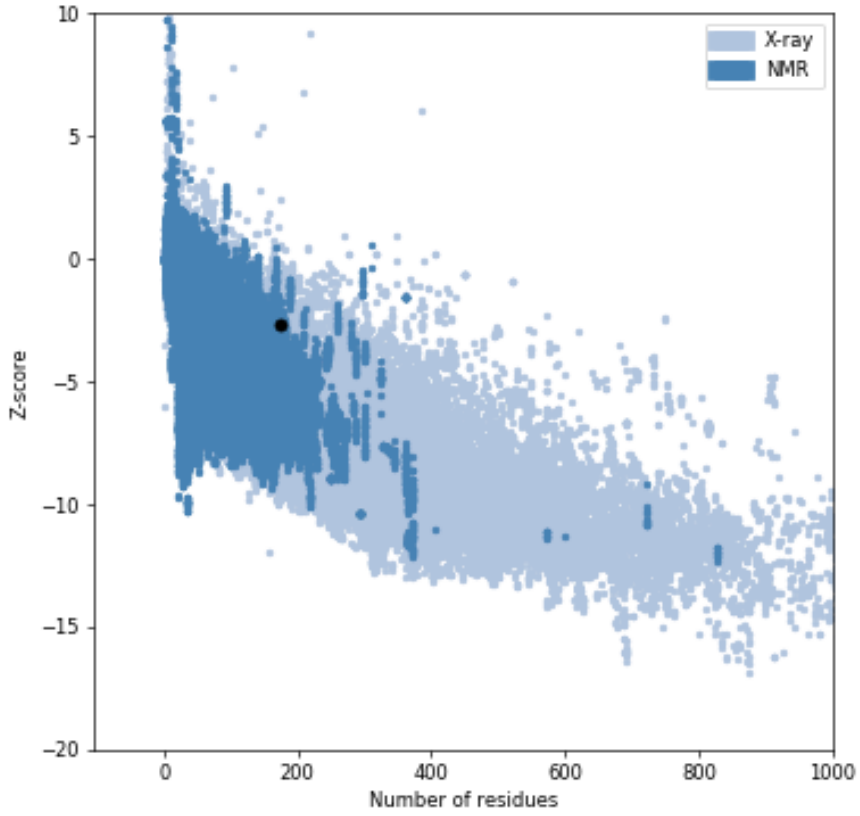


Figura 22. Z-Score de EhYY1. El eje X representa el número de residuos, mientras que el eje Y muestra el Z-score. En la posición de 175 residuos, el Z-score es -3.47 y corresponde a EhYY1 (punto negro) En la esquina superior derecha, se emplean colores distintos para representar los datos experimentales de Rayos-X y RMN.

Un Z-score negativo, como el valor obtenido de -3.47 en nuestro modelo, indica que la estructura tiene una energía potencial baja y, por lo tanto, favorable. Esto sugiere que el modelo tridimensional generado es de alta calidad y óptimo en términos de conformidad con las estructuras experimentales conocidas.

6.6 Análisis estructural de EhYY1

En el factor EhYY1 se identificaron cuatro dedos de zinc en las posiciones 26-55, 56-83, 84-111 y 112-141 (Tabla 6), los cuales conforman el dominio de unión al DNA de la familia de factores de transcripción YY1 y se caracterizan por tener dos láminas beta antiparalelas que forman una horquilla y una hélice alfa, que se denota como ββα, dicha estructura secundaria se estabiliza por interacciones hidrofóbicas y por la coordinación de un átomo de zinc dada por dos residuos de cisteína y dos de histidina (Cys2His2) (Kuida, 2019). Además, hemos detectado la presencia de sitios de fosforilación, SUMOilación y acetilación mismos que pueden encontrar en HYY1 (Figura 23).

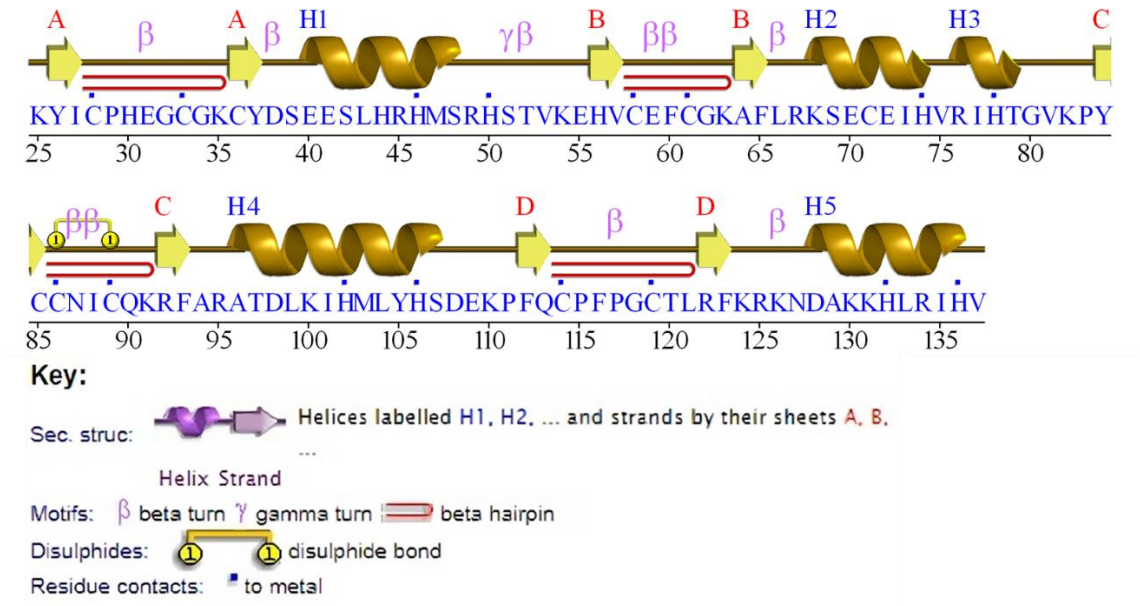


Figura 23. Estructura secundaria EhYY1. Se muestran las estructura secundaria de EhYY1 asi como los aminoácidos que la componen, en la parte inferior de la imagen se muestra la simbología de las estructuras presentadas.

Los elementos estructurales identificados para EhYY1 a través de un escaneo en PDBsum (Figura 23) incluyen:

Dedos de Zinc: EhYY1 contiene cuatro dedos de zinc en las posiciones 26-55, 56-83, 84-111 y 112-141. Los dedos de zinc son dominios estructurales que pueden unirse específicamente al DNA y están involucrados en la regulación de la expresión génica.

Estructuras de Hojas Beta: La proteína presenta 4 hojas beta en su estructura secundaria. Las hojas beta son estructuras plegadas que forman láminas en la estructura tridimensional de la proteína.

Horquillas Beta y Protuberancias Beta: Además de las hojas beta, EhYY1 incluye 4 horquillas beta y 2 protuberancias beta. Estas son características comunes en las proteínas que contribuyen a su conformación tridimensional única.

Láminas beta: Se observa la presencia de 8 de ellas, pueden estar involucradas en interacciones proteína-proteína, ya que las regiones de láminas beta pueden actuar como superficies de unión para otras proteínas o ligandos.

Hélices: Se identifican 5 hélices en la estructura secundaria de EhYY1. Las hélices son estructuras enrolladas que también desempeñan un papel importante en la conformación y función de las proteínas.

Interacciones Hélices-Hélices: Se han observado 8 interacciones hélices-hélices en la proteína. Estas interacciones pueden ser cruciales para estabilizar la estructura global de EhYY1.

Giros Beta y Giro Gamma: La proteína cuenta con 10 giros beta y 1 giro gamma en su estructura secundaria. Estos giros son regiones que conectan diferentes elementos de estructura secundaria y permiten la flexibilidad y adaptación de la proteína.

Puentes Disulfuro: Se ha detectado un puente disulfuro en EhYY1. Los puentes disulfuro son enlaces covalentes entre dos residuos de cisteína y pueden contribuir a la estabilidad de la estructura proteica.

La presencia de estos elementos estructurales sugiere una proteína altamente organizada y funcionalmente importante, especialmente en términos de su capacidad de unión al DNA y su posible papel en la regulación génica

Tabla 10. Resumen de la estructura secundaria EhYY1. Se muestra en la tabla la estructura y su cantidad correspondientes.

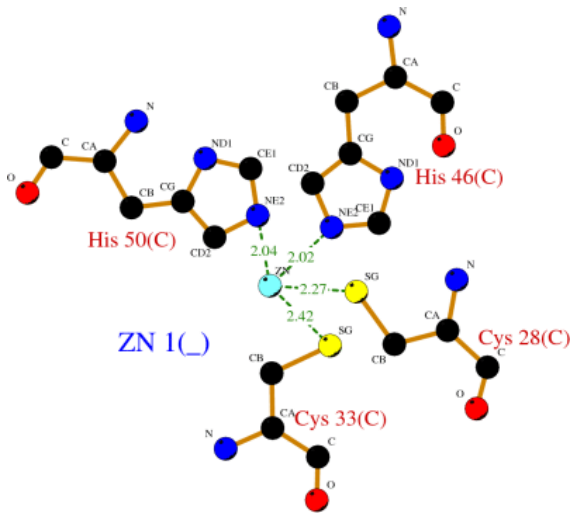
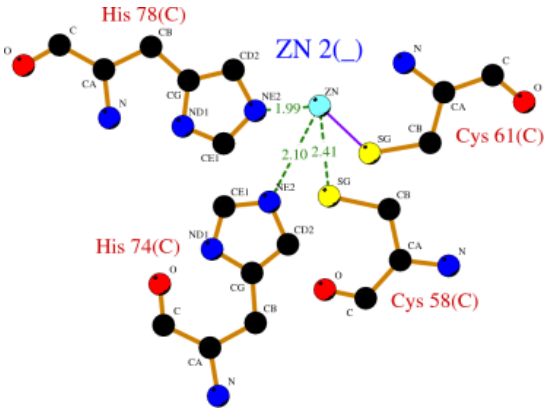
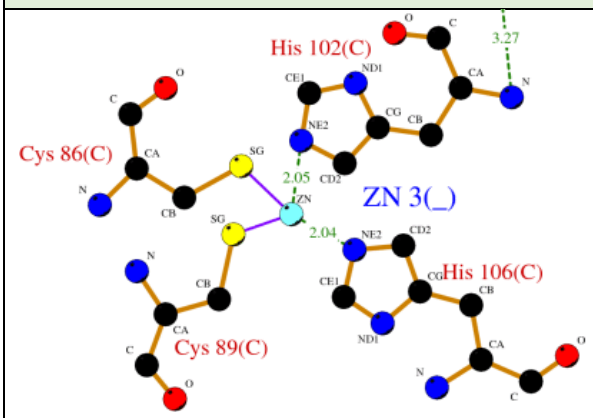
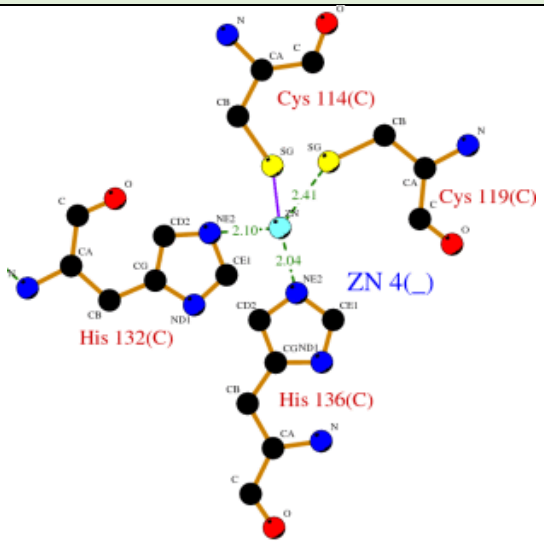
Estructura	Cantidad
Hojas Beta	4
Horquillas Beta	4
Protuberancias Beta	2
Laminas Beta	8
Hélices	5
Interacciones hélices-hélices	8
Giro Beta	10
Giro Gamma	1
Puentes disulfuro	1

.

Es importante destacar que la predicción de la estructura secundaria se realizó solo para el sitio de unión a DNA de EhYY1 y se basa en datos de PDBsum. Sin embargo, se señala que otras regiones de la proteína pueden tener estructuras más flexibles o en loop, lo que hace que su predicción sea más desafiante con métodos convencionales ya no tienen una estructura rígida.

Además, PDBsum generó diagramas de los aminoácidos que conforman los cuatro dedos de zinc presentes en EhYY1. Estos dedos son del tipo C2H2, característicos de los factores de transcripción YY1 y se destacan en los diagramas en color rojo. En contraste, el zinc (Zn) al que están unidos se muestra en color azul.

Tabla 11. Dedos de zinc EhYY1. Se muestran un diagrama de los cuatro dedos de zinc asi como la posición de las cisteínas e histidinas que forman los dedos C2H2.

Zn 1	Zn 2
	
Cys 28/ Cys 33 / His 46 / His 50	Cys 58/ Cys 61 / His 74 / His 78
Zn 3	**Zn 4**
	
Cys 86 / Cys 89 / His 102 / His 106	Cys 114 / Cys 119 / His 132 / His 136

Adicionalmente, se llevó a cabo un análisis de la secuencia de EhYY1 para identificar algunos sitios potenciales de los procesos postraduccionales más importantes como son la fosforilación, la SUMOilación y la acetilación.

La fosforilación es un proceso biológico fundamental en el que se agrega un grupo fosfato (PO4) a una proteína. Esta adición de fósforo puede tener un impacto significativo en la función de la proteína, activándola, desactivándola o alterando su función de alguna manera. Los aminoácidos que son especialmente propensos a la fosforilación son la serina, la treonina y la tirosina, ya que contienen grupos hidroxilo (-OH) que pueden ser fosforilados por enzimas llamadas quinasas. Este análisis permitió identificar posibles sitios en la secuencia de EhYY1 donde la fosforilación podría ocurrir y desempeñar un papel en la regulación de su actividad biológica (Spiegato, 2023).

Tabla 12. Sitios de fosforilación. Se muestra 5 columnas, posición, aminoácido, secuencia (rojo la ubicación del aminoácido dentro de la secuencia), Score y la quinasa.

SITIOS DE FOSFORILACIÓN				
POSICIÓN	AA	SECUENCIA	SCORE	QUINASA
3	S	MESQIER	0.554	ATM
			0.528	CKII
			0.516	DNAPK
13	Y	VNIIYDKFI	0.552	INSR
18	S	DKFISKKHT	0.934	UNSP
			0.653	PKC
22	T	SKKHTTGKY	0.720	PKC
23	T	KKHTTGKYI	0.759	PKC
26	Y	TTGKYICPH	0.735	UNSP
			0.515	INSR
39	S	KCYDSEESL	0.982	UNSP
42	S	DSEESLHRH	0.851	UNSP
48	S	HRHMSRHST	0.957	UNSP

			0.579	PKA
51	S	MSRHSTVKE	0.971	UNSP
			0.546	PKC
			0.536	PKA
52	T	SRHSTVKEH	0.987	UNSP
			0.889	PKC
69	S	FLRKSECEI	0.993	UNSP
			0.546	RSK
79	T	VRIHTGVKP	0.682	UNSP
97	T	FARATDLKI	0.777	PKC
107	S	MLYHSDEKP	0.837	UNSP
120	T	FPGCTLRFK	0.508	PKC
149	S	TELMSPHHS	0.806	UNSP
			0.586	CDK5
			0.557	P38MAPK
			0.506	GSK3
167	S	IRLESYGIM	0.538	PKA
			0.504	RSK
168	Y	RLESYGIMY	0.757	UNSP

Los resultados del análisis revelaron la presencia de 19 sitios potenciales de fosforilación en la secuencia de EhYY1. El "Score" de predicción proporciona un valor en un rango de 0.000 a 1.000, y las puntuaciones superiores a 0.500 indican predicciones positivas para las quinasas que podrían fosforilar los aminoácidos en esos sitios. La distribución de estos sitios se puede observar con mayor detalle en Apéndice I.

También, se llevó a cabo la búsqueda de sitios de SUMOilación en EhYY1. La SUMOilación implica la adición de una pequeña proteína llamada SUMO (Small Ubiquitin-related Modifier) a una proteína objetivo y puede tener una serie de efectos diversos que dependen de la naturaleza de la proteína modificada. Entre las consecuencias de la SUMOilación se incluyen la regulación de la transcripción génica, la modulación de la proliferación celular, la participación en la reparación del DNA, la regulación de la degradación de proteínas y la localización nuclear de las proteínas (Verger, 2003).

Tabla 13. Sitios de SUMOilación EhYY1. Se presentan cuatro columnas, posición, péptido (rojo los aminoácidos con los que interacciona SUMO), score, P-value y tipo

Sitios de SUMOilación				
Posición	Péptido	Score	P-value	Tipo
162 - 166	FKAVPKQLIRLESYGIMYH	36.919	0.077	SUMO Interacción

Se detecto la presencia de un sitio de interacción SUMO en la posición 162-166 con un alto score y un P-value de < 0.5 es decir los resultados son estadísticamente significativos, es altamente probable la presencia de un sitio de SUMOilación en EhYY1.

Otro de los procesos postraduccionales más importantes es la acetilación, es decir la adición de grupos acetilo a residuos de lisina (K). Esto puede afectar la estructura de cromatina y la actividad de las proteínas histonas, la acetilación de proteínas no histonas puede modificar la actividad de las proteínas, ya sea activándolas o inhibiéndolas. Esto puede afectar diversas funciones celulares, como la señalización, la regulación de la transcripción y la estabilidad de las proteínas.

Tabla 14.Sitios de acetilación para EhYY1. Se presentan cinco columnas, posición, péptido (color rojo el aminoácido donde se realiza este proceso), HAT (histona acetiltransferasa), score y cutoff.

Sitios de acetilación				
Posición	Péptido	HAT	Score	Cutoff
139	HLRIHVKKIHGNVTE	KAT8	9.2	7.222

Se identificó un sitio de acetilación en EhYY1 en la posición 139, y la enzima responsable de este proceso de acetilación es la histona acetiltransferasa (HAT) denominada "KAT8". Con un punto de corte de 7.222 y un score de 9.2, es altamente probable la existencia de este sitio de acetilación.

A continuación, se muestra esquema de las posiciones de los distintos procesos postraduccionales de EhYY1:

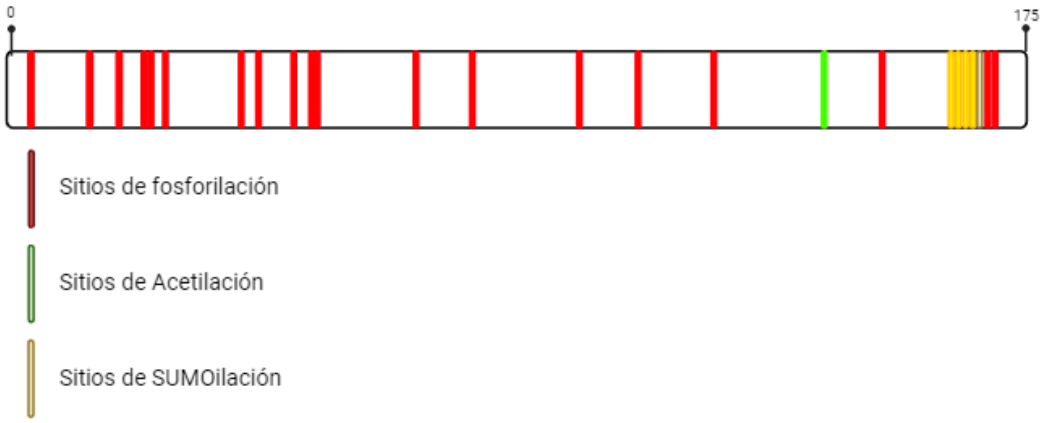


Figura 24. Ubicación de procesos postraduccionales en EhYY1. Aminoácidos 0 y 175 en las esquinas superiores. Sitios de fosforilación (rojo), sitos de SUMOilación (amarillo) y sitios de acetilación (verde).

6.7 Obtención de secuencia consenso de unión a DNA

Debido a la existencia de múltiples secuencias de unión a YY1 documentadas, se tomó la decisión de generar una secuencia consenso a partir de estas (Figura 25). Se procedió al alineamiento múltiple de estas secuencias, permitiendo identificar los nucleótidos más comúnmente conservados. Esta secuencia consenso resultante será empleada en los ensayos de acoplamiento molecular (Docking) que se detallarán posteriormente.

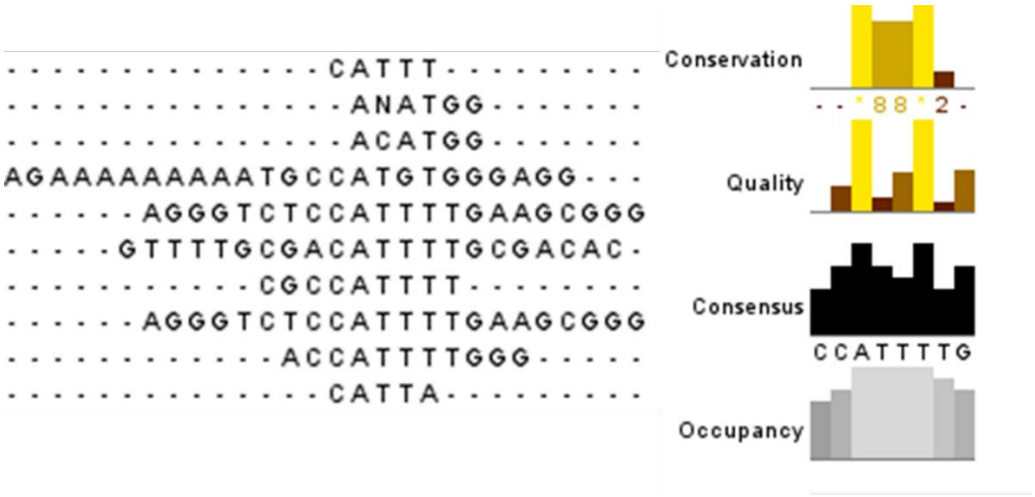


Figura 25. Alineamientos secuencia consenso de unión a DNA para YY1. En la sección izquierda se muestra el alineamiento de secuencias de unión a DNA de YY1, en la sección derecha los datos del alineamiento como, conservación, calidad, consenso y ocupación.

La secuencia generada a partir del alineamiento tiene una longitud de 8 pares de bases y es "**CCATTTTG**". Para una visualización más clara, se utilizó un logotipo de secuencia que representa gráficamente el alineamiento. En este gráfico, la altura de los nucleótidos en cada columna refleja la frecuencia relativa de cada ácido nucleico en esa posición.

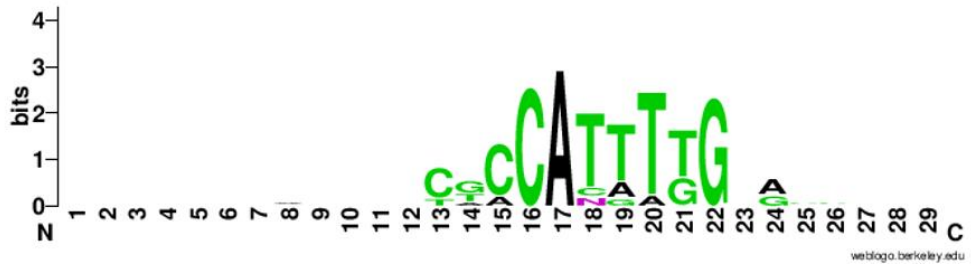


Figura 26. Logotipo de secuencia consenso YY1. Muestra con letras de mayor tamaño el nucleótido más conservador en cada posición.

6.8 Docking molecular Proteína-DNA

Con la finalidad de analizar la interacción Proteína-DNA entre el factor EhYY1 y su secuencia consenso, realizamos ensayos de Docking molecular, con el modelo previamente obtenido (Figura 20) y la secuencia de unión para YY1 (11 pb) localizada en el promotor del gen *EhPgp5* de *E. histolytica* (- 170 hasta - 136 pb) (Figura 27).

Figura 27. Modelo de la secuencia de unión a DNA. En rojo se muestra los 11 pb modelados de la secuencia YY1.

Los resultados mostraron la unión del factor EhYY1 a su secuencia consenso (Figura 28) con una energía de unión de ΔG 460 kcal/mol Esta unión se da a través de la interacción de los aa (5' – 3') H102, F93, R92, K82, A94 R76, I77, R96, R49 y R67 con ocho nucleótidos (CCATTTTG).

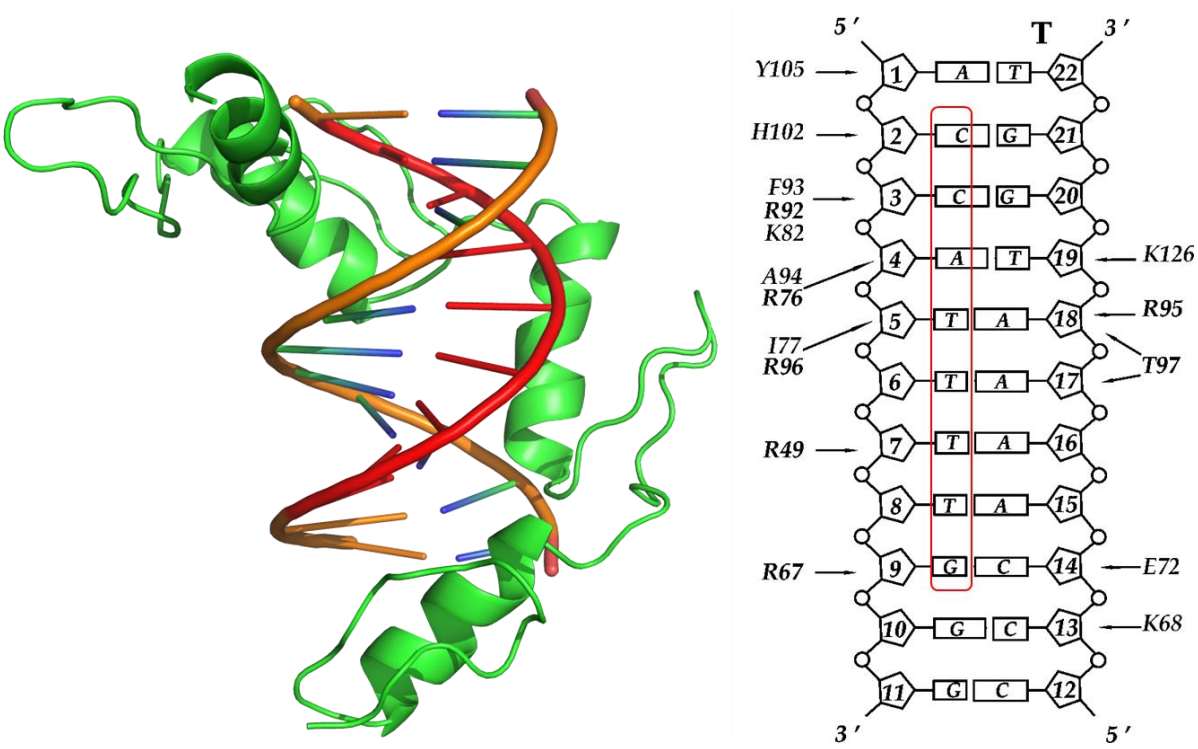


Figura 28. Docking molecular EhYY1- EhPgp5. En la parte izquierda se muestra la interacción de EhYY1 con su secuencia promotora en EhPgp5, de lado derecho una representación de las interacciones de los aminoácidos con sus respectivos nucleótidos. En rojo se encuentra resalta la secuencia consenso de unión a DNA.

Se observa la interacción de los aminoácidos de EhYY1 con la secuencia promotora de *EhPgp5*, que consta de 11 pares de bases, en particular con los 8 pares de bases que conforman nuestra secuencia consenso de unión al DNA "CCATTTTG" predicha previamente. Además, se nota la interacción con los aminoácidos K126, R95, T97 y E72 en la cadena de DNA complementaria.

Tabla 15. Interacciones de EhYY1 con el promotor de EhPgp5. Se presenta una tabla resumen que detalla las distancias entre los aminoácidos y los nucleótidos para ambas cadenas de DNA. En negrillas la secuencia consenso de unión a DNA.

pb (5' – 3')	AMINOÁCIDO	DISTANCIA (Å)	pb (3' – 5')	AMINOÁCIDO	DISTANCIA (Å)
A (1)	TYR (105)	2.2	T (22)		
C (2)	HIS (102)	1.8	G (21)		
C (3)	PHE (93)	2.8	G (20)		
	ARG (92)	2.9			
	LYS (82)	2.6			
A (4)	ALA (94)	2.8	T (19)	LYS (126)	2.3
	ARG (76)	2.9			
T (5)	ILE (77)	2.9	A (18)	ARG (95)	2.3
	ARG (96)	2.3			
T (6)			A (17)	ARG (95)	2.3
				THR (97)	3.4
T (7)	ARG (49)	2.8	A (16)	THR (97)	3.2
T (8)			A (15)		
G (9)	ARG (67)	2.5	C (14)	GLU (72)	2.9
G (10)			C (13)	LYS (68)	2.5
G (11)			C (12)		

Finalmente, es relevante mencionar que este acoplamiento evidencía similitudes en los nucleótidos con los que EhYY1 y HYY1 interaccionan. La estructura cristalográfica de HYY1 (1UBD) muestra un marcado parecido en la segunda citosina de la región de unión al DNA [5′- C**C**ATTTTG - 3′]. En este punto, las interacciones compartidas con arginina y fenilalanina son notablemente coincidentes (como se detalla en la tabla de interacciones de HYY1 en el apéndice 2 y 3).

Con el propósito de determinar si el factor EhYY1 puede también interactuar con la secuencia consenso de YY1 en humanos, llevamos a cabo un acoplamiento molecular. De manera notable, EhYY1 demostró la capacidad de interactuar con los mismos 8 nucleótidos que conforman el núcleo de la secuencia HYY1 (humana), junto con los aminoácidos H46, R49, F65, E70, I73 y H74, el modelo generado tuvo un ΔG de 433 Kcal/mol.

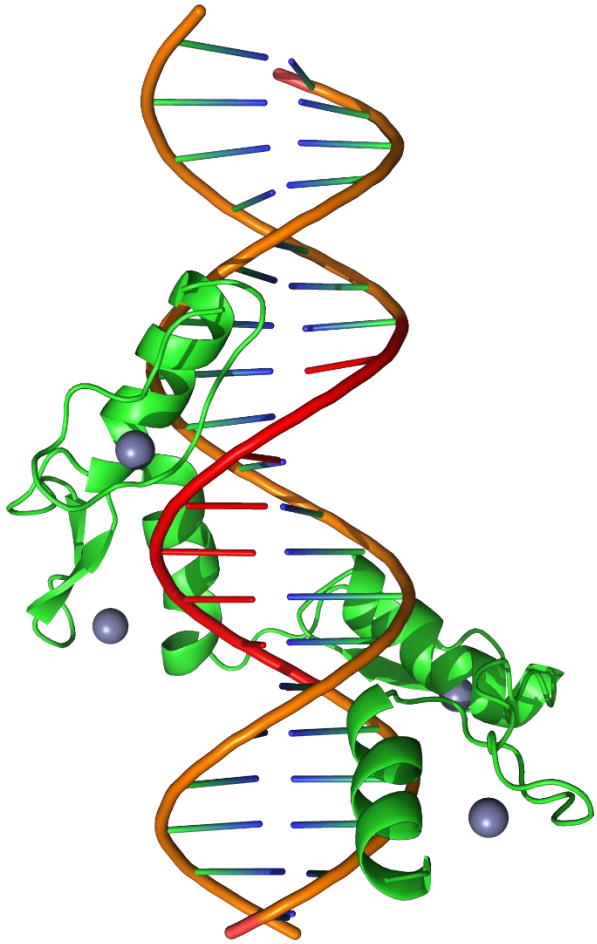


Figura 29. Docking molecular EhYY1- DNA humano. Se muestra la interacción de EhYY1 acoplada a la secuencia de DNA obtenida de 1UDB. En rojo se encuentra resaltada la secuencia consenso de unión a DNA "CCATTTTG".

Nota: Aunque se menciona como "DNA humano", la secuencia de DNA obtenida del cristal proteico (1UBD) corresponde en realidad a la proteína HYY1 (humana) unida a la región promotora P5 de adenovirus. Esta elección se debe a que es la única estructura cristalizada de HYY1 disponible que además se une a una región DNA y no niega la capacidad de HYY1 para identificar la secuencia.

Tabla 16. Interacciones de EhYY1 con el promotor de EhPgp5. Se presenta una tabla resumen que detalla las distancias entre los aminoácidos y los nucleótidos para ambas cadenas de DNA. En negrillas la secuencia consenso de unión a DNA.

pb (5'–3')	AMINOÁCIDO	DISTANCIA (Å)	pb (3'–5')	AMINOÁCIDO	DISTANCIA (Å)
A (1)			T (40)		
G (2)			C (39)		
G (3)			C (38)		
G (4)			C (37)		
T (5)			A (36)		
C (6)			G (35)	GLU (41)	1.6
T (7)			A (34)	ARG (45)	2.5
C (8)			G (33)		
C (9)			G (32)	SER (69)	2.9
A (10)			T (31)		
T (11)	HIS (46)	2.9	A (30)	ALA (96)	2.8
	ARG (49)	2.9			
T (12)			A (29)	THR (97)	1.9
				LYS (100)	2.1
T (13)	PHE (65)	2.3	A (28)		
	GLU (70)	2.4			
T (14)	ILE (73)	2.9	A (27)	LYS (126)	2.8
	HIS (74)	2.4			
G (15)			C (26)		
A (16)	PHE (93)	2.2	T (25)		
A (17)			T (24)		
G (18)	ASP (128)	2.9	C (23)		
C (19)	HIS (132)	2.9	G (22)		
G (20)			C (21)		

Más aún al evaluar si el factor YY1 de humano reconoce la secuencia de unión de YY1 de la amiba, observamos que esta interacción se vuelve a presentar con los mismos 8 pares de bases (Figura 30). Estos resultados muestran que *E. histolytica* presenta una proteína YY1 y además que, tanto el sitio de unión al DNA de los factores YY1 de este parásito y de humano se conservan, así como los nucleótidos que forman la secuencia consenso en el DNA.

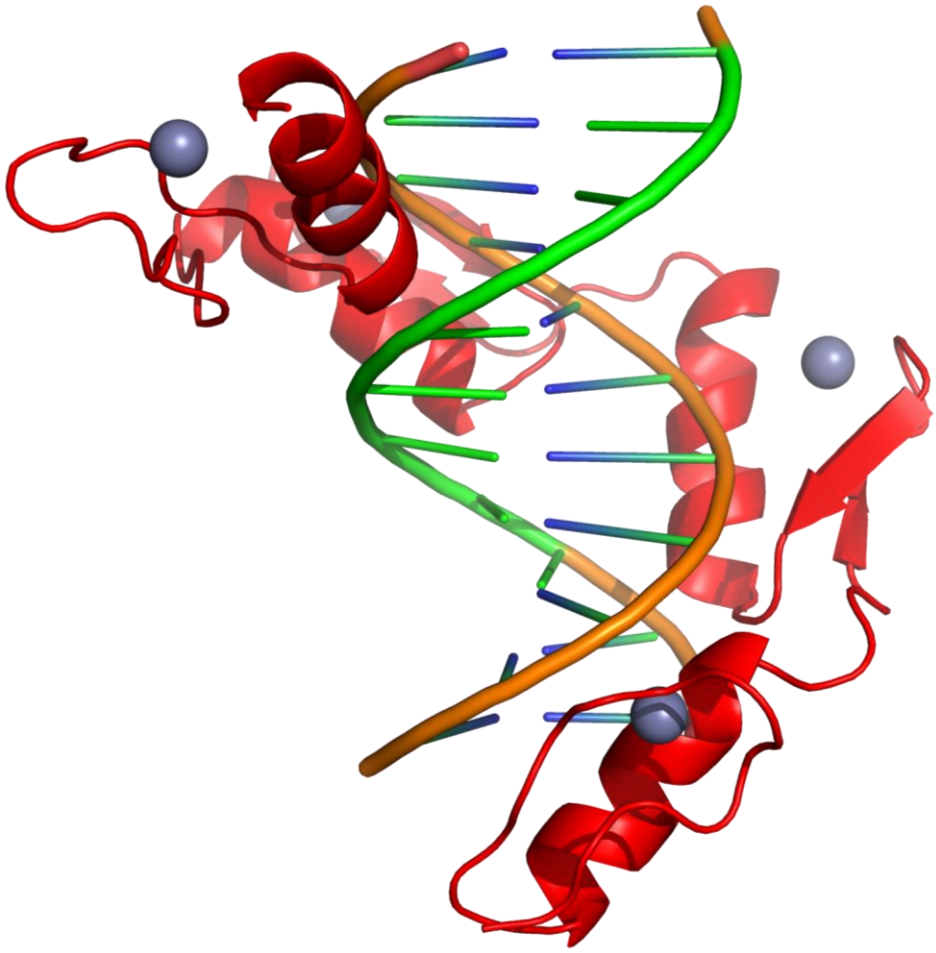


Figura 30. Docking molecular HYY1- Secuencia del promotor EhPgp5. Se muestra la interacción de HYY1(humano) acoplada a la secuencia de DNA de E. histolytica. En verde se encuentra resalta la secuencia consenso de unión a DNA "CCATTTTG".

Los resultados revelaron una ΔG de 240 Kcal/mol para el modelo generado, además de las interacciones T398, S397, F368 y D367 en la secuencia consenso de unión al DNA.

Tabla 17. Interacciones de HYY1- Secuencia del promotor EhPgp5. Se presenta una tabla resumen que detalla las distancias entre los aminoácidos y los nucleótidos para ambas cadenas de DNA. En negrillas la secuencia consenso de unión a DNA.

pb (5' – 3')	AMINOÁCIDO	DISTANCIA (Å)	pb (3' – 5')	AMINOÁCIDO	DISTANCIA (Å)
A (1)			T (22)	HIS (403)	2.7
C (2)			G (21)	ASN (399)	2.4
C (3)			G (20)	HIS (373)	2.6
A (4)			T (19)	PHE (364)	3.0
				ASN (369)	2.4
T (5)	THR (398)	2.9	A (18)	LEU (345)	2.7
T (6)	SER (397)	2.4	A (17)		
T (7)			A (16)		
T (8)	PHE (368)	2.9	A (15)		
G (9)	ASP (367)	2.9	C (14)	THR (319)	2.8
G (10)	SER (338)	2.6	C (13)		
G (11)	SER (337)	2.8	C (12)		
	HIS (318)	2.9			

Además de ello, al superponer los factores de transcripción de EhYY1 y HYY1 utilizando Raptor X, se obtuvo un porcentaje de homología entre ellas del 98.6%. Esto indica que no solo comparten aminoácidos con propiedades similares, sino que también presentan una estructura altamente coincidente.

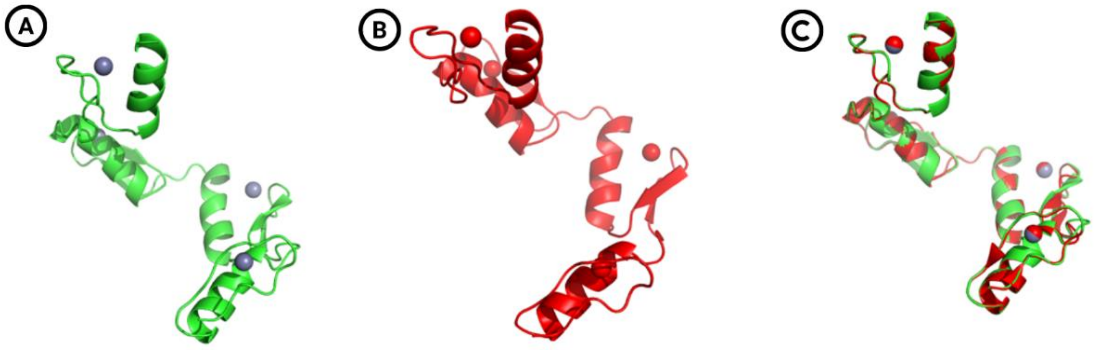


Figura 31.Comparación estructural EhYY1 y HYY1. En verde factor de transcripción EhYY1 (A), Rojo factor de transcripción HYY1 (B), superposición de ambos factores (C).

La imagen resultante de la superposición ilustra cómo ambos factores encajan en gran parte de su estructura. La mayoría de los residuos de aminoácidos de una proteína se alinean y superponen con los de la otra. Ambos factores muestran una similitud casi idéntica. Las áreas con colores diferentes indican zonas donde las proteínas difieren en su estructura tridimensional o no se superponen perfectamente, con solo unos pocos ángstroms de diferencia.

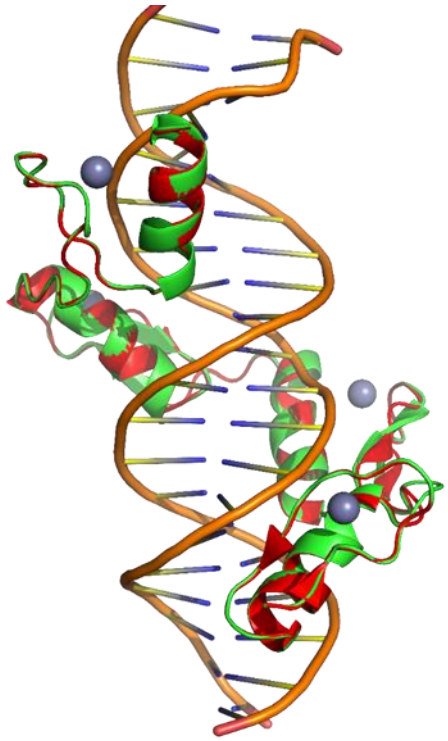


Figura 32. Superposición HYY1 y su DNA con EhYY1. En color verde EhYY1 y en rojo HYY1.

EhYY1 exhibe una estructura casi idéntica a HYY1, reflejada en su similitud en la unión al DNA. Esta semejanza se destaca en la interacción de EhYY1 con HYY1 y su DNA, donde ambas estructuras interactúan con los mismos nucleótidos del core.

6.9 Mutaciones a la secuencia de unión a DNA

Se ha investigado la unión del factor YY1 a su secuencia consenso. Estos estudios han identificado que la unión depende de cinco pares de bases cruciales “CATTT” que están presentes en nuestra secuencia consenso de 8 pb [5′- C**CATTT**TG - 3′] (Ye J.,1994). Dada esta información, se plantea la pregunta de ¿qué sucede cuando se alteran estos pares de bases y cuál es su impacto en la unión de YY1?. Por esta razón, se llevaron a cabo tres tipos de mutaciones; intercambio, inversion y scramble en estos cinco nucleótidos cruciales. Posteriormente se realizó un ensayo de acoplamiento molecular Docking entre las secuencias mutadas y EhYY1.

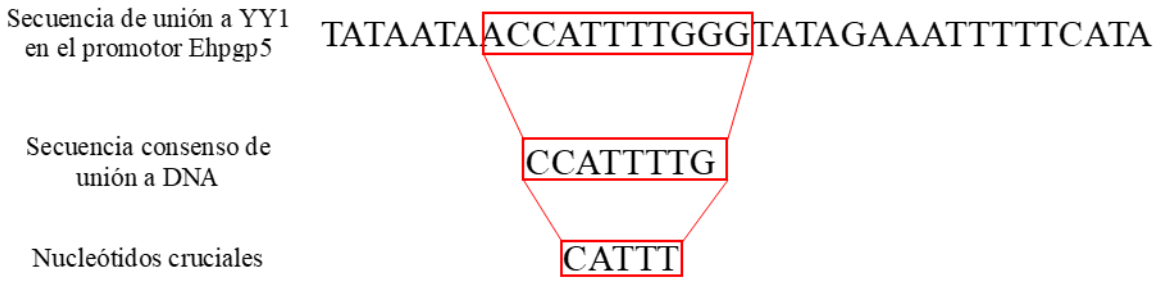


Figura 33. Nucleótidos cruciales en la secuencia consenso de unión a DNA de EhYY1.

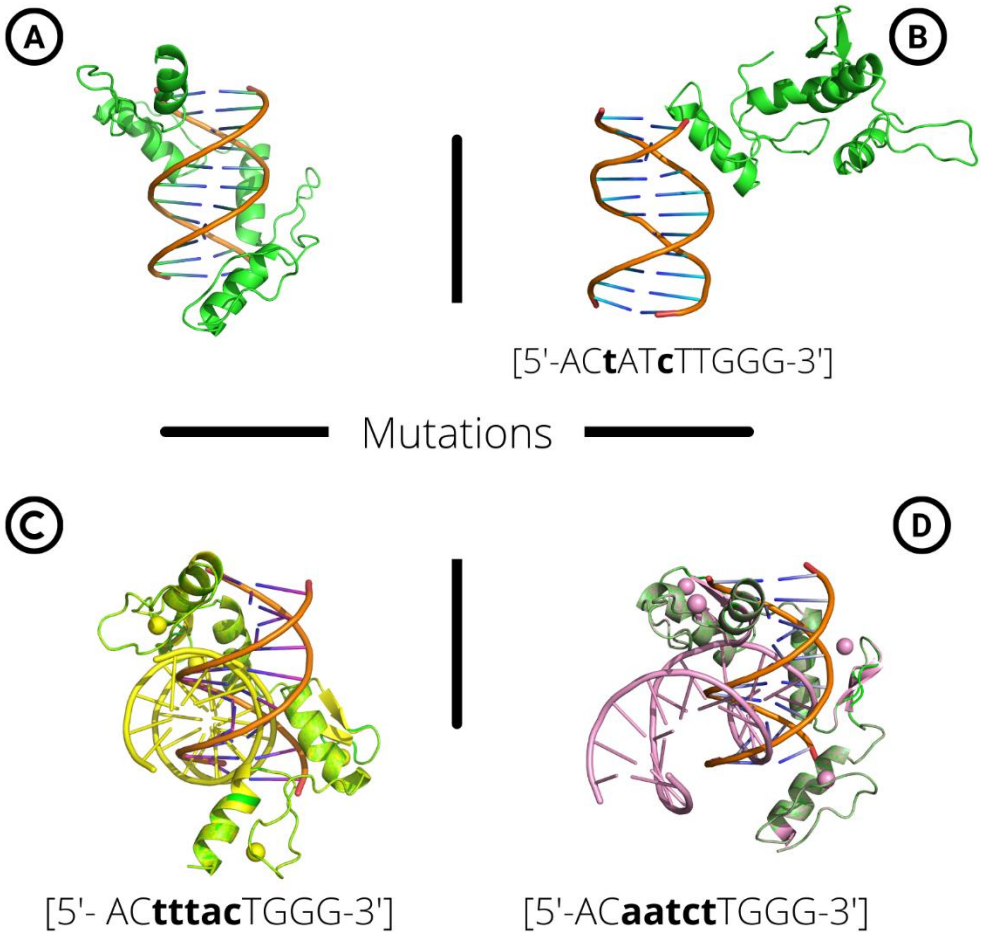


Figura 34. Acoplamiento molecular entre mutaciones a la secuencia de unión a DNA y EhYY1. (A) Secuencia wilde type. (B) Mutación de intercambio. (C) La mutación por inversión se destaca en amarillo, mientras que se emplea el acoplamiento wilde type como punto de referencia verde. (D) La mutación scramble se observa en color rosado, en color verde se encuentra el acoplamiento wilde type como referencia.

Intercambio: en esta mutación se seleccionaron dos posiciones de la secuencia wilde type [5′-AC**C**AT**T**TTGGG-3′] el segundo nucleótido "C" intercambia posición con el segundo nucleótido "T". Para obtener la secuencia [5′-ACtATcTTGGG-3′]. Se optó por utilizar estos nucleótidos porque tanto EhYY1 como HYY1 comparten las mismas interacciones con aminoácidos en su unión al DNA. La mutación llevó a la pérdida de la capacidad de reconocimiento de la secuencia por parte de EhYY1 y a la consiguiente pérdida de interacción, lo que denota la importancia de la citocina en la unión adecuada.

Inversion: Una región de la secuencia de DNA gira sin eliminar o agregar nucleótidos teniendo como resultado la secuencia [5′-ACtttacTGGG-3′]. Se opto por esta mutación dado su alta frecuencia en el proceso de replicación de DNA. En el acoplamiento Docking, la secuencia mutada se visualiza en amarillo y la secuencia tipo salvaje en verde. En ambos modelos superpuestos, la proteína EhYY1 se mantiene inmóvil para destacar el desplazamiento del DNA. El Docking reveló que la secuencia mutada conservó su

interacción con EhYY1, aunque lo hizo mediante un cambio conformacional, ya que los aminoácidos reconocieron nucleótidos distintos a los presentados en el acoplamiento wilde type.

Scramble: Implican la reorganización aleatoria de nucleótidos de DNA obteniendo la secuencia [5′-ACaatctTGGG-3′]. Se opto por esta mutación ya que permite identificar nucleótidos criticos en la unión a DNA que no habían sido considerados en las mutaciones anteriores. Al igual que lo ocurrido con la mutación de inversión el Docking evidenció que la secuencia mutada conservó su interacción con EhYY1, aunque lo logró mediante una modificación en su conformación, debido a que los aminoácidos reconocieron nucleótidos distintos a los presentados en el acoplamiento.

De igual forma se realizaron mutaciones en la secuencia de unión a DNA humana (Figura 35) donde se tomaron 11 pb [5′-TCCATTTTGAA-3′] a la que se le realizaron 3 tipos de mutaciones en los nucleótidos criticos “CATTT”, scramble: [5′-TCaatctTGAA-3′], inversión: [5′-TCtttacTGAA-3′] e intercambio: [5′-TCtATcTTGAA- 3′], posteriormente se realizó un ensayo de acoplamiento molecular Docking con la proteína YY1 humana.

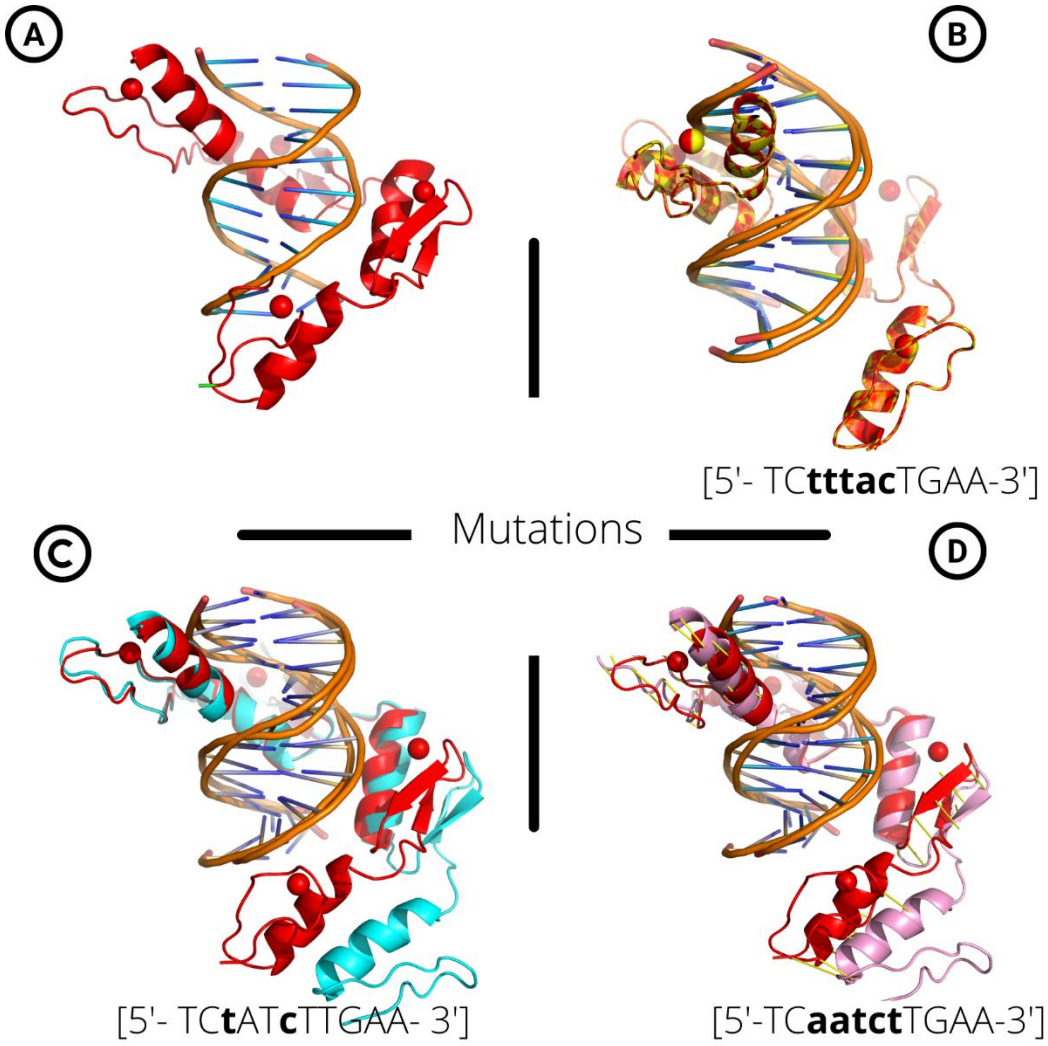


Figura 35. Acoplamiento molecular entre mutaciones a la secuencia de unión a DNA y HYY1. (A) Secuencia wilde type. (B) La mutación por inversión se destaca en amarillo, mientras que se emplea el acoplamiento wilde type como punto de referencia rojo. (C) Se nuestra la mutación de intercambio en azul y el acoplamiento wilde type en rojo. (D) La mutación scramble se observa en color rosado, en color rojo el acoplamiento wilde type.

El análisis de Docking reveló que la secuencia mutada por inversión mantuvo su interacción con HYY1, pero con una unión atípica, desplazando la unión de HYY1 hacia 3´. Tanto la mutación de intercambio como la scramble mostraron una interacción similar, HYY1 reconoció la secuencia de DNA, pero desplazándose unos pares de bases hacia 3´.

Los resultados *In silico* sugieren que EhYY1 es más susceptible a las mutaciones que HYY1. En la mutación de intercambio, EhYY1 perdió por completo su capacidad de unión a su secuencia de unión, mientras que HYY1 la conserva. En las mutaciones por inversión y scramble, EhYY1 experimenta un cambio radical en su unión a la secuencia de unión, reconociendo otros nucleótidos, mientras que en HYY1 solo se desplaza unos pocos pares de bases hacia 3´.

VII DISCUSIÓN

La trascripción es un evento fundamental en todos los organismos y los factores de transcripción son moléculas clave en este mecanismo. El factor de transcripción YY1 presenta relevancia al ser un factor con una actividad dual, es decir como activador y como represor transcripcional (Rizkallah, 2009). Por lo que su presencia en *E. histolytica* fue estudiada en el presente trabajo.

El análisis de la secuencia obtenida mediante el alineamiento múltiple de nueve organismos distintos permitió identificar una posible secuencia YY1 en el genoma de *E. histolytica*, la cual presenta una longitud de 528 pares de bases (pb) y codifica para una proteína compuesta por 175 aminoácidos. El número de aa de las proteínas YY1 observados en organismos como *E. histolytca*, *B. rerio*, *X. levis* y *H. sapiens* va de 175 a 414 aa (GeneCards, 2017) es decir que EhYY1 es la proteína la más corta registrada hasta la fecha. Pese a esto la secuencia se encuentra altamente conservada, aunque se trate de organismos muy distantes como mamíferos, anfibios, peces y protistas.

La homología de EhYY1 con respecto al resto de las especies (*H. sapiens, M. musculus, X. laevis, D. rerio, M. caroli, C. jacchus, T. gelada, S. apella y P. troglodytes*) es de 66% mientras que la identidad es de un 37.17 %. Es decir que un 66% de los aminoácidos pese a no ser los mismos mantiene características fisicoquímicas similares, y que un 37.17% de la secuencia cuenta con exactamente los mismos aminoácidos, estos porcentajes nos muestran las similitudes en la estructura y función de YY1 en diversos organismos. Los porcentajes de identidad de especies filogenéticamente distantenes no suelen ser muy altos, para ponerlo en perspectiva las proteínas Rab en *E. histolytica* tienen entre un 40 - 50 % de identidad con la proteína Rab de mamíferos. Debido a estos datos sabemos que la proteína encontrada bajo el nombre "*Entamoeba histolytica* HM-1:IMSS hypothetical protein" se trata de la proteína EhYY1 que hasta este momento no había sido caracterizada.

Por su parte *H. sapiens*, *M. musculus*, *X. laevis*, *B. rerio*, *M. caroli*, *C. jacchus*, *T. gelada*, *S. apella* y *P. troglodytes* comparten entre toda una homología del 100 % y una identidad que va desde 68 a 99% al ser más emparentados filogenéticamente como se pudo observar en el análisis de máxima verosimilitud.

Ampliando la búsqueda, se encontró que EhYY1 no solo está presente en la cepa *E. histolytica HM-1: IMSS*, sino también en las cepas *HM-3: IMSS*, *KU27*, así como en las especies *SAW760*, *Laredo* e *IP1*. Los alineamientos EhYY1 en Entamoeba muestran secuencias de entre 175 y 176 aminoácidos (siendo E. invadens quien presenta un aminoácido más) todas con homologías de 100 % e identidades que van desde 100 % hasta 69 % (Saito-Nakano, 2021), lo que sugiere que este factor de transcripción debe de tener una relevancia en el control transcripcional también en el género *Entamoeba*.

En cuanto al porcentaje de abundancia, es decir la cantidad relativa de cada aminoácido en la secuencia, resulta ser muy semejante en todas las especies de Entamoeba, donde 4 de ellas *HM-1:IMSS, HM-3:IMSS, KU27* y nuestra proteína modelada EhYY1 comparten en un 100 % su secuencia y estas a su vez comparten similitudes en su homología con *SAW760*. La diferencia más destacada reside en los porcentajes de histidinas e isoleucinas en *Laredo* y *E. invadens IP1* ya que tienen los porcentajes menos abundantes de todas las secuencias.

A pesar de la longitud relativamente corta de la secuencia de Ehyy1 en comparación con otras especies, su alta conservación a nivel de secuencia y estructura proteica indica su importancia funcional en procesos biológicos fundamentales. El análisis comparativo muestra similitudes significativas en la estructura y función de YY1 entre organismos filogenéticamente distantes, lo que sugiere su conservación evolutiva y su papel esencial en la regulación génica a lo largo de la evolución. Además, la presencia de EhYY1 en diferentes cepas de *E. histolytica* y en otras especies de Entamoeba resalta su amplia distribución y conservación en este género de parásitos, subrayando su importancia biológica y su potencial como objetivo terapéutico.

En la evaluación de los diversos parámetros fisicoquímicos entre las secuencias de YY1, se observa que el punto isoeléctrico (pI) para *H. sapiens, M. musculus, X. laevis, D. rerio, M. caroli, C. jacchus, T. gelada, S. apella y P. troglodytes* se encuentra en un rango de pH ácido, entre 5.67 y 5.85. En relación con el índice GRAVY, todas estas especies presentan valores negativos, oscilando entre -0.610 y -0.782, lo que indica la naturaleza hidrofílica de la proteína. Los índices de inestabilidad, ubicados por encima de 40, predicen que la proteína no es estable fuera del ambiente celular, mientras que los altos índices alifáticos, de 61.23 a 70, sugieren una buena estabilidad térmica de la proteína. Por otro lado, las proteínas EhYY1 de Entamoeba exhiben un pI básico y presentan un índice GRAVY con características hidrofílicas. Su índice de inestabilidad indica que no son estables fuera del ambiente celular, y su índice alifático muestra valores altos, entre 62.29 y 71.36, lo que sugiere una buena estabilidad térmica de la proteína. En consecuencia, podemos decir que YY1 se caracteriza como un factor de transcripción con puntos isoeléctricos ácidos o básicos dependiendo de la especie. Además, es hidrofílica, no presenta estabilidad fuera del ambiente extracelular y exhibe una buena estabilidad térmica. Sin embargo, no hay a la fecha estudios que hayan analizado estas características en este tipo de factor de transcripción.

En la estructura primaria de EhYY1 se identificaron 4 dominios de dedos de zinc de tipo C2H2, localizados en el extremo carboxilo terminal entre los aminoácidos 26 y 141. En esta región, *E. histolytica* presenta una mayor homología con otras especies que también exhiben 4 dominios de dedos de zinc C2H2 en su extremo carboxilo terminal, comprendidos entre los aminoácidos 296 y 412. Estos dedos de zinc son particularmente importantes ya que están involucrados en la unión al DNA. En cuanto a la estructura secundaria de EhYY1, se modeló únicamente el dominio de unión al DNA, el cual consta de 113 aminoácidos. En este dominio se identificaron interacciones hélice-hélice, giros beta y gamma, puentes disulfuro y

horquillas beta, destacándose 4 hojas β y 5 hélices α por su contribución a la conformación y función de EhYY1. La presencia de estos elementos estructurales sugiere una proteína altamente organizada y funcionalmente compleja, especialmente en términos de su capacidad de unión al DNA.

En comparación con su homóloga humana, el dominio de la proteína HYY1 está conformado por 114 aminoácidos y muestra 3 hojas β plegadas y 6 hélices α. Un análisis de homología en Raptor X reveló un porcentaje de homología entre ambas proteínas del 98.6%. A pesar de que la homología en la estructura primaria entre ambas proteínas es solo del 37.17%, esto se debe a la omisión de la región de N-terminal, donde la homología era baja y no presentaba propiedades de unión al DNA. En resumen, ambas proteínas muestran una alta homología en su estructura secundaria, lo que sugiere una conservación funcional significativa a pesar de las diferencias en su secuencia primaria.

Además de los dedos de zinc, un análisis detallado de la secuencia de EhYY1 reveló la presencia de dominios TRIAD, FYVE y FCS. Sin embargo, es importante señalar que estos hallazgos presentan un nivel de confianza bajo. Este nivel bajo de confianza generalmente abarca una zona de incertidumbre en la que pueden estar presentes muy pocos positivos verdaderos, y es difícil distinguirlos de los falsos positivos. Por lo tanto, aunque estos dominios fueron detectados, es necesario llevar a cabo investigaciones adicionales para confirmar su autenticidad y entender su función potencial en el contexto de EhYY1.

La familia de proteínas de dedos de zinc TRIAD contiene un motivo tripartito de tres dedos dobles de zinc, el primero de los cuales, RING1, es un típico Dedo RING con una firma C3HC4 de residuos conservados de cisteína e histidina. El segundo (In-Between-Ring, IBR, BetweenRING o DRIL) y el tercero (RING2) son diferentes a los RING pero comparten una similitud notable, que se manifiesta en un espaciado similar de cisteínas y algunos residuos conservados (Borden, 2004).

Por su parte FEYV es un dominio rico en cisteína de aproximadamente 70 aminoácidos que coordina dos iones Zn2 +. El dominio FYVE se encuentra en varias proteínas eucariotas no nucleares implicadas en distintas funciones celulares, incluido el transporte de vesículas, la transducción de señales y la regulación citoesquelética (D'Arrigo, 1996).

Finalmente, el dedo de zinc de tipo FCS, está presente en varias proteínas del grupo polycomb. Los datos moleculares sugieren que este dominio puede unirse al RNA de una manera no específica de secuencia (Zhang, 2004).

Adicionalmente, se investigaron posibles sitios de modificación postraduccional en la secuencia de EhYY1. Las modificaciones postraduccionales son cambios químicos que ocurren en una proteína después de su síntesis por los ribosomas. Estas alteraciones incluyen una amplia gama de procesos, como la fosforilación, SUMOilación, acetilación, entre otros, y son fundamentales para regular la actividad, estabilidad, localización celular e interacciones de la proteína con otras moléculas.

Se identificaron 19 sitios potenciales de fosforilación a lo largo de la secuencia de EhYY1, ubicados en los aminoácidos serina, treonina y tirosina, que contienen grupos hidroxilo (-OH) susceptibles de fosforilación. La adición de grupos fosfato puede influir considerablemente en la función de la proteína. Por ejemplo, la fosforilación de HYY1 regula su participación en varios procesos celulares clave, como la unión al ADN, la transcripción y la interacción con otras proteínas. Dependiendo del sitio fosforilado, este proceso puede activar o inhibir la capacidad de YY1 para interactuar con elementos reguladores del ADN, como ocurre en la regulación del ciclo celular, la respuesta al estrés, la apoptosis y la hipertrofia cardíaca.

En cuanto a la acetilación, se identificó un único sitio potencial en la posición 139 de EhYY1. Este tipo de modificación puede activar o inhibir proteínas no histónicas. En el caso de HYY1, se ha observado que puede reclutar HATs (como p300 y PCAF) y HDACs para modificar histonas, influyendo en la compactación de la cromatina y en la accesibilidad de los factores de transcripción. Además, la interacción de YY1 con p300 puede facilitar el cambio de represor a activador en la regulación génica. La acetilación en el dominio de unión al ADN (C-terminal) disminuye la capacidad de YY1 para unirse a regiones promotoras específicas, como el promotor P5 del adenovirus.

Por otro lado, se detectó un posible sitio de SUMOilación en las posiciones 162-166 de EhYY1. Este proceso implica la unión covalente de la proteína SUMO. Se ha demostrado que la ligasa SUMO E3, PIASy, SUMOila exógenamente a YY1 en la lisina 288 en células COS-7 (fibroblastos de mono). Además, PIASy colocaliza con YY1 en el núcleo, estabilizándolo in vivo, lo que incrementa su vida media y reduce su actividad transcripcional en los promotores de cdc6, c-myc y ezh2 (Ari Kassardjian, 2012) . Estos descubrimientos subrayan la complejidad de la regulación de las proteínas en *E. histolytica* y sugieren una red intrincada de mecanismos de control que pueden influir en su función biológica podrían potencialmente ser explotadas en el desarrollo de nuevas estrategias terapéuticas contra la amebiasis.

En el analisis filogenético el árbol generado a partir de 20 especies se divide en tres grandes clados, en el primero se agrupan todos los factores de transcripción YY1, esta rama contiene a mamíferos primates y roedores quienes muestran un alto grado de conservación para YY1. En el segundo las proteínas EhYY1 de Entamoeba y en el tercer grupo se compone de 5 ramificaciones de proteínas de dedos de zinc en donde se muestra una gran diversidad de organismos, bacterias, peses y arácnidos. Podemos observar que ambos grupos de proteínas (ZFP y YY1) derivan de un mismo ancestro común, sugiriendo que EhYY1 puede pertenecer a los dos, en el árbol se agrupan en las ramas inferiores las ZFP y en una rama superior todos los factores de transcripción YY1, lo que nos indica la separación evolutiva de ambos grupos proteicos. EhYY1 se encuentra en una posición intermedia esto sugiere que la proteína se encuentra en una etapa de transición, además esta no posee los dominios conservados KRAB y BTB que se existen en las ZFP (Bellefroid, 1991), en cambio presenta un alto grado de

conservación en el dominio de unión al DNA conformado por los cuatro dedos de zinc, tanto en posición como en tamaño propio de los factores de transcripción YY1.

La evidencia sugiere que tanto las proteínas ZFP como YY1 comparten un ancestro común, respaldado por la presencia de dominios de dedos de zinc en nuestra proteína en estudio. Sin embargo, el análisis filogenético muestra una clara separación evolutiva, con las ZFP en ramas inferiores y los YY1 en una rama superior. Esto indica una divergencia en la evolución de ambos grupos, a pesar de compartir un ancestro común. Interesantemente, EhYY1 se encuentra en una posición intermedia entre estos dos grupos en el árbol filogenético, sugiriendo que esta proteína podría estar en una etapa de transición, posiblemente debido al grado evolutivo de *E. histolytica*.

Para el modelado tridimensional de EhYY1, se optó por utilizar la herramienta de inteligencia artificial AlphaFold, dado que no existe una plantilla homóloga completa de YY1 obtenida experimentalmente. Este enfoque generó un modelo completo de EhYY1, compuesto por 175 aminoácidos. Al evaluar la calidad del modelo mediante el Test IDDT, se encontró que el extremo carboxilo terminal muestra un porcentaje de confianza "muy bajo" debido a la presencia de estructuras de tipo “Loop” altamente flexibles en esa zona, las cuales pueden adoptar diversas conformaciones. Por otro lado, la región de dedos de zinc, crucial para la estructura y función de EhYY1, se modeló con una confianza "muy alta", indicando una representación precisa.

En los ensayos de acoplamiento entre EhYY1 y su secuencia de unión al DNA, se empleó un modelo diferente al generado en AlphaFold, utilizando el modelado por homología y modelando únicamente el sitio de unión al DNA de EhYY1 (113 aa). Se empleó una plantilla de YY1 humana de esta región, lo que brindó una mayor confiabilidad a la técnica. Al evaluar la calidad del modelo con un diagrama de Ramachandran, se observó que el 87% de los aminoácidos se encuentran en “Regiones más favorecidas” y un 0% en “Regiones no Permitidas”, lo que indica que su ubicación no presenta impedimentos estéricos. Además, al evaluarlo con Z-Score se obtuvo un puntaje de -3.47, lo que sugiere una baja energía potencial en la estructura, aspecto importante dado que las estructuras proteicas tienden a buscar el equilibrio termodinámico modificando su estructura para alcanzar una baja energía. En conjunto, estos datos revelan que la estructura generada de EhYY1 no es simplemente un modelo computacional especulativo, sino que, debido a la baja presencia de impedimentos estéricos y a su baja energía potencial, se considera un modelo realista y preciso.

La secuencia de unión a DNA de YY1 se encuentra reportada sin embargo esta varía según la fuente que se consulte, debido a esto se recabo todas las secuencias disponibles para YY1 obteniendo una secuencia consenso, esta comprende el nucleótido más común en cada posición en un alineamiento de secuencia múltiple (MSA). La búsqueda en distintas bases de

datos dio como resultado 10 secuencias diferentes reportadas para la region promotora de YY1, incluyendo la secuencia promotora encontrada en *EhPgp5*. Todas ellas con una cantidad de nucleótidos variables. Al realizar el gráfico de la secuencia se puede observar que 8 son los nucleótidos que presentan mayor frecuencia [5′- CCATTTTG - 3′] A su vez se sabe que la región compuesta por 5 nucleótidos CATT (A / T), es necesaria para la actividad del promotor, y esta región probablemente sirve como una secuencia de reconocimiento central para un factor de transcripción celular (J-Ye, 1996). Lo visto en las secuencias reportadas sostiene esta hipótesis puesto que todas las secuencias tienen conservados o bien con ligeras variantes estos nucleótidos en su secuencia.

A través de un acoplamiento molecular entre el factor de transcripción EhYY1 y su secuencia consenso de unión al DNA de 11 pares de bases, encontrada en el promotor del gen *EhPgp5*, se demostró su interacción Proteína-DNA, con una energía de unión de ΔG de 460 kcal/mol, involucrando 8 nucleótidos [5′- CCATTTTG - 3′] que coinciden con nuestra secuencia predicha. Además, con el objetivo de determinar si el factor EhYY1 puede también interactuar con la secuencia consenso de YY1 en humanos, se llevó a cabo otro acoplamiento molecular. En este caso, EhYY1 demostró la capacidad de interactuar con los mismos 8 nucleótidos, además de presentar una energía de unión bastante similar, con un ΔG de 433 kcal/mol.

Por otro lado, la proteína HYY1 humana, obtenida mediante métodos experimentales de cristalografía, presenta una energía de unión de 316 kcal/mol con su DNA. Se observa que tanto en el Docking EhYY1-Secuencia de unión al DNA como en la proteína HYY1 humana experimental, el sitio de unión coincide en la región de 8 pares de bases propuesta. Los aminoácidos que comparten interacciones con el DNA son arginina, histidina, fenilalanina y treonina. Además, hay una similitud notable en la segunda citosina de la región de unión al DNA [5′-C**C**ATTTTG - 3′], donde se comparten las interacciones con arginina y fenilalanina. La similitud en los valores de ΔG en ambas interacciones son muy similares y positivos esto sugiere que tanto EhYY1 como HYY1 humana presentan una afinidad comparable por su secuencia de unión al DNA y que ambas requieren un aporte energético similar para llegar a superar las fuerzas de repulsión entre la proteína y su DNA.

Mas aun, al investigar si el factor YY1 humano reconoce la secuencia de unión de EhYY1 en *E. histolytica,* se encontró que esta interacción se vuelve a presentar con los mismos 8 pares de bases y tiene una ΔG de 240 kcal/mol según el modelo generado. Estos resultados confirman la presencia de una proteína YY1 funcional en *E. histolytica* y resaltan la conservación tanto del sitio de unión al DNA de los factores YY1 en este parásito como en los humanos, así como de los nucleótidos que forman la secuencia consenso en el DNA. La unión reciproca de ambas proteínas EhYY1 a DNA humano y HYY1 a promotor *EhPgp5* nos permitió comprobar dos hipótesis la primera de ellas es el parentesco estructural de ambas, así como las propiedades fisicoquímicas semejantes que comparten. La segunda la determinación del sitio de unión a DNA y la importancia de la secuencia [5′- CCATTTTG - 3′] ya que ambas se unen a esta.

Con base a los estudios realizados, se ha identificado que la unión entre EhYY1 y HYY1 depende de cinco pares de bases cruciales "CATTT", presentes en la secuencia consenso de unión a DNA. Esta observación plantea la pregunta sobre qué sucede cuando se alteran estos pares de bases y cuál es su impacto en la unión de YY1. Por lo tanto, se llevaron a cabo tres tipos de mutaciones: intercambio, inversión y scramble, en estos cinco nucleótidos cruciales.

En las mutaciones realizadas en EhYY1-Seucuencia de unión a DNA al realizar la mutación de intercambio se seleccionaron dos posiciones de la secuencia wilde type [5′-ACCATTTTGGG-3′] el segundo nucleótido "C" intercambia posición con el segundo nucleótido "T". Para obtener la secuencia [5′-ACtATcTTGGG-3′]. Se optó por utilizar estos nucleótidos porque tanto EhYY1 como HYY1 comparten las mismas interacciones con aminoácidos en su unión al DNA. La mutación llevó a la pérdida de la capacidad de reconocimiento de la secuencia por parte de EhYY1 y a la consiguiente pérdida de interacción, lo que denota la importancia de la citocina. Al realizar la inversión de los 5 pares de bases [5′-ACtttacTGGG-3′] el Docking reveló que la secuencia mutada conservó su interacción con EhYY1, aunque lo hizo mediante un cambio en conformacional, ya que los aminoácidos reconocieron nucleótidos distintos a los presentados en el acoplamiento wilde type. De manera similar en la mutación de tipo scramble destaco la plasticidad de la interacción proteína-DNA, mostrando que, aunque se produjeron cambios conformacionales, EhYY1 aún pudo interactuar con la secuencia mutada, aunque de manera modificada.

En cuanto a las mutaciones realizadas en HYY1- DNA humano el intercambio de C por T en la secuencia de unión del DNA humano [5′-TCtATcTTGAA- 3′] presento cambios en la interacción de la proteína con el DNA, recorriendo su unión unos pares de bases hacia el extremo 3′. Resulta interesante ver como a diferencia de su homólogo en amiba, la HYY1 no perdió su interacción con el DNA, esto sugeriría que YY1 es menos susceptible a las mutaciones en su secuencia consenso. De manera similar la secuencia mutada por inversión y scramble mantuvieron su interacción con HYY1, pero con una unión atípica, desplazando la unión de YY1 hacia 3´, lo que probablemente impida su función correcta o completamente la inhiba porque no presenta una unión estable, ni una conformación adecuada que le permita interaccionar con la maquinaria basal de la transcripción.

Los resultados obtenidos *In silico* sugieren que EhYY1 es más susceptible a las mutaciones que la proteína humana esto se puede deber a que si bien las proteínas poseen características fisicoquímicas muy similares no son las mismas. En la mutación de intercambio, se observa que EhYY1 perdió por completo su capacidad de unión a su secuencia de unión, mientras que HYY1 conserva esta capacidad. En las mutaciones por inversión y scramble, EhYY1 experimenta un cambio radical en su conformación para unirse a su secuencia de DNA, de manera que los aminoácidos reconocen nucleótidos distintos, mientras que HYY1 no tuvo un cambio abrupto, solo se observa un ligero desplazamiento de unos pocos pares de bases hacia el extremo 3'.

VIII CONCLUSIONES

Se caracterizó estructuralmente la proteína de dedos de zinc EhYY1 presente en *E. histolytica*.

La proteína YY1 está altamente conservada en otras en otras especies de *Entamoeba; E. invadens, E. laredo, E. histolytica SAW760, E. histolytica HM-3 IMSS y E. histolytica KU27*.

La proteína EhYY1 obtenida por homología presenta scores que validan su modelado además tiene una alta homología en su region carboxilo terminal donde presenta 4 dedos de zinc al igual que HYY1 y una homología estructural de 98.6 % con esta misma proteína.

El factor EhYY1 se une a su secuencia consenso [5′- CCATTTTG - 3′] a través de su dominio de unión al DNA.

Las proteínas EhYY1 y HYY1 se unen a la misma secuencia consenso con ΔG 460 kcal/mol y ΔG de 433 Kcal/mol respectivamente lo que nos confirma su parentesco estructural y fisicoquímico.

El analisis filogenético nos permitió identificar que EhYY1 se encuentra en una posición intermedia entre factores de transcripción y proteínas de dedos de zinc sugiriendo que la proteína se encuentra en una etapa de transición, EhYY1 tiene ausente dominios KRAB y BTB que se existen en las ZFP en cambio presenta un alto grado de conservación en el dominio de unión al DNA conformado por los cuatro dedos de zinc, tanto en posición como en tamaño propio de los factores de transcripción YY1.

Otros parentescos encontrados es que ambas proteínas presentan valores de ΔG positivos, es decir requieren de energía para poder llevarse a cabo, en cuanto a las interacciones con el DNA comparten muchos aminoácidos como arginina, histidina, fenilalanina y treonina.

En cuanto a las mutaciones estas afectan de manera significativa a EhYY1, inhibiendo su interacción con el DNA o interaccionando de una manera completamente diferente a lo esperado. A diferencia de HYY1 donde existe solo un desplazamiento en la interacción de la proteína hacia 3′.

IX PERSPECTIVAS

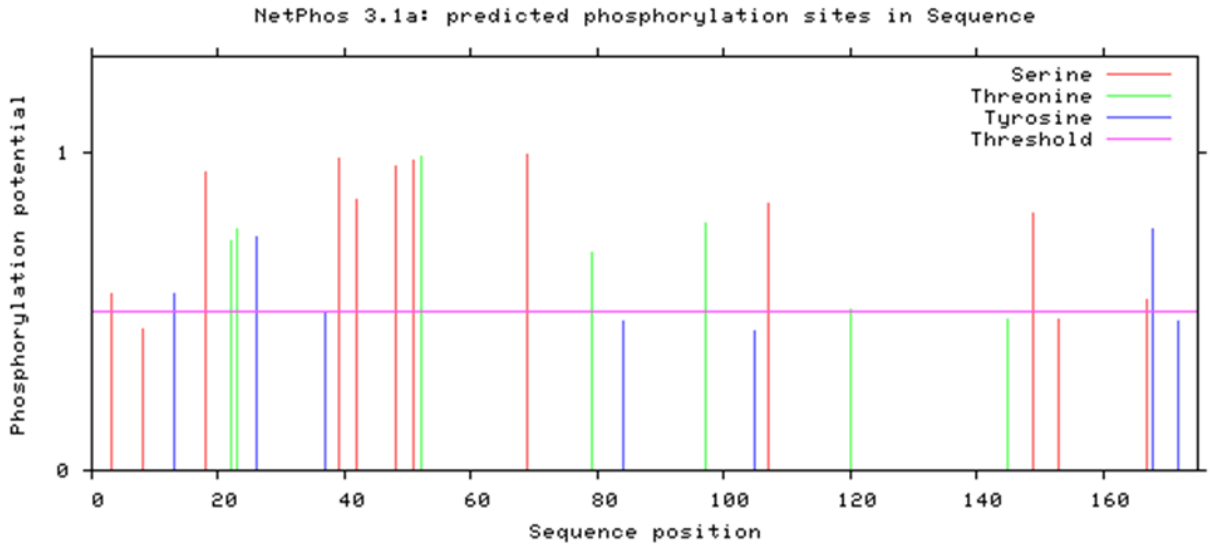
Considerando los resultados obtenidos mediante los ensayos de acoplamiento molecular, es imperativo proceder con análisis *in vitro* para determinar la afinidad de la proteína EhYY1 hacia su secuencia consenso de unión a DNA y a las secuencias mutadas. Este análisis se llevará a cabo a través de ensayos de retardamiento, los cuales se fundamentan en la modificación de la migración del DNA debido a la unión de la proteína.

Para lograr esto, es crucial realizar la clonación del gen Ehyy1 utilizando un vector de expresión con el propósito de producir la proteína recombinante EhYY1. Además, se llevará a cabo la generación de anticuerpos específicos contra EhYY1 en ratones. Estos anticuerpos se emplearán en un ensayo de inmunodetección con el fin de localizar la presencia de EhYY1 en *E. histolytica*.

Evaluar mediante ensayos funcionales el papel del factor EhYY1 en el control transcripcional de *E. histolytica*.

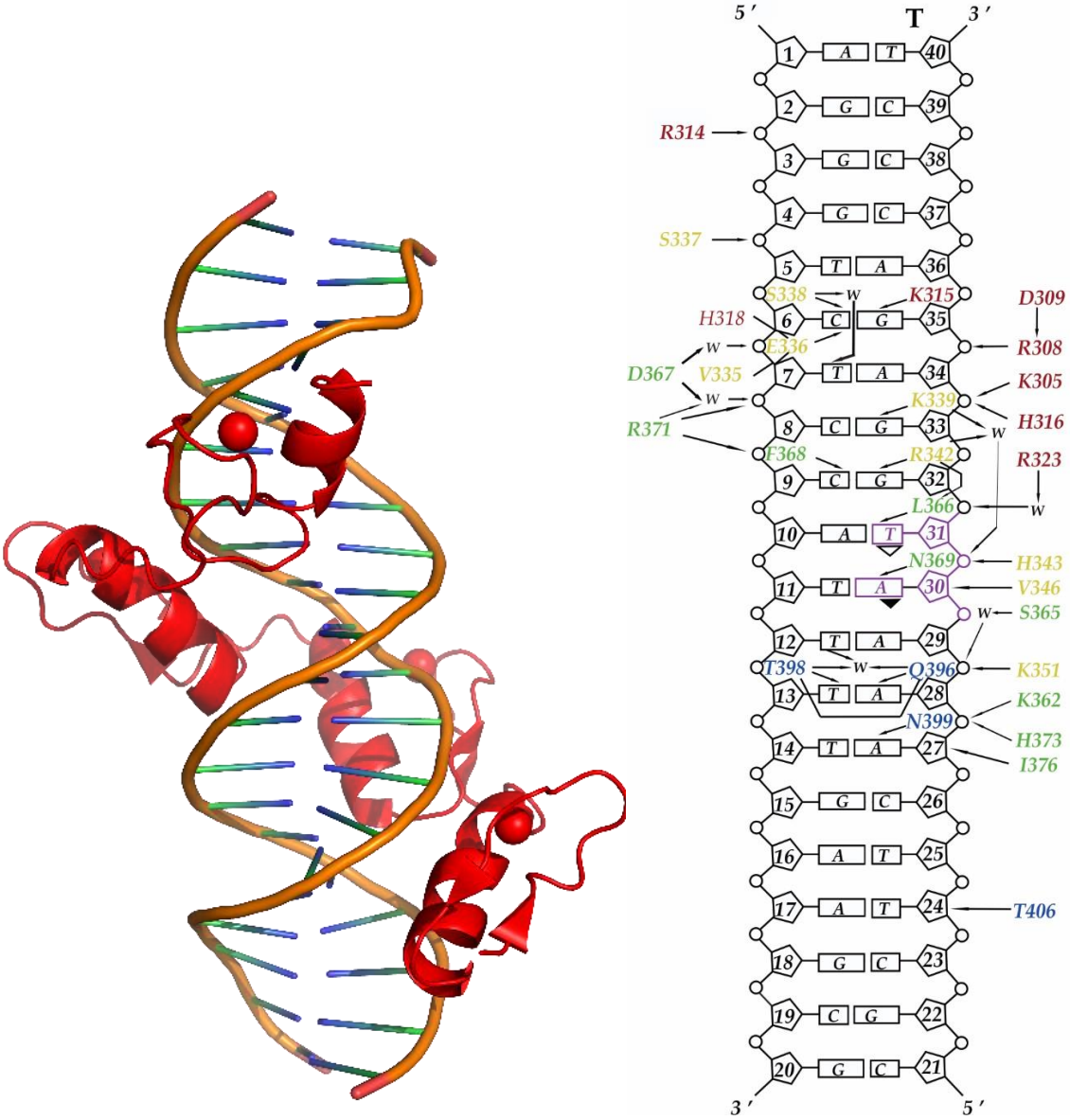
X APÉNDICE

Sitios de fosforilación predichos en la secuencia EhYY1



Apéndice 1. Sitios de fosforilación predichos para la secuencia EhYY1. En el eje x se muestra una escala del "Score" del 0 al 1 partido horizontalmente por una línea rosa "Threshold" que señala los valores mayores a 0.500. En el eje Y se muestran las posiciones de los aminoácidos serina (rojo), Treonina (verde) y Tirosina (azul).

Estructura de cristalográfica de HYY1



Apéndice 2. Estructura cristalográfica HYY1 (1UBD) e interacciones de HYY1 con su secuencia de unión a DNA. En rojo HYY1

Tabla de interacciones Proteína – DNA de cristal HYY1

Pb (5’ – 3’)	AMINOÁCIDO	DISTANCIA (Å)	Pb (3’ – 5’)	AMINOÁCIDO	DISTANCIA (Å)
G (2)	ARG (314)	3.4	G (35)	LYS (315)	-
T (5)	SER (337)	3.0	A (34)	LYS (305)	3.1
C (6)	SER (338)	3.4		HIS (316)	2.8
	GLU (336)	2.9		ARG (308)	3.7
	HIS (318)	-	G (33)	LYS (339)	2.7
T (7)	VAL (335)	-			
C (8)	ARG (371)	3.2	G (32)	ARG (342)	3.8
C (9)	ARG (371)	3.3	T (31)	LEU (366)	3.6
	PHE (368)	3.2		HIS (343)	3.8
T (13)	THR (398)	3.0	A (30)	ASN (369)	2.9
				VAL (346)	3.9
			A (29)	LYS (351)	-
			A (28)	GLN (396)	2.8
				LYS (362)	3.8
				HIS (373)	2.7
			A (27)	ILE (376)	2.8
				ASN (399)	3.8
			T (24)	THR (406)	3.2

Apéndice 3. Tabla de interacciones Proteína – DNA de cristal HYY1

XI BIBLIOGRAFÍA

Abhyankar, M.M, Hochreiter, A.E., Hershey, J., Evans, C., Zhang, Y., Crasta, O., Sobral, B.W.S., Mann, B.J., Petri Jr, W.A., Gilchrist, C.A. (2008). Characterization of an Entamoeba histolytica high-mobility group box protein induced during intestinal infection. *Eukaryotic Cell, 7*(9), 1565-1572. doi:https://doi.org/10.1128/EC.00123-08

Arellano-Aguilar, G. (2017). Breve historia del absceso hepático amebiano, con un caso ilustrativo. *Revista Gastroenterologia Mexico*, 344-348. doi:10.1016/j.rgmx.2016.05.007

Ari Kassardjian. (2012). *Regulation of the Transcription Factor YY1 by Phosphorylation.* FLORIDA: FLORIDA STATE UNIVERSITY. Obtenido de https://repository.lib.fsu.edu/islandora/object/fsu:183503/datastream/PDF/view

Arranz, Jéssica Arribas. (2015). Los factores de transcripción TBX e YY1 en cáncer. Función y regulación de TBX15. Expresión de YY1 en cáncer de tiroides. *Universidad Autónoma de Barcelona*, 26-27.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M. L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. *EMBO, 22*(6), 1347–1358. doi:https://doi.org/10.1093/emboj/cdg124

Austen, M., Lüscher B. and Lüscher-Firzlaff J. (1996). Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protei. *J. Biol. Chem., 272*(3), 272(3):1709-1717. doi:https://doi.org/10.1074/jbc.272.3.1709

Bellefroid, E., Poncelet D., Lecocq P., Revelant O., Martial J. (1991). The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. U. S. A, 89*(9), 88. doi:https://doi.org/10.1073/pnas.88.9.3608

Berg, M.J., L. Stryer, J Tymoczko. (2008). *Bioquimica* (5 ed.). Barcelona, España: Reverté.

Beyrouthy, Maroun J. (2007). The Subcellular Localization of the Transcription Factor YY1 during Cellular Life and Death. *Florida State University*, 1-8. doi:http://purl.flvc.org/fsu/fd/FSU_migr_etd-1402

Bioinformatics Resource Center . (12 de 07 de 2023). *AmoebaDB*. Obtenido de AmoebaDB: https://amoebadb.org/amoeba/app

Borden, Capili A.D. Edghill E.L. Wu K. (2004). Structure of the C-terminal RING finger from a RING-IBR-RING/TRIAD motif reveals a novel zinc-binding domain distinct from a RING. *J. Mol. Biol.*, 340:1117-1129. doi:https://doi.org/10.1016/j.jmb.2004.05.035

Cabello, R. Romero. (2008). *Aspectos históricos de la amibiasis* (1 ed.). Ciudad de México, México: Sociedad Mexicana de Parasitología ; Instituto para el Desarrollo Integral de la Salud.

Cavalier-Smith, Thomas. (2004). Only six kingdoms of life. *Proceedings. Biological sciences, 271*(1545), 1251–1262. doi:https://doi.org/10.1098/rspb.2004.2705

CDC. (2020). Biosafety in Microbiological and Biomedical laboratories. *National Institutes of Health*, 228-233.

Center for Disease Control. (15 de octubre de 2019). *Center for Disease Control*. Obtenido de https://www.cdc.gov: https://www.cdc.gov/dpdx/amebiasis/index.html

Chavez, Lilia Yvonne Tappia. (2008). Estudio del factor de transcripcion EhYY1 de Entamoeba. *Instituto Politecnico Nacional*, 20-23.

Cruz-Vera, J., Clara, L., Hernández-Kelly, R., Méndez, J.A., Pérez-Salzar, E., Ortega, A. (2003). Collagen-induced STAT family members activation in Entamoeba histolytica trophozoites. *FEMS Microbiology Letters, 229* (2), 203-209. doi:https://doi.org/10.1016/S0378-1097(03)00822-X

D'Arrigo, Stenmark H. Aasland R. Toh B.-H. (1996). Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.*, *271*(39), 24048-24054. doi:https://doi.org/10.1074/jbc.271.39.24048

DeepMind y EMBL-EBI. (1 de 1 de 2023). *AlphaFold*. Obtenido de DeepMind and EMBL-EBI: https://alphafold.ebi.ac.uk/

EBI. (01 de 01 de 2022). *Multiple Sequence Alignment*. Obtenido de Clustal Omega: https://www.ebi.ac.uk/Tools/msa/

EBI. (1 de 1 de 2023). *PDBsum*. Obtenido de European Molecular Biology Laboratory : https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum

Espinosa-Cantellano, M; Palomo, A. Marinez. (1987). The pathogenesis of amoebiasis. *Clinical microbiology reviews, 13*(2), 111-118. doi:https://doi.org/10.1128/CMR.13.2.318

Fonte, Luis. (2001). Tratamiento de amebiasis. *Centro de Ingeniería Genética y Biotecnología*, 1-21. Obtenido de https://elfosscientiae.cigb.edu.cu/PDFs/Muestras/amebisis-cap9.pdf

Fundación iO. (1 de Enero de 2019). *Fundación iO*. Obtenido de Fundacion iO: https://fundacionio.com/salud-io/enfermedades/parasitos/entamoeba-histolytica/

Gilchrist, C.A, Holm, C.F., Hughes, M.A., Schaenman, J.M., Mann, B.J., Petri Jr, W.A. (2001). Identification and characterization of Entamoeba histolytica Upstream Regulatory Element 3 sequence-specific DNA-binding Protein containing EF-hand Motifs. *The Journal of Biological Chemistry, 276*(15), 11838-11843. doi:https://doi.org/10.1074/jbc.M007375200

Gilchrist, C.A, Houpt E., Trapaidze N., Fei, Z., Crasta, O., Asgharpour, A., Evans, C., Martino-Catt, S., Baba, D.J., Stroup, S., Hamano, S. (2006). Impact of intestinal colonization and invasion on the Entamoeba histolytica transcriptome. *Molecular and Biochemical Parasitology, 147*(2), 163-176. doi:https://doi.org/10.1016/j.molbiopara.2006.02.007

Gómez, C., Ramirez M.E., Galvez C.M., Medel O., Rodríguez M.A. (2010). Regulation of Gene Expression in Protozoa Parasites. *Journal of Biomedicine and Biotechnology, 2010*, 726045, 1-24. doi:https://doi.org/10.1155/2010/726045

Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene, 25*(8), 1125–1142. doi:https://doi.org/10.1038/sj.onc.1209080

GPS-PAIL 2.0. (23 de 12 de 2014). *GPS-PAIL 2.0*. Obtenido de Acetylation Processes: http://pail.biocuckoo.org/

GPS-SUMO 2.0. (01 de 01 de 2022). *https://sumo.biocuckoo.cn/*. Obtenido de GPS-SUMO 2.0: https://sumo.biocuckoo.cn/

Hegazi M, Patel E, El-Deek B. (2013). Prevalence and characters of Entamoeba histolytica infection in Saudi infants and children admitted with diarrhea at 2 main hospitals at south Jeddah: a re-emerging serious infection with unusual presentation. *Braz J Infect Dis, 17*(1), 32–40. doi:https://doi.org/10.1016/j.bjid.2012.08.021

Hernández de la Cruz, O., Marchat, L.A., Guillén, N., Weber, C., López-Rosas, I., Díaz-Chávez, J., Herrera, L., Rojo-Domínguez, A., Orozco, E., López-Camarillo, C. (2016). Multinucleation and Polykaryon Formation is Promoted by the EhPC4 Transcription Factor in Entamoeba histolytica. *Scientific Reports, 6*, 19611. doi:https://doi.org/10.1038/srep19611

Hogart, Amber; Lichtenberg, Jens ; Subramanian , S Ajay; Anderson, Stacie; Bodine, David M; Margulies, Elliott. (2011). Genome-Wide DNA Methylation Profiling of Hematopoietic Stem and Progenitor Cells Reveals Over-Representation of ETS Transcrition Factor Binding Sites. *Blood, 118*(21), 118-211. doi:https://doi.org/10.1182/blood.V118.21.211.211

Huazhong University of Science and Technology. (1 de 1 de 2019). *HDOCK SERVER*. Obtenido de HDOCK SERVER: http://hdock.phys.hust.edu.cn/

Hyde-DeRuyscher, R., Jennings E. and Shenk T. (1995). DNA binding sites for the transcriptional activator/repressor YY1. *Nucl. Acid. Res, 23*(21), 4457-4465. doi:https://doi.org/10.1093/nar/23.21.4457

INSST. (3 de Febrero de 2022). *Instituto Nacional de Seguridad y Salud en el Trabajo*. Obtenido de INSST: https://www.insst.es/agentes-biologicos-basebio/parasitos/entamoeba-histolytica

Institute European Bioinformatics. (23 de 08 de 2023). *EMBOSS Cons*. Obtenido de emboss_cons: https://www.ebi.ac.uk/Tools/msa/emboss_cons/

J-Ye., X Zhang, Z Dong. (1996). Characterization of the human granulocyte-macrophage colony-stimulating factor gene promoter: an AP1 complex and an Sp1-related complex transactivate the promoter activity that is suppressed by a YY1 complex. *Mol Cell Biol., 16*(1), 157. doi:https://doi.org/10.1128/MCB.16.1.157

Kantor, Micaella. (19 de Noviembre de 2018). *Entamoeba Histolytica: Updates in Clinical Manifestation, Pathogenesis, and Vaccine Development*. Obtenido de https://www.hindawi.com/journals/cjgh/2018/4601420/: https://doi.org/10.1155/2018/4601420

Kawano-Sugaya, T., Izumiyama, S., Yanagawa. (2020). Near-chromosome level genome assembly reveals ploidy diversity and plasticity in the intestinal protozoan parasite Entamoeba histolytica. *BMC Genomics*, 813.

Krap, G. (2005). Biologia celular y molecular, conceptos y experimentos. En G. Krap, *Biologia celular y molecular, conceptos y experimentos* (pág. 1). Mexico: McGraw Hill Interamericana.

Kuida, Chen, Yao Lu,Ke Shi,Daniel B. Stovall,Dangdang Li,Guangchao Sui. (2019). Functional analysis of YY1 zinc fingers through cysteine mutagenesis. *FEBS Letters, 593*(12), 1392 - 1402. doi:https://doi.org/10.1002/1873-3468.13431

Lee M.-H., Lahusen, T., Wang, R.-H., Xiao, C., Xu, X., Hwang, Y.-S., He, W.-W., Shi, Y., Deng, C.-X. (2012). Yin Yang 1 positively regulates BRCA1 and inhibits mammary cancer formation. *Oncogene*, 31: 116-127.

Liljas, L. (2013). Consensus Sequences. *Brenner's Encyclopedia of Genetics*, 163-164. doi:https://doi.org/10.1016/B978-0-12-374984-0.00325-9

Loftus, Anderson I.,Davies R., Alsmark U., SamuelsonJ., Amedeo P.,Roncaglia P. (2005). The genome of the protist parasite Entamoeba histolytica. *Nature, Vol. 433*(7028), 865–868. doi:https://doi.org/10.1038/nature03291

Luke, MP, Sui G, Liu H, Shi Y. (2006). Yang 1 physically interacts with Hoxa11 and represses Hoxa11-dependent transcription. *The Journal of biological chemistry, 281*(44). doi:https://doi.org/10.1074/jbc.M606584200

Luna-Arias, J.P, Hernández-Rivas, R., De Dios-Bravo, G., García, J., Mendoza, L., Orozco, E. (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. doi:https://doi.org/10.1099/13500872-145-1-33

Manual MSD. (1 de Abril de 2020). *Amebiasis*. Obtenido de msdmanuals: https://www.msdmanuals.com/es-mx/professional/enfermedades-infecciosas/protozoos-intestinales-y-microsporidias/amebiasis

Marchat, L.A, Pezet-Valdez, M., López-Camarillo, C., Orozco, E. (2003). histolytica: expression and DNA binding of CCAAT/enhancer-binding proteins are regulated through the cell cycle. *Experimental Parasitology, 103*(1-2), 82-87. doi:https://doi.org/10.1016/s0014-4894(03)00064-x

Mariani, V., Biasini M, Barbato A, Schwede T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England), 29*(21), 2722-2728. doi:https://doi.org/10.1093/bioinformatics/btt473

Martínez Labat, Juan Pablo. (2019). Entamoeba histolytica-dispar. En *Manual de Parasitología* (1 ed., Vol. 1, pág. 178). Cuautitlán Izcalli, Estado de México: Facultad de Estudios Superiores Cuautitlán, UNAM.

Mendoza, L., Orozco, E., Rodríguez, M.A., García-Rivera, G., Sánchez, T., García, E., Gariglio, P. (2003). Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53. *Microbiology (Reading, England), 144*(Pt. 14), 885-893. doi:https://doi.org/10.1099/mic.0.25892-0

Nei, Saitou. (1 de 1 de 1987). *megasoftware*. Obtenido de Phylogenetic Inference: https://www.megasoftware.net/mega1_manual/Phylogeny.html

NetPhos - 3.1. (1 de 1 de 2023). *Generic phosphorylation sites in eukaryotic proteins*. Obtenido de NetPhos - 3.1: https://services.healthtech.dtu.dk/services/NetPhos-3.1/

Nieto, A., Pérez Ishiwara DG, Orozco E, Sánchez Monroy V, Gómez García C. (2017). A Novel Heat Shock Element (HSE) in Entamoeba histolytica that Regulates the Transcriptional Activation of the EhPgp5 Gene in the Presence of Emetine Drug. *Frontiers in cellular and infection microbiology, 7*, 492. doi:https://doi.org/10.3389/fcimb.2017.00492

Olivos-García. (2011). Amibiasis: mecanismos moleculares de la patogenicidad de Entamoeba histolytica. *Revista de la Facultad de Medicina (México)*, 10-20. doi:ISSN 0026-1742

Omim. (16 de 8 de 2023). *omim.org*. Obtenido de TRANSCRIPTION FACTOR YY1; YY1: https://omim.org/entry/600013

OPS. (2003). ZOONOSIS Y ENFERMEDADES TRANSMISIBLES COMUNES AL HOMBRE Y A LOS ANIMALES. . *ORGANIZACIÓN PANAMERICANA DE LA SALUD*, 423.

Orphanet. (16 de 08 de 2023). *Orphanet*. Obtenido de Orphanet versión 5.54.0: https://www.orpha.net/consor/cgi-bin/Disease_Genes.php?lng=ES&data_id=25043&MISSING%20CONTENT=YY1&search=Disease_Genes_Simple&title=YY1

Panda, S., Chandra G. (2012). Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformation, 8*(18), 891-896. doi:https://doi.org/10.6026/97320630008891

PatchDock. (1 de 1 de 2019). *PatchDock*. Obtenido de PatchDock: https://bioinfo3d.cs.tau.ac.il/PatchDock/

PDBsum. (1 de 1 de 2019). *Bioinformatics Institute European*. Obtenido de PDBsum: https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/

Pearson, R.J, Morf, L., Singh, U. (2013). Regulation of H2O2 stress-responsive genes through a novel transcription factor in the protozoan pathogen Entamoeba histolytica. *The Journal of Biological Chemistry, 288*(6), 4462-4474. doi:https://doi.org/10.1074/jbc.M112.423467

Peinado, J., Fermín Toribio Meléndez-Valdés. (2023). Medida del pH: Disoluciones reguladoras. *Campus Universitario de Rabanales*, 1-6. Obtenido de https://www.uco.es/dptos/bioquimica-biol-mol/pdfs/07%20MEDIDA%20pH.pdf

Pérez, G., Gómez C., López-Bayghen E., Tannich E., and Orozco E. (1998). Transcriptional Analysis of the EhPgp5 Promoter of Entamoeba histolytica. *The Journal of biological chemistry, 273*(13), 7285–7292. doi:https://doi.org/10.1074/jbc.273.13.7285

Pettersen, EF., Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. (1 de 1 de 2019). *UCSF Chimera*. Obtenido de UCSF Chimera: https://www.cgl.ucsf.edu/chimera/download.html

ProtParam tool. (05 de 09 de 2023). *Expasy.org*. Obtenido de https://web.expasy.org/protparam/: https://web.expasy.org/protparam/

Ramachandran. (1963). Protein Structure and Crystallography. *Science, 141*(3577), 288-291. doi:https://doi.org/10.1126/science.141.3577.288

Ravdin, J. I:. (2000). *Amebiasis, Series on tropical medicine* (1 ed ed., Vol. 2). (e. e. Hoffman, Ed.) London, England: Imperial College Press. doi:https://doi.org/10.1002/1096-9896(200106)194:2<270::AID-PATH863>3.0.CO;2-U

Rizkallah, R., Hurt MM. (2009). Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. *Molecular biology of the cell, 20*(22), 4766–4776. doi:https://doi.org/10.1091/mbc.e09-04-0264

Rojas, M. (2017). *Efecto del silenciamiento del gen Ehhstf7 de Entamoeba histolytica.* Ciudad de México: Escuela Nacional de Medicina y Homeopatía, Instituto Politécnico Nacional, Tesis de licenciatura.

S, Pérez-Mora. (2024). Entamoeba histolytica: In Silico and In Vitro Oligomerization of EhHSTF5 Enhances Its Binding to the HSE of the EhPgp5 Gene Promoter. *Int J Mol Sci*, 11;25(8):4218. doi:10.3390/ijms25084218

Saito-Nakano, Y., Wahyuni R, Nakada-Tsukui K, Tomii K, Nozaki T. (2021). Rab7D small GTPase is involved in phago-, trogocytosis and cytoskeletal reorganization in the enteric protozoan Entamoeba histolytica. *Cellular microbiology, 23*(1), 1-15. doi:https://doi.org/10.1111/cmi.13267

Samarawickrema NA, Brown DM, Upcroft JA, Thammapalerd N, Upcroft P. (1997). Involvement of superoxide dismutase and pyruvate:ferredoxin oxidoreductase in mechanisms of metronidazole resistance in Entamoeba histolytica. *Antimicrob Chemother*, 833. Obtenido de doi: 10.1093/jac/40.6.833.

Santiago, F. S., Ishii, H., Shafi, S., Khurana, R., Kanellakis, P., Bhindi, R., Ramirez, M. J., Bobik, A., Martin, J. F., Chesterman, C. N., Zachary, I. C., Khachigian, L. M. (2007). Yin Yang-1 inhibits vascular smooth muscle cell growth and intimal thickening by repressing p21(WAF1/Cip1) transcription and p21(WAF1/Cip1)-Cdk4-cyclin D1 assembly. *Circ. Res.*, *101*(2), 101: 146-155. doi:https://doi.org/10.1161/CIRCRESAHA.106.145235

Schaenman, JM., Gilchrist, C.A., Mann, B.J., Petri Jr, W.A. (2001). Identification of Two Entamoeba histolytica sequence-specific URE4 enhancer-binding proteins with homology to the RNA-binding Motif RRM. *The Journal of Biological Chemistry*, *276*(2), 1602-1609. doi:https://doi.org/10.1074/jbc.M006866200

Schier, A.C, Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & Development, 34*(7-8), 465-488. doi:https://doi.org/10.1101/gad.335679.119

Secretaria de Salud. (2023). Amebiasis Intestinal. *boletin epidemiologico*, 19-20.

SEMARNAT. (18 de Julio de 2022). *Nuevos casos de amebiasis*. Obtenido de http://web.archive.org/web/20220728121253/http://dgeiawf.semarnat.gob.mx:8080/ibi_apps/WFServlet?IBIF_ex=D1_SAMBIENTAL01_02&IBIC_user=dgeia_mce&IBIC_pass=dgeia_mce&NOMBREENTIDAD=*&NOMBREANIO=*: https://www.gob.mx/cms/uploads/attachment/file/13289/sem49.pdf

Shi Y., Seto, E., Chang, L.-S., Shenk, T. (2001). Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell*, 67: 377-388.

Shi, Wei. (2006). Frequency distribution of TATA Box and extension sequences on human promoters. *BMC bioinformatics, 7*(suppl 4), 1 - 10. doi:https://doi.org/10.1186/1471-2105-7-S4-S2

Shirley, Watanabe K, Moonah S. (2019). Significance of amebiasis: 10 reasons why neglecting amebiasis might come back to bite us in the gut. *PLoS Negl Trop Dis*, *13*(11), 11-13, e0007744. doi:https://doi.org/10.1371/journal.pntd.0007744

Spiegato. (01 de 01 de 2023). *Spiegato*. Obtenido de ¿Qué es la fosforilación de proteínas?: https://spiegato.com/es/que-es-la-fosforilacion-de-proteinas

Sui G., Affar, E. B., Shi, Y., Brignone, C., Wall, N. R., Yin, P., Donohoe, M., Luke, M. P., Calvo, D., Grossman, S. R., Shi, Y. (2004). Yin Yang 1 is a negative regulator of p53. *Cell*, 117: 859-872.

Swiss Institute of Bioinformatics. (05 de 09 de 2023). *PROSITE*. Obtenido de https://prosite.expasy.org/: https://prosite.expasy.org/

SWISS-MODEL. (1 de 1 de 2019). *expasy*. Obtenido de expasy: https://swissmodel.expasy.org/

Thijs, C. J. Verheul, Levi van Hijfte, Elena Perenthaler, Tahsin Stefan Barakat. (2020). The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Frontiers in Cell and Developmental Biology*, 1-8.

University of Chicago. (1 de 1 de 2019). *raptorx*. Obtenido de raptorx: http://raptorx.uchicago.edu/

Vanacova, S., Liston, D.R., Tachezy, J., Johnson, P.J. (2003). Molecular biology of the amitochondriate parasites, Giardia intestinalis, Entamoeba histolytica and Trichomonas vaginalis. *International Journal for Parasitology, 33*(3), 235-255. doi:https://doi.org/10.1016/s0020-7519(02)00267-9

Velázquez, F. Anaya. (2008). Entamoeba histolytica vs Entamoeba dispar. *En Amibiasis en el siglo XXI*, 29-38.

Verger, A. , Perdomo J, Crossley M. (2003). Modification with SUMO. A role in transcriptional regulation. *EMBO, 4*(2), 137-42. doi:https://doi.org/10.1038/sj.embor.embor738

VEuPathDB. (1 de 1 de 2019). *AmoebaDB*. Obtenido de AmoebaDB: https://amoebadb.org/amoeba/app

Wai, DC, Shihab M, Low JK, Mackay JP. (2016). The zinc fingers of YY1 bind single-stranded RNA with low sequence specificity. *Nucleic Acids Res.*, (19):9153-9165. doi:https://doi.org/10.1093/nar/gkw590

WebLogo. (1 de 1 de 2019). *WebLogo*. Obtenido de WebLogo: https://weblogo.berkeley.edu/

WHO. (2003). Zoonoses and Communicable Diseases Common to Man and Animals. *ORGANIZACIÓN MUNDIAL DE LA SALUD, 3*, 3-7. Obtenido de https://www3.paho.org/hq/dmdocuments/2010/ZoonosesVol-3.pdf

Wiederstein & Sippl. (1 de 1 de 2007). *ProSA-web*. Obtenido de ProSA-web: https://prosa.services.came.sbg.ac.at/prosa.php

Wiederstein, M. & Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*(Web Server issue), 407-410. doi:https://doi.org/10.1093/nar/gkm290

Yamada, Y., Wang, H.Y., Fukuzawa, M., Barton, G., Williams, J.G. (2008). A new family of transcription factors. *Development (Cambridge, England), 135*(18), 3093-3101. doi:https://doi.org/10.1242/dev.026377

Yang, W M; Yao Y L, Sun J M, Davie J R. (1997). Isolation and characterization of cDNAs corresponding to an additional member of the human histone deacetylase gene family. *J. Biol. Chem, 272*(44), 28001–28007. doi:https://doi.org/10.1074/jbc.272.44.28001

Yao, Y. L., Yang, W. M., and Seto, E. (2001). Regulation of transcription factor YY1 by acetylation and deacetylation. *Mol. Cell Biol., 27*(17), 5979–5991. doi:https://doi.org/10.1128/MCB.21.17.5979-5991.2001

Ye, J., Young HA, Ortaldo JR, Ghosh P. (1994). Identification of a DNA binding site for the nuclear factor YY1 in the human GM-CSF core promoter. *Nucleic Acids Res., 22*(25), 5672–5678. doi:https://doi.org/10.1093/nar/22.25.5672

Zhang, H. Christoforou, Aravind L. Emmons S.W. van den Heuvel S. Haber D.A. (2004). The C. elegans Polycomb Gene sop-2 Encodes an RNA Binding Protein. *Mol. Cell, 14*(6), 14:841-847. doi:https://doi.org/10.1016/j.molcel.2004.06.001