



Casa abierta al tiempo

Universidad Autónoma Metropolitana

Departamento de Química

“Modelado molecular y caracterización energética de complejos Rho-GEF de Entamoeba histolytica y búsqueda de inhibidores potenciales de la proteína EhGEF1”

T E S I S

QUE PRESENTA

I.Q. Juan Antonio Sarmiento Muro

Para obtener el grado de

MAESTRO EN CIENCIAS (QUÍMICA)

Asesor:

Dr. Arturo Rojo Domínguez

México, D.F.

Abril de 2007

Agradecimientos

A Dios por estar en cada instante de mi vida y ser todo.

De forma muy especial a las personas que nunca han dejado de amarme y que más amo y respeto: mis padres.

A mis hermanos que son mi corazón: Zaira, Miguel y Héctor.

A todos mis amigos que han formado parte de mi vida y me han hecho crecer.

Agradezco al Dr. Arturo por su apoyo en la elaboración del proyecto, por su paciencia y sobretodo por estar ahí como un amigo.

Agradezco y admiro a cada uno de los profesores que participaron en mi formación durante los estudio de la maestría en Química.

El presente trabajo se realizó en el área de Biofísicoquímica del departamento de Química de la universidad Autónoma Metropolitana unidad Iztapalapa, bajo la dirección del Doctor Arturo Rojo Domínguez. Fue realizado gracias a la beca para estudio de maestría otorgado por el CONACYT. Número de registro 188113.

Resumen: *Entamoeba histolytica* es uno de los principales parásitos humanos que causan enfermedades en todo el mundo. Las GTPasas tipo Rho han sido implicadas en la reorganización del citoesqueleto de actina, transcripción de algunos genes y progresión del ciclo celular en células eucariontes tales como *E. histolytica*, lo que sugiere junto con otras evidencias que estas proteínas se encuentran implicadas en procesos de patogenicidad y de supervivencia del parásito. Se ha demostrado que la proteína EhGEF1 activa a dos GTPasas llamadas EhRacG y EhRho1. En este trabajo se generaron a través del procedimiento de modelado molecular por homología las estructuras tridimensionales de los complejos que se forman durante el proceso de activación de ambas GTPasas. Estos modelos permitieron la caracterización estructural y energética del reconocimiento entre EhGEF1 y las dos GTPasas, lo que nos permitió: 1) determinar la contribución de residuos específicos de EhGEF1 para la formación y estabilización de los complejos, 2) hacer un análisis del proceso de activación para las proteínas Rho y 3) obtener detalles del mecanismo de reconocimiento y selectividad de EhGEF1 sobre EhRacG y EhRho1, que es la suma de diferentes contribuciones y efectos: la alta identidad en secuencia, la presencia de aminoácidos esenciales para el reconocimiento y la presencia de mutaciones específicas cerca de la interfaz. Aprovechando la generación de los modelos, adicionalmente se realizó la búsqueda de la primera generación de inhibidores específicos para la proteína EhGEF1.

Índice temático

Introducción	1
Capitulo I. Antecedentes, problema de investigación y objetivos	12
1.1 Antecedentes	12
1.2 Problema a investigar	13
1.3 Objetivos del proyecto	14
1.3.1 Objetivos generales	14
1.3.2 Objetivos particulares	15
Capitulo II. Métodos	17
2.1 Validación del procedimiento de modelado molecular por homología para los sistemas Rho-GEF	17
2.2 Generación de los modelos moleculares de los complejos que se forman durante el proceso de activación de EhRacG y EhRho1	18
2.2.1 Obtención de las secuencias de aminoácidos y estructuras tridimensionales que pudieran servir como molde	19
2.2.2 Alineamiento múltiple entre las secuencias problema y las secuencias molde	19
2.2.3 Elección del molde	20

2.2.4	Construcción de los modelos.....	21
2.3	Evaluación de la calidad estereoquímica de los modelos.....	22
2.4	Análisis de la contribución de los residuos de la superficie interfacial..	23
2.5	Análisis comparativo de las energías de unión de los complejos usando un campo de fuerza.....	24
2.6	Búsqueda de la base molecular del reconocimiento y selectividad.....	25
2.7	Búsqueda de moléculas para la inhibición de la proteína EhGEF1.....	26
2.7.1	Procedimiento de diseño de fármacos con base en la estructura	26
2.7.2	Procedimiento del anclaje molecular sobre la proteína EhGEF1.....	27
Capítulo III.	Herramientas computacionales y sus aspectos teóricos.....	30
3.1	Buscadores de secuencias y estructuras tridimensionales	30
3.2	Herramientas de alineamiento de secuencias de aminoácidos.....	31
3.3	Modelado molecular por homología.....	33
3.4	Herramienta para el análisis de la contribución de los residuos de interfaz.....	36

3.5 Herramienta de anclaje molecular.....	37
Capitulo IV. Análisis y discusión de resultados obtenidos.....	40
4.1 Resultados de la validación del procedimiento de modelado molecular por homología para sistemas Rho-GEF.....	40
4.2 Resultados de la búsqueda de secuencias y estructuras tridimensionales que pudieran servir como molde.....	44
4.3 Resultado de los alineamientos múltiples entre las secuencias problema y las secuencias molde.....	45
4.4 Elección de la estructura molde.....	45
4.5 Análisis de las características estructurales de la familia de las GTPasas Rho y de la familia de las Dbl GEF.....	47
4.6 Análisis de los modelos del proceso de activación de EhRho1 por EhGEF1.....	51
4.6.1 Modelo EhRho1-GDP-EhGEF1.....	52
4.6.2 Modelo EhRho1-EhGEF1.....	54
4.6.3 Modelo EhRho1-GTP-EhGEF1.....	54
4.7 Análisis de los modelos del proceso de activación de EhRacG por EhGEF1.....	56

4.7.1	Modelo EhRacG-GDP-EhGEF1.....	56
4.7.2	Modelo EhRacG-EhGEF1.....	59
4.7.3	Modelo EhRacG-GTP-EhGEF1.....	60
4.8	Análisis y discusión del cálculo de las energías de unión estimadas con el campo de fuerza CHARMM27.....	61
4.9	Contribución del nucleótido y el Mg ²⁺ a la interacción en los complejos ternarios.....	62
4.10	Discusión del mecanismo de reconocimiento y selectividad.....	64
4.11	Resultados de la búsqueda de moléculas líderes para la inhibición de la proteína EhGEF1.....	66
4.11.1	Elección del sitio de anclaje en EhGEF1.....	67
4.11.2	Resultado del <i>docking</i> de compuestos sobre la cavidad elegida	68
	Conclusiones generales y perspectivas.....	73
	Perpectivas.....	75
	Referencias bibliográficas.....	76

Índice de figuras

Fig. 1. Ciclo vital de <i>E. histolytica</i>.....	2
Fig. 2. Algunas funciones en las que se encuentran implicadas las GTPasas Rho....	4
Fig. 3. Movimiento celular independiente y coordinado.....	5
Fig. 4. Ciclo de regulación de GTPasas Rho.....	7
Fig. 5. Mecanismo propuesto para la activación de las GTPasas tipo Rho.....	8
Fig. 6. Ejemplo de una salida del programa Alanine Scanning.....	24
Fig. 7. Dependencia de la similitud en estructura (RMS) con el porcentaje de identidad en secuencia.....	33
Fig. 8. Estructura tridimensional de las proteínas GEFs.....	48
Fig. 9. Proteína RhoA que nos permite visualizar la estructura tridimensional característica de las GTPasas Rho.....	49
Fig. 10. Alineamiento de secuencias de GTPasas pequeñas.....	50
Fig. 11. Diagrama esquemático del mecanismo general de expulsión de GDP.....	51
Fig. 12. Etapas del mecanismo propuesto para el proceso de activación de EhRho1 por EhGEF1.....	52
Fig. 13. Cambio conformacional del switch 1.....	53

Fig. 14. Etapas del mecanismo propuesto para el proceso de activación de EhRacG por EhGEF1.....	57
Fig. 15. Mecanismo de expulsión del GDP del complejo EhRacG-GDP-EhGEF1.....	58
Fig. 16. Cavity elegida para el docking en la proteína EhGEF1.....	67
Fig. 17. Anclaje molecular.....	71

Índice de tablas

Tabla 1. Términos de la función de la energía libre.....	39
Tabla 2. Valores de $\Delta\Delta G$ de algunos residuos en la estructura cristalográfica, el modelo 1 y el promedio de diez confórmeros.....	41
Tabla3. Coeficientes de correlación.....	42
Tabla 4. Porcentaje de identidad en secuencia para la proteína EhRacG y EhRho1 con las GTPasas molde.....	46
Tabla 5. Porcentaje de identidad en secuencia para la proteína EhGEF1 con las GEFs molde.....	46
Tabla 6. Valores de $\Delta\Delta G$ de algunos residuos importantes en la interfaz para los complejos de EhRho1.....	55
Tabla 7. Valores de ΔG de disociación para los complejos de EhRho1.....	56

Tabla 8. Valores de ΔG de disociación para los complejos de EhRacG.....	59
Tabla 9. Valores de $\Delta\Delta G$ de algunos aminoácidos importantes en la interfaz.para los complejos de EhRacG.....	60
Tabla 10. Energías de unión calculadas con el campo de fuerza.....	62
Tabla 11. Energías de unión sin nucleótido.....	63
Tabla 12. Desglosamiento de las energías de unión en las contribuciones GTPasa-GEF y nucleótido(Mg^{2+})-GEF en los complejos ternarios.....	63
Tabla 13. Valores de $\Delta\Delta G$ de algunos residuos de EhGEF1 que contribuyen de forma importante a la intefaz.....	64
Tabla 14. Valores de $\Delta\Delta G$ de algunos aminoácidos importantes de EhGEF1 para la especificidad.....	66
Tabla 15. Compuestos con las mejores energías de unión sobre la proteínas EhGEF1.....	70

Índice de gráficas

Gráfica 1. Valores de $\Delta\Delta G$ de los residuos de interfaz para el modelo 1 contra los valores de $\Delta\Delta G$ de residuos de interfaz para la estructura cristalográfica.....	43
Gráfica 2. Valores promedio de $\Delta\Delta G$ de los diez modelos contra los valores de $\Delta\Delta G$ de los residuos de interfaz para la estructura cristalográfica.....	44

Índice de ecuaciones

Ecuación 1. Energía de interacción.....38

Ecuación 2. Contribución de la entalpía a la energía libre de unión.....39

Abreviaciones	
Dbl-GEF	Guanine nucleotide exchange factor; responsable de la activación de las GTPasas Rho.
GAP	GTPasa-activating protein; proteínas estimuladoras de la actividad GTPasa.
GDI	Guanine nucleotide-dissociation inhibitor; inhibidores de la disociación de los nucleótidos de guanina.
GDP	Guanine diphosphate.
GTP	Guanine triphosphate.
DH	Dbl homology domain, dominio presente en las proteínas GEF.
PH	Pleckstrin homology domain, dominio presente en las proteínas GEF.
ITSN	Intersectin; factor intercambiador de GDP/GTP para RhoA y Cdc42.
LARG	<u>L</u> eukaemia- <u>a</u> ssociated <u>R</u> ho <u>G</u> EF; factor intercambiador de GDP/GTP para RhoA.
PDZ RhoGEF	<u>P</u> ostsynaptic-density protein <u>d</u> isc large and <u>z</u> one occludens-1; factor intercambiador de GDP/GTP para RhoA.

Introducción

El parásito *Entamoeba histolytica* es un organismo unicelular que pertenece a la familia Entamoebidae del orden Amoebida, subfilo Sarcodina, superclase Rhizopoda de protozoos formadores de pseudópodos, de la clase Lobosea. La *E. histolytica* puede existir en dos formas: el quiste y el trofozoíto [1].

La *E. histolytica* es el agente causante de la amibiasis, y se define morfológicamente como un eucarionte unicelular con trofozoítos uninucleados de 10 a 50 μm de diámetro y quistes de 10 a 15 μm . El núcleo es vesicular, esférico, con una membrana recubierta de pequeños gránulos de cromatina [28].

La infección por *Entamoeba histolytica* ocurre por ingestión de quistes maduros presentes en alimentos, agua o manos contaminados con heces. El desenquistamiento ocurre en el intestino delgado, los trofozoítos son liberados y migran al colon (Fig. 1). Los trofozoítos se reproducen por fisión binaria y producen quistes que se eliminan en las heces. Debido a la protección de su pared celular, los quistes pueden sobrevivir días o semanas en el medio externo y son los responsables de la transmisión. Los trofozoítos en el medio externo mueren rápidamente. De forma alternativa, en muchos casos permanecen confinados en la luz intestinal de los infectados, quienes son entonces portadores asintomáticos o eliminadores de quistes. En algunos individuos los trofozoítos invaden la mucosa intestinal, o bien invaden el hígado, pulmones, cerebro, etc. llegando a esos órganos a través de la circulación (Fig. 1). La amibiasis intestinal afecta, y es con más frecuencia letal, en los extremos de la vida, mientras que el absceso hepático causado por *E. histolytica* es más frecuente en varones y tiene una alta mortalidad [29].

Se estima que el 10% de la población mundial está infectada por *E. histolytica*, lo que resulta en aproximadamente 50 millones de casos de amibiasis invasora y hasta 100,000 muertes por año. La prevalencia de infección amibiana puede ser tan alta como 50% en ciertas áreas de los países en desarrollo. Los estudios serológicos en la Ciudad de México

indicaron que hasta el 9% de la población estaba infectada con *E. histolytica*. Tasas aún más elevadas de infección amibiana ocurren en lugares tales como la India, África Occidental, Lejano Oriente y Centroamérica. La prevalencia de la infección amibiana depende de hábitos culturales, edad, nivel de saneamiento y condición socioeconómica [1].

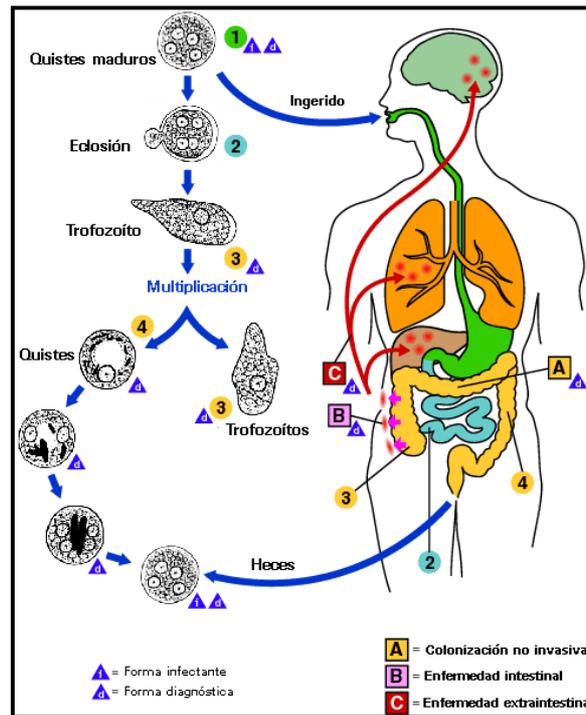


Fig. 1 Ciclo vital de *E. histolytica*. [1]

La virulencia de *E. histolytica* está íntimamente relacionada con su capacidad motil para penetrar en el tejido humano, su poderosa actividad citotóxica y su actividad fagocítica sobre las células humanas. El contacto cercano entre el parásito y la célula eucariótica blanco es mediado por moléculas específicas de ambas superficies celulares, sugiriendo que las interacciones parásito-célula tienen un rol muy importante en la virulencia. La actividad citolítica, su movilidad y la fagocitosis se basan en la polimerización altamente dinámica de actina. Estas funciones celulares son esenciales para la invasión del tejido y requieren de una reorganización del citoesqueleto de la célula para que se lleven a cabo. Muchos estudios han implicado a las GTPasas Rho en los procesos de reorganización del citoesqueleto. [31].

Muchas de las características del comportamiento de las células eucarióticas, tales como las células de *E. histolytica*, son reguladas por la familia de las GTPasas tipo Rho y sus blancos efectores sobre el citoesqueleto y la polimerización de fibras de actina, esta familia de proteínas pertenecen a la superfamilia de las proteínas Ras [2]. Debido a los procesos celulares tan importantes en los que se encuentran involucradas estas proteínas en las células eucariontes, incluyendo a *E. histolytica*, es de suma importancia conocer y caracterizar los mecanismos moleculares por los cuales estas proteínas realizan sus funciones celulares.

Características generales de la superfamilia de las proteínas Ras o GTPasas pequeñas

Las GTPasas son proteínas que funcionan como interruptores moleculares que usan una estrategia bioquímica relativamente simple (intercambio de GDP-GTP) para controlar una gran variedad de procesos complejos en las células eucariontes. Los miembros que pertenecen a estas proteínas son: la superfamilia de las proteínas Ras, las proteínas G heterotriméricas y factores involucrados en síntesis de proteína [19]. Éstas se encuentran en dos estados conformacionales: uno unido al GTP (estado activo) y otro unido al GDP (estado inactivo). Las GTPasas pequeñas o superfamilia de las Ras son un tipo de GTPasas de particular interés en la biología celular, debido a que se encuentran involucradas en muchos aspectos del comportamiento de las células eucariontes. Las GTPasas pequeñas se dividen en cinco familias principales: Ras, Rho, Rab, Arf y Ran [19]. Todas las proteínas Ras comparten un grupo conservado de dominios G denominados G1 (GXXXXGKS/T), G2 (T), G3 (DXXGQ/H/T), G4 (N/TKXD) y G5 (S/CAK/L/T) relacionados con la unión e hidrólisis de los nucleótidos [22].

Algunos de los procesos celulares en los que participan cada una de las familias pertenecientes a la superfamilia de las Ras son los siguientes: la familia de las proteínas Ras controla procesos tales como la expresión génica, proliferación y diferenciación celular, la mayoría de las proteínas correspondientes a esta familia son oncoproteínas y se encuentran involucradas en aproximadamente el 30% de los casos de cáncer en humanos. La familia de

las proteínas Rho regula la reorganización del citoesqueleto de actina, progresión del ciclo celular y transcripción de algunos genes. La familia de las proteínas Rab regula el transporte vesicular intracelular y el tráfico de proteínas entre los diferentes organelos durante procesos de endocitosis y secreción; es la familia con la mayor cantidad de miembros conocidos. La familia de las proteínas Ran son las GTPasas pequeñas más abundantes en la célula, la función más conocida en la que se encuentran involucradas es el transporte núcleo-citoplasma de RNA y proteínas. Por último, las proteínas de la familia de las Arf se encuentran también involucradas en la regulación del transporte vesicular [19].

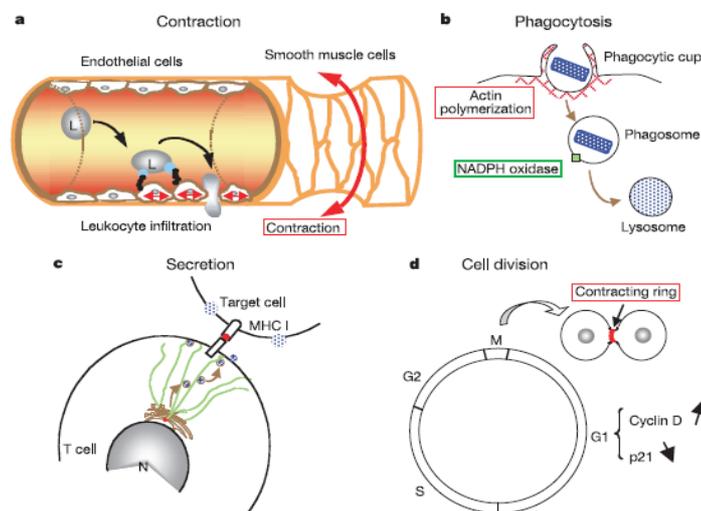


Fig. 2. Algunas funciones en las que se encuentran implicadas las GTPasas Rho. a) Contracción muscular (para regular la circulación sanguínea) y contracción de células endoteliales (para permitir el movimiento de los leucocitos a través de los tejidos). b) Fagocitosis. c) Secreción. La reorientación del microtubo en las células T citotóxicas les permite unirse a las células invasoras y realizar su función. d) División celular. Varias proteínas Rho se encuentran activadas durante la fase G1. [2]

Familia de las proteínas Rho

Como ya se ha mencionado la familia de las GTPasas tipo Rho comprende una de las ramas principales de la superfamilia Ras, son proteínas pequeñas de aproximadamente 21 kDa. En base a su identidad en secuencia, motivos estructurales, y función biológica la familia de las Rho puede dividirse en seis subfamilias que exhiben propiedades similares pero no idénticas. Éstas son las subfamilias: relacionada con Cdc42, la relacionada con Rac1, la

relacionada con RhoA, la Rnd, la RhoBTB y la recientemente descrita subfamilia Miro [20]. Las proteínas Rho funcionan como interruptores bi-moleculares adoptando diferentes estados conformacionales en respuesta a la unión de GDP o GTP. La activación de la familia de las proteínas Rho es mediada a través de receptores de la superficie celular, incluyendo receptores de citoquinas, receptores de tirosina quinasa, receptores de adhesión, así como receptores acoplados a la proteína G [2].

El gen Rho fue identificado en 1985 pero no fue sino hasta 1992 cuando se conocieron algunas de las funciones celulares en las que participan las GTPasas Rho. Se encontró que las mutantes de Rho y Rac humanas activadas constitutivamente inducían el ensamble de actina contráctil, filamentos de miosina y formación de protuberancias llamadas “lamelipodios” [2]. Más tarde, se demostró que Cdc42 promueve la formación de actina y la extensión de la membrana celular en forma de dedos, llamados “filopodios”. Las proteínas Rho son reguladores de la reorganización del citoesqueleto de actina, además se encuentran implicadas en la progresión del ciclo celular, transcripción de algunos genes y en procesos celulares tales como adhesión, migración, transporte vesicular, fagocitosis, citocinesis, contracción y extensión neurítica, morfogénesis celular, polarización de la célula, crecimiento y supervivencia celular [2] (Figs. 2 y 3).

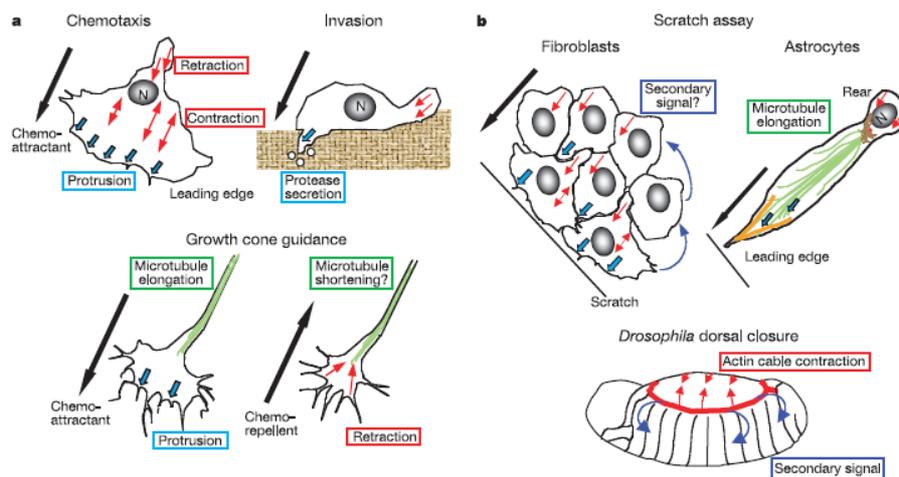


Fig. 3. Movimiento celular independiente y coordinado. [2]

Las proteínas Rho funcionan como interruptores binarios, activos en su forma unida a GTP e inactivos en la forma unida a GDP. El ciclo entre el estado activo y el inactivo es controlado principalmente por tres clases distintas de proteínas: (1) factores intercambiadores de GDP/GTP (Dbl GEFs), los cuales catalizan el intercambio de GDP por GTP; (2) proteínas estimuladoras de la actividad GTPasa (GAPs), las cuales estimulan la hidrólisis intrínseca de GTP de las proteínas Rho promoviendo la formación de la proteína en estado inactivo unida a GDP; y (3) inhibidores de la disociación de los nucleótidos de guanina (GDIs), los cuales inhiben a las proteínas Rho mediante el bloqueo del proceso de intercambio de nucleótidos (Fig. 4) [5]. Todas las GEFs para Rho comparten un dominio DH (dominio de homología a Dbl) de cerca de 200 aminoácidos, que lleva a cabo la actividad catalítica de intercambio de nucleótidos. También poseen un dominio PH (dominio de homología a pleckstrina) adyacente al dominio DH, que media las interacciones proteína-proteína y proteína-fosfolípido [3].

Una vez activadas las GTPasas Rho (unidas a GTP) interactúan con sus blancos celulares o proteínas efectoras para que la señal pueda seguir río abajo y se obtenga una respuesta biológica. Se han identificado al menos 30 efectores para GTPasas Rho entre ellos Roka, Rok β , PAK1, PAK2, PAK3, Dia1, Dia2 y WASP [21]. Los motivos de unión a GTPasas Rho son: CRIB, REM y RKH asignados de acuerdo a su identidad en secuencia. El mecanismo más común de activación de efectores por GTPasas Rho parece ser que es la interrupción de interacciones de auto inhibición intramolecular de los efectores [21].

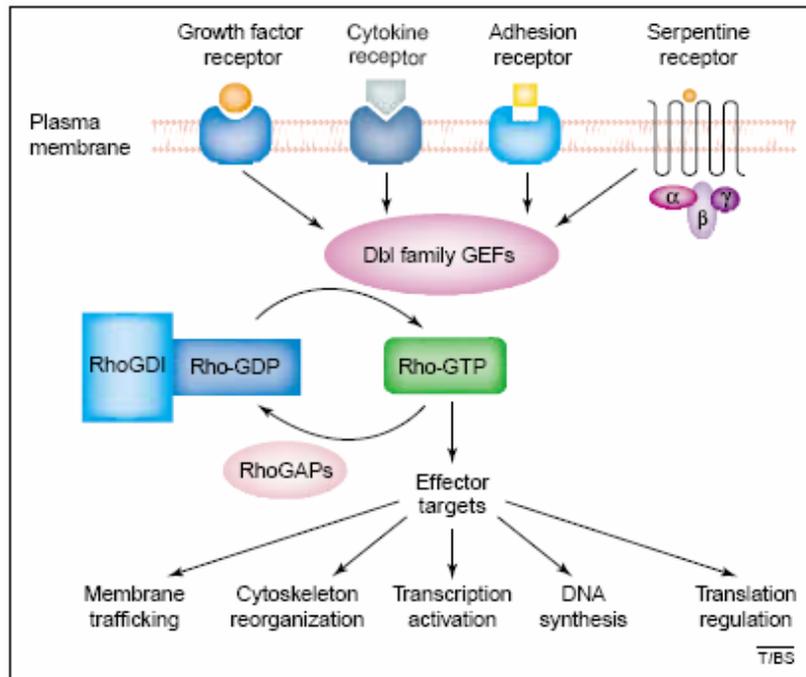


Fig. 4. Ciclo de regulación de las GTPasas Rho. [5]

El mecanismo propuesto para la activación de las proteínas Rho comprende varios pasos. Primero la GEF forma un complejo de baja afinidad con la Rho previamente unida a GDP. El GDP es disociado de dicho complejo, con lo cual la afinidad entre las dos proteínas aumenta, este complejo no se acumula en la célula porque es rápidamente disociado por la posterior unión de GTP (Fig. 5). Las GEFs son también capaces de disociar GTP *in vitro*; aunque *in vivo*, la concentración de GTP es mayor que la de GDP y la interacción de la forma unida a GTP con los efectores dirige la reacción en la dirección GDP por GTP. Esto implica que la GTPasa libre tanto del nucleótido como de la GEF, virtualmente no exista en la célula. Por lo anterior, puede decirse que las GEFs tienen una actividad bioquímica dual: desestabilizar la interacción de la proteína Rho con el GDP y estabilizar a la proteína libre del nucleótido, de tal forma que la Rho prácticamente nunca se encuentra en forma aislada [7].

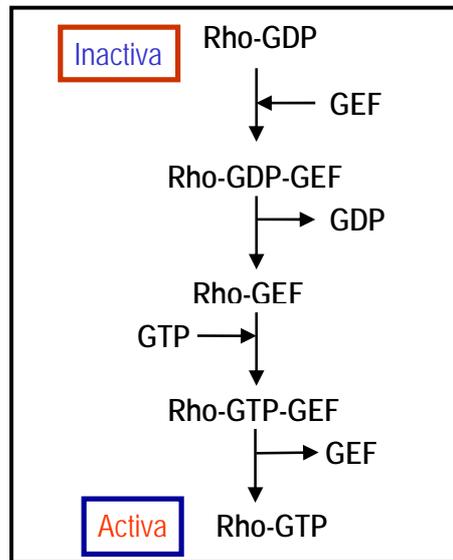


Fig. 5. Mecanismo propuesto para la activación de las GTPasas tipo Rho.

Implicación de las GTPasas tipo Rho y sus reguladores (GEFs, GAPs y GDIs) en el cáncer

Las GTPasas Rho, como ya se mencionó anteriormente, desempeñan un papel importante en el control del crecimiento y la proliferación celular. Se sabe que estas proteínas estimulan la progresión de la fase G-1 del ciclo celular y la síntesis de ADN, así como vías de señalización que activan gran variedad de factores de transcripción nuclear. Además, se ha demostrado que las proteínas Rho poseen un potencial transformante en algunas líneas celulares y muchos GEFs para Rho, Rac y Cdc42 son oncogénicos. Asimismo, Rho y Rac constitutivamente activadas incrementan el crecimiento celular e inducen crecimiento independiente y formación de tumores. Por otro lado, Rac, Rho y Cdc42 son esenciales para la transformación maligna inducida por Ras. Existen datos experimentales que demuestran que la coexpresión de Rac1 y RhoA con Raf-1 provoca una actividad transformante cooperativa. De forma similar, la coexpresión de Rac1 y RhoA activados con Ras oncogénico causa un gran incremento de la transformación morfológica. En cambio, la coexpresión de mutantes dominantes negativas de Rac1, RhoA y Cdc42 reduce la actividad transformante de Ras en fibroblastos de roedores [18]. Tales observaciones indicarían que

las proteínas Rho probablemente contribuyen a las acciones transformantes de Ras; sin embargo, los cambios morfológicos observados en las células transformadas podrían ser causados, en parte, por desregulación de la función de las proteínas Rho, en cuya regulación participan las proteínas GEFs, GAPs y GDIs, lo que indicaría que estas últimas proteínas se encuentran implicadas en los procesos oncológicos [10].

Implicación de las GTPasas Rho y sus reguladores en la patogenicidad bacteriana y viral

Algunas bacterias patógenas han aprovechado los beneficios potenciales que, para el patógeno, posee la activación de las vías de señalización de algunas proteínas Rho. Eso ha causado que se hayan desarrollado durante su proceso evolutivo diversos factores de virulencia que modifican el ciclo de activación y desactivación de las GTPasas Rho [28]. Se ha demostrado que la enterobacteria *Salmonella spp.* posee dos proteínas estructuralmente muy relacionadas (SopE y SopE2) que se comportan como GEFs para Rac1 y Cdc42. La activación de estas GTPasas origina cambios en el citoesqueleto de actina que facilitan la internalización de la bacteria en las células epiteliales del intestino. Una vez en el interior de la célula hospedadora, *Salmonella* contribuye al regreso al estado basal del citoesqueleto liberando SptP, una proteína GAP para Rac1 y Cdc42 que facilita el cese de la señalización inducida por éstas. Así, el patógeno evita los daños potenciales que podría producir a su célula hospedadora. De forma análoga, *Yersinia spp.* y *Pseudomonas aeruginosa* también poseen proteínas que actúan como GAPs para las algunas GTPasas Rho (YopE y ExoS, respectivamente) [28].

Otras bacterias patógenas han desarrollado toxinas capaces de modificar directamente a las GTPasas Rho, lo que contribuye a la activación o inactivación de éstas [27]. La exoenzima C3 de *Clostridium botulinum*, así como otras exoenzimas con actividad de transferasa, sintetizadas por *Costridium limosum*, *Bacillus cereus* y *Staphylococcus aureus*, inducen la ADP-ribosilación de RhoA, RhoB y RhoC. Este proceso no afecta a la unión de RhoA a sus efectores ni a la actividad GAP, pero impide la interacción con los GEFs y bloquea la disociación de los GDIs. De esta manera, la GTPasa se acumula en el citosol en su forma

inactiva, lo que impide su interacción con las proteínas efectoras. Algunas toxinas producidas por *Clostridium spp.* dan lugar a la glucosilación de residuos situados en el *switch I* de las GTPasas Rho [27]. Esta modificación impide la interacción de las GTPasas con sus efectores, bloquea la actividad GTPasa tanto intrínseca como inducida por GAPs, e inhibe el intercambio de nucleótidos producido por los GEFs. Además, la glucosilación impide la interacción de las GTPasas con los GDIs, manteniéndolas permanentemente unidas a la membrana plasmática e impidiendo la translocación de las GTPasas no modificadas presentes en la misma célula. Los factores CNF1 y CNF2, producidos por formas enteropatógenicas de *Escherichia coli*, y la toxina DNT, sintetizada por *Bordetella spp.*, poseen respectivamente actividad desamidasa y desamidasa/transglutaminasa sobre el residuo Gln 63 de RhoA. Esta modificación bloquea la actividad GTPasa de RhoA, tanto intrínseca como inducida por GAPs, dando lugar a su activación constitutiva y aumentando así su poder invasor sobre las células humanas [27].

Las GTPasas de la familia Rho también son utilizadas durante el proceso infeccioso de diversos virus. Quizá el caso más llamativo en este sentido sea v-Nef, una proteína codificada por el virus de la inmunodeficiencia adquirida que es esencial para mantener niveles altos de viremia y para la progresión de la enfermedad en la célula hospedadora. v-Nef interacciona con diversos elementos pertenecientes a la ruta de señalización de las GTPasas Rho, uniéndose y activando a las GEFs de la familia Vav y al efector de Rac1 PAK. La interacción simultánea de v-Nef con las proteínas Vav y PAK favorece la activación de esta cinasa mediante Rac1, el sustrato de Vav que es arrastrado hacia la proximidad de PAK vía su interacción con el dominio catalítico de dicho GEF. Similarmente, el dominio citoplasmático de la glicoproteína gp41 del VIH tipo 1, la cual está implicada en la replicación y citopatogenicidad del virus, inactiva a la proteína p115-RhoGEF (activador de la proteína RhoA). Como resultado de la pérdida de la función de la proteína RhoA se puede facilitar la virulencia del VIH-1 debido a la perturbación de la interacción de las células T con otras células y de la migración e inhibición de la proliferación de las células T [3].

Como hemos podido ver en la introducción de la tesis, el parásito *E. histolytica* causa problemas de salud pública sobre todo en países subdesarrollados. Una de las bases fundamentales de su patogenicidad es su capacidad motil para penetrar en el tejido humano y además, como cualquier célula eucarionte, muchas de sus funciones celulares, sobre todo aquellas en la que participa la reorganización del citoesqueleto de actina, son controladas por la familia de proteínas Rho y sus proteínas reguladoras. La descripción de estos procesos a nivel molecular y atómico, en especial de las interacciones que permiten la activación de las proteínas Rho de la amiba, son de suma importancia en el conocimiento de las funciones celulares que controlan estas proteínas en la células de *E. histolytica*, así como para la búsqueda de nuevas proteínas diana para el diseño de fármacos. Es importante destacar que para tal motivo no se cuenta con la estructura tridimensional de estas proteínas. En la introducción también se han mencionado la implicación de las proteínas Rho y sus reguladores en el cáncer, la invasión viral y bacteriana. El conocimiento de las interacciones de estas proteínas en la amiba podría ser usado como modelo para el estudio de estos procesos patológicos en el ser humano.

Capítulo I. Antecedentes, problema de investigación y objetivos.

En este capítulo se expone el problema al que se plantea dar respuesta, por lo cual es necesario saber de donde se parte y qué conocimiento se tiene del sistema de investigación. Para ello se presenta un apartado de antecedentes y los objetivos generales y particulares del proyecto.

1.1. Antecedentes

Muchos estudios han demostrado que existen diferentes tipos de GTPasas pequeñas en *E. histolytica*; esto sugiere que la función y regulación de estas moléculas están íntimamente relacionadas con los mecanismos patógenos y supervivencia del parásito [6]. *E. histolytica* es capaz de ingerir microorganismos en el intestino grueso, así como células rojas de la sangre y células blancas muertas durante la invasión del tejido. La fagocitosis juega un rol esencial en el crecimiento y constituye una de las claves de la virulencia del parásito. Se ha demostrado que hay varios pasos implicados en la fagocitosis: interacción receptor-ligando de las superficies celulares, activación de señales que comienzan con la reorganización del citoesqueleto de actina y el tráfico de membrana. Se ha demostrado mediante la expresión de dominantes negativos que RacA está implicada en la reorganización del citoesqueleto durante la fagocitosis de bacterias, eritrocitos y leucocitos en *E. histolytica* [24], mientras que RacG se encuentra involucrada en la formación del uroide [29]. Aunque no se conoce exactamente la función celular que desempeña la proteína EhRho1, pero debido a su alto porcentaje en identidad en secuencia con la proteína RhoA (aprox. 72 %) de humano, que controla el ensamblaje de las fibras de estrés de actina y la adhesión focal, puede sugerirse que EhRho1 podría estar involucrada en procesos celulares semejantes en *E. histolytica* [6]. Existen otras proteínas de la superfamilia de las Ras que se encuentran involucradas en la fagocitosis como son: EhRab1, las isoformas de Rab7 (Rab7A, Rab7B, Rab7C, Rab7D, Rab7E), RacC y Rap2, involucradas principalmente en el tráfico vesicular [24].

Recientemente un grupo de investigación del Departamento de Biomedicina Molecular del CINVESTAV identificó el primer factor intercambiador de nucleótidos de guanina en *E. histolytica*; el cual fue denominado EhGEF1. Esta proteína activa a las GTPasas: EhRacG y EhRho1 [6].

En trabajos recientes se han obtenido los modelos moleculares de los complejos binarios que se forman en el proceso de activación de EhRacG y EhRho1 (EhRacG-EhGEF1 y EhRho1-EhGEF1) que permitieron realizar una caracterización estructural de la interacción entre las proteínas en ambos complejos. Se utilizó como herramienta para obtener las estructuras tridimensionales el modelado molecular por homología, debido a que no se cuenta con las estructuras cristalográficas ni de resonancia magnética nuclear. Es importante destacar que los modelos de los complejos ternarios (GEF y GTPasa unidos a nucleótidos) no se han generado hasta el momento mediante el procedimiento de modelado molecular por homología, además éstos no son accesibles experimentalmente debido a que son inestables y rápidamente tienden a disociarse.

1.2. Problema de investigación

Con base en lo descrito en los párrafos anteriores, el problema que se abordará es generar los modelos para cada una de las etapas propuestas del proceso de activación de dos GTPasas de *E. histolytica*: EhRacG y EhRho1 ambas activadas por el mismo factor intercambiador de nucleótidos de guanina, EhGEF1. Primeramente los modelos permitirán realizar una caracterización energética del contacto proteína-proteína en cada uno de los complejos formados, tanto los binarios como los ternarios. Este análisis energético permitirá proponer cuáles de los residuos en la superficie interfacial contribuyen de forma importante al reconocimiento y la estabilización. Este conocimiento podría ser usado para generar dominantes negativas que permitan inferir las funciones celulares en las que participan EhGEF1, EhRacG y EhRho1. Además con los modelos ternarios se realizará una caracterización estructural y energética de estos complejos, de los cuales no se contaba con

su geometría. Con ayuda de los modelos moleculares se podrán proponer detalles del posible mecanismo de reconocimiento de EhGEF1 sobre una determinada GTPasa.

Adicionalmente, la generación de los modelos moleculares nos permitirá buscar, mediante técnicas computacionales de acoplamiento molecular, moléculas orgánicas pequeñas que se unan con afinidad y selectividad en una cavidad de EhGEF1 donde se encuentren algunos residuos esenciales para el reconocimiento y estabilización entre EhGEF1 y una de las dos GTPasas: EhRacG y EhRho1, con la intención de que esta unión impida la formación de los complejos. Estas moléculas orgánicas pueden servir como una sonda molecular que nos permita conocer la actividad bioquímica de EhGEF1 en el parásito.

Al inhibir la formación de los complejos, debido a la molécula orgánica, se podría apagar la vía de señalización que permite la activación de las GTPasas EhRacG y EhRho1, lo que nos permitiría inferir las funciones celulares en las que se encuentran involucradas y qué tan importantes son para el parásito ambas GTPasas.

Si al inhibir a la proteína EhGEF1, se demuestra que se afectaron mecanismos importantes para la supervivencia del parásito y que además no hay un mecanismo homeostático del parásito para contrarrestar el efecto de la inhibición, las moléculas orgánicas que actúan como inhibidores de EhGEF1 podrían ser usadas en una primera etapa de la búsqueda de un nuevo fármaco antiamebiano.

1.3. Objetivos del proyecto

1.3.1 Objetivos generales

El objetivo general de este trabajo es conocer detalles de la relación estructura función del proceso de activación de las proteínas EhRacG y EhRho1 por EhGEF1 a través del modelado molecular por homología y del análisis estructural y energético de los siguientes complejos:

- Para EhRacG: EhRacG-GDP-EhGEF1, EhRacG-EhGEF1 y EhRacG-GTP-EhGEF1
- Para EhRho1: EhRho1-GDP-EhGEF1, EhRho1-EhGEF1 y EhRho1-GTP-EhGEF1

A los modelos se les realizará una caracterización energética mediante métodos computacionales, además serán usados para estimar cuáles son los residuos más importantes que participan en el reconocimiento molecular y la estabilización de cada uno de los complejos formados en el mecanismo propuesto para la activación de EhRacG y EhRho1, esto se hará mediante un barrido computacional de alaninas.

La generación de los modelos y la posterior identificación de los aminoácidos esenciales para el reconocimiento y estabilización de las proteínas, permitirá realizar una búsqueda por técnicas computacionales de moléculas orgánicas pequeñas con afinidad hacia una región de la superficie interfacial de GEF importante para el reconocimiento con las GTPasas, cuya unión impida la formación de los complejos. Un análisis estructural de los compuestos identificados como posibles inhibidores podría ser usado para generar ciertos cambios químicos en los compuestos con el fin de aumentar la afinidad y selectividad de estas moléculas sobre la GEF. Ésta podría ser una metodología que guíe el diseño racional de inhibidores de GEF y posteriormente la posible creación de una biblioteca de inhibidores.

1.3.2. Objetivos particulares

- Generar cada uno de los modelos moleculares para los complejos del mecanismo propuesto para la activación de las GTPasas: EhRacG y EhRho1, incluyendo los complejos ternarios que contienen GDP y GTP.
- Caracterizar energéticamente la superficie interfacial de cada uno de los complejos.

- Proponer detalles del mecanismo molecular del reconocimiento y selectividad de EhGEF1 con ambas GTPasas de la amiba.
- Buscar sitios de anclaje potenciales que sirvan para el diseño de inhibidores de EhGEF1.
- Realizar la búsqueda racional de moléculas orgánicas que pudieran actuar como inhibidores potenciales de la proteína EhGEF1.

Capítulo II. Métodos.

El procedimiento que se ha realizado para llegar a los objetivos que se han planteado se presenta en este capítulo, sin hacer énfasis en el detalle de los programas que se han utilizado, para no distraer la secuencia de la metodología. Para la descripción de las herramientas computacionales se ha dedicado el capítulo III.

2.1. Validación del procedimiento de modelado molecular por homología para los sistemas Rho-GEF

Con el fin de determinar la precisión del modelo molecular para sistemas Rho-GEF se estudió la correlación entre un complejo cristalográfico de la familia de las proteínas Rho y las proteínas GEF y la estructura del mismo complejo generada mediante el procedimiento de modelado molecular por homología.

La validación consistió específicamente en generar el modelo molecular del complejo Cdc42-ITSN, cuya estructura tridimensional ya había sido determinada mediante la técnica de cristalografía de rayos X (código PDB 1K11), a partir de otro complejo cristalográfico cuyas proteínas que lo forman tienen homología con las proteínas Cdc42 y ITSN; éste sirvió como plantilla para el procedimiento de modelado molecular por homología. La estructura que se utilizó como molde fue RhoA-Dbs (código PDB 1LB1). Cabe señalar que en este proceso de validación, el tipo de proteínas y los porcentajes de identidad entre el molde y la proteína por modelar son similares a los de la proteína de amiba y su molde.

Un criterio cuantitativo de comparación entre la estructura cristalográfica y la obtenida mediante el modelado hizo uso del servidor Alanine Scanning (www.bakerlab.org), que permitió determinar la contribución de cada uno de los residuos de la superficie de contacto GTPasa-GEF tanto en la estructura experimental como en la modelada. El resultado esperado era que la identificación de los residuos de la interfase GEF-GTPasa coincidieran

en ambas estructuras. Además se esperaba que cada uno de estos aminoácidos contribuyera de manera semejante a la estabilización de la interfaz.

Se modeló el complejo formado por las proteínas que se encuentran caracterizadas en 1KI1 a partir de la estructura 1LB1 con el programa MOE 2005.06. En el procedimiento de modelado molecular por homología se generan varios conformeros (en este caso 1,000) debido a los diferentes rotámeros posibles de las cadenas laterales que se reemplazan durante el proceso, y a las distintas conformaciones de los fragmentos de cadena principal que se reemplazan para contender con las inserciones y pérdidas de residuos. Se seleccionaron diez de los modelos intermedios que se distribuyen de manera uniforme en el intervalo de valores del índice de calidad de empaçado. A cada uno de estos diez conformeros se les realizó una minimización de energía además de un barrido computacional de alaninas. A la estructura cristalográfica también se le realizó un barrido virtual de alaninas la cual sirvió como control de comparación. El hecho de tomar varios modelos intermedios era determinar si había una correlación de los resultados con la calidad de empaçado que presenta cada uno de los modelos intermedios.

También se realizó una simulación de la dinámica molecular sobre los diez modelos intermedios elegidos con el propósito de mejorar la correspondencia con la estructura cristalográfica. De los conformeros generados en la dinámica molecular se eligió una muestra representativa de cuatro y se les efectuó un barrido virtual de alaninas a cada uno, los resultados se compararon con los obtenidos del barrido computacional de la estructura cristalográfica.

2.2. Generación de los modelos moleculares de los complejos que se forman durante el proceso de activación propuesto para EhRacG y EhRho1

El procedimiento que se siguió para la generación de los modelos de los tres complejos que se forman durante el proceso de activación de cada una de las GTPasas (para EhRacG: EhRacG-GDP-EhGEF1, EhRacG-EhGEF1 y EhRacG-GTP-EhGEF1 y para EhRho1:

EhRho1-GDP-EhGEF1, EhRho1- EhGEF1 y EhRho1-GTP-EhGEF1) fue similar al descrito en la sección anterior. A continuación se presentan detalles del trabajo que se realizó para la generación de los seis modelos.

2.2.1. Obtención de las secuencias de aminoácidos y estructuras tridimensionales que pudieran servir como moldes

Las secuencias de aminoácidos de las proteínas de interés EhGEF1, EhRho1 y EhRacG de *E. histolytica* fueron obtenidas del sitio NCBI (<http://www.NCBI.nlm.nih.gov/>).

El primer paso en el procedimiento de modelado molecular por homología es asignar a la proteína o proteínas de estructura desconocida a una familia de proteínas ya descrita. En nuestro caso deseamos realizar el modelo de complejos, en los cuales una de las proteínas pertenece a la familia de las proteínas Rho; la otra proteína pertenece a la familia de las Dbl GEF las cuales son directamente responsables de la activación de las GTPasas tipo Rho en células eucariontes. A continuación se buscaron en el Protein Data Bank (<http://www.rcsb.org>) complejos binarios formados por proteínas de estas familias que pudieran servir como molde, para ser usados en el procedimiento de modelado molecular por homología.

2.2.2. Alineamiento múltiple entre las secuencias problema y las secuencias molde

Una vez que se obtuvieron las secuencias de aminoácidos de las proteínas problema y las secuencias de aminoácidos de las proteínas de los complejos molde se realizó el alineamiento de EhRho1 y EhRacG con las GTPasas tipo Rho de los moldes. Este alineamiento nos permitió determinar cuáles son las posiciones conservadas y cuáles las variables. Además, el alineamiento permite determinar el porcentaje en identidad en secuencia entre las proteínas, el cual nos indica si el procedimiento de modelado puede ser aplicado y permite elegir a la mejor estructura de molde. Todos los alineamientos se realizaron con el programa CLUSTAL W que se encuentra disponible en la red

(<http://www.ebi.ac.uk/clustalw/>). También se realizó el alineamiento múltiple entre la secuencia de aminoácidos de la proteína EhGEF1 y las secuencias de las proteínas GEFs de los complejos molde.

Alineamiento basado en la estructura

Debido a que el alineamiento de secuencias es un paso crucial en el modelado molecular de proteínas, frecuentemente se realiza un realineamiento o ajuste manual, usando como información adicional el conjunto de las estructuras tridimensionales de los complejos molde. Lo primero fue realizar una superposición de todas las estructuras tridimensionales de las proteínas tipo Rho y de las proteínas GEF, y con los alineamientos múltiples se determinaron qué inserciones y eliminaciones son más convenientes, tomando en cuenta los aspectos estructurales de la estabilidad de las estructuras secundarias y terciarias en proteínas, las propiedades fisicoquímicas de los aminoácidos, la cercanía de los aminoácidos que se están alineando y la frecuencia de sustitución de cierto aminoácido por otro en las proteínas conocidas. Además este procedimiento permite visualizar cuáles son las regiones variables y conservadas en estas familias de proteínas, lo que finalmente ayudó a determinar cual es el mejor molde para ser usado.

2.2.3. Elección del molde

Después de haber realizado los alineamientos múltiples y topológicos entre las GTPasas Rho y las GEFs de las estructuras cristalográficas y las correspondientes de amiba (EhRho1, EhRacG y EhGEF1), se tiene que tomar la decisión de elegir un molde principal, los otros moldes no quedan eliminados debido a que se usan como plantillas secundarias que permiten modelar ciertas zonas en las cuales el modelo principal no presenta una alta similitud en secuencia con la proteína problema.

Un primer criterio que se tomó en cuenta para la elección del molde fue el porcentaje en identidad en secuencia que se presentó entre EhRacG y EhRho1 con las GTPasas molde.

Además del porcentaje en identidad en secuencia de EhGEF1 con las GEFs molde. Otro aspecto importante en la elección fue la resolución reportada para cada estructura cristalográfica, en el procedimiento de modelado molecular por homología se considera que una resolución de las estructuras plantilla menor a 2.5 Å es aceptable [30]. Finalmente se eligió como molde principal al complejo que presentó el mayor porcentaje en identidad en secuencia y la mayor resolución cristalográfica.

Una aportación importante de este trabajo será la construcción de los complejos ternarios. En el PDB no se encuentran disponibles complejos Rho-GEF unidos a nucleótido debido a que son inestables y no pueden ser cristalizados, pero se buscaron estructuras cristalográficas de proteínas de la familia de las proteínas Rho unidas a los nucleótidos GDP y GTP. Estos archivos nos servirán para tener las coordenadas iniciales que ambos nucleótidos poseen respecto a las GTPasas.

2.2.4. Construcción de los modelos

Se construyeron los modelos de las GTPasas EhRho1 y EhRacG en complejo únicamente con el dominio DH (esto es, sin el dominio adyacente PH), debido a que en ensayos *in vitro* de intercambio de nucleótidos con el dominio DH o con el módulo DH-PH de EhGEF1 no hay diferencia en la actividad bioquímica de la GEF sobre las GTPasas. Tal hecho sugiere que en este sistema el dominio PH no participa en el reconocimiento molecular entre las proteínas [6]. Para generar las estructuras se utilizó el programa MOE y se generaron 1,000 modelos para cada complejo, de entre ellas se seleccionaron diez que cubrieran uniformemente el intervalo de índices de empaquetamiento calculado por el mismo programa, y que fueron posteriormente minimizados y usados para su posterior análisis.

Para generar los modelos de los complejos ternarios, el procedimiento que se siguió fue similar al realizado para los modelos binarios, pero se requería además de conocer las coordenadas iniciales de los nucleótidos GDP y GTP. Para ello se usaron las estructuras cristalográficas de las proteínas unidas GTP y GDP. Se partió del modelo del complejo

binario correspondiente y sobre él se colocó uno de los nucleótidos en la posición observada en las GTPasas libres de GEF. A continuación se mantuvieron fijas las coordenadas de los nucleótidos y se realizó un proceso de minimización de energía de las dos proteínas que forman el complejo. Este procedimiento permite estimar el efecto del nucleótido sobre la porción proteica del complejo ternario.

2.3. Evaluación de la calidad estereoquímica de los modelos

Una vez construidos los complejos mediante el procedimiento de moldeado molecular por homología, es importante demostrar que las características estructurales de los modelos generados son razonables con respecto al conocimiento que se tiene de la estructura de las proteínas en la naturaleza. Esto es, se han realizado investigaciones acerca de la estructura tridimensional de las proteínas permitiendo observar principios básicos en la estructura proteica.

Algunos criterios para el análisis de la calidad de los modelos son: que la mayoría de los ángulos ϕ y ψ , se encuentren en regiones permitidas del mapa de Ramachandran y que los enlaces peptídicos sean planos. Adicionalmente, los átomos polares que se encuentren en el interior de la proteína deben formar los correspondientes puentes hidrógeno, debe haber un ambiente apropiado para los residuos hidrofóbicos e hidrofílicos y no deben existir contactos inadecuados átomo-átomo, ni huecos en el interior de la estructura. Si se encuentra alguna irregularidad, la estructura completa o ciertas regiones son sujetas a un refinamiento local o global con el fin de reajustar el modelo.

Los archivos de las coordenadas de los modelos fueron enviados al sitio *iMolTalk* - Structural Bioinformatics Toolkit, disponible en la red (<http://imoltalk.org>). En el Mapa de Ramachandran obtenido en ese sitio web se evaluó el número de residuos en la región no permitida, que nos indica la existencia de ángulos ϕ y ψ anormales, que son reflejo de la distorsión en la cadena proteica. Las coordenadas también fueron enviadas al sitio MolProbity que se encuentra disponible en la red (<http://kinemage.biochem.duke.edu>) y

que entre otras cosas permite identificar si hay algunas cadenas laterales que presenten conformaciones que sean poco comunes, posiblemente como resultado de la tensión local en la proteína.

2.4. Análisis de la contribución de los residuos de la superficie interfacial

Un procedimiento experimental dirigido a conocer la contribución de algunas cadenas laterales a la función de una proteína es llamado barrido de alaninas (Alanine Scanning). La idea es generar mutantes puntuales de cada aminoácido de interés, demostrar que la proteína se pliega correctamente y realizar un experimento que evalúe el grado de pérdida en la función que se asigna a la mutación creada. Para el caso de contactos proteína-proteína existe un barrido virtual de alaninas (<http://robetta.bakerlab.org/alaninescan>), en donde se determina la serie de aminoácidos que contribuyen al contacto entre dos cadenas de proteína, se genera un modelo molecular para cada uno de esos residuos en donde éste ha sido reemplazado por alanina y, finalmente, se realiza una evaluación del cambio en la energía de la asociación del complejo [8].

A cada uno de los diez conformeros elegidos de cada modelo construido se le realizó un barrido computacional de alaninas, lo cual nos permitió determinar qué residuos contribuyen energéticamente al contacto entre las proteínas de los complejos. El valor de $\Delta\Delta G$ (diferencia entre el ΔG del complejo silvestre y el ΔG de complejo mutante) que se reporta en este trabajo para cada residuo interfacial es el promedio de los resultados del barrido independiente que hace Alanine Scanning sobre cada conformero. Consideramos durante el análisis (de acuerdo a los autores del programa) que una mutación por alanina es significativa cuando tiene un valor de $\Delta\Delta G$ por arriba de 1 kcal/mol [8]. El valor de $\Delta\Delta G$ puede considerarse como la contribución energética de la cadena lateral de ese residuo a la estabilización de la interfaz del complejo.

La salida del programa Alanine Scanning es una lista (Fig. 6) que incluye los valores de $\Delta\Delta G$ de cada uno de los residuos de la superficie interfacial del complejos, la suma de cada

uno de los valores de $\Delta\Delta G$ representa la contribución de las cadenas laterales al ΔG de disociación del complejo, o con signo opuesto el ΔG de unión.

#Res	Cadena	aa	$\Delta\Delta G$ (kcal/mol)
			(complejo)
146	A	Gln	7.86
147	A	Arg	0.20
150	A	Arg	3.75
153	A	Leu	0.46
154	A	Leu	0.28
187	A	Asn	0.76
36	B	Val	1.11
37	B	Phe	0.04
38	B	Glu	0.37
39	B	Asn	8.55
40	B	Tyr	0.61
ΔG_d			23.99

Fig. 6. Ejemplo de una salida del programa Alanine Scanning.

2.5. Análisis comparativo de las energías de unión de los complejos usando un campo de fuerza

Una vez que se cuenta con los modelos para los complejos formados durante el proceso de activación de EhRho1 y EhRacG se realizó una comparación energética entre los complejos de ambas proteínas. Para ello se evaluó la energía de unión de cada uno de los complejos usando un campo de fuerza disponible en el programa MOE (CHARMM27, el cual está parametrizado para proteínas y nucleótidos).

El cálculo de la energía de unión se realizó de manera independiente sobre cada uno de los diez conformeros elegidos y minimizados para cada uno de los complejos, y el valor que se comparó fue el promedio obtenido para cada complejo, como consecuencia de los resultados obtenidos en la validación. El objetivo de la comparación fue poder determinar que GTPasa forma los complejos más estables, además de identificar si los complejos binarios son más estables que los ternarios.

Con el objetivo de determinar el efecto de los nucleótidos a larga distancia sobre los complejos, éstos fueron borrados de las estructuras tridimensionales de los complejos ternarios junto con el ion Mg^{+2} , sin permitir un reajuste molecular de las dos proteínas remanentes y se reevaluó la energía de unión en los complejos usando el mismo campo de fuerza disponible en el programa MOE. De manera similar a otros cálculos de este trabajo, la reevaluación se realizó de forma independiente sobre los diez conformeros que se seleccionaron y los valores que se muestran en el capítulo de resultados son el promedio de los diez resultados individuales.

2.6. Búsqueda de la base molecular del reconocimiento y selectividad

Como ya se ha mencionado a lo largo del documento, EhGEF1 reconoce tanto a EhRacG como a EhRho1. El alineamiento entre ambas proteínas muestra que existe una alta identidad en secuencia, aproximadamente del 80%, pero esto no podría explicarnos del todo que EhGEF1 reconozca a ambas proteínas. Se sabe que las regiones β_2 y β_3 de las GTPasas son unas de las primeras en ser reconocidas por las GEFs, lo que sugiere que los contactos en estas regiones dictan en gran medida la selectividad de determinada GEF por ciertas GTPasas. Con base en esto, y haciendo uso de los resultados del barrido computacional de alaninas, se buscaron residuos en las regiones β_2 y β_3 de EhRacG y EhRho1 que contribuyeran de forma importante a la estabilización de la interfaz [11].

El barrido computacional de alaninas también nos permitió identificar residuos de GEF1 que contribuyen a la formación de los complejos de EhRacG pero no para los de EhRho1 o

viceversa, es decir a los residuos responsables de la selectividad. Además se encontraron residuos de GEF1 que contribuyen a la formación de los complejos de ambas GTPasas pero con una diferencia significativa en sus valores energéticos al unirse a una u otra de las GTPasas. Este tipo de residuos también serían candidatos a ser responsables de la selectividad por una GTPasa u otra.

2.7. Búsqueda de moléculas líder para la inhibición de la proteína EhGEF1

Una vez que se presenta un estímulo en la membrana celular de la amiba para que se lleve a cabo algún proceso celular en el que participa la reorganización del citoesqueleto de actina, la proteína EhGEF1 reconoce a EhRho1 o a EhRacG y las activa. Una molécula que interactúe fuertemente con la zona de EhGEF1 que reconoce a las GTPasas, puede impedir la formación del complejo y entonces actuar como un inhibidor del proceso de activación. Esta molécula puede utilizarse como sonda molecular que permita inferir la vía metabólica en la que participa EhGEF1. Si resulta que EhGEF1 es fundamental para la invasión, desarrollo y supervivencia del parásito, estas moléculas podrían ser usadas como guía para el diseño de un fármaco antiamebiano. Debe aclararse en este punto que para el desarrollo de un fármaco todavía faltaría probar su capacidad para inhibir tanto *in vitro* como *in vivo*, además de todas las pruebas farmacológicas (farmacodinámica, toxicidad, etc.) y clínicas. Obtener un fármaco de la manera descrita anteriormente, comenzando con la identificación de una proteína diana y luego buscar o generar un posible ligando que inhiba su actividad es llamado “diseño de fármacos con base en la estructura”.

2.7.1. Procedimiento de diseño de inhibidores con base en la estructura

Si se conoce la base biológica de una enfermedad, por ejemplo la invasión de un parásito en el ser humano que causa una infección, o de un desarreglo metabólico, es posible diseñar un medicamento, utilizando un mecanismo de aproximación al proceso patológico. Cuando se conoce este proceso en su base molecular y se puede definir qué moléculas se encuentran implicadas y son fundamentales para el desarrollo de la patología, es posible diseñar un

medicamento que interactúe con las moléculas responsables, de tal forma que modifique su función y se modifique de esta forma la patología.

Lo primero, es definir la base molecular del proceso patológico, para lo cual es preciso conocer los diversos pasos implicados en el proceso fisiológico, que con lleva la realización de una función normal, y el conocimiento de qué pasó exactamente para que se presente la situación patológica. El punto clave es conocer cuales son las moléculas responsables de la enfermedad, frecuentemente son macromoléculas: enzimas, receptores, proteínas de señalización o ácidos nucleicos [15].

En ocasiones, una patología se reduce por el bloqueo de la acción de un efector sobre un receptor celular o en otros casos, por ejemplo en la invasión de un parásito, bloqueando la acción de alguna enzima indispensable para que el parásito pueda invadir al huésped.

El diseño de fármacos con base en la estructura consiste en conocer la geometría de una macromolécula que participa en un estado patológico, y con esto buscar o generar un posible ligando que inhiba su actividad. La enfermedad es entonces combatida al inhibir la función de la macromolécula participante [16].

2.7.2. Procedimiento del anclaje molecular sobre la proteína EhGEF1

Para realizar la búsqueda de posibles inhibidores de una proteína se puede usar una herramienta computacional llamada “docking”. El anclaje molecular o docking es la identificación de la posición y orientación más favorable de una molécula pequeña en el sitio de reconocimiento de una proteína. Un compuesto que interactúa o se une fuertemente a una proteína asociada con una enfermedad, puede inhibir su función y actuar como un fármaco. Realizar docking requiere de una estimación apropiada de la energía de interacción entre moléculas y de un algoritmo eficiente para la búsqueda de los diferentes modos de la unión u orientaciones entre el ligando y el receptor [15]. El procedimiento de

docking sobre la proteína EhGEF1 se realizó en tres etapas principales que se presentan a continuación.

2.7.2.A. Elección del sitio de anclaje

El sitio donde se realiza el anclaje del ligando sobre el receptor es por lo general una cavidad en la superficie del receptor, en donde algunos átomos pertenecientes a un grupo funcional del ligando pueden establecer interacciones favorables con el receptor.

Para el caso de la EhGEF1, se utilizó una herramienta llamada “Site Finder” que se encuentra disponible en MOE. Ésta buscó todas las posibles cavidades que podrían servir como sitio de anclaje. En particular nos interesaban las cavidades de la interfaz que interacciona con las proteínas EhRho1 y EhRacG. Algunos criterios que nos permitieron decidir cual podría ser el mejor sitio de unión fueron: a) el tamaño de la cavidad; entre mayor sea el número de residuos habrá una mayor cantidad de átomos que puedan servir como anclaje para los ligandos ya sea por la formación de puentes de hidrógeno o por las interacciones electrostáticas, b) el número de aminoácidos no conservados en la cavidad; la presencia de estos residuos permitirá encontrar ligandos específicos para la cavidad aumentando la selectividad, c) y cuál cavidad tenía aminoácidos que contribuyeran de forma importante a la formación de los complejos; si se interfiere con la interacción de residuos claves para la formación y estabilización de los complejos habrá una mayor posibilidad de que los ligandos sirvan como inhibidores potenciales.

2.7.2.B. Base de datos de ligandos

Se utilizó una base de datos de más 50,000 compuestos con potencial farmacológico ofrecida por la compañía Tripos (<http://www.tripos.com>), llamada LeadQuest. La serie de ligandos se representa como una colección de puntos en el espacio que corresponde a los átomos o grupos de átomos.

Cuando se realiza la simulación del anclaje molecular, se debe tener el interés no sólo de probar la molécula en su conformación de mínimo de energía global sino en todas aquellas conformaciones que pueden estar pobladas de forma significativa, hasta unas cuantas kcal/mol por encima del mínimo global. Para localizar todas estas conformaciones de baja energía definidas por una compleja hipersuperficie de energía potencial es necesario realizar una búsqueda conformacional. En nuestro caso se realizó el docking con las moléculas rígidas, aunque en otros trabajos paralelos se tienen ya resultados utilizando varios conformeros para cada una de las moléculas de ligando.

2.7.2.C. Ejecución del anclaje molecular

Una vez que se identificó el sitio de anclaje en el receptor y con la base de datos de ligandos preparada, se procedió a ejecutar el anclaje molecular mediante el programa Dock-MOE. Se eligió una metodología de colocación del ligando sobre el sitio de interacción, en este caso “Alpha Triangle”. Una vez finalizado el anclaje molecular de las 50,000 moléculas, la base de datos fue ordenada de acuerdo al valor de energía de unión, lo cual nos permitió seleccionar aquellos con las energías más bajas, es decir, los que tienen mayor afinidad por el sitio.

Capítulo III. Herramientas computacionales y sus aspectos teóricos.

El procedimiento realizado para llegar a los objetivos que se han planteado en el proyecto fue computacional, por ello es importante revisar de una manera general las bases teóricas del funcionamiento y uso de las herramientas computacionales utilizadas.

Existe una gran variedad de programas tanto comerciales como académicos que nos permiten realizar cada una de las tareas descritas en el capítulo II. Para la búsqueda de secuencias tenemos varios sitios disponibles en la red tales como; NCBI (National Center for Biotechnology Information), Uniprot (The Universal Protein Resource) y PIR (Protein Information Resource). La búsqueda de estructuras cristalinas de proteínas se puede hacer en el PDB (Protein Data Bank). Las herramientas para realizar el alineamiento entre las secuencias de las proteínas molde y las proteínas problema son entre otras; FASTA, BLAST (Basic Local Alignment Search Tool) y CLUSTALW. El barrido virtual de alaninas puede efectuarse en la página Web del laboratorio del Dr. David Baker (<http://www.bakerlab.org>). Para la construcción de los modelos y la simulación del acoplamiento se usó el programa MOE 2005.06 (<http://www.chemcomp.com>), además MOE también puede realizar alineamientos tanto por pares de secuencias como múltiples.

3.1. Buscadores de secuencias y estructuras tridimensionales

Los buscadores de secuencias listados en ExPASy, buscan sobre grandes bases de datos como NCBI, Uniprot y PIR. Estos recursos se encuentran disponibles de manera gratuita en la red y ofrecen una gran cantidad de información acerca de biomoléculas. El uso particular que nosotros le dimos fue en la búsqueda de la secuencia de las proteínas EhGEF1, EhRacG y EhRho1 y de la secuencia del complejo que nos sirvió como molde para la construcción de los modelos. Cuando se desea hacer una búsqueda en alguno de estos sitios

se debe conocer el código que se asigna a la proteína o complejo que se busca. Si no se cuenta con el código, se puede usar alguna palabra relacionada a dicho complejo o proteína.

La búsqueda de las estructuras cristalinas de los complejos que podrían servir como molde se realizó en el PDB, el cual es un depósito mundial para el procesamiento y distribución de datos tridimensionales de macromoléculas biológicas, en particular de proteínas. Es importante destacar que esta parte es de suma importancia para la construcción de un modelo adecuado debido a que se deben buscar y encontrar complejos con la mayor identidad en secuencia respecto a los complejos que se desean modelar.

3.2. Herramientas de alineamiento de secuencias de aminoácidos

Un alineamiento de secuencias biológicas consiste en establecer un segmento o segmentos entre ellas donde el número de coincidencias de sus residuos sea máximo. Una consideración fundamental es que las cadenas que sean alineadas deben estar relacionadas, es decir compartir un ancestro en común; o visto de otra forma, ser estructuralmente equivalentes. El alineamiento puede hacerse para un par de secuencias o sobre muchas de ellas, el cual es llamado alineamiento múltiple. También se pueden realizar alineamientos globales o locales; estos últimos encuentran los segmentos de dos secuencias con el mejor alineamiento entre ellas. A continuación se describen los aspectos básicos de los programas de alineamiento usados.

Alineamientos de secuencias usando el programa MOE 2005.06

Los alineamientos son calculados al desplazar una secuencia sobre la otra para optimizar el puntaje basado en la similitud de los residuos (obtenido de aplicar una matriz de sustitución de aminoácidos a cada par de residuos alineados) y disminuciones del puntaje debido a inserciones o pérdidas de residuos. La disminución de puntaje es impuesta por introducción y extensión de inserciones en una secuencia con respecto a la otra. El valor final de la función optimizada está referido como un puntaje de alineamiento [13].

Programa FASTA

Compara una secuencia de ADN o de proteína contra todas las secuencias de una base de datos o también compara pares de secuencias. FASTA puede ser muy específico cuando se usa para identificar regiones largas de baja similitud, especialmente para secuencias con una alta divergencia. El sitio de Internet donde se puede localizar el programa es: <http://www.ebi.ac.uk/fasta33/>.

Programa BLAST

El programa compara secuencias de aminoácidos o nucleótidos con todas las secuencias de una base de datos y calcula el peso estadístico de los apareamientos. BLAST puede ser usado para inferir relaciones evolutivas y funcionales entre secuencias. Su sitio Web es: <http://www.ncbi.nlm.nih.gov/BLAST/>.

Programa CLUSTALW

Es un programa que realiza alineamientos múltiples de secuencias, calcula el mejor emparejado de las secuencias seleccionadas y las acomoda de tal forma que se pueda observar las identidades, similitudes y diferencias. También permite observar relaciones evolutivas. Su sitio Web: <http://www.ebi.ac.uk/clustalw/>.

Un paso crítico en el desarrollo de modelos por homología es el alineamiento de las secuencias problema con las homólogas. Algunos factores que hay que considerar cuando se realiza un alineamiento son: (1) cuales algoritmos son los más convenientes para el alineamiento de secuencias (algunos mostrados anteriormente), (2) cual método de puntaje se aplica, (3) y cómo asignar las disminuciones de puntaje por inserciones o pérdidas de residuos.

3.3. Modelado molecular por homología

El modelado molecular por homología se fundamenta en la idea de que la estructura tridimensional se conserva más que la secuencia de aminoácidos. De esta forma, si se cuenta con la estructura cristalográfica o determinada por RMN de una proteína con similitud en secuencia a una proteína con estructura desconocida, esa estructura puede usarse como molde o plantilla en el procedimiento de modelado.

Cuando se tienen proteínas cuyas secuencias de aminoácidos poseen una identidad mayor al 30%, sus estructuras tridimensionales tendrán una alta similitud estructural, y ésta aumentará conforme su identidad en secuencia sea mayor. La similitud en estructura se puede cuantificar con el RMS de la distancia entre C_{α} de aminoácidos equivalentes en las dos proteínas en comparación, por debajo de 20% de identidad en secuencia se puede ver (Fig. 7) que hay una gran variación en la similitud estructural. Con base en esto, para poder aplicar el modelado molecular por homología es conveniente usar como plantilla una proteína con una identidad en secuencia mayor al 30% respecto a las secuencias con estructura desconocida.

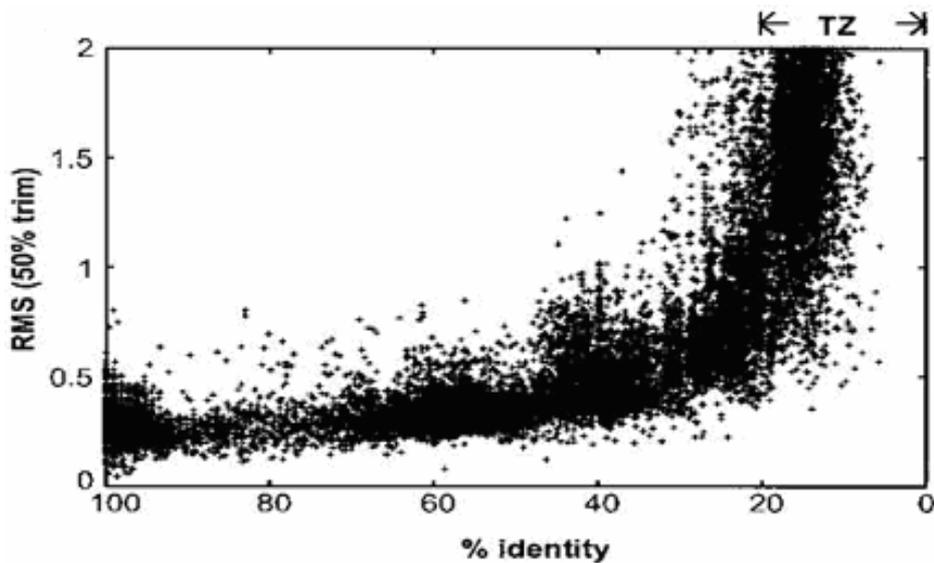


Fig. 7. Dependencia de la similitud en estructura (RMS) con el porcentaje de identidad en secuencia.

Para realizar el modelado por homología se utilizó el programa MOE que contiene la subrutina llamada MOE-Homology que es la que ejecuta el modelado como se describe enseguida:

Lo primero que realiza el programa es identificar las posiciones conservadas y variables entre el complejo problema y el molde. En este punto es importante tener un alineamiento topológico múltiple entre las diferentes proteínas que forman a los complejos usados como moldes, porque nos permite determinar regiones con estructura claramente conservada en esta familia de proteínas y qué inserciones y eliminaciones son más convenientes [13].

Las regiones en donde la superposición entre los complejos presenta una alta similitud estructural se denominan regiones estructurales conservadas (SCRs), mientras que aquellas en las que hay marcadas diferencias estructurales son llamadas regiones estructurales variables (SVRs) y corresponden principalmente a giros en las partes externas de las proteínas. Enseguida se generan las coordenadas para las SCRs copiando directamente las coordenadas del molde. Para generar las coordenadas de las SVRs de la cadena principal existen dos casos: si entre los moldes secundarios se encuentra algún asa con la misma longitud y secuencia de aminoácidos, sus coordenadas pueden ser copiadas; si no es así, se aplica un método de búsqueda de asas en una biblioteca para localizar aquellas que se ajusten lo mejor posible al segmento que se desea modelar. Para el modelado de las cadenas laterales de las SVRs se usa una biblioteca de rotámeros, en donde se encuentran los confórmeros de cadenas laterales observados en estructuras cristalográficas [13].

La combinación de diferentes asas y rotámeros permite la construcción de una amplia variedad de estructuras para cada sistema por modelar, a estas estructuras les llamaremos confórmeros, pues a pesar de tener exactamente la misma composición química divergen entre sí sólo por la orientación de sus cadenas laterales y el impacto de esta diferente orientación en el proceso de minimización. El programa tiene la opción de elegir un número especificado de las posibles estructuras. Una vez generados los confórmeros de los complejos se les realizó una minimización de energía.

Minimización de energía

La función algebraica que describe la energía potencial del sistema en función de las posiciones de los átomos recibe el nombre de campo de fuerza. Esta función está formada por la suma de varios términos que incluyen la distorsión de los enlaces covalentes respecto a valores de referencia (distancia, ángulos y rotaciones) e interacciones por pares entre átomos no unidos covalentemente (van der Waals, electrostáticas, y enlaces de hidrógeno). La energía es función del conjunto de coordenadas cartesianas, que especifican las posiciones de todos los átomos implicados, de una serie de parámetros energéticos, de las cargas atómicas y de la constante dieléctrica del medio (ϵ). Existen diferentes campos de fuerza usados para la minimización de energía para proteínas, entre ellos se encuentran AMBER, CVFF, CHARMM y GROMOS que son los más usados [13]. En particular para los cálculos realizados sobre los sistemas que se modelaron se utilizó el campo de fuerza CHARMM 27, el cual se encuentra parametrizado para ser usado no sólo en cálculos donde se involucren proteínas, sino también nucleótidos, condición indispensable para generar los complejos ternarios dado que incluyen GDP y GTP.

La minimización de la energía de sistemas macromoleculares se llevaba a cabo por cálculos iterativos sucesivos en los que se somete a la conformación inicial del sistema a un proceso de optimización geométrica, ya sea parcial o completa. Todos los parámetros que definen la geometría del sistema se modifican en pequeños incrementos hasta que la estructura alcanza un mínimo energético local. Aunque algunos algoritmos hacen uso de la derivada de la función (gradiente conjugado), o incluso de la segunda derivada (Newton-Rapshon), para guiar la minimización, ninguno de estos métodos puede garantizar que se ha encontrado la estructura con la energía más baja posible, el llamado mínimo global. La minimización puede proceder bien en coordenadas internas (enlaces y ángulos) o, más frecuentemente para proteínas, en coordenadas cartesianas (cada átomo esta caracterizado por unas coordenadas x , y , z , y los átomos se mueven con pequeños incrementos a lo largo de estos ejes).

3.4. Herramientas para el análisis de la contribución de los residuos de interfaz

Con el fin de determinar la contribución de cada aminoácido de la superficie interfacial a la energía libre de asociación en los complejos se les realizó un barrido computacional de alaninas con el programa Alanine Scanning.

La entrada del programa consiste de las coordenadas de la estructura tridimensional de un complejo proteína-proteína y la especificación de cuales son las cadenas cuyo contacto se desea estudiar, enseguida se realiza la etapa de preparación donde se adicionan los átomos de hidrógeno a la estructura, se optimiza la red de puentes de hidrógeno y se definen automáticamente a los residuos interfaciales.

Un residuo interfacial se define como aquél que tiene al menos un átomo a una distancia menor o igual 4 Å de uno de los átomos que pertenecen a la otra proteína del complejo. Adicionalmente se consideran residuos interfaciales a los que tienen su C_{β} a menos de 8 Å de un C_{β} de la proteína compañera en el complejo [8].

Entonces, el programa reemplaza cada uno de los residuos interfaciales por residuos de alanina, y calcula el efecto de la mutación sobre la energía libre de asociación del complejo. Para calcular la energía libre el programa usa una función que consiste de una combinación lineal de varios términos contribuyentes a la energía libre, entre ellos un término atractivo y uno repulsivo representados por un potencial tipo Lennard-Jones, un potencial que evalúa la energía de contactos por puente de hidrógeno de la cadena lateral con la cadena principal y uno de contactos por puente de hidrógeno de la cadena lateral con la cadena lateral, además de un término que simula el efecto de hidratación del solvente sobre las superficies [8].

La salida es una lista de aminoácidos interfaciales con la estimación de su contribución a la unión intercadenas (generalmente estabilizante), el valor que indica el grado de estabilización es el $\Delta\Delta G$ de unión (diferencia entre el ΔG de asociación en el complejo

silvestre y el calculado en la mutante puntual), tomando como dato de la literatura que una mutación por alanina desestabiliza significativamente cuando se predice un valor de $\Delta\Delta G$ causado por su mutación por alanina por arriba de 1 kcal/mol [8].

3.5. Herramienta de anclaje molecular

La herramienta utilizada para realizar el docking sobre la proteína EhGEF1 fue MOE-Dock que es una subrutina que se encuentra disponible en el programa computacional MOE. Realizar docking requiere de una estimación apropiada de la energía de interacción entre moléculas y de un algoritmo eficiente para la búsqueda de los diferentes modos de la unión u orientaciones entre el ligando y el receptor.

El punto inicial consiste en tener la estructura tridimensional de la macromolécula sobre la cual se desea hacer “docking”, así como la estructura del ligando que se pretende unir o asociar, ambas estructuras pueden ser tomadas de bases de datos, para las proteínas ya se ha mencionado al PDB, para los ligandos existe el Cambridge Crystallographic Database y diversas bibliotecas de compuestos con potencial farmacológico, entre otras [16]. Alternativamente, puede modelarse el receptor, el ligando o ambos componentes.

Algunos de los factores que se tiene que considerar para el realizar el docking son:

- a) Conocimiento del sitio de unión óptimo en la macromolécula. En muchos casos el lugar de anclaje es el propio sitio activo, aunque en ocasiones el sitio de docking puede ser diferente al sitio activo, lo que supone exploraciones adicionales.
- b) Determinación de la orientación relativa del ligando con respecto al sitio del receptor. Por lo general los métodos computacionales buscan aquellas orientaciones que dan lugar a las energías de interacción más favorables.
- c) Caracterización de la conformación óptima del ligando en el sitio de unión.

d) Evaluación de los posibles cambios conformacionales en el sitio del receptor.

Las técnicas más comunes de docking tratan inicialmente al receptor como si fuera completamente rígido, la orientación del ligando respecto al receptor se va ajustando de forma iterativa por movimientos de traslación y rotación, mientras que las conformaciones del ligando se generan mediante giros de los enlaces rotables. Esta maniobra de acoplamiento se simplifica considerablemente si se cuenta con la información sobre las características químicas de las moléculas. Los cálculos energéticos proporcionan asimismo una medida de la energía de interacción, y sirven como guía para elegir la orientación preferida de una molécula respecto a la otra. La función más simple usada para evaluar la energía de interacción está compuesta de un término estérico y otro electrostático, ignorando la contribución entre átomos que pertenecen a una misma molécula (Ec. 1) [13].

$$E_{\text{int}} = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right) \quad [1]$$

MOE divide el procedimiento de docking en los siguientes pasos

Análisis conformacional: los ligandos son tratados de una manera flexible mediante la rotación de los enlaces con libertad de giro. Alternativamente, las conformaciones del ligando pueden ser provistas en una base de datos de confórmeros donde cada una de ellas se trata posteriormente de manera rígida.

Si las conformaciones del ligando no son provistas por una base datos, el programa las generará a partir de una sola estructura tridimensional aplicando un conjunto de ángulos de torsión para los enlaces giratorios. Para ligandos pequeños, una búsqueda sistemática es conducida para generar todas las combinaciones de ángulos posibles, en el caso de ligandos más complejos o flexibles, se produce un muestreo aleatorio de confórmeros.

Colocación: se genera una colección miles de formas en las que el ligando se puede unir a la proteína usando cada una de las conformaciones del ligando. En este caso se eligió “Alpha Triangle”, este método genera las orientaciones por la superposición de tres átomos del ligando sobre tres puntos de anclaje óptimo predichos en el sitio del receptor.

Metodología de puntaje: se usa una función de puntaje para evaluar y mejorar las orientaciones del ligando en el sitio, así como para determinar cuál de los candidatos propuestos tiene la menor energía de unión que corresponde a una mayor afinidad potencial [13].

Para evaluar cada orientación del ligando sobre el sitio, MOE usa una función de puntaje que estima la contribución de la entalpía a la energía libre de unión usando una combinación lineal, (ecuación 2 y Tabla 1) [13].

$$H = C_{hb}f_{hb} + C_{ion}f_{ion} + C_{mlig}f_{mlig} + C_{hh}f_{hh} + C_{hp}f_{hp} + C_{aa}f_{aa}\dots\dots\dots[2]$$

Donde:

f es un factor de peso calibrado por los autores del paquete.

C es cada una de las diferentes contribuciones a la entalpía.

Término	Tipo de interacción
Hb	Por puente de hidrógeno
ion	Iónica
mlig	Entre metales
Hh	Hidrofóbica
Hp	Entre átomos hidrofóbicos y polares
Aa	Entre dos átomos

Tabla 1. Términos de la función de la energía libre.

Capítulo IV. Análisis y discusión de resultados obtenidos.

En este capítulo se presentan los resultados de la validación del procedimiento de modelado molecular por homología para sistemas Rho-GEF. A continuación se describen los resultados de la búsqueda de moldes y la elección de uno para modelar los seis complejos de EhGEF1 que son los sujetos de estudio de este trabajo. También se describen los resultados y la discusión del análisis estructural y energético de cada uno de los modelos generados, se discute el mecanismo de reconocimiento y selectividad. Finalmente se presentan los resultados y discusión de la simulación del anclaje molecular, esto es, la búsqueda de moléculas orgánicas con potencial de inhibición sobre la proteína EhGEF1 con posibilidad de actuar como moléculas líder para el diseño de sondas moleculares, e incluso de fármacos antiamicibianos.

4.1. Resultados de la validación del procedimiento de modelado molecular por homología para sistemas Rho-GEF

El objetivo de la validación fue tener la posibilidad de comparar la estructura de un complejo GEF-GTPasa, generada mediante el procedimiento de modelado molecular por homología, con la obtenida mediante cristalografía de rayos X. En específico, se modeló el complejo Cdc42-ITSN a partir del complejo RhoA-Dbs (1LB1). La comparación y análisis del barrido computacional de alaninas realizado tanto a las estructuras modeladas como a la cristalográfica, indican lo siguiente:

- (1) En los modelos generados se pueden identificar prácticamente los mismos aminoácidos que forman la zona interfacial que se reconocen a partir de la estructura cristalográfica.
- (2) Existe una gran semejanza estructural entre las estructuras tridimensionales de los complejos obtenidos mediante el procedimiento de modelado molecular por homología y la estructura cristalográfica del complejo, con una diferencia medida por su RMSD entre 1.1 y

1.4 Å. Esta es la precisión esperada entre los modelos que se construirán y la estructura que se pudiera determinar experimentalmente.

(2) No se encuentra correlación entre la función de empaqueo del programa MOE (que propone cuál de los modelos construidos tiene la mejor calidad) y la capacidad predictiva de los modelos.

(3) La simulación de dinámica molecular realizada algunos de los conformeros no aumentó su correlación con los resultados obtenidos del barrido de alaninas para la estructura cristalográfica. Por esta razón únicamente se realizará minimización de energía a los modelos generados.

(4) Un promedio de los valores energéticos que arroja el programa Alanine Scanning ($\Delta\Delta G$) obtenidos a partir del análisis de varios modelos (en este caso siete) para el mismo complejo GTPasa-GEF, representa de una manera más adecuada a los resultados que se obtienen del análisis de la estructura cristalográfica en lo referente a la contribución de los aminoácidos de la superficie interfacial (Tabla 2).

Residuos	$\Delta\Delta G$ (kcal/mol)		$\Delta\Delta G$ (kcal/mol) Confórmero 1
	Estructura cristalográfica	Promedio (diez conf.)	
Y13	2.16	1.22 (± 0.83)	1.10
E16	2.95	3.51 (± 0.37)	3.68
L148	1.31	1.27 (± 0.27)	1.07
M151	1.30	1.09 (± 0.35)	0.94
I160	0.44	0.49 (± 0.13)	0.61
T234	2.04	1.16 (± 0.30)	0.97
V235	0.98	1.12 (± 0.12)	1.07

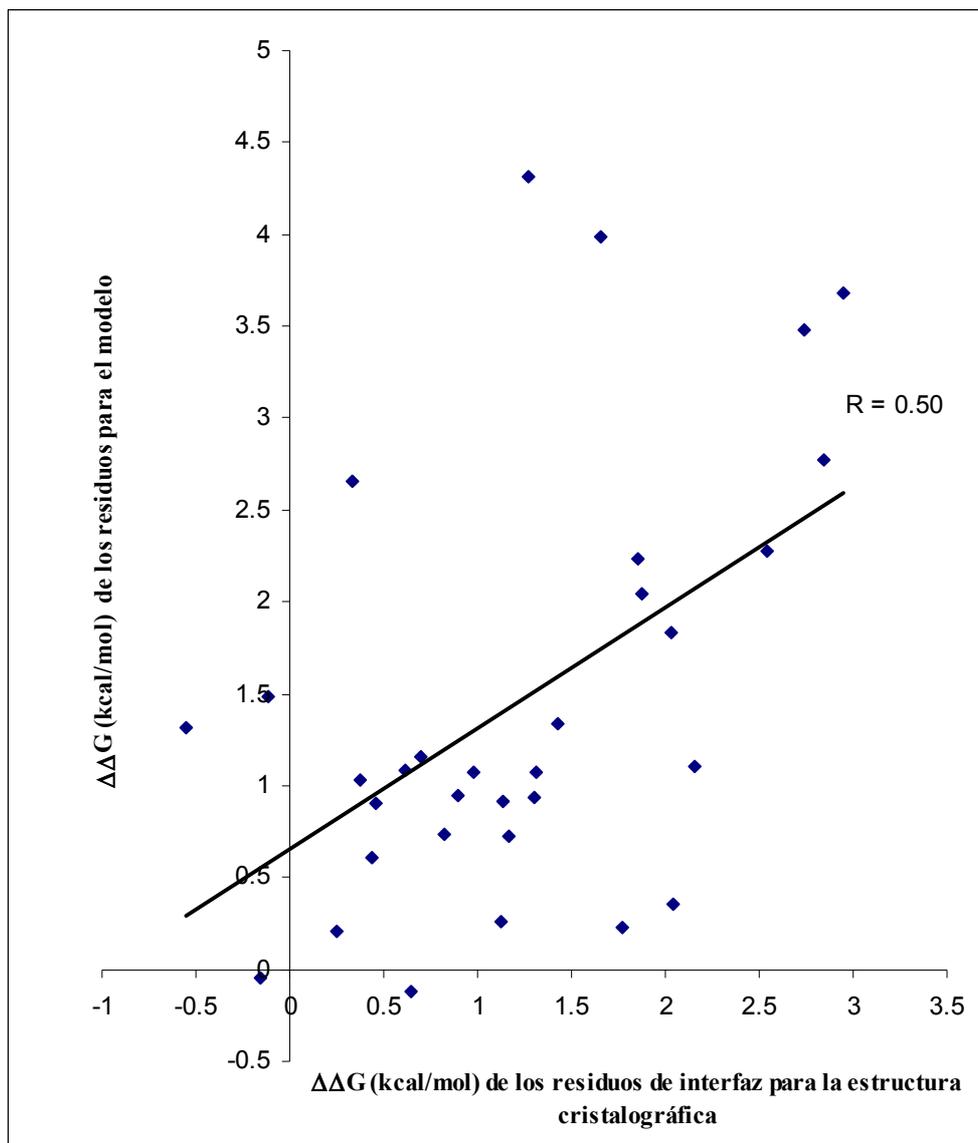
Tabla 2. Valores de $\Delta\Delta G$ de algunos residuos en la estructura cristalográfica, el modelo 1 y el promedio de diez siete conformeros que nos permite concluir el punto (3).

El procedimiento de validación además de indicarnos que el modelado por homología puede ser aplicado con confianza a los sistemas Rho-GEF de la amiba, también nos permitió diseñar la estrategia a seguir para el manejo de la información del barrido

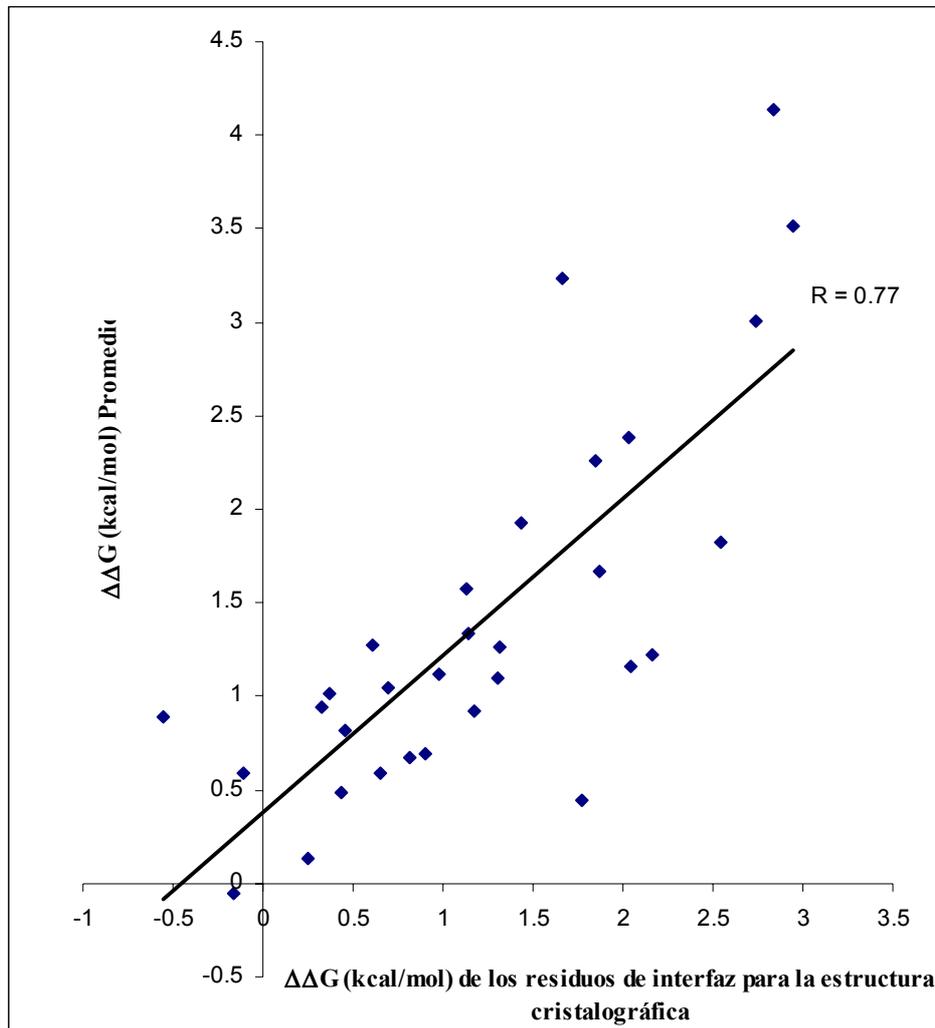
computacional de alaninas, en lo referente a tomar el promedio de valores de $\Delta\Delta G$ de diez de los diferentes conformeros de cada complejo modelado. Estos conformeros se seleccionaron de acuerdo a una distribución uniforme de los 1,000 conformeros para cada complejo, ordenados según el criterio de empaqueo de MOE. Esta estrategia de selección permite determinar si existe una correlación entre la calidad predictiva del modelo y su empaqueo. En el Gráfico 1 se pueden observar los valores de $\Delta\Delta G$ de los residuos de interfaz para el conformero 1 contra los valores de $\Delta\Delta G$ de los residuos de interfaz para la estructura cristalográfica. Se obtienen resultados similares si se usan los valores de $\Delta\Delta G$ de los residuos de interfaz de los otros conformeros. En el Gráfico 2 se muestran en el eje de las abscisas los valores de $\Delta\Delta G$ para los mismos residuos en la estructura cristalográfica pero ahora en el eje de las ordenadas se representa el promedio de los valores de $\Delta\Delta G$ obtenidos independientemente de los siete conformeros. Se puede observar claramente que cuando se utiliza el promedio de los valores se tiene una mayor correlación entre los valores de $\Delta\Delta G$ de la estructura cristalográfica y los de los modelos Tabla 3.

	Coefficiente de Correlación (R)
Conformero 1	0.50
Conformero 2	0.61
Conformero 3	0.47
Conformero 4	0.61
Conformero 5	0.67
Conformero 6	0.59
Conformero 7	0.44
Conformero 8	0.51
Conformero 9	0.48
Conformero 10	0.62
Promedio	0.77

Tabla 3. Coeficientes de correlación presentes entre los resultados obtenidos de la estructura cristalográfica y cada conformero, así como el que se presenta usando un promedio de los diez conformeros.



Gráfica 1. Valores de $\Delta\Delta G$ de los residuos de interfaz para el modelo 1 contra los valores de $\Delta\Delta G$ de los residuos de interfaz para la estructura cristalográfica.



Gráfica 2. Valores promedio de $\Delta\Delta G$ de los siete conformeros contra los valores de $\Delta\Delta G$ de los residuos de interfaz para la estructura cristalográfica.

4.2. Resultados de la búsqueda de secuencias y estructuras tridimensionales que pudieran servir como molde

Las secuencias de aminoácidos de EhRacG, EhRho1 y EhGEF1 fueron encontradas en el sitio NCBI con códigos: AAC24704, P3102 y AAQ18153, respectivamente. Por otra parte, se encontraron un total de 5 archivos de complejos GTPasa-GEF que pueden servir como plantilla para la construcción de los modelos moleculares, estos son: Rac1-Tiam1, Cdc42-

ITSN, RhoA-Dbs, RhoA-LARG y RhoA-PDZRhoGEF; sus códigos en el PDB son: 1FOE, 1KI1, 1LB1, 1X86 y 1XCG, respectivamente.

4.3. Resultado de los alineamientos múltiples entre las secuencias problema y las secuencias molde

Los alineamientos de EhRho1 y EhRacG con las tres GTPasas Rho de los posibles moldes (Rac1, Cdc42 y RhoA) nos permitieron determinar cuáles son las posiciones conservadas y variables entre estas proteínas. Para EhRacG se puede observar que existe una mayor identidad en secuencia (Tabla 4) respecto a EhRho1, pero ambas presentan una identidad de al menos 69% con las proteínas molde lo cual nos indica que sus estructuras tridimensionales serán muy similares y que el procedimiento de modelado molecular por homología puede ser aplicado con toda confianza.

También se realizó el alineamiento entre la EhGEF1 y las GEFs de las estructuras que pueden servir de molde (Tabla 5). Para este alineamiento existe una mayor cantidad de regiones variables, con una identidad en secuencia de 19 a 28 %, aunque con la presencia de residuos conservados simultáneamente en todas las secuencias. Cabe señalar que entre las proteínas GEF con estructura conocida existe un porcentaje de identidad de 19% a 34%, y que a pesar de la relativamente baja similitud en secuencia, ellas presentan divergencias en su estructura tridimensional con valores de RMSD entre 1.9 y 2.6 Å, lo cual nos permite confiar que puede hacerse modelado por homología para esta familia con alta conservación estructural a pesar de los porcentajes bajos mostrados en la tabla 5.

4.4. Elección de la estructura molde

De las cinco estructuras encontradas, un primer complejo que fue eliminado como plantilla es 1FOE debido a que en estudios previos se utilizó como molde para generar la estructura de la proteína EhRho1 se tenía el antecedente que ciertos residuos en esta proteína eran importantes para el reconocimiento con la proteína EhGEF1. Al realizar el estudio

computacional para determinar qué residuos en el modelo contribuían al contacto proteína-proteína hubo algunos residuos que no mostraron serlo siendo que la evidencia experimental indicaba lo contrario, al realizar un análisis estructural se descubrió que en IFOE había una variación muy importante en la orientación de ciertos aminoácidos y esto impidió que se modelara adecuadamente esta zona en la EhRho1. Para las proteínas de los complejos restantes en las Tablas 4 y 5 se presentan los porcentajes de identidad en secuencia con las proteínas de interés.

	RhoA	Cdc42
EhRacG	81%	87%
EhRho1	71%	69%

Tabla 4. Porcentaje de identidad en secuencia entre las proteínas EhRacG y EhRho1 y las GTPasas molde.

	Dbp	ITSN	LARG	PDZRhoGEF
EhGEF1	27%	28%	24%	19%

Tabla 5. Porcentaje de identidad en secuencia entre la proteína EhGEF1 y las GEFs molde.

Se puede observar que para el caso de las GTPasas, EhRacG presenta un mayor porcentaje en identidad en secuencia con Cdc42 mientras que EhRho1 con RhoA. Para el caso de EhGEF1 se tienen con mayor porcentaje ITSN y Dbp. Por lo tanto los complejos que podrían servir como molde son 1KI1 (Cdc42-ITSN) y 1LBI (RhoA-Dbp). Otro aspecto importante en la elección fue la resolución cristalográfica de los moldes, en el procedimiento de modelado molecular por homología se considera que una resolución de las estructuras plantilla menor a 2.5 Å es aceptable [30], la resolución de los complejos es: 1KI1 2.30 Å, 1LBI 2.81 Å, 1X86 3.22 Å, 1XCG 2.50 Å. Debido a que los mejores en porcentaje en identidad en secuencia son los complejos 1LBI y 1KI1, pero el complejo

1KI1 es el que tiene la mejor resolución, esta estructura fue elegida como plantilla para la generación de los seis modelos.

En el PDB no se encuentran disponibles complejos Rho-GEF unidos a nucleótido para generar los modelos ternarios. Sin embargo existe la estructura cristalográfica de Cdc42 (libre de GEF) unida al nucleótido GDP, cuyo código PDB es 1AN0. Por otra parte, para modelar a los complejos con GTP se encontró a RhoA (también libre de GEF) unida a un análogo de GTP, con código 1A2B. Estos dos archivos nos sirvieron como molde de las coordenadas que ambos nucleótidos tienen respecto a las GTPasas para modelar los complejos ternarios. Es muy importante enfatizar aquí, como ya se había mencionado antes, que los complejos ternarios no son fáciles de preparar para ser cristalizados debido a su inestabilidad. Los complejos Rho-GDP-GEF liberan al nucleótido rápidamente, mientras que los Rho-GTP-GEF tienen baja afinidad por la GEF. Dada la inaccesibilidad experimental para obtener sus estructuras, el producir los modelos de estos complejos ternarios es una de las contribuciones más importantes de este trabajo.

4.5. Análisis de las características estructurales de la familia de las GTPasas Rho y de la familia de las Dbl GEFs

Características estructurales generales de las proteínas Dbl GEFs

Todas las GEFs comparten un dominio DH (dominio de homología a Dbl) de cerca de 200 aminoácidos, que lleva a cabo la actividad catalítica de reconocer a una GTPasa específica e inducir el intercambio de nucleótidos. También poseen un dominio PH (dominio de homología a pleckstrina) de aproximadamente 100 aminoácidos, adyacente al dominio DH, que tiene como función anclar a la proteína a la membrana y en algunos sistemas participa en la actividad de intercambio. Las estructuras tridimensionales de los dominios DH, que han sido determinadas, tanto libres como en complejo con GTPasas Rho, muestran estar constituidas principalmente por 10 a 15 hélices α , arregladas de forma antiparalela para

formar un atado rectangular helicoidal que ha sido denominado como una “chaise longue” (silla larga) [3].

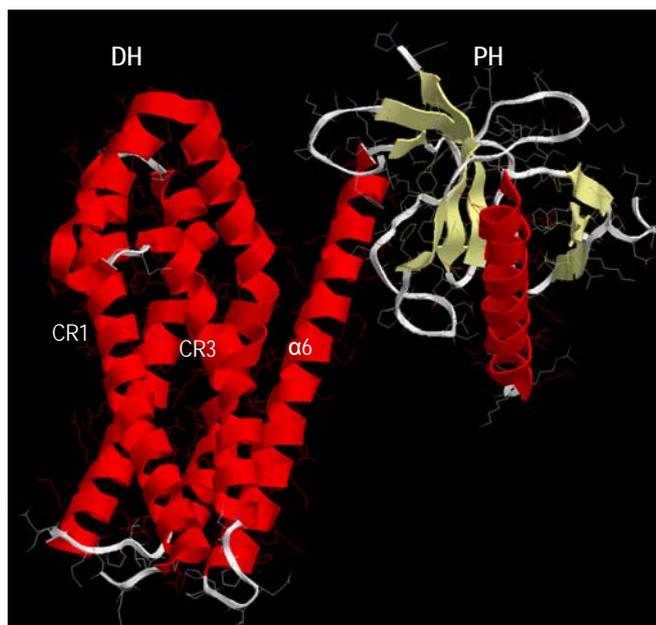


Fig. 8. Estructura tridimensional de las proteínas GEFs. En esta imagen se puede visualizar los dominios presentes en esta familia de proteínas (DH a la izquierda y PH, arriba a la derecha), así como regiones importantes para el contacto con sus GTPasas blanco (CR1, CR3 y α_6).

El dominio DH tiene tres regiones conservadas CR1 (entre los residuos 183-215), CR3 (entre los residuos 310-340) y la hélice 6α (349-372), las cuales se empaquetan para formar el centro del dominio. Estas regiones constituyen la mayor parte de la superficie interfacial de interacción con las GTPasas (Fig. 8) [3].

Características estructurales generales de las GTPasas Rho

Las GTPasas tipo Rho están formadas por una hoja β de seis hebras, donde cada una está conectada por una hélice α o por giros. La hoja β está formada por la asociación

antiparalela de dos hebras β (B2 y B3) y la asociación paralela de cinco hebras (B3, B1, B4, B5, B6). Las GTPasas Rho cuentan con 5 hélices α (A1, A3, A3', A4 y la A5) y tres hélices 3_{10} (H1, H2 y H3) (Fig. 9) [1].

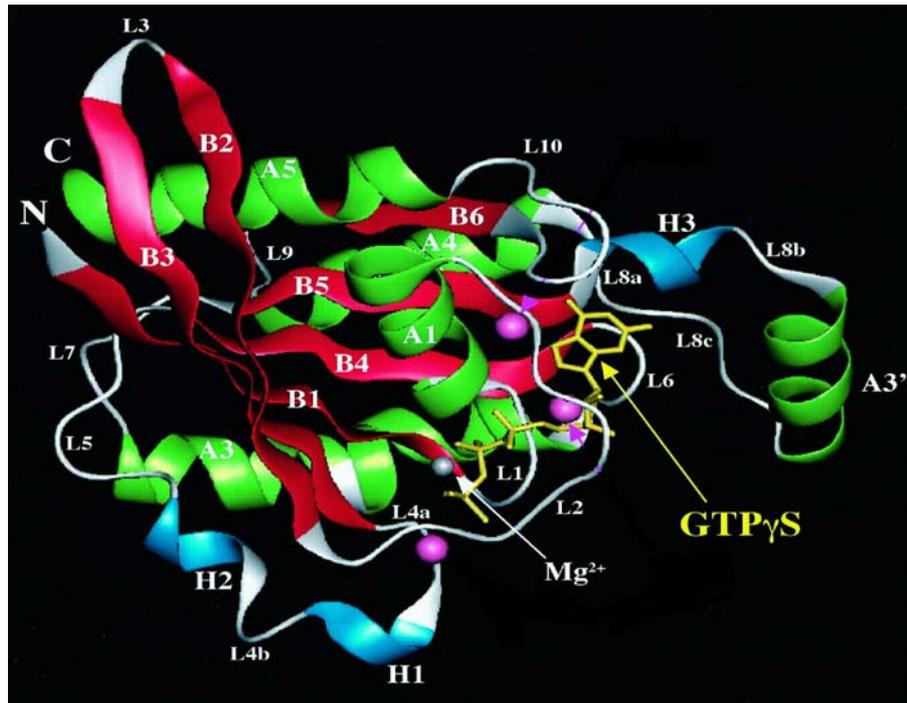


Fig. 9. Proteína RhoA que nos permite visualizar la estructura tridimensional característica de las GTPasas Rho [4].

Las regiones de importancia en la interacción con la GEF son: el *switch 1* formado por los residuos 28-38 (entre la hélice A1 y la hebra B2) y el *switch 2* que se encuentra entre la hebra B3 y la B4 que corresponde a los residuos 61-79. El sitio de unión al nucleótido está localizado entre los aminoácidos 13-20, además interaccionan con el nucleótido en forma general los residuos Y34, G62, D120, S160, A161, K162, la numeración de los residuos del *switch 1* y 2 y el bolsillo del nucleótido varían de acuerdo al tipo de GTPasas como se puede ver en la Fig. 10 [4]. Todas las GTPasas pequeñas tienen el bolsillo de unión a nucleótido y Mg^{2+} . El nucleótido (GDP o GTP) unido a la GTPasa modula la conformación

de las regiones *switch*. El Mg^{2+} es requerido para aumentar la afinidad de los nucleótidos de guanina por la GTPasa [26].

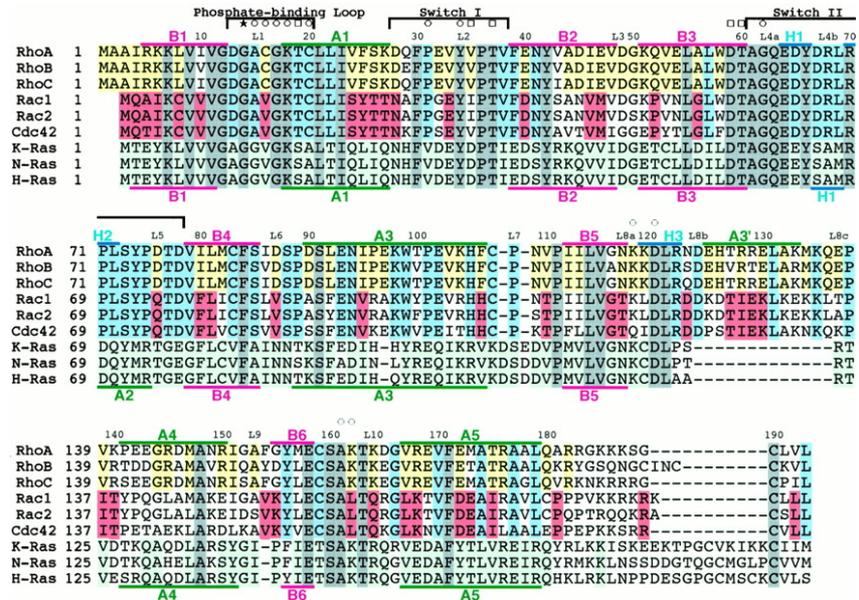


Fig. 10. Alineamiento de secuencias de GTPasas pequeñas. Este alineamiento nos permite visualizar los residuos que forman las regiones *switch* 1 y *switch* 2, así como los elementos de estructura secundaria que se presentan comúnmente en esta superfamilia de proteínas. Las hélices α son representadas con la letra A (A1-A5), las hebras β con la letra B (B1-B6) y las hélices 3_{10} con la letra H (H1-H3) [4].

Interfaz entre los dominios DH y GTPasas Rho

El dominio DH interactúa directamente con las regiones *switch* de las GTPasas tipo Rho. El *switch* 1 interactúa con CR1 y CR3; Existe un residuo altamente conservado en la familia de las Dbl GEF en la región CR1 (Glu 190) crucial para la formación del complejo y la actividad de intercambio del nucleótido. El *switch* 2 hace contacto predominantemente con CR3 y algunas porciones de la hélice $\alpha 6$. Existen dos residuos en el dominio DH (Lys 329 en la región CR3, muy conservado en la familia, y el residuo semi-conservado Asn 366) que también tienen una interacción significativa con el *switch* 2 en el mecanismo de intercambio [3]. Estas interacciones causan un reacomodo de las regiones *switch* que altera significativamente el bolsillo de unión del nucleótido, con lo cual el GDP es expulsado. En

particular, el *switch* 1 se mueve a lo largo del bolsillo de unión al nucleótido, perturbando algunas de las interacciones con el nucleótido y el Mg^{2+} , mientras que el reacomodo del *switch* 2 causa efecto similar (Fig. 11) [3].

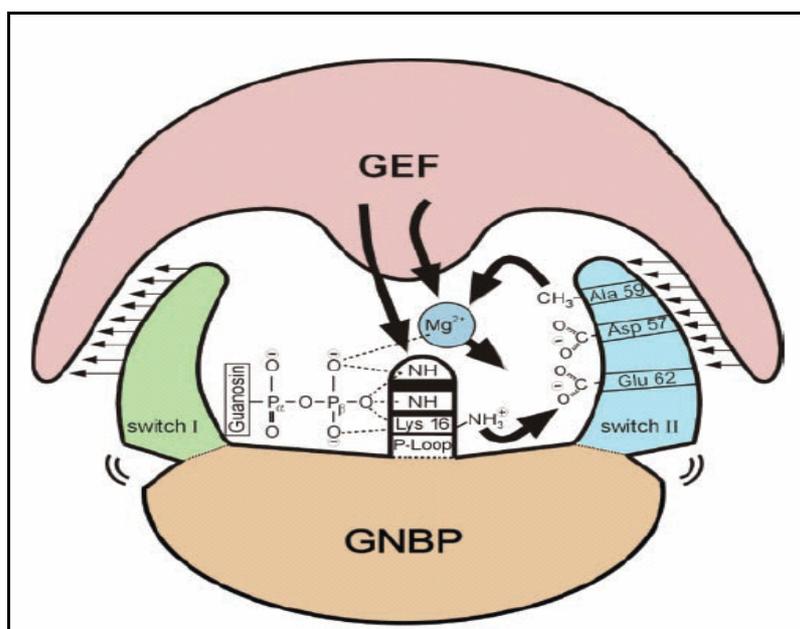


Fig. 11. Diagrama esquemático del mecanismo general de expulsión de GDP por un GEF. GNBP abreviatura de guanine nucleotide-binding protein. La numeración coincide con las proteínas Rac y Cdc42 del alineamiento de la Fig. 10.

4.6. Análisis de los modelos del proceso de activación de EhRho1 por EhGEF1

De acuerdo a los resultados de la evaluación de la calidad estereoquímica de los diez conformeros para cada uno de los tres complejos formados entre las proteínas GEF1 y Rho1 y los tres complejos formados entre las proteínas GEF1 y RacG, presentaron entre 1.9% y 2.2% de residuos en las regiones prohibidas del mapa de Ramachandran y entre 9% y 10% de rotámeros inusuales, comparado con 1.7% y 5.3% respectivamente para la estructura molde (1KI1). En todos los casos los aminoácidos de menor calidad estructural estuvieron lejos de la región interfacial. Esto nos indica que los análisis de contacto proteína-proteína realizados sobre los modelos son confiables.

4.6.1. Modelo EhRho1-GDP-EhGEF1

El barrido computacional de alaninas realizado a este modelo nos permitió identificar residuos conservados que típicamente tienen una contribución importante en la superficie interfacial de complejos Rho-GEF, estos son: Glu 190, Arg 329 y Asn 366 con valores de $\Delta\Delta G$ de 2.26, 4.01 y 1.70 kcal/mol, respectivamente. Un residuo no conservado en la familia que resultó importante para este complejo es Tyr 183 ($\Delta\Delta G= 1.53$ kcal/mol), y dado que no se encuentra conservado entre especies este valor alto sugiere que podría tratarse de un residuo que contribuye en la formación de complejos específicos para amiba.

Debido a la alta identidad en secuencia en la familia de las GTPasas nos hace suponer que los mecanismos de expulsión del nucleótido GDP, ingreso del GTP y disociación de la GEF son similares a los propuestos en la literatura [3].

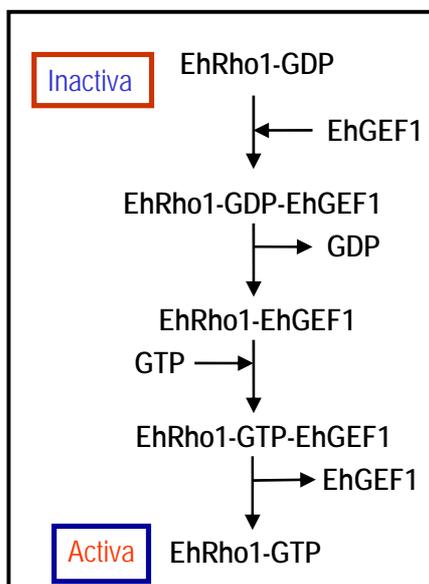


Fig. 12. Etapas del mecanismo propuesto para el proceso de activación de EhRho1 por EhGEF1.

El complejo ternario EhRho1-GDP-EhGEF1 es el primero que se forma durante el proceso de activación de la EhRho1 (Fig. 12). Una vez que se presenta un estímulo en la membrana

celular de la amiba para que se efectúe un proceso celular en el que se requiere la reorganización del citoesqueleto de actina, la EhRho1 cargada de GDP es reconocida por EhGEF1 con el propósito de activar a EhRho1 (reemplazando GDP por GTP). Para ello la EhGEF1 debe reconocer específicamente a sus GTPasas blanco, en este caso EhRho1, y el GDP debe ser disociado. Se encuentran tres residuos de EhRho1 que interaccionan con el GDP: Cys 35, Ile 45 y Thr 52.

Un residuo clave de EhGEF1 en la región conservada 1 (CR1) es el Glu 190 que interacciona con Tyr 49, Val 53 y Thr 52 de la región *switch* 1 de EhRho1 formando tres puentes de hidrógeno en el que intervienen los grupos amino de la Thr 52 y la Val 53, así como el grupo hidroxilo de la Tyr 49. Tales interacciones ocasionan un cambio conformacional en el giro correspondiente al *switch* 1 (Fig. 13). Esto hace que Ile 45 y Cys 35 se alejen del GDP, disminuyendo notablemente la estabilidad del nucleótido, sobre todo porque la alteración del *switch* 1 provoca el movimiento de la Cys 35 lejos del fosfato α , disminuyendo así la afinidad por GDP, lo cual es crucial para la actividad de intercambio.

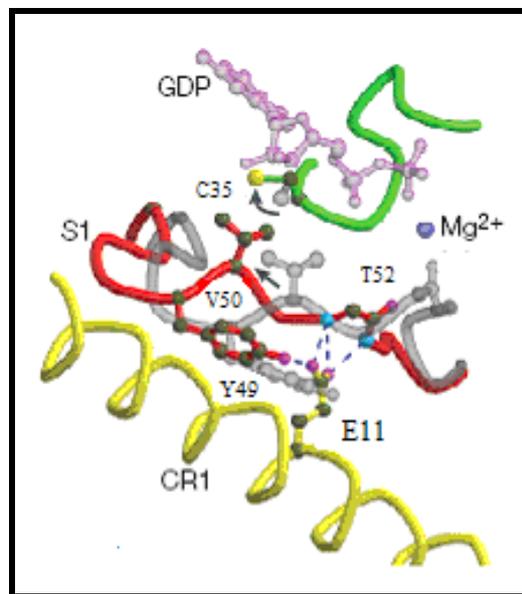


Fig. 13. Cambio conformacional del *switch* 1(en rojo).

Por su lado la región CR3 presenta una gran cantidad de residuos importantes en la interfaz, en particular la Arg 329 de EhGEF1 interacciona fuertemente con la Ala 76 del *switch 2* reposicionándola cerca del Mg^{2+} esto impide que el ion interactúe de manera adecuada con el nucleótido. En conjunto. Los cambios estructurales traen consigo que el GDP se desestabilice y sea expulsado del complejo para entonces dejar el complejo binario EhRho1-EhGEF1 (Fig. 12).

4.6.2. Modelo EhRho1-EhGEF1

El complejo EhRho1-EhGEF1 se presenta cuando en el complejo ternario EhRho1-GDP-EhGEF1 el GDP es expulsado. En comparación con el complejo ternario, EhRho1-EhGEF1 es más estable. Los aminoácidos que aparecen con un $\Delta\Delta G$ mayor a 1 kcal/mol al ser mutados por alaninas, es decir que contribuyen a la estabilización del complejo, son primordialmente los mismos que en el complejo EhRho1-GDP-EhGEF1, aunque para residuos que contribuyen de forma importante a la estabilización de la interfaz se presenta un ligero aumento en sus valores de $\Delta\Delta G$ (Tabla 6), además hay dos residuos que aparecen con valores $\Delta\Delta G$ mayores a 1 kcal/mol, que no aparecían anteriormente. Estos son la Tyr 279 (1.47 kcal/mol) y la Lys 287 (1.00 kcal/mol). Los resultados descritos para estos residuos permiten sugerir que son los responsables de que el complejo binario sea más estable respecto a los complejos binarios, en concordancia con lo que se indica en la literatura [7].

4.6.3. Modelo EhRho1-GTP-EhGEF1

La formación del complejo EhRho1-GTP-EhGEF1 es la etapa final en el proceso de activación propuesto para EhRho1 (Fig. 12). En este complejo la GEF es disociada, la EhRho1 queda en su forma activa cargada de GTP y puede interactuar entonces con sus moléculas efectoras.

Los resultados del barrido virtual de alaninas muestran una disminución en el $\Delta\Delta G$ de residuos que se sabe contribuyen de manera importante a la interfaz, entre ellos se encuentran Glu 190, Arg 329 y Asn 366. En la Tabla 6 se muestran las energías de contacto de estos residuos comparadas en los tres complejos modelados. Las cantidades mostradas son el promedio de los valores de $\Delta\Delta G$ calculadas para los diez diferentes conformeros intermedios previamente seleccionados y minimizados. Entre paréntesis se presentan las desviaciones estándar correspondientes. Se puede observar que en el tercer complejo se presentan los valores más bajos de $\Delta\Delta G$ para los residuos descritos, lo cual sugiere que el complejo EhRho1-GTP-EhGEF1 es menos estable que EhRho1-GDP-EhGEF1. Dichos valores son concordantes con la idea de que este último complejo ternario es inestable y de él se disocia GEF [7].

Complejos	Glu 190 (kcal/mol)	Arg 329 (kcal/mol)	Asn 366 (kcal/mol)
EhRho1-GDP-EhGEF1	2.7 (± 0.70)	4.0 (± 0.17)	1.4 (± 0.11)
EhRho1-EhGEF1	3.1 (± 0.50)	3.5 (± 0.90)	1.6 (± 0.18)
EhRho1-GTP-EhGEF1	2.1 (± 0.30)	3.1 (± 0.80)	1.3 (± 0.15)

Tabla 6. Valores de $\Delta\Delta G$ de algunos residuos importantes en la interfaz para los complejos de EhRho1.

Como ya se había mencionado, si se suman los valores de $\Delta\Delta G$ de cada uno de los residuos de la superficie interfacial se puede estimar el ΔG de disociación o bien, cambiando el signo, el ΔG de unión; al menos la contribución de las cadenas laterales al proceso. En la Tabla 7 se muestran los valores del ΔG de disociación para cada uno de los complejos que se forman en el proceso de activación de EhRho1. Los valores también sugieren que el complejo más estable es el binario y que el más inestable es el unido a GTP lo que se traduce en la disociación de la EhGEF1. Nuevamente nuestros resultados están en concordancia con la información presente en la literatura, que marca que los complejos ternarios son inestables y que el binario es el más estable [7]. Probablemente por esta razón sólo se han logrado cristalizar complejos binarios.

Complejos	ΔG de disociación (kcal/mol)
EhRho1-GDP-EhGEF1	57.8 (± 7.58)
EhRho1-EhGEF1	74.1 (± 7.56)
EhRho1-GTP-EhGEF1	32.2 (± 3.95)

Tabla 7. ΔG de disociación de los complejos de EhRho1.

En el análisis estructural realizado al modelo, encontramos al residuo Cys 35 en una posición cercana al fosfato α (aprox. a 1.69 Å), lo cual era de esperarse debido a que este residuo es importante en la estabilización del nucleótido. Los átomos de oxígeno del fosfato γ del GTP causan una gran interacción con del *switch 2* atrayéndolo hacia él debido a que hay una mayor cercanía respecto al complejo con GDP. Esto trae consigo que algunos contactos importantes como el que había entre la Ala 76 de EhRho1 y la Arg 329 de CR3 de EhGEF1 se pierdan y así la GEF se disocie del complejo.

4.7. Análisis de los modelos del proceso de activación de EhRacG por EhGEF1

De manera adicional a los tres complejos de GEF1 con Rho1, para los complejos formados entre las proteínas GEF1 y RacG, los aminoácidos de menor calidad estructural estuvieron lejos de las regiones interfaciales. Esto nos indica que los análisis energéticos realizados en la interfaz son confiables.

4.7.1. Modelo EhRacG-GDP-EhGEF1

En EhRacG se encuentran los residuos conservados entre las GTPasas Rho: Cys 18, Phe 28 y Thr 35, quienes en la literatura presentan una interacción importante con el fosfato α del GDP y el cofactor Mg^{2+} , lo que nos sugiere que el GDP se mantiene unida a la EhRacG por interacciones similares a las que se presentan en esta familia de proteínas. Adicionalmente se realizaron medidas de las distancias entre cada uno de los tres residuos y los átomos del

GDP y Mg^{2+} (2.4 Å, 2.42 Å, 3.9 Å, respectivamente), permitiéndonos corroborar lo antes mencionado.

El barrido computacional de alaninas nos permitió identificar aminoácidos conservados que típicamente tienen una contribución importante a la superficie interfacial de complejos Rho-GEF, estos son: Glu 190, Arg 329 y Asn 377 con valores de $\Delta\Delta G$ de 2.15, 2.93 y 1.37 kcal/mol respectivamente. Al igual que en EhRho1 se encontró la Tyr 183 ($\Delta\Delta G=1.4$ kcal/mol) que no se encuentra conservada entre especies lo cual nos indica que podría ser importante para la estabilización de los complejos en amiba.

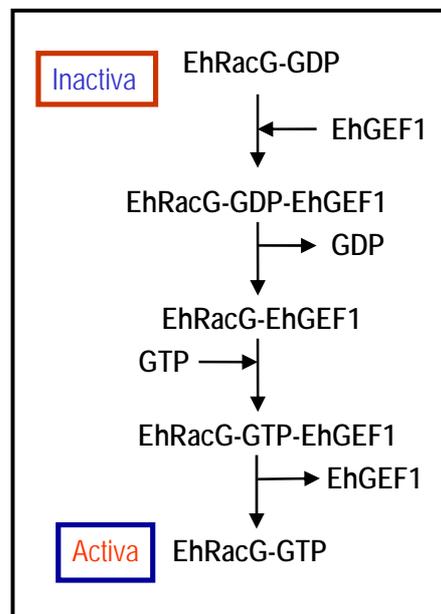


Fig. 14. Etapas del mecanismo propuesto para el proceso de activación de EhRacG por EhGEF1.

El complejo EhRacG-GDP-EhGEF1 es el primero que se forma en el mecanismo propuesto durante el proceso de activación de la EhRacG (Fig. 14). De igual modo una vez que se presenta un estímulo en la membrana celular de la amiba para llevar a cabo un proceso celular en el que interviene la reorganización del citoesqueleto de actina, la EhRacG

cargada de GDP es reconocida por EhGEF1 con el propósito de activarla, reemplazando el GDP por el GTP y que la señal que inició en la membrana celular pueda viajar río abajo.

De igual manera debido a la alta conservación estructural de estas proteínas suponemos que el mecanismo de expulsión de GDP es similar al descrito en la literatura [3, 11]. Un residuo en la CR1, el Glu 190 interacciona con los grupos amino de la Thr 35 y la Val 36 de la región *switch* 1 de la EhRacG y el grupo hidroxilo de la Tyr 32 formando dos puentes de hidrógeno, ocasionando un cambio conformacional de este giro. Esto hace que Phe 28 y Cys 18 se alejen del GDP, disminuyendo notablemente la estabilidad del nucleótido, sobre todo porque los cambios estructurales del *switch* 1 provocan que la Ile 33 induzca un cambio conformacional de la Cys 18 alejándola del fosfato α , lo cual es crucial para la actividad de intercambio, además Thr 35 que se encuentra interaccionando con el Mg^{2+} se aleja de él (Fig. 15).

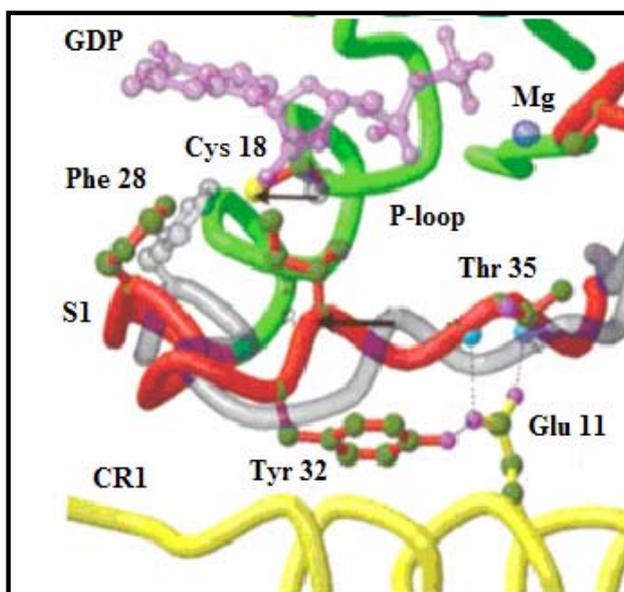


Fig. 15. Mecanismo de expulsión del GDP del complejo EhRacG-GDP-EhGEF1. En rojo se muestra el *switch* 1 (S1) de EhRacG1 y en amarillo la región conservada 1 (CR1) de EhGEF1 [11].

Por su parte en la región CR3 existen varios residuos importantes en la estabilización de la interfaz, en particular la Arg 329 de EhGEF1 interacciona fuertemente con la Ala 59 de

EhRacG causando que ésta se acerque e interactúe con el Mg^{2+} impidiéndole un adecuado contacto con el GDP, todo esto trae consigo que el nucleótido se desestabilice y sea expulsado del complejo para entonces dejar el complejo binario EhRacG-EhGEF1.

4.7.2. Modelo EhRacG-EhGEF1

De acuerdo al mecanismo planteado una vez que el GDP es expulsado del complejo ternario EhRacG-GDP-EhGEF1 se tiene el complejo binario EhRacG-EhGEF1 (Fig. 14). En comparación con el complejo ternario, EhRacG-EhGEF1 es más estable, esto lo podemos observar en la Tabla 8 en la que se presentan los valores de ΔG de disociación para los tres complejos que se forman durante del proceso de activación siendo el binario el que tiene el valor de ΔG de disociación más grande.

Los aminoácidos que aparecen con un $\Delta\Delta G$ mayor a 1 kcal/mol al ser mutados por alaninas, es decir que contribuyen significativamente a la estabilización del complejo, son primordialmente los mismos que en el complejo EhRho1-GDP-EhGEF1, aunque para algunos de los aminoácidos más importantes que estabilizan al complejo (Glu 190, Arg 329, Asn 366 y Gln 325, aminoácidos de la EhGEF1) los valores de $\Delta\Delta G$ son los mayores (Tabla 9). Lo anterior nos confirma que este complejo es más estable que los complejos unidos a GDP y GTP. Las cantidades mostradas en la Tabla 9 son el promedio de los valores de $\Delta\Delta G$ calculados para los diferentes modelos intermedios previamente seleccionados y minimizados, entre paréntesis se presenta la desviación estándar.

Complejos	ΔG de disociación (kcal/mol)
EhRacG-GDP-EhGEF1	50.6 (± 6.01)
EhRacG-EhGEF1	56.6 (± 6.50)
EhRacG-GTP-EhGEF1	50.3 (± 4.32)

Tabla 8. Valores de ΔG de disociación para los complejos del proceso de activación de EhRacG.

Complejos	Glu 190 (kcal/mol)	Arg 329 (kcal/mol)	Asn 366 (kcal/mol)
-----------	-----------------------	-----------------------	-----------------------

EhRacG-GDP-EhGEF1	2.1 (± 0.10)	3.0 (± 0.30)	1.6 (± 0.20)
EhRacG-EhGEF1	2.6 (± 0.15)	4.3 (± 0.20)	1.9 (± 0.40)
EhRacG-GTP-EhGEF1	2.0 (± 0.20)	1.7 (± 0.10)	1.4 (± 0.20)

Tabla 9. Valores de $\Delta\Delta G$ de algunos aminoácidos importantes en la superficie interfacial.

4.7.3. Modelo EhRacG-GTP-EhGEF1

La formación del complejo EhRacG-GTP-EhGEF1 se presenta en la etapa final del proceso de activación de EhRacG (Fig. 14). En este complejo la GEF es disociada, la EhRacG queda en su forma activa cargada de GTP y puede interactuar con sus moléculas efectoras. En la Tabla 8 se puede observar cómo el ΔG de disociación para este complejo es el más bajo, lo que nos indica que es el más inestable de los tres como se espera de acuerdo al mecanismo descrito en la figura 14. Los resultados del barrido virtual de alaninas muestran una disminución en el $\Delta\Delta G$ de residuos que contribuyen de manera importante al reconocimiento, entre ellos se encuentran Glu 190, Arg 329 y Gln 325. En la Tabla 9 se muestran los valores del $\Delta\Delta G$ de estos residuos comparados en los tres complejos de EhRacG modelados. Se puede observar que en el tercer complejo se presentan los valores más bajos de $\Delta\Delta G$ para los residuos, lo cual indica que el complejo EhRacG-GTP-EhGEF1 es menos estable que EhRacG-GDP-EhGEF1, lo que se traduce finalmente en la disociación de la EhGEF1.

En el análisis estructural el residuo Cys 18, que corresponde a la Cys 35 en EhRho1, aparece en una posición más cercana al fosfato α del GTP, lo cual era de esperarse debido a que este residuo es importante en la estabilización del nucleótido. Los átomos de oxígeno del fosfato γ del GTP causan una gran interacción con el *switch 2* atrayéndolo hacia él, debido a que hay una gran cercanía (aprox. 2.0 Å). Esto trae consigo que algunos contactos importantes como el que había entre la Ala 59 y la Arg 329 de CR3 de la EhGEF1 se pierdan y de esta forma la GEF sea disociada del complejo [11].

4.8. Análisis y discusión del cálculo de las energías de unión estimadas con el campo de fuerza CHARMM 27

El cálculo de la energía de interacción entre GEF y las GTPasas se realizó de manera independiente sobre cada uno de los diez conformeros minimizados para cada complejo y el valor que se reporta es el promedio. Lo que se observa nuevamente es que los complejos binarios para ambas proteínas (EhRho1-EhGEF1 y EhRacG-EhGEF1) tienen los valores energéticos más negativos lo que nos confirma que estos complejos son los más estables en ambos sistemas, mientras que las energías de unión de los complejos unidos a GTP son las más grandes, siendo éstos los más inestables. Esto concuerda con los resultados obtenidos del barrido computacional de alaninas. Los resultados de las energías de unión para cada uno de los complejos de ambas proteínas se pueden visualizar en la Tabla 10, con sus desviaciones estándar entre paréntesis.

En los modelos generados, EhRho1 siempre forma complejos más estables con EhGEF1 que los formados por EhRacG, lo que se puede deducir al comparar las energías unión para cada uno de los complejos, siendo las de EhRho1 las más pequeñas (Tabla 10). El hecho de que la proteína EhRho1 forme complejos más estables nos hace suponer que quizás en la célula haya una mayor concentración de la proteína EhRacG. Hasta donde conocemos, no existen medidas experimentales de la afinidad de RacG y Rho1 por GEF, por lo que nuestra propuesta de que la afinidad de Rho1 por GEF es mayor que la de RacG es una contribución predictiva al conocimiento de estos sistemas. En la Tabla 10 se han desglosado las contribuciones a la energía de unión, que son la electrostática y la van der Waals. Puede apreciarse que la energía electrostática tiene la mayor contribución a la energía total.

Energía de unión (kcal/mol)	Rho1-GEF	Rho1-GDP-GEF	Rho1-GTP-GEF
van der Waals	-54.42 (±7.8)	-61.40 (±10.1)	-64.20 (±6.9)
Electrostática	-487.79(±58.0)	-472.91 (±41.0)	-455.39 (±30.8)
Total	-542.22 (±57.0)	-534.31 (±39.0)	-519.59 (±30.2)
	RacG-GEF	RacG-GDP-GEF	RacG-GTP-GEF
van der Waals	-55.78 (±7.5)	-58.40 (±9.8)	-57.58 (±10.4)
Electrostática	-463.83 (±26.9)	-430.89 (±33.12)	-424.04 (±39.0)
Total	-519.61 (±27.0)	-489.29 (±37.38)	-481.62 (±39.5)

Tabla 10. Energías de unión calculadas con el campo de fuerza. En la parte superior se presentan las energías de unión para los complejos formados entre las proteínas EhGEF1 y EhRho1; en la parte inferior las energías de unión para los formados entre EhGEF1 y EhRacG.

4.9. Contribución del nucleótido y el Mg²⁺ a la interacción en los complejos ternarios.

Como ya se describió en la sección 2.5 del capítulo II, con el fin de determinar el efecto del nucleótido sobre la región interfacial de los complejos, éste fue eliminado de las estructuras tridimensionales sin permitir un reajuste molecular y se reevaluó la energía de unión entre EhGEF1 y las GTPasas. Los resultados se presentan en la Tabla 11 y de igual manera se han desglosado las contribuciones a la energía total. De manera similar a los resultados discutidos previamente, esta reevaluación se realizó de forma independiente sobre los diez conformeros que se tienen y los valores que se muestran en la Tabla 11 son el promedio de esos resultados, con la desviación estándar entre paréntesis.

Realizando una comparación de los valores de energía de unión para los complejos con nucleótido que se presentan en la Tabla 10 y los complejos a los que se eliminó el nucleótido, presentados en la Tabla 11, se puede observar que al eliminar el GDP en los complejos tanto de EhRho1 como de EhRacG, el cambio en la energía de unión es de aproximadamente -8 kcal/mol, mientras que al eliminar el GTP el cambio es de aproximadamente 15 kcal/mol. Lo anterior sugiere que el nucleótido GTP contribuye a la

estabilización GTPasa-GEF, mientras que el que el nucleótido GDP interviene desfavorablemente al reconocimiento. La diferencia en comportamiento se debe en gran medida a que el GTP tiene un grupo fosfato adicional al GDP, lo que se traduce en más cargas que pueden establecer interacciones electrostáticas con los aminoácidos que lo rodean además de que se tiene una carga negativa extra. Estas interacciones adicionales son las que provocan cambios estructurales tales que la GEF1 es disociada en los complejos ternarios EhRacG-GTP-EhGEF1 y EhRho1-GTP-EhGEF1.

Energía de unión (kcal/mol)	Rho1-GDP-GEF sin GDP	Rho1-GTP-GEF sin GTP
van der Waals	-61.15 (±10.0)	-62.35 (±6.9)
Electrostática	-479.35 (±41.6)	-440.78 (±30.0)
Total	-540.50 (±39.6)	-503.13 (±29.2)
	RacG-GDP-GEF sin GDP	RacG-GTP-GEF sin GTP
van der Waals	-51.40 (±13.0)	-56.03 (±10.3)
Electrostática	-446.19 (±46.7)	-411.75 (±37.4)
Total	-497.59 (±38.2)	-467.78 (±36.4)

Tabla 11. Energías de unión sin nucleótido.

En la interfaz de los complejos ternarios hay dos contribuciones que son la interacción GTPasa-GEF y la interacción nucleótido (Mg^{2+})-GEF. La eliminación del nucleótido y el ion magnesio permitió desglosar la energía de unión en ambas contribuciones (Tabla 12).

Rho1-GDP-GEF1	Rho1-GDP-GEF1 sin GDP	Contribución GDP(Mg)-GEF1	Rho1-GTP-GEF1	Rho1-GTP-GEF1 Sin GTP	Contribución GTP(Mg)-GEF1
-534 kcal/mol =	-541 kcal/mol	+7 kcal/mol	-519 kcal/mol =	-503 kcal/mol	-16 kcal/mol
RacG-GDP-GEF1	RacG-GDP-GEF1 sin GDP	Contribución GDP(Mg)-GEF1	RacG-GTP-GEF1	RacG-GTP-GEF1 Sin GTP	Contribución GTP(Mg)-GEF1
-489 kcal/mol =	-497 kcal/mol	+8 kcal/mol	-482 kcal/mol =	-467 kcal/mol	-15 kcal/mol

Tabla 12. Desglose de las energías de unión en las interacciones GTPasa-GEF y nucleótido(Mg^{2+})-GEF en los complejos ternarios.

Haciendo una comparación entre las energías de unión de los complejos binarios que son los más estables con las contribuciones GTPasa-GEF a la energías de unión en los complejos binarios, podemos observar que en el caso de las contribuciones GTPasa-GEF para los complejos unidos a GDP sus energías de unión son muy cercanas a las energía de unión de los complejos binarios, lo que en cierta forma concuerda con el mecanismo de activación ya que lo que sucede en esta etapa es la expulsión del GDP y no una disociación de las proteínas. Para el caso de las contribuciones GTPasa-GEF de los complejos unidos GTP estos presentan las energías de unión más pequeñas presentando la mayor diferencia con las energías de unión de los complejos binarios, lo que también concuerda con el mecanismo de activación ya que lo que sucede en esta etapa es una disociación entre las proteínas.

4.10. Discusión del mecanismo de reconocimiento y selectividad

Se analizaron los residuos de la superficie interfacial que contribuyen a la estabilización de los complejos de ambas proteínas. En la Tabla 13 se presentan los valores $\Delta\Delta G$ de tres residuos se sabe contribuyen de forma importante a la formación y estabilización de complejos Rho-GEF. Para los complejos de ambas proteínas estos aminoácidos tienen valores de $\Delta\Delta G$ mayores a 1 kcal/mol lo que indica según los autores del programa que tienen una contribución significativa en la interacción con EhGEF1 y esto podría permitir el reconocimiento de ambas proteínas y su posterior activación.

Complejos	Glu 190 (kcal/mol)	Arg 329 (kcal/mol)	Asn 366 (kcal/mol)
EhRho1-GDP-EhGEF1	2.7 (± 0.70)	4.0 (± 0.17)	1.4 (± 0.11)
EhRacG-GDP-EhGEF1	2.1 (± 0.10)	3.0 (± 0.30)	1.6 (± 0.20)
EhRho1-EhGEF1	3.1 (± 0.50)	3.5 (± 0.90)	1.6 (± 0.18)
EhRacG-EhGEF1	2.6 (± 0.15)	4.3 (± 0.20)	1.9 (± 0.40)
EhRho1-GTP-EhGEF1	2.1 (± 0.30)	3.1 (± 0.80)	1.3 (± 0.15)
EhRacG-GTP-EhGEF1	2.0 (± 0.20)	1.7 (± 0.10)	1.4 (± 0.20)

Tabla 13. Valores de $\Delta\Delta G$ de algunos residuos de EhGEF1 que contribuyen a la formación y estabilización de los complejos.

Se sabe que las regiones β_2 y β_3 de las GTPasas son las primeras en ser reconocidas por las GEFs [11], lo que sugiere que los contactos en estas regiones dictan en gran medida la selectividad de determinada GEF por ciertas GTPasas. Haciendo uso de los resultados del barrido computacional de alaninas se buscaron residuos en las regiones β_2 y β_3 de EhRacG y EhRho1 que contribuyeran de forma importante a la estabilización de la superficie interfacial, el procedimiento fue descrito más detalladamente en la sección 2.6.

Los resultados de la búsqueda mostraron que para ambas proteínas las hebras β_2 y β_3 tienen como importantes para la estabilización del contacto interfacial los mismos dos residuos que son para EhRacG: Asn 39 y Trp 56 y los correspondientes para EhRho1 Asn 56 y Trp 73. Estos residuos son semiconservados y en cierto grado podrían ser responsables de la especificidad pero existen GTPasas con los mismos residuos que no son reconocidas por la EhGEF1, como por ejemplo EhRacA. La explicación para tal hecho es que posibles mutaciones cerca de las regiones β_2 y β_3 pueden causar que las cadenas laterales adquieran una diferente conformación que impida que sea reconocida por la GEF aunque en otra posición presente la misma secuencia de aminoácidos. Esto nos indica que la base del reconocimiento de la EhGEF1 sobre las GTPasas es la suma de varias contribuciones y efectos incluyendo desde reemplazo de aminoácidos en la secuencia hasta la presencia o ausencia de ciertos residuos cercanos a la región interfacial que provocan cambios estructurales en las cadenas laterales alterando o impidiendo el reconocimiento molecular.

El barrido computacional de alaninas también nos permitió identificar residuos de GEF1 que contribuyen de forma importante a la formación en los complejos de EhRacG pero no así de los de EhRho1 o viceversa. Además, algunos aminoácidos contribuyen de forma importante en ambas proteínas pero presentan una diferencia significativa en sus valores de $\Delta\Delta G$. Algunos residuos se presentan en la Tabla 14 (de igual manera los valores de $\Delta\Delta G$ son el promedio de los diez conformeros elegidos y minimizados, en paréntesis se presenta la desviación estándar del conjunto de valores promediados). Estos resultados son importantes debido a que uno de los objetivos del proyecto es proponer aminoácidos esenciales en el reconocimiento de ambas proteínas por EhGEF1 y proponer mutantes que

funcionen como dominantes negativas. Estas últimas tienen el propósito de poder conocer y distinguir las funciones celulares en las que se encuentran involucradas dichas GTPasas de la amiba. Algunos aminoácidos de EhGEF1 interaccionan fuertemente con ambas GTPasas, pero los aminoácidos que sólo interaccionan de forma importante con alguna de las dos proteínas podrían ser mutados para generar dominantes negativas de EhGEF1. Si estas mutantes se transfectan a amibas, su presencia competiría con la enzima endógena por la formación del complejo con esa proteína pero sin activar la GTPasa, esto apagaría parcialmente la vía de señalización y se podrían observar efectos en el fenotipo debido al proceso celular que se afectó.

Complejos	Thr 321 (kcal/mol)	Lys 331 (kcal/mol)	Leu 283 (kcal/mol)	Leu 332 (kcal/mol)	Gln 325 (kcal/mol)
EhRho1-GDP-EhGEF1	0.13 (\pm 0.60)	1.0 (\pm 0.30)			5.7 (\pm 2.2)
EhRacG-GDP-EhGEF1	2.0 (\pm 0.50)	0.62 (\pm 0.40)			4.0 (\pm 0.71)
EhRho1-EhGEF1	0.0 (\pm 0.20)		0.69 (\pm 0.18)		4.3 (\pm 0.80)
EhRacG-EhGEF1	1.7 (\pm 0.58)		1.1 (\pm 0.20)		6.2 (\pm 1.1)
EhRho1-GTP-EhGEF1	0.25 (\pm 0.20)		0.77 (\pm 0.32)	1.0 (\pm 0.20)	
EhRacG-GTP-EhGEF1	1.2 (\pm 0.32)		1.0 (\pm 0.20)	0.78 (\pm 0.10)	

Tabla 14. Valores de $\Delta\Delta G$ de algunos aminoácidos de EhGEF1 importantes para la especificidad.

4.11. Resultados de la búsqueda de moléculas líderes para la inhibición de la proteína EhGEF1

Una molécula que interactúe fuertemente con la zona de EhGEF1 que reconoce a las GTPasas, puede impedir la formación del complejo y entonces actuar como un inhibidor del proceso de activación. Esta molécula puede utilizarse como sonda molecular que permita inferir la vía metabólica en la que participa EhGEF1, si resulta que EhGEF1 es fundamental para la invasión, desarrollo y supervivencia del parásito, éstas moléculas podrían ser usadas adicionalmente como guía para el diseño de un fármaco antiamibiano. Para identificar moléculas con capacidad inhibitoria se utilizó la técnica computacional llamada *docking*. A continuación se presentan los resultados del procedimiento de anclaje molecular sobre la proteína EhGEF1.

4.11.1. Elección del sitio de anclaje en EhGEF1

Con la herramienta “Site Finder” disponible en MOE se buscaron las posibles cavidades que pudieran servir como sitio de docking. Se encontraron dos cavidades en la superficie interfacial de EhGEF1 que interactúa con las GTPasas, algunos criterios nos permitieron decidir cual de ellos podría ser el mejor sitio de unión (cuál de las dos cavidades era más grande, cuál tenía el mayor número de aminoácidos no conservados respecto a otras GEF en general con el propósito de lograr especificidad hacia GEF1 y cuál cavidad tenía aminoácidos que contribuyeran de forma importante a la formación de los complejos). Con base en los criterios el sitio que se eligió es el formado por los siguientes residuos: Val 187, Glu 190, Leu 191, Thr 194, Ser 197, Tyr 198, Gln 201, Phe 314, Ser 318, Phe 319, Ile 321, Thr 322, Gln 325, Arg 326, Arg 329, Leu 332 y Asp 336. Esta cavidad está formada de 18 residuos de los cuales 8 son residuos no conservados en general y contiene tres residuos esenciales para la formación de los complejos (residuos 190, 326 y 329) (Fig. 16).

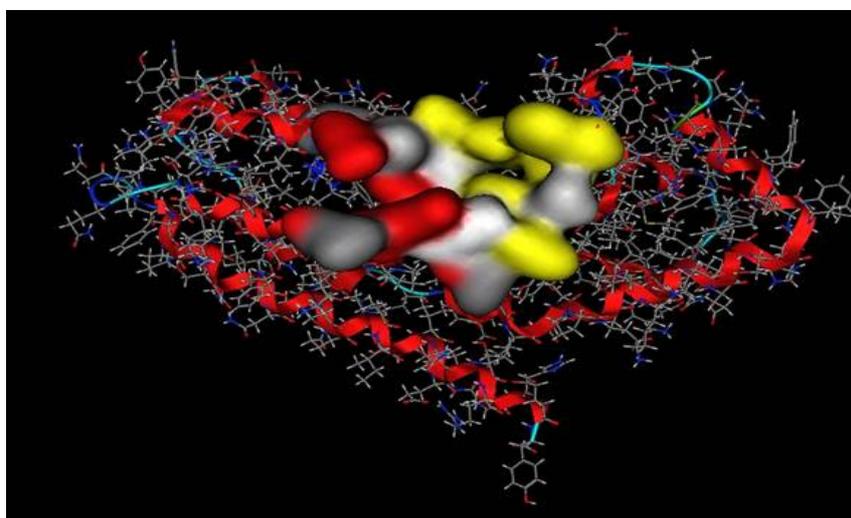


Fig. 16. Cavidad elegida para el docking en la proteína EhGEF1. Los residuos mostrados en amarillo son los conservados, los blancos no conservados y los rojos son residuos que experimentalmente han mostrado ser importantes para la formación y estabilización de los complejos con las GTPasas.

4.11.2. Resultado del *docking* de compuestos sobre la cavidad elegida

Una vez que se realizó el barrido virtual rígido de la base de más de 50,000 compuestos, se prosiguió a ordenar el archivo de salida de acuerdo al puntaje calculado para cada uno. Éste contiene cada una de las moléculas de la base sobre el sitio de anclaje en su mejor posición y orientación así como el valor energético asociado a ese complejo. Esto permitió identificar aquellos que tienen los puntajes más negativos, lo que nos indica que son los de mayor afinidad potencial por el sitio de anclaje.

Para los compuestos con las energías de unión más bajas se realizó un análisis estructural con el fin de poder determinar que aminoácidos de la EhGEF1 son los más afectados por la posición del ligando y qué tanto se va impedir la formación del complejo de tal manera que la molécula pueda servir como un ligando eficaz para la inhibición de EhGEF1 y que tenga las cualidades que se buscan en cualquier inhibidor: afinidad y selectividad. El análisis se realizó para los diez compuestos con energías de unión más bajas. Éstos fueron etiquetados con los números del uno al diez y se presentan en la Tabla 15, con su valor de energía asociado a la unión con la proteína EhGEF1 en orden descendente.

Compuesto	Puntaje	Compuesto	Puntaje
1	-5.8401	6	-5.6596
2	-5.8002	7	-5.6336
3	-5.7785	8	-5.6195
4	-5.7437	9	-5.6172
5	-5.7242	10	-5.6123

Tabla 15. Compuestos con las mejores energías de unión sobre la proteínas EhGEF1. Código de colores: amarillo; átomos de azufre, verde; flúor, gris fuerte; carbono, gris; hidrógeno, azul; nitrógeno, rojo; oxígeno.

El análisis de la orientación de los diez compuestos nos permitió identificar qué residuos de EhGEF1 son los que participan en la interacción con los compuestos anclados. En la Fig. 17 se muestra la orientación y posición de menor energía del compuesto 1 en el sitio de anclaje de la proteína EhGEF1, un análisis estructural similar se realizó para los compuestos restantes. Los residuos de EhGEF1 que interaccionan fuertemente con la mayoría de los diez ligandos son la Arg 326 y la Arg 329. En general las interacciones electrostáticas son las que mantienen unidos los compuestos a EhGEF1 y para algunos compuestos se presentan puentes de hidrógeno principalmente con la Arg 329, lo cual es importante debido a que los puentes de hidrógeno no sólo proveen energía de anclaje a los ligandos, sino que se sabe que son causantes de especificidad por la dependencia geométrica de este tipo de interacción. El análisis de los residuos también nos permitió identificar cuales aminoácidos de las GTPasas perderían interacciones importantes con EhGEF1 debido a la presencia de las moléculas orgánicas. Estos aminoácidos son principalmente para EhRacG: Asn 39, Val 36, Glu38, y Trp 56, y los correspondientes a la EhRho1. En particular los residuos 39 y 56 de las GTPasas son importantes para la formación de los complejos debido a que en los resultados del barrido de computacional de alaninas presentan valores de $\Delta\Delta G$ de aprox. 5.8 kcal/mol y 2.2 kcal/mol respectivamente y además se encuentran en las hebras β_2 y β_3 , las cuales se sabe que son las primeras regiones en ser reconocidas por GEF [3]. Lo anterior permite suponer que la eliminación de estos contactos podría impedir la formación de los complejos.

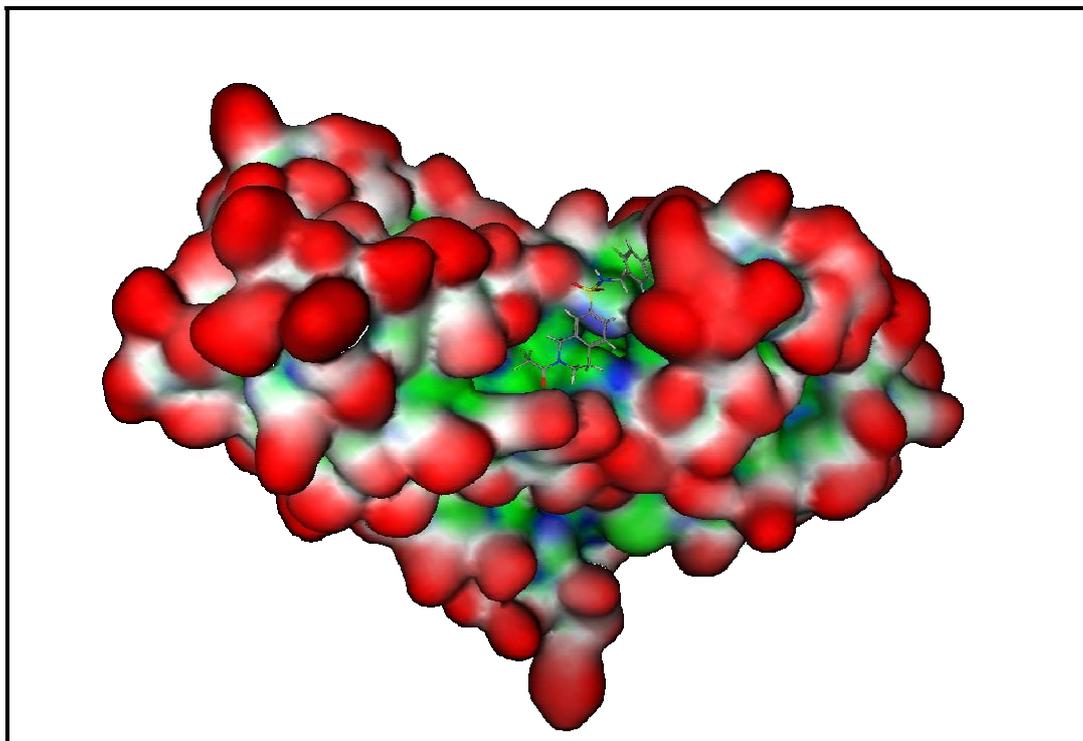


Fig. 17. Anclaje molecular. En esta imagen podemos apreciar la mejor orientación del compuesto 1, esto es, el caso de mayor afinidad por EhGEF1.

Es importante retomar la idea de que los compuestos encontrados son potenciales inhibidores de la proteína EhGEF1, faltaría determinar si experimentalmente esto sucede. Si es así, estas moléculas de EhGEF1 podrían ser usadas como sondas moleculares que pudieran bloquear *in vivo* la función de esta GEF y permitir la determinación del papel metabólico de la proteína en el parásito. Más aún, si resulta que esta proteína se encuentra involucrada en procesos fundamentales para el parásito, estos compuestos podrían ser la base para la búsqueda de moléculas líder para el diseño de un nuevo fármaco antiamebiano. Para tal motivo se tendrían que realizar estudios adicionales en donde se demuestre que al inhibir la proteína EhGEF1 el parásito no tiene una vía alterna o un proceso homeostático que contrarreste el efecto de la inhibición de EhGEF1, además de realizar estudios *in vivo* con *E. histolytica* para determinar que realmente las moléculas tienen las características fisicoquímicas para atravesar la membrana celular de la ameba y llegar a la proteína blanco.

Una vez realizados estos estudios, entonces se podrían comenzar con experimentos en donde ahora se evalúen las propiedades farmacológicas de los compuestos para poder ser usados en humanos, como son la selectividad, afinidad, concentración local, absorción, toxicidad, etc. y por último las pruebas preclínicas y clínicas.

En este trabajo la principal intención es proponer un protocolo para identificar moléculas con potencial para inhibir alguna de las GEF. La idea del fármaco sólo se plantea como una posibilidad a mediano o largo plazo debido a que la inhibición de una proteína diana es la base actual del “diseño de fármacos en base a la estructura”. Pero se debe tener en mente que entre el hecho de encontrar ligandos con actividad inhibitoria potencial y obtener un nuevo fármaco existe un largo y muy complejo camino.

Conclusiones generales y perspectivas

Conclusiones

- ❖ 1. En los análisis energéticos, se obtuvo una mayor correlación con datos calculados sobre estructuras cristalográficas al utilizar promedios extraídos de varios conformeros (diez en nuestro caso) para cada modelo generado.
- ❖ 2. Los valores de las energías de unión para los complejos generados es coincidente con la información del mecanismo propuesto para el proceso de activación de las GTPasas descrito en la literatura. Esto nos da confianza de que es correcto el análisis molecular realizado a los complejos obtenidos.
- ❖ 3. Los tres residuos conservados Glu 190, Arg 329 y Asn 366 tienen una contribución importante en la interacción de la EhGEF1 con EhRacG y EhRho1. Según la literatura, estos residuos son esenciales para la formación de complejos Rho-GEF en general.
- ❖ 4. Se encontró un grupo de residuos no conservado de EhGEF1 importantes para el reconocimiento debido a que no se encuentran conservados entre especies, estos podrían ser importantes para la formación de complejos GEF-GTPasa específicamente en amiba.
- ❖ 5. Se encontraron residuos que contribuyen a la estabilización y formación de los complejos de EhRho1 y de igual manera para EhRacG.
- ❖ 6. Los complejos formados por la proteína EhRho1 son más estables que los formados por EhRacG. Esto predice que la constante de unión será mayor para el complejo que incluye a la primera de las GTPasas.

- ❖ 7. EhGEF1 reconoce a sus GTPasas blanco, EhRacG y EhRho1, no por la participación de un único residuo, sino como resultado de la suma de pequeños efectos: la alta identidad en secuencia, la presencia de aminoácidos esenciales para el reconocimiento y la presencia de mutaciones específicas cerca de la superficie interfacial.
- ❖ 8. La energía electrostática es la que presenta la mayor contribución al contacto interproteína en los complejos formados durante la activación de EhRacG y EhRho1 por EhGEF1. Lo anterior sugiere que un cambio en el ambiente electrostático afectará directamente la capacidad de formación de los complejos.
- ❖ 9. Se encontraron compuestos orgánicos que pueden servir como posibles inhibidores de la actividad de la proteína EhGEF1.

En resumen las contribuciones de este trabajo son: haber realizado el primer análisis estructural mediante técnicas computacionales para el mecanismo de activación propuesto para las proteínas EhRacG y EhRho1; así como la generación de modelos de especies químicas no accesibles experimentalmente, como es el caso de los complejos ternarios. Una contribución adicional es el haber desarrollado un planteamiento metodológico para el modelado y análisis de complejos Rho-GEF y en general de complejos proteína-proteína. También se realizaron propuestas de mutaciones que podrían permitir conocer más acerca de la fisiología de la amiba y finalmente la identificación de los primeros inhibidores potenciales de la proteína EhGEF1 de *E. histolytica*.

Perspectivas

- ❖ 1. El conocimiento adquirido sobre los sistemas GTPasa-GEF de este trabajo puede ser usado para estudiar sistemas equivalentes en amiba, e incluso los correspondientes en humano, algunos de las cuales se han propuesto como involucrados en el cáncer.
- ❖ 2. El conocimiento adquirido de las proteínas GEFs y GTPasas puede ser usado para estudiar mecanismos patogénicos de algunos microorganismos donde están involucradas biomoléculas de estas familias de proteínas.
- ❖ 3. Los diez inhibidores seleccionados pueden ser ensayados para determinar experimentalmente su capacidad para evitar la formación de los complejos en amiba. Esto los convertiría en posibles compuestos líder para el diseño de fármacos
- ❖ 4. A algunos de los mejores compuestos inhibidores se les puede realizar modificaciones con el fin de aumentar su afinidad por la proteína EhGEF1.
- ❖ 5. El docking para la base de compuestos se puede realizar de una manera flexible, no sólo para las moléculas orgánicas sino también para la cavidad de la proteína EhGEF1.

Referencias bibliográficas

- [1]. Ravdin J., 1995. Amebiasis. Clin Infect Dis. **20 (6)**, 1453-1466.
- [2]. Etienne-Manneville S., y Hall A., 2002. Rho GTPases in cell biology. Nature **420 (6916)**, 629-635.
- [3]. Rossman K.L., Der C.J. y Sondek J., 2005. GEF means go: turning on Rho GTPases with guanine nucleotide-exchange factors. Nat Rev Mol Cell Biol. **6 (2)**, 167-80.
- [4]. Ihara K., Muraguchi S., Kato M., Shimizu T y Kuroda S., 1998. Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue. J Biol Chem. **273 (16)**, 9656-66.
- [5]. Zheng Y., 2001. Dbl family guanine nucleotide exchange factors. Trends Biochem Sci. **26 (12)**, 724-32.
- [6]. Aguilar-Rojas A., Almaraz-Barrera M. de J., Krzeminski M., Robles-Flores M., Hernandez-Rivas R., Guillen N., Maroun RC. y Vargas M., 2005. *Entamoeba histolytica*: inhibition of cellular functions by overexpression of EhGEF1, a novel Rho/Rac guanine nucleotide exchange factor. Exp Parasitol. **109 (3)**, 150-62.
- [7]. Cherfils J. y Chardin P., 1999. GEFs: structural basis for their activation of small GTP-binding proteins. Trends Biochem Sci. **24 (8)**, 306-11.
- [8]. Kortemme T., Kim D. E., and Baker D., 2004. Computational alanine scanning of protein-protein interfaces. Sci. STKE, **2004 (219)**, 12.

- [9]. Lovell S. C., Davis I. W., Arendall III W. B. y Bakker P. I., 2003. Structure validation by C_{α} geometry: ϕ , ψ and C_{β} deviation. *Proteins* **50** (3), 437-450.
- [10]. Cerione R.A. y Zheng Y., 1996. The Dbl family of oncogenes. *Cell Biol.* **8** (2), 216-22.
- [11]. Worthylake D. K., Rossmann K. L. y Sondek J., 2000. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408** (6813), 682-688.
- [12]. Aghazadeh B., Zhu K., Kubiseski T. J., Liu G. A., Pawson T., Zheng Y. y Rosen M. K., 1998. Structure and mutagenesis of Dbl homology domain. *Nat Struct Biol.* **5** (12), 1098-1107.
- [13]. User Guide of MOE 2005.06, Chemical Computing Group, <http://www.chemcomp.com/>.
- [14]. Rossmann K.L., 2005. Larger than Dbl: new structural insights into RhoA activation. *Trends in Biochemical Sciences* **30** (4),163-165.
- [15]. Kuntz I.D., 1992. Structure-based strategies for drug design and discovery. *Science* **257** (5073), 1078-82.
- [16]. Kuntz I.D., Meng E.C. y Shoichet B.K., 1992. Structure-based molecular design. *Acc. Chem. Res.* **27** (5), 117-123.
- [17]. Worthylake D.K., Rossmann K.L. y Sondek J., 2004. Crystal structure of the DH/PH fragment of Dbs without bound GTPase. *Structure* **12** (6), 1078-1086.

- [18]. Lin R., Cerione R.A. y Manor D., 1999. Specific contributions of the small GTPases Rho, Rac and Cdc 42 to Dbl transformation. *J Biol Chem.* **274 (33)**, 23633-23641.
- [19]. Wennerberg K., Rossam K.L. y Der C.J., 2005. The Ras superfamily at a glance. *J. Cell Sci.* **118 (5)**, 843-846.
- [20] Wennerberg K. y Der C. J., 2004. Rho-family GTPases: it's not only Rac and Rho (and I like it). *J Cell Sci.* **177 (8)**, 1301-1312.
- [21] Bishop A. L. and Hall A., 2000. Rho GTPases and their effector proteins. *Biochem. J.* **348 (2)**, 241-255.
- [22] Vetter I. R. and Wittinghofer A., 2001. The guanine nucleotide-binding switch in three dimensions. *Science* **294 (5545)**, 1299-1304.
- [23] Zheng Y. and Quilliam L.A., 2003. Activacion of Ras superfamily of small GTPases. *EMBO Rep.* **4 (5)**, 463-468.
- [24] Okada M., Houston C.D. y Barbara J. M., 2004. Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica*. *Eukaryot Cell* **4 (4)**, 827-831.
- [25] Gao Y., Dickerson B., Guo F., Zheng J. and Zheng Y., 2004. Rational design and characterization of a Rac GTPases-specific small molecule inhibitor. *PNAS USA* **101 (20)**, 7618-7623.
- [26] Zhang B., Zhang Y., Wang Z., and Zheng Y., 2000. The role of Mg²⁺ cofactor in the guanine nucleotide exchange and GTP hydrolysis reaction of Rho family GTP-binding proteins. *J Biol Chem.* **275 (33)**, 25299-25307.

- [27] Guillén N., P. Boquet and P. Sansonetti. 1998. The small GTP-binding protein RacG regulates uroid formation in the protozoan parasite *Entamoeba histolytica*. *J. Cell Sci.* **111** (12), 1729-1739.
- [28] Stanley S. 2003. Amoebiasis. *Lancet* **361** (4), 1025-1034.
- [29] Sehgal D., Bhattacharya A., Bhattacharya S. 1996. Pathogenesis of infection by *Entamoeba histolytica*. *J. Biosci.* **21** (3), 423-432.
- [30] Mannhold R., Kubinyi H., Timmerman H., *Bioinformatics From Genomes to Drugs*. 2002: Wiley-VCH.
- [31] Voig H., Guillén N. 1999. New insights into the role of the cytoskeleton in phagocytosis of *Entamoeba histolytica*. *Cell Microb.* **1**(3), 195-203.
- [32] Kitchen D. B., Decornez H., Furr J. R., Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev. Drug Discov.* **3** (5), 935-949.