



**UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO**

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS  
QUÍMICAS**

**ESTUDIOS DE MODELADO MOLECULAR DE BENCIMIDAZOLES ACTIVOS CONTRA  
*Entamoeba histolytica* Y DISEÑO DE ANÁLOGOS ANTIAMEBIANOS**

TESIS  
PARA OPTAR POR EL GRADO DE

**DOCTOR EN CIENCIAS**

PRESENTA

FABIAN HARVEY LÓPEZ VALLEJO



TUTOR: Dr. RAFAEL CASTILLO BOCANEGRA  
AÑO: 2007



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

A mis padres, Mario y Deyanira  
por la oportunidad de vida

La libertad no hace felices a los hombres, los hace simplemente  
hombres

Manuel Azaña Díaz

## Agradecimientos

Mi más sincero agradecimiento al Dr. Rafael Castillo por brindarme la oportunidad de trabajar en su grupo de investigación, por su apoyo enorme durante el tiempo que duró mis estudios de doctorado, por su preocupación para que tuviera la mejor formación académica y profesional, por los buenos momentos de esparcimiento y sobre todo por su amistad.

Agradezco los grandes aportes por parte de mi comité tutelar durante todo el período académico. Gracias al Dr. Gabriel Eduardo Cuevas González Bravo y al Dr. Enrique Ángeles Anguiano por sus críticas constructivas que ayudaron a fortalecer el trabajo de investigación.

Al Dr. Daniel Alejandro Fernández Velasco por permitirme trabajar en su laboratorio y por su ayuda en todo lo que tuvo que ver con los estudios de inactivación de la enzima triosafosfato isomerasa de *E. histolytica* (EhTIM). Al Dr. Alfredo Téllez por sus sugerencias, al Biomédico Luis Téllez por su gran ayuda en la purificación de la EhTIM.

A la Dra. Lilián Yépez y la Biol. Amparo Tapia por la dirección y realización, respectivamente, de los ensayos de actividad biológica contra *E. histolytica*.

A todos los compañeros del laboratorio 122 del Departamento de Farmacia que sintetizaron los compuestos usados en esta investigación.

Al Dr. Alexander Tropsha y su grupo de investigación por proporcionar los programas necesarios usados en el estudio *kNN-QSAR*. Agradezco al Dr. Artur Olson y su grupo de investigación por proporcionar el programa AutoDock para los estudios de acoplamiento molecular

Al Dr. Sergio Rodríguez por su valiosa ayuda e iniciación en esto del diseño de fármacos asistido por computadora. Especialmente al Dr. José Luis Medina Franco por sus innumerables aportes y críticas certeras a este trabajo, principalmente, por su paciencia, su ejemplo y su amistad incondicional.

Al Dr. Gabriel Eduardo Cuevas, Dr. Francisco Miguel Castro, Dr. Arturo Rojo Domínguez, Dr. Hugo Alejandro Jiménez Vázquez, Dr. Rogelio Gregorio Pereda Miranda y el Dr. Daniel Alejandro Fernández Velasco, miembros del jurado que revisaron y aportaron observaciones valiosas en el manuscrito de esta tesis.

A la Dra. Maria Alicia Hernández por sus oportunas anotaciones, apoyo moral y su valiosa amistad. Al Dr. Francisco Hernández Luis por sus valiosas observaciones en los seminarios.

A la familia Gamiño, Fabiola, Fernando y Naty por su apoyo incondicional de siempre y su valiosa amistad. A Mario y su familia por su amistad.

A la Dirección General de Estudios de Posgrado (DGEP) por el apoyo económico recibido por la beca DGAPA de intercambio académico durante los dos primeros años de los estudios de Maestría y por los años siguientes de beca de los estudios de Doctorado. Agradezco a los proyectos de CONACYT No. G-34851-M y V43629M y a los proyectos DGAPA No: IN202101 y IX244704.

Esta tesis se desarrolló en el laboratorio 122 del Conjunto E de la Facultad de Química de la Universidad Nacional Autónoma de México, bajo la dirección del **Dr. Rafael Castillo Bocanegra**.

Parte del trabajo de esta tesis se desarrolló en el Laboratorio de Fisicoquímica de Proteínas del Departamento de Bioquímica de la Facultad de Medicina de la Universidad Nacional Autónoma de México, bajo la dirección del **Dr. Daniel Alejandro Fernández Velasco**.

<b>Resumen</b>	i
<b>Abstract</b>	iii
<b>Introducción</b>	1
<b>1. Antecedentes</b>	3
1.1 El estado actual de la amebiosis	3
1.2. Diseño de fármacos	5
1.3. Diseño de fármacos asistido por computadora (DiFAC)	6
1.3.1. Estrategias del DiFAC en el descubrimiento y optimización de líderes	7
1.3.1.1. Diseño de fármacos en base a la estructura	7
1.3.1.1.1. Diseño <i>de novo</i>	8
1.3.1.1.2. Acoplamiento molecular ( <i>docking</i> )	9
1.3.1.2. Diseño de fármacos en base al ligando	12
1.3.1.2.1. Relaciones cuantitativas estructura actividad (QSAR)	12
1.3.2. Algunos casos de éxito en el DiFAC	16
<b>2. Objetivos</b>	19
<b>3. Metodología</b>	21
3.1. QSAR-3D	21
3.1.1. Modelado de las moléculas	21
3.1.2. Metodología de CoMFA	21
3.1.3. Validación de los modelos de CoMFA	22
3.2. QSAR-2D	23
3.2.1. Desarrollo de los modelos <i>k</i> NN-QSAR	23
3.2.2. Algoritmo general del método <i>k</i> NN-QSAR	24
3.2.3. Búsqueda de compuestos activos en la base de datos del Instituto Nacional del Cáncer (NCI) de los Estados Unidos	26
3.3. Purificación, inactivación y estudios de acoplamiento molecular de la triosafosfato isomerasa de <i>E. histolytica</i> .	28
3.3.1. Purificación de la enzima recombinante triosafosfato isomerasa de <i>E. histolytica</i> .	28



3.3.2. Determinación de la actividad enzimática de la EhTIM	29
3.3.3. Acoplamiento molecular ( <i>Docking</i> ) de los compuestos inactivantes con la EhTIM	30

## Resultados y discusión

<b>4. Parte I. Relaciones cuantitativas estructura actividad en tres dimensiones (QSAR-3D) de compuestos derivados de bencimidazol con actividad amebicida: Un análisis comparativo de campos moleculares</b>	<b>35</b>
4.1. Desarrollo de los modelos de CoMFA	35
4.1.1. Selección de las formas tautoméricas relevantes para la actividad biológica	38
4.1.1.1. Estudio de CoMFA basado en el tautómero de más baja energía	39
4.1.1.2. Estudio CoMFA basado en las formas tautoméricas mejor predichas	41
4.1.2. Desarrollo de modelos de CoMFA estadísticamente robustos	45
4.2. Validación de los modelos de CoMFA	46
4.3. Mapas de contorno de los modelos de CoMFA	48
4.4. Conclusiones	50
<b>5. Parte II. Estudio de las relaciones cuantitativas estructura actividad en dos dimensiones (QSAR-2D) de compuestos derivados del bencimidazol con actividad amebicida y búsqueda de compuestos activos en bases de datos</b>	<b>53</b>
5.1. Desarrollo de los modelos QSAR-2D	53
5.1.1. Algoritmo <i>k</i> NN-QSAR	53
5.1.2. Robustez de los modelos <i>k</i> NN-QSAR	54
5.1.3. Validación de los modelos <i>k</i> NN-QSAR	56
5.2. Predicción de la actividad amebicida para compuestos antiparasitarios sintetizados en el laboratorio L-122	58
5.3. Búsqueda de compuestos activos en la base de datos del Instituto Nacional del Cáncer (NCI)	61
5.4. Conclusiones	63

<b>6. Parte III. Estudios de inactivación, acoplamiento molecular y diseño de nuevos agentes inactivantes de la triosafosfato isomerasa de <i>E. histolytica</i> (EhTIM).</b>	66
6.1. La triosafosfato isomerasa (TIM) como blanco para el diseño de fármacos	66
6.2. Estructura general de la TIM	67
6.2. Catálisis de la TIM	69
6.3. Interfaz de la TIM como blanco para el diseño de fármacos	70
6.4. Purificación de la EhTIM	71
6.5. Evaluación de compuestos orgánicos como agentes inactivantes de la EhTIM	72
6.7. Exploración de los posibles sitios de unión de los derivados de bencimidazol en la EhTIM usando el acoplamiento molecular <i>in silico</i>	74
6.8. Estudio de acoplamiento molecular de los derivados de bencimidazol <b>1-5</b> sobre la TIM de humano (HsTIM) y selectividad de los compuestos por la EhTIM	85
6.9. Diseño de nuevas moléculas con alta afinidad por la interfaz de la EhTIM como posibles agentes inactivantes	86
6.10. Conclusiones	94
<b>7. Conclusiones generales</b>	97
<b>Referencias</b>	99
<b>Anexo (Publicación)</b>	

## RESUMEN

En este trabajo se aplicaron diferentes métodos del diseño de fármacos asistido por computadora (DiFAC) para la búsqueda y diseño de nuevos compuestos antiamebianos. Los métodos computacionales empleados se seleccionaron de acuerdo a la información experimental disponible en cada caso, esto es, se aplicaron métodos basados en el ligando, como el análisis cuantitativo de la relación estructura-actividad en dos y tres dimensiones (QSAR 3D y 2D) y, en la estructura del receptor, conocido como acoplamiento molecular o *docking*. Para el diseño basado en el ligando se empleó una base de datos de diversos derivados del bencimidazol con actividad *in vitro* contra *E. histolytica*.

Aplicando el método QSAR-3D, conocido como **análisis comparativo de campos moleculares (CoMFA)**, se logró definir cuáles fueron las formas tautoméricas con mayor relevancia para la actividad biológica de algunos compuestos presentes en la base de datos. Una vez seleccionados estos tautómeros, se desarrollaron nuevos modelos de CoMFA usando diferentes tipos de cargas atómicas determinadas usando dos métodos de cálculo: funcionales de la densidad (DFT) y el semiempírico PM3. Los modelos obtenidos mostraron el mismo patrón de sustitución, siendo C-2 y C-5 el de mayor relevancia para la actividad biológica.

Otro de los métodos basados en el ligando que se empleó fue el método QSAR-2D, conocido como **el k vecino más cercano (kNN)**. Con este método se lograron desarrollar 41 modelos QSAR, y éstos se usaron para hacer una búsqueda virtual de nuevos compuestos activos; para esto, se empleó la base de datos del Instituto Nacional del Cáncer de los Estados Unidos (NCI), la cual cuenta con alrededor de 250,000 compuestos. En esta búsqueda se encontraron doce compuestos activos *in vitro* contra *E. histolytica*. Algunos de éstos fueron derivados del bencimidazol, y de otros sistemas de anillos tales como el benzotriazol, el indazol y la indazolona.

Para el diseño basado en el receptor, se purificó la enzima triosafosfato isomerasa de *E. histolytica* (EhTIM) y luego se probó el efecto de inactivación de una serie de 95 compuestos, la mayoría derivados del bencimidazol. Se encontró que cinco de ellos, a una concentración de 250  $\mu\text{M}$ , inactivaron parcialmente a la EhTIM. A partir de estos resultados, se empleó el acoplamiento molecular para estudiar los posibles sitios y modos de unión de los cinco compuestos con la EhTIM. Con los resultados obtenidos se diseñó una serie nueva de compuestos, en su mayoría derivados de bencimidazol, los cuales presentaron mejores energías de acoplamiento calculadas que los cinco compuestos

experimentales. Se espera que los nuevos compuestos diseñados tengan un mayor efecto de inactivación sobre la EhTIM.

## ABSTRACT

In this work different computer assisted drug design methods were employed in order to search and design new active compounds against *Entamoeba histolytica*. The methods used were selected according to available experimental information. The ligand-based drug design was performed by quantitative structure-activity relationships 3D and 2D methods (QSAR 3D and QSAR 2D), whereas for the structure-based drug design, molecular docking was employed

In order to develop some QSAR models, a database of benzimidazole derivatives active against *E. histolytica* was used. A comparative molecular field analysis (CoMFA) study was employed to determine the tautomeric forms more significant for the biological activity. Once the bioactive tautomeric forms were defined, some new CoMFA models were developed using different types of atomic charges. Two levels of calculations -density functional theory and semiempiric PM3 calculations- were applied to estimate atomic charges, and the models thus developed were compared. The final CoMFA models developed showed the same substitution pattern, being the C-2 and C-5 the most relevant for the biological activity.

Another QSAR method employed in this work was the well known *k*- nearest neighbor (*k*NN). Using this method, it was possible to carry out a database mining employing the National Cancer Institute (NCI) database. This database has around 250000 experimental compounds. The results of the database mining showed that 12 compounds had biological activity against *E. histolytica*. The active compounds mined from the NCI database had structures derived from benzimidazole, benzotriazole, indazole and indazolone.

For the structure-based drug design, the enzyme triosephosphate isomerase from *E. histolytica* (EhTIM) was purified and the inactivation effect caused by a series of compounds synthesized in our laboratory (95 compounds) was tested. Five compounds, tested at 250  $\mu$ M, showed a partial inactivation effect on EhTIM. From these results, the molecular docking was used in order to explore the possible binding sites and the manner in which they can bind to EhTIM. The results of these studies allowed the design of novel compounds; most of them were benzimidazole derivatives. These compounds had better affinity than the five experimental compounds. We expect that the novel compounds designed have a bigger inactivation effect than the experimental compounds.

## Introducción

Las enfermedades parasitarias causadas por protozoos se conocen desde hace cientos de años. Entre las más comunes, las enfermedades intestinales han sido un motivo de preocupación principalmente en países en desarrollo. La entamoebosis es una de estas parasitosis causada por el protozoo *Entamoeba histolytica* (ameba). En particular, la entamoebosis es una enfermedad que ataca principalmente a la población más vulnerable, la infantil y de la tercera edad. A pesar de que existen en el mercado medicamentos para el control de esta parasitosis, tales como los heterociclos nitrados, éstos producen severos efectos colaterales y en algunos casos se ha demostrado un efecto de resistencia. Consecuentemente, la búsqueda o descubrimiento de nuevos agentes terapéuticos antiprotozoarios es de suma importancia.

El descubrimiento de un fármaco, y su posterior desarrollo, son dos fases que implican un trabajo interdisciplinario arduo y complejo. Este tipo de investigación en la industria farmacéutica, en los grupos de investigación académicos de diferentes universidades y en los centros de investigación es una tarea difícil.

El diseño racional de un fármaco, en la fase de descubrimiento, puede llevarse a cabo usando diferentes estrategias. Una de las estrategias más recientes, y en creciente desarrollo, es el uso de diversos métodos computacionales; lo que se conoce como el diseño de fármacos asistido por computadora (DiFAC). Esta estrategia constituye una poderosa herramienta para el diseño de nuevas moléculas con una actividad biológica específica. Los métodos de diseño usados en el DiFAC pueden dividirse categóricamente en dos: diseño basado en el ligando y diseño basado en la estructura del receptor. El diseño basado en el ligando es potencialmente útil cuando sólo se tiene información estructural de una serie de pequeñas moléculas y de sus correspondientes valores de actividad biológica. El diseño basado en la estructura del receptor comúnmente se usa cuando existe información, ya sea experimental o teórica, de la estructura tridimensional de la macromolécula receptora implicada en determinado proceso biológico o enfermedad.

En el laboratorio 122 del Departamento de Farmacia de la Facultad de Química (UNAM) se han sintetizado un gran número de moléculas derivadas del bencimidazol. Estas moléculas han presentado actividad antiprotozoaria y antihelmíntica en ensayos *in*

*vitro*. Por su elevado número, forman una base de datos (biblioteca de moléculas) suficiente para llevar a cabo estudios en base al ligando y desarrollar modelos cuantitativos de relación estructura actividad, conocidos como estudios QSAR (*Quantitative Structure-Activity Relationships*). De tal forma que es posible usar la información estructural y de actividad biológica para desarrollar modelos QSAR. Cuando se dispone de la información estructural acerca de los blancos es posible usar métodos con base en la estructura. Uno de los métodos más usados es el acoplamiento molecular (*docking*). Este método permite predecir el modo de unión y calcular el cambio en la energía libre de unión entre un ligando y el receptor.

# Capítulo 1

## Antecedentes

### 1.1. El estado actual de la entamoebosis

Son múltiples los parásitos que pueden alojarse en el ser humano causando diferentes tipos de enfermedades. Dentro de éstas, las enfermedades intestinales constituyen un serio problema de salud pública, primordialmente, en países en vía de desarrollo (Byington *et al*, 1997). Una de las enfermedades intestinales más comunes, que ataca principalmente a la población infantil y a las personas de la tercera edad, es la entamoebosis\* (Kassai *et al*, 1988). Esta enfermedad es producida por el protozoo *Entamoeba histolytica*, conocido comúnmente como amiba o ameba; *E. histolytica* es un microorganismo anaeróbico y fermentativo, sin embargo, se ha reportado que es un aerobio facultativo o tolerante, y puede crecer en un ambiente con 5% de oxígeno (Upcroft y Upcroft, 2001). *E. histolytica* se puede encontrar en dos estadios: en forma de trofozoíto y de quiste. El quiste puede sobrevivir en el agua y los alimentos, y constituye la forma infectiva por ingestión. Éste mide de 10 a 18 micras, es redondeado, posee una cubierta gruesa y resistente al medio que le rodea. Los quistes pueden permanecer vivos durante semanas o meses en un medio externo y se diseminan por agua, artrópodos, alimentos y objetos contaminados (Botero y Restrepo, 1998). En un medio neutro o ligeramente alcalino, el quiste se convierte en trofozoíto metaquiste el cual madura hasta trofozoíto en el intestino grueso del ser humano (Upcroft y Upcroft, 2001), muchas veces sin causarle molestias al huésped (Zavala y Sánchez, 2001). El trofozoíto o forma vegetativa mide de 20 a 40 micras de diámetro; cuando éste está móvil emite unseudópodo, el cual es unidireccional, y le sirve para su movimiento. Regularmente, el trofozoíto se encuentra en la luz del colon o invadiendo la mucosa de la pared intestinal por acción de una proteína de adherencia o lectina, la cual tiene una gran afinidad por la galactosa que se encuentra de forma abundante en las células del colon; allí *E. histolytica* se reproduce por simple división binaria. La penetración a la mucosa se favorece por enzimas líticas que generan lesiones en las superficies de las células (Botero y Restrepo, 1998). El resultado de esta infección es la destrucción tisular de la mucosa intestinal como consecuencia de la acción

---

\*En esta tesis se utiliza la nomenclatura propuesta por la Federación Mundial para el Avance de la Parasitología Veterinaria y la Federación Mundial de Parasitología para denominar las enfermedades parasitarias.



enzimática, mecánica y traumática de los trofozoítos. Es en este punto donde se presentan los síntomas de la entamoebosis intestinal aguda (Zavala y Sánchez, 2001). La infección por *E. histolytica* puede llegar a ser extraintestinal, donde un 20% de los parásitos puede invadir otros órganos produciendo una entamoebosis hepática, pleuropulmonar, pericárdica, esplénica, renal y cerebral, entre otras. Se estima que un 20% de los abscesos extraintestinales resultan fatales (Upcroft y Upcroft, 2001).

Cerca del 90% de las personas que tienen alojada a la *E. histolytica* son asintomáticas, en donde el parásito se encuentra en forma de quiste (Bansal *et al*, 2006). Se estima que alrededor de 500 millones de personas en el mundo están infectadas con este parásito, y cerca de un 10% (48 millones) de éstos padecen los síntomas clínicos. La entamoebosis causa la muerte de entre 70,000 y 100,000 personas cada año en el mundo (Upcroft y Upcroft, 2001). Hoy en día existen tratamientos para el control de la entamoebosis; siendo el metronidazol, un derivado nitrado del imidazol, el fármaco de elección usado en su tratamiento. Existen además otros derivados del nitroimidazol, como el tinidazol, el secnidazol y el ornidazol, que son igualmente efectivos. Este tipo de compuestos presentan efectos secundarios severos y se ha demostrado, en el caso del metronidazol, casos de resistencia debido al tratamiento indiscriminado e incompleto de la entamoebosis (Johnson, 1993; Upcroft y Upcroft, 1993; Voolmann y Boreham, 1993). Algunos estudios han mostrado que el metronidazol es potencialmente carcinogénico en humanos debido a que éste causa daños a nivel del ADN. Se ha demostrado que es mutagénico en sistemas bacteriales, es genotóxico en células humanas y es carcinogénico en los animales (Menéndez *et al*, 2001; Bendesky *et al*. 2002).

Otros fármacos alternos se han usado para el tratamiento de la entamoebosis, tales como: el furoato de diloxanida, la diyodohidroquinina, el paromomicin, la cloroquina y la emetina; sin embargo, la emetina es muy raramente usada debido a su alta toxicidad (Knight, 1980).

Estudios sobre la actividad antiprotozoaria de los bencimidazol 2-carbamatos (BC), tales como el albendazol, el mebendazol, el flubendazol y el febendazol demostraron que estos compuestos antihelmínticos también inhiben el crecimiento *in vitro* de los parásitos *Trichomonas vaginalis* (Sears y O'Jare, 1988) y *Giardia intestinales* (Cedillo-Rivera y Muñoz, 1992). De igual forma, se ha demostrado clínicamente, que estos BC no son efectivos contra *E. histolytica* (Chávez *et al*, 1992). La actividad biológica contra *T. vaginalis* y *G. intestinales* se ha atribuido principalmente a la capacidad que tienen los BC de inhibir la polimerización de la tubulina; de tal forma que éstos tienen un efecto sobre el

citoesqueleto (Reynoldson *et al*, 1992). Trabajos recientes reportados por el grupo de investigación del laboratorio 122, del Departamento de Farmacia (UNAM) y la Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias del Centro Médico Nacional (IMSS) han mostrado que algunos derivados del bencimidazol, no carbamatos, resultaron activos *in vitro* contra *E. histolytica*. Incluso, algunos de estos compuestos presentaron una mayor actividad que el metronidazol (Valdez *et al*, 2002; Navarrete-Vázquez *et al*, 2001; Sánchez, 1999; Navarrete-Vázquez *et al*, 2003; Navarrete-Vázquez *et al*, 2006). Los resultados de los trabajos anteriores culminaron con la obtención de un conjunto de moléculas o base de datos que han permitido estudiar de manera cualitativa la relación entre la estructura de estos derivados del bencimidazol, no carbamatos, con la actividad biológica en *E. histolytica*. La base de datos obtenida resulta ser un excelente punto de partida para estudios que relacionen la estructura con la actividad biológica en *E. histolytica* de forma cuantitativa. Estos estudios son básicamente apoyados por programas de cómputo y tienen como objeto encontrar nuevas moléculas activas.

Por otro lado, en el laboratorio de Ingeniería de Proteínas del Departamento de Bioquímica, de la Facultad de Medicina, se ha logrado obtener y purificar la enzima recombinante triosafosfato isomerasa de *E. histolytica* (Landa *et al*, 1997). Debido al importante papel que juega esta enzima en la glicólisis en microorganismos anaerobios, se considera como un excelente blanco para la búsqueda de nuevos fármacos (Gómez-Puyou *et al*, 1995).

## 1.2. Diseño de fármacos

El descubrimiento de un fármaco comienza al encontrar un compuesto líder. Hasta los años 70, el descubrimiento de un compuesto líder dependía esencialmente de procesos u observaciones al azar, casualidad, por ensayo de grandes cantidades de moléculas (bases de datos) o por aislamiento de productos naturales. Desde entonces, el diseño de nuevas moléculas activas está basado en un conocimiento racional. De allí que existan cuatro estrategias que permiten encontrar nuevos compuestos líderes (Wermuth, 2003). La primera estrategia se basa en la modificación estructural de moléculas activas ya existentes. En esta estrategia se pueden usar diferentes metodologías: alargamiento de cadenas, hibridación molecular, optimización bioisostérica, etc. Una segunda estrategia consiste en probar sistemáticamente la actividad biológica de una gran cantidad de compuestos seleccionados aleatoriamente. En esta estrategia los experimentos de alta capacidad (HTS, *High-throughput screening*) y la química combinatoria juegan un papel

importante. Una tercera estrategia comprende una explotación más exhaustiva de la información biológica, de manera que la observación en estudios etnofarmacológicos, observación sobre los efectos colaterales de algunas medicinas, el uso de medicamentos largamente conocidos, las observaciones hechas sobre animales, plantas y microorganismos pueden aportar información valiosa para obtener nuevos líderes. Finalmente, una cuarta estrategia comprende un diseño racional basado en el conocimiento molecular involucrado en un proceso patológico. En esta última estrategia el DiFAC ha cobrado vital importancia, principalmente, en la medida que aumenta la velocidad y la capacidad de proceso de las nuevas computadoras (Wermuth, 2003).

### **1.3. Diseño de fármacos asistido por computadora (DiFAC)**

El DiFAC es una disciplina del diseño racional de fármacos que se enfoca en la búsqueda y la optimización de nuevos compuestos líderes con una actividad biológica específica. El DiFAC se define como el uso de diversos métodos computacionales para relacionar la actividad biológica de un compuesto con su estructura (Hopfinger, 1995). Los métodos computacionales se basan en el postulado de que las moléculas biológicamente activas actúan por interacción con una biomacromolécula-receptor (blanco), las cuales son principalmente proteínas o ácidos nucleicos, involucrados en algún proceso patológico. La naturaleza de estas interacciones es principalmente de tipo estérico, electrostático, hidrofóbico y de puentes de hidrógeno (Veselovsky y Ivanov, 2003). Sin embargo, el DiFAC no es un método directo para obtener nuevos fármacos; por lo tanto debe partir de algún tipo de información experimental.

Uno de los primeros estudios que dieron origen al DiFAC se atribuye a los trabajos de Hansch-Fujita en 1963. Hansch correlacionó la actividad biológica de una serie de derivados análogos con ciertas características electrónicas e hidrofóbicas propias de estos derivados. Los investigadores pudieron derivar una ecuación que permitió observar de manera cuantitativa esta correlación (Hansch *et al*, 1963).

Con el creciente desarrollo de las computadoras en los años 70, comenzaron a surgir diversos programas de cómputo que permitían visualizar las moléculas en tres dimensiones y calcular de manera teórica algunas de sus propiedades. Esto dio origen al modelado molecular, el cual se define como *algo virtual que requiere el uso de una computadora y programas para dibujar, describir o evaluar cualquier aspecto de las propiedades de la estructura de una molécula* (Penssak, 1989) Como resultado de esto, el modelado molecular provee una excelente herramienta para construir, visualizar, analizar

y almacenar modelos moleculares que pueden ayudar a interpretar la relación que existe entre la estructura de un compuesto con su actividad biológica (Cohen *et al*, 1990).

### **1.3.1. Estrategias del DiFAC en el descubrimiento y optimización de líderes**

La estrategia a seguir en el DiFAC depende básicamente de la información estructural disponible sobre el blanco o el ligando. Esta información puede ser experimental o teórica (en el caso del modelado por homología de la estructura de los blancos (Wieman *et al*. 2004)). De ahí que se utilicen dos estrategias en el DiFAC. Cuando se conoce la estructura del blanco o receptor, comúnmente se usan métodos derivados del **diseño en base a la estructura** como estrategia de diseño; y cuando no se conoce la estructura del blanco, pero se conoce la estructura y actividad biológica de una serie de moléculas que tienen la misma actividad biológica, se aplica la estrategia del **diseño en base al ligando** (Ooms, 2000). Lo anterior implica que los ligandos o moléculas activas deben unirse en una sola conformación del farmacóforo y al mismo sitio receptor. Cuando no existe información acerca del blanco, ni de las moléculas responsables de la actividad biológica, es necesario generar información experimental.

#### **1.3.1.1. Diseño de fármacos en base a la estructura**

El creciente desarrollo de disciplinas como la biología molecular, la genómica y la proteómica ha dejado un aumento en el conocimiento estructural de una gran cantidad de blancos. Prueba de esto ha sido el aumento en el número de estructuras depositadas en la base de datos del *Protein Data Bank* (a abril de 2007 cuenta con 42,627 estructuras, <http://www.rcsb.org/pdb/>). Existen varias formas de obtener la estructura tridimensional, siendo la más importante la cristalografía de rayos X, en segundo lugar la resonancia magnética nuclear (RMN) y la criomicroscopía electrónica (Saibil, 2000). Este último método tiene la desventaja de que la resolución de la estructura es mucho más baja que la obtenida por experimentos de rayos X y RMN.

Cuando no se dispone de métodos experimentales para determinar la estructura tridimensional de una macromolécula, existe como alternativa el obtener ésta a partir de una estructura que tenga una alta similitud en la secuencia de aminoácidos con la estructura del blanco desconocido. Ahora bien, aplicando métodos computacionales es posible obtener un modelo aproximado de la estructura tridimensional de la macromolécula desconocida. A esto se le conoce como modelado por homología (Wieman *et al*, 2004; Hillisch *et al*, 2004).

Una vez que se conoce la estructura tridimensional del blanco, el siguiente paso es buscar el sitio de unión. Un blanco puede tener varios sitios donde un ligando puede interactuar. Los sitios de unión pueden ser sitios catalíticos, donde el ligando compite por el sitio de unión; a estos ligandos se les llama inhibidores competitivos. De igual forma, un ligando puede unirse a un sitio diferente al sitio catalítico. A estos ligandos se les conoce como inhibidores no competitivos. En ocasiones resulta ser fácil encontrar el sitio de unión, pues la estructura del blanco pudo haber sido resuelta unida al ligando natural o un inhibidor. En algunos casos, algunas proteínas pueden cocrystalizar con compuestos orgánicos pequeños (solventes); esto permite encontrar sitios hidrofóbicos en la superficie del blanco (Gao *et al*, 1999). Cuando el sitio de unión no es conocido, pueden emplearse métodos computacionales para determinarlo. Una opción es usar algún programa de acoplamiento molecular y hacer una búsqueda sobre toda la superficie del blanco (*blind docking*) y determinar el sitio de unión por comparación de las energías libres de enlace. Esto es conocido como un acoplamiento molecular ciego. El programa Grid también es una buena alternativa, pues éste sugiere probables sitios de unión (Goodford, 1985).

Cuando el sitio de unión es identificado se pueden aplicar dos métodos para diseñar moléculas activas. Uno de ellas es el diseño *de novo*. Éste es llamado así, pues la molécula es construida dentro del sitio de unión a partir de fragmentos, previamente acoplados al sitio de unión, y luego unidos por átomos o pequeños fragmentos (Schneider y Fechner, 2005). Otro método es el acoplamiento molecular automatizado o *docking*, donde una molécula construida previamente se somete a ciertos procedimientos que permiten predecir la conformación, la orientación y la energía con que ésta se une al sitio del receptor; lo anterior produce una propuesta para uno o varios modelos del complejo ligando-receptor (modelo de unión). Una vez caracterizado el modelo de unión, es entonces posible proponer variaciones estructurales en el ligando para mejorar su afinidad por el sitio de unión (Taylor *et al*, 2002).

#### **1.3.1.1.1. Diseño *de novo***

El diseño *de novo* es una tarea difícil, sobre todo cuando se sabe que el espacio de búsqueda es enorme (cantidad de estructuras viables). En la primera etapa del diseño *de novo* es necesario reducir considerablemente el espacio de búsqueda. Esto consiste en aplicar un conjunto de restricciones primarias que pueden ser aplicadas sobre el receptor. En estas restricciones las interacciones del sitio receptor forman la base del diseño. Las

interacciones son de tipo no covalente, tales como estéricas, electrostáticas, hidrofóbicas y de puentes de hidrógeno.

Una segunda etapa consiste en caracterizar el sitio de la interacción. Algunos programas, como DOCK (Kuntz *et al*, 1982; Swing *et al*, 2001), pueden ser usados para calcular el volumen o espacio accesible al ligando. Posteriormente, se colocan átomos o fragmentos en él ya bien caracterizado sitio de unión. Debido a que muchas soluciones son posibles, es necesario asignarle a cada solución una puntuación con el objeto de decidir cual es más promisorio; esta tarea la realiza una función de evaluación (*scoring function*). Ésta mide la afinidad del fragmento por la fracción del sitio de unión. Existen varios programas que pueden hacer diseño *de novo*, tales como: LUDI (Böhm, 1992), SPROUT (Guillet *et al*, 1995), LeapFrog (Tripos) entre otros. Estos programas pueden usar campos de fuerza o métodos empíricos para medir la afinidad. Cuando los átomos o fragmentos seleccionados por la función de evaluación han tomado una posición energéticamente favorable, el siguiente paso es el ensamble de éstos. Para el ensamble se pueden usar átomos, sistemas de anillos o cadenas carbonadas, los cuales se unen de forma covalente a los fragmentos prediseñados. Al final, se evalúa la afinidad del ligando completamente ensamblado usando una función de evaluación (Schneider y Fechner, 2005). Se han reportado recientemente casos donde el diseño *de novo* ha sido exitoso (Ji *et al*, 2003; Honma *et al*, 2001; Rogers-Evans *et al*, 2004).

#### **1.3.1.1.2. Acoplamiento molecular (*docking*)**

Es necesario comprender como se llevan a cabo los procesos de reconocimiento molecular fármaco-receptor para entender cómo es el mecanismo de acción de los fármacos. Esto no es un proceso sencillo, pues es necesario saber cuál es la magnitud y las fuerzas de interacción que intervienen, y la conformación que adoptan tanto el receptor como el ligando para que sean complementarios. Por tanto, la estructura tridimensional del complejo ligando-receptor obtenida experimentalmente, constituye una información valiosa para comprender los mecanismos de acción y es la base de partida para el diseño de nuevas moléculas activas.

Las interacciones que resultan de este acoplamiento pueden conducir a una serie de eventos moleculares; por ejemplo una reacción catalítica, la estabilización de un estado de transición, alteración de la actividad catalítica o bloqueo del sitio activo del receptor. El entendimiento de estos procesos constituye la base de la farmacología molecular.

El acoplamiento molecular o *docking* es un procedimiento que permite identificar el modo aproximado de unión (modelo de unión) de un ligando en el sitio de unión del receptor. A partir del modelo es posible diseñar nuevas moléculas. También puede ser usado para la evaluación *in silico* de bases de datos virtuales en la búsqueda de moléculas activas. Actualmente, el acoplamiento molecular es un método muy usado en el descubrimiento y optimización de compuestos líderes (Halperin *et al*, 2002; Waszkowycz, 2002; Schneidman-Duhovny *et al*, 2004).

Para hacer *docking* es necesario contar con la estructura tridimensional del receptor, como en el diseño *de novo*, y con la estructura del ligando o una librería virtual. La forma de representar a la proteína y al ligando, así como la forma de evaluar el espacio de búsqueda conformacional del ligando en el sitio de unión dependen del programa que se use. Existen tres formas para representar al receptor que son usadas por los programas computacionales de acoplamiento molecular: La primera es una representación atómica, la cual se utiliza en conjunto con una función de energía potencial (Taylor y Burnett, 2000). Una segunda representación se basa en superficies moleculares, aunque es más comúnmente usada para el estudio de las interacciones proteína-proteína (Norel *et al*, 1994; Norel *et al*, 1999). Una tercera representación utiliza una rejilla o malla tridimensional dividida por puntos espaciados regularmente. Estos puntos guardan información acerca de la contribución energética del receptor para la unión de un ligando. Las contribuciones energéticas son usualmente estéricas y electrostáticas (ecuaciones 1.1 y 1.2) (Goodford, 1985). Este último método comúnmente se usa en muchos de los programas actuales de acoplamiento molecular.

$$E_{vdW} = \sum_{i=1}^{n \text{ átomos}} [A_{ij}r_{ij}^{-12} - C_{ij}r_{ij}^{-6}] \quad (1.1)$$

$$E_C = \sum_{i=1}^{n \text{ átomos}} \frac{q_i \bullet q_j}{\epsilon \bullet r_{ij}} \quad (1.2)$$

Al igual que en el diseño *de novo*, en un estudio de acoplamiento molecular es necesario identificar el sitio de unión. En ocasiones es posible obtener la información de los sitios de unión o de los sitios alostéricos a partir de las estructuras cristalográficas obtenidas de la base de datos del *Protein Data Bank*; donde el receptor o una proteína se cocrystaliza con un inhibidor o con su ligando natural.

Un estudio de acoplamiento molecular requiere de dos etapas: La primera consiste en la búsqueda del espacio conformacional accesible al ligando. En esta etapa se pueden utilizar diferentes algoritmos de búsqueda. Estos algoritmos permiten encontrar las posibles orientaciones y conformaciones de unión del ligando al sitio de unión en el receptor (Krovat *et al*, 2005; Halperin *et al*, 2002). En la segunda etapa se usan funciones de evaluación, las cuales utilizan como registro de entrada la posición y conformación del ligando interactuando con los átomos en el sitio de unión y retorna como registro de salida un número. Este número representa la probabilidad que tiene el ligando, en la posición y conformación del registro de entrada, de unirse al sitio de unión. La mayoría de las funciones de evaluación, las cuales usan diversos programas de acoplamiento molecular, utilizan campos de fuerza. Éstos calculan la energía libre de unión del complejo ligando-receptor. Entre más negativa sea la energía, más afinidad tendrá el ligando por el receptor (Sotriffer *et al*, 2003). Aunque hay que tener precaución, pues dada la naturaleza de la evaluación es posible obtener resultados errados debidos a falsos positivos o falsos negativos.

Es necesario saber hasta que punto los resultados que se obtienen de un estudio de acoplamiento molecular son confiables, esto es, con qué grado de exactitud se pueden reproducir los resultados experimentales o en dado caso predecirlos. El resultado de cualquier estudio de acoplamiento molecular es un modelo de unión ligando-receptor y es necesario validarlo. Una forma de validarlo es tomar una estructura tridimensional de algún complejo ligando-receptor obtenida experimentalmente y reproducir el modo de unión. Entre menor sea la diferencia del modelo obtenido y la estructura experimental, mejor será el programa para predecir otras estructuras de complejos de otros ligandos con el mismo receptor.

Una de las aplicaciones para la búsqueda de moléculas activas, cuyo empleo va en aumento, es el uso del acoplamiento molecular en la evaluación de bases de datos extensas (Sotriffer *et al*, 2003; Shoichet *et al*, 2002), en el cual se pueden evaluar miles de moléculas *in silico* en cortos periodos de tiempo. La base de datos Zinc (Irwin y Shoichet, 2005) es una biblioteca virtual de carácter público, la cual contiene más de 4.6 millones de compuestos (<http://blaster.docking.org/zinc/>, 2006). Esta base de datos fue desarrollada en la Universidad de California en San Francisco (UCSF). Los archivos que contiene esta base de datos pueden ser usados directamente por el programa DOCK (Kuntz *et al*, 1982; Swing *et al*, 2001) para hacer su evaluación *in silico*. La gran ventaja que tiene esta base



de datos es que todos los compuestos relacionados allí se encuentran comercialmente disponibles.

### **1.3.1.2. Diseño de fármacos en base al ligando**

A mediados del siglo XIX se tuvo la sospecha que la actividad bloqueadora neuromuscular del curare era atribuida al grupo amonio cuaternario de algunas sales orgánicas, de igual manera, se sugirió que la actividad hipnótica de ciertos alcoholes alifáticos estaba relacionada con el peso molecular. Esto fue una evidencia de que la actividad biológica estaba estrechamente relacionada con la estructura. De tal forma que la actividad biológica podía ser modulada con los cambios en la estructura de las moléculas (Claramun, 2001). Cuando la variación de la actividad biológica, como consecuencia de los cambios estructurales de las moléculas, se determina de forma cuantitativa, recibe el nombre de Relaciones Cuantitativas Estructura-Actividad. Conocido en el idioma inglés como **QSAR (Quantitative Structure-Activity Relationships)**.

#### **1.3.1.2.1. Relaciones cuantitativas estructura actividad (QSAR)**

Hansch y Fujita establecieron que el efecto de los sustituyentes sobre la fuerza de interacción entre un fármaco y su receptor es una combinación aditiva de los efectos de los sustituyentes del fármaco sobre un modelo de interacción intermolecular; las interacciones que asumieron fueron de tipo estérico, electrostático e hidrofóbico, principalmente. La magnitud de estas interacciones puede ser descrita cuantitativamente por ciertos parámetros o propiedades fisicoquímicas. Este método es conocido como extratermodinámico o de Hansch (Hansch *et al*, 1963; Hansch y Fujita, 1964)

Los métodos **QSAR** poseen dos características: primero, son métodos que transforman la estructura química de una molécula en un grupo de propiedades que pueden ser descritas cuantitativamente; estas propiedades son conocidas como descriptores. Los descriptores relevantes son aquellos que tienen significatividad para la actividad biológica. La segunda característica es que establecen una relación cuantitativa entre los descriptores relevantes y las propiedades biológicas atribuidas a sus moléculas (Martin, 1981). La relación entre los descriptores y la actividad biológica puede representarse por medio de una ecuación matemática que los relacione proporcionalmente (Ecuación 1.3).

$$\text{Log}(1/C) = k_1 \log P + k_2 s + \dots + k_n X + \dots + k_{n+1} \quad (1.3)$$

C = Mínima dosis efectiva

$\log P$  = Coeficiente de partición octanol-agua (descriptor 1)

s = Constante del sustituyente de Hammet (descriptor 2)

X = Algún otro descriptor

$k_n$  = Constante ajustada por la regresión lineal

Al tiempo que Hansch publica su primer trabajo sobre el método extratermodinámico, Free y Wilson publican el trabajo sobre su método. En este último, las moléculas no son descritas por propiedades fisicoquímicas, en su lugar, se usa un grupo de indicadores para denotar la ausencia o presencia de ciertos grupos en las moléculas. Wilson supuso que la adición de un grupo o un sustituyente en una posición particular de una molécula se refleja en un cambio cuantitativo en la actividad biológica (Free y Wilson, 1964). Esta suposición es un análisis más próximo a una relación estructura-actividad cualitativa. Esta relación puede denotarse por la ecuación 1.4.

$$\text{Log}(1/C) = x + \sum_{ij}^{mn} a_{ij} G_{ij} \quad (1.4)$$

C = Mínima dosis efectiva

i = Número de la posición de la sustitución

j = Número del sustituyente a una posición en particular

m = Número total de posiciones de sustitución

n = Número de sustituyentes

$a_{ij}$  = Indica la presencia (1.0) o ausencia (0.0) del sustituyente ij

$G_{ij}$  = Valor de la contribución del grupo y es definida por una regresión lineal múltiple

Los métodos QSAR han ido evolucionando de acuerdo al tipo de descriptores que usan y de acuerdo al método matemático usado para correlacionar la estructura con la actividad biológica. Actualmente, pueden clasificarse de acuerdo a esta evolución.

De acuerdo al método usado para correlacionar la estructura con la actividad biológica, los estudios QSAR pueden clasificarse como lineales y no lineales. Los métodos lineales se pueden clasificar a su vez de acuerdo al tipo de gráficos moleculares usados para caracterizar o determinar los descriptores. Así, el método extratermodinámico usa gráficas moleculares en dos dimensiones para derivar sus descriptores. Esto quiere decir que los descriptores derivados no requieren de la información de cómo están los átomos de la molécula dispuestos en el espacio para la interacción con su receptor. Por lo tanto, estos métodos se conocen como métodos QSAR en dos dimensiones (QSAR-2D).

Existen otros métodos donde las gráficas moleculares en tres dimensiones son indispensables para el cálculo de los descriptores, de manera que es necesario conocer la disposición espacial de los átomos de la molécula. Éstos son conocidos como métodos QSAR en tres dimensiones (QSAR-3D) (Oprea y Waller, 1997; Greco, 1997).

Uno de los métodos QSAR-3D más reportados en la literatura científica es el conocido análisis comparativo de campos moleculares (**Comparative Molecular Field Analysis, CoMFA**). CoMFA se basa principalmente en el hecho de que todas las interacciones entre el fármaco y el receptor son de tipo no covalente, principalmente electrostáticas y estéricas. Mediante un adecuado muestreo de los campos estéricos y electrostáticos se pueden definir los campos y las regiones que tienen gran significatividad para la actividad biológica. Este programa fue desarrollado por Cramer y colaboradores (Cramer *et al*, 1988). De igual forma, se han desarrollado otros programas para hacer estudios QSAR-3D, entre ellos están el análisis comparativo de índices de similitud (**Comparative Molecular Similarity Indices Analysis, CoMSIA**) (Klebe *et al*, 1994) y un programa de reciente aparición es el análisis comparativo de sitios moleculares activos (**Comparative Molecular Active Site Analysis, CoMASA**) (Kotani y Higashiura, 2004) Otro programa en fase de desarrollo es el CAMOL o campos moleculares, el cual está siendo desarrollado en Brasil (<http://www.ufpe.br/farmacologia/zaldini/>). Los estudios QSAR presentados hasta el momento usan regresiones lineales múltiples o análisis de mínimos cuadrados parciales para correlacionar los descriptores (la estructura) con la actividad biológica, por tanto estos métodos son lineales.

Pueden usarse otros métodos de correlación que no son lineales; un ejemplo es el método de clasificación del  $k$  vecino más cercano aplicado a los estudios QSAR (***k*-Nearest Neighbour, kNN-QSAR**) (Sharaf *et al*, 1986). Este método evalúa la similitud entre moléculas de acuerdo a ciertos descriptores, y las moléculas que son similares se sitúan en un espacio multidimensional que constituye un farmacóforo hipotético. La actividad biológica de una molécula se determina como el promedio ponderado de sus  $k$  vecinos más cercanos, siendo  $k$  un número entero típicamente entre 1 y 5 (Zheng y Tropsha, 2000).

Uno de los pasos más importantes en cualquier estudio QSAR es la validación, debido a que las correlaciones que se encuentran podrían ser producto de la casualidad. La validación trata de evaluar la capacidad que tiene un modelo QSAR para predecir con exactitud la actividad biológica o alguna otra propiedad que se esté correlacionando de un grupo de moléculas que no fueron usadas para desarrollar el modelo. Existen dos formas

de validación. El primer método es la validación interna o cruzada. Para la validación interna lo más usado es el método de dejar uno afuera (*Leave-One-Out, LOO*). Éste consiste en dejar fuera un compuesto a la vez y desarrollar  $n$  modelos como  $n$  compuestos existan. En cada modelo desarrollado un compuesto diferente es excluido y con su correspondiente ecuación de correlación se calcula la actividad biológica del compuesto excluido. Finalmente se grafica las actividades calculadas contra las experimentales y se obtiene un valor de  $R^2$  (coeficiente de determinación), comúnmente conocido como  $q^2$ . El valor de  $q^2$  se ha usado comúnmente como criterio de la capacidad de un modelo para predecir la actividad biológica y entre mayor sea el valor de  $q^2$  mayor es la capacidad de predicción del modelo.

Cuando se desarrolla un modelo QSAR se usa un grupo de moléculas o una base de datos al cual se le llama grupo de entrenamiento y es común dividir la base de datos original en dos grupos. El primero es el grupo de entrenamiento y el segundo grupo, uno más pequeño y no menor a 5 moléculas (Golbraikh y Tropsha, 2002), es llamado grupo de prueba. El grupo de prueba sirve para hacer la validación externa, es decir, se predice la actividad biológica de las moléculas del grupo de prueba usando el modelo desarrollado con el grupo de entrenamiento. Haciendo la correlación entre los valores de actividad biológica calculados versus los experimentales, se obtiene un valor de coeficiente de determinación  $R^2$ ; este valor es una medida de la capacidad de predicción del modelo. Dado que no hay una correlación entre los valores de  $q^2$  de la validación interna o cruzada con los valores de  $R^2$  de la validación externa, se han definido algunos criterios adicionales para demostrar efectivamente la capacidad de predicción de cualquier modelo QSAR. De manera que el empleo de  $q^2$  como un criterio único de validación no es suficiente (Golbraikh y Tropsha, 2002).

Al igual que en el diseño basado en la estructura, la búsqueda en bases de datos también ayuda a identificar nuevas moléculas líderes. La búsqueda en bases de datos ha venido en aumento gracias a los avances en el desarrollo de computadoras cada vez con mayor capacidad de procesamiento, de manera que es posible evaluar una gran cantidad de moléculas en tiempos muy cortos. La búsqueda de compuestos activos en bases de datos puede conducirse sobre la base de la similitud de un compuesto activo incluido en un modelo QSAR. Ésta puede determinarse en un espacio multidimensional donde se grafican los valores de los descriptores. La similitud se evalúa en cada molécula en la base de datos como el valor de la distancia euclidiana entre una molécula del grupo de entrenamiento y una molécula de la base de datos en un espacio multidimensional del

descriptor. El *k*NN-QSAR constituye una excelente herramienta en la búsqueda de moléculas activas en bases de datos virtuales (Medina-Franco *et al*, 2005; Shen *et al*, 2004).

Existen ciertas ventajas y desventajas de los métodos QSAR-2D respecto a los QSAR-3D y viceversa (Tabla 1.1).

**Tabla 1.1.** Ventajas y desventajas de los métodos QSAR en dos y tres dimensiones

QSAR-3D	QSAR-2D
<b>Ventajas</b>	
Puede predecir la actividad biológica de otras moléculas	Puede predecir la actividad biológica de otras moléculas
No es necesario interpretar la ecuación de correlación	Requiere poco tiempo de cómputo
Sugiere automáticamente moléculas más activas	No se necesita asumir la posición en el sitio del receptor
Explica directamente las características estructurales necesarias para la actividad biológica	Se pueden evaluar bases de datos extensas
<b>Desventajas</b>	
Es necesario asumir la posición en el sitio receptor	Solo moléculas de estructura similar pueden ser analizadas
Se necesita conocer o asumir un modelo de farmacóforo	No sugiere automáticamente moléculas más activas
Requiere mayor tiempo de cómputo	Es necesario interpretar la ecuación

### 1.3.2. Algunos casos de éxito en el DiFAC

El DiFAC ha jugado un papel muy importante en la búsqueda y la optimización de moléculas líderes. Algunos de los medicamentos que hoy día se encuentran en el mercado han utilizado algún método del DiFAC. Uno de los primeros ejemplos de un fármaco diseñado usando el QSAR-2D es la norfloxacin, un agente antibacteriano derivado de la quinolona. Este medicamento presenta un espectro antibacteriano más amplio que el del ácido nalidíxico, el cual tiene estructura similar (Koga *et al*, 1980).

Estudios QSAR jugaron también un papel importante en el descubrimiento del Donepezil, un potente inhibidor de acetilcolinesterasa (AChE), el cual es empleado en el tratamiento de pacientes con la enfermedad de Alzheimer (Cardozo *et al*, 1992; Cardozo *et al*, 1992a; Kawkami *et al*, 1996). Otros ejemplos de casos de éxito, donde los estudios QSAR han jugado un papel importante en la optimización de moléculas líderes, se reportan por Fujita (Fujita, 1990).

El diseño de fármacos basado en la estructura también ha rendido sus frutos en el descubrimiento de nuevos medicamentos. El Zanamivir es un medicamento antiviral usado para el tratamiento de la gripe. Éste es un potente inhibidor de la neuraminidasa,

enzima que se encuentra en la superficie del virus que produce la influenza. Este último constituye el ejemplo más conocido de diseño basado en la estructura (Von *et al*, 1993). Otros ejemplos donde el uso de algún método del DiFAC o modelado molecular ha resultado en medicamentos comerciales pueden verse en la Tabla 1.2 (Medina-Franco *et al*, 2007).

**Tabla 1.2.** Medicamentos comerciales desarrollados usando algún método del DiFAC

Compuesto	Nombre comercial	Tratamiento	Método	Ref.
Losortan	Cozaar®	Hipertensión	Modelado molecular y QSAR	Duncia <i>et al</i> , 1990
Dorzolamida	Truspot®	Glaucoma	Diseño basado en el receptor y cálculos <i>ab initio</i>	Baldwin <i>et al</i> , 1989; Greer <i>et al</i> , 1994
Eprosortan	Teveten®	Hipertensión	Modelado molecular	Weinstock <i>et al</i> , 1991; Keenan <i>et al</i> , 1993
Zolmitriptan	Zoming®	Migraña	Modelado molecular y desarrollo del farmacóforo	Bucknham <i>et al</i> , 1995
Nelfinavir	Viracept®	Sida	Diseño basado en el receptor	Kaldor <i>et al</i> , 1997
Indinavir	Crixivan®	Sida	Diseño basado en el receptor	Dorsey <i>et al</i> , 1994

Como puede observarse, el DiFAC ha tenido un alto impacto en el desarrollo de nuevos medicamentos; su aplicación ha sido no sólo importante en la optimización farmacodinámica sino también en la optimización farmacocinética. De tal forma que algunos métodos del DiFAC, tanto basado en la estructura como en el ligando, han sido usados para explorar las propiedades ADME/Tox (absorción, distribución, metabolismo, excreción y toxicidad).

La importancia del estudio de las propiedades ADME/Tox radica en que casi el 50% de los fármacos fallan debido a una eficacia inaceptable por una baja biodisponibilidad, debido a unas propiedades biofarmacéuticas o farmacocinéticas pobres (Li, 2001).

El modelado molecular y métodos del DiFAC se han usado de igual forma para construir modelos que permitan predecir si un compuesto puede tener o no buenas propiedades farmacocinéticas. Esto ha llevado al desarrollo de una gran cantidad de programas computacionales basados en el estudio de las propiedades tridimensionales de las moléculas. (Li, 2002; Waterbeemd y Gifford, 2003; Turner y Glass, 2003).

Uno de los parámetros más importantes en la predicción de las propiedades ADME/Tox ha sido el coeficiente de partición octanol-agua (logP). Numerosos estudios QSAR o de relaciones cuantitativas estructura-propiedad QSPR (*Quantitative Structure-*

*Property Relationships*) han permitido la predicción de estas propiedades (Hansch *et al*, 2004).

Algunos de estos programas usan a menudo reglas empíricas para predecir las propiedades ADME/Tox aunque por ser empíricas no son menos importantes. Estas reglas fueron definidas por Lipinski y son conocidas como la regla de los cinco o de Lipinski (Lipinski *et al*, 1997; Lipinski *et al*, 2001).

En el presente trabajo de tesis se emplearon diversos métodos del DiFAC, tanto del diseño basado en el ligando como del diseño basado en la estructura, para la búsqueda y diseño de nuevas moléculas con actividad amebicida.

# Capítulo 2

## Objetivos

### Objetivo general

Buscar y diseñar compuestos con actividad amebicida a través de los métodos de diseño de fármacos asistido por computadora.

### Objetivos Específicos

- Desarrollar modelos cuantitativos de relación estructura-actividad en tres dimensiones, con alta capacidad de predicción de la actividad biológica, usando la metodología de análisis comparativo de campos moleculares (CoMFA).
- Validar los modelos de CoMFA obtenidos.
- Diseñar nuevos derivados del bencimidazol activos contra *E. histolytica* a partir de los estudios de CoMFA.
- Desarrollar modelos cuantitativos de relación estructura-actividad en dos dimensiones usando el método de “El *k* vecino más cercano (*k*NN-QSAR)”.
- Realizar una búsqueda automatizada de compuestos con actividad amebicida usando la base de datos del Instituto Nacional del Cáncer de los Estados Unidos (NCI) mediante los modelos *k*NN-QSAR desarrollados.
- Probar la actividad biológica contra *E. histolytica* de los compuestos seleccionados del NCI.
- Obtener, purificar y caracterizar la enzima triosafosfato isomerasa de *E. histolytica* (EhTIM).



- Realizar ensayos de inactivación enzimática sobre EhTIM con compuestos derivados del bencimidazol y algunos compuestos intermediarios de sus síntesis.
- Estudiar los modelos de acoplamiento molecular “*docking*” ligando-receptor de la EhTIM con los compuestos inactivantes y analizar las interacciones en el sitio de unión del ligando.
- Estudiar los modelos de acoplamiento molecular de los compuestos inactivantes con la TIM de humano (HsTIM) y comparar las energías libres de unión encontradas con las energías libres de unión de los mismos compuestos con la EhTIM.
- A partir de los modelos de acoplamiento molecular de los compuestos inactivantes, diseñar nuevas moléculas potencialmente inactivantes de la EhTIM y determinar su selectividad respecto a la HsTIM.

# Capítulo 3

## Metodología

### 3.1. QSAR-3D

#### 3.1.1. Modelado de las moléculas

La construcción y el modelado de todas las estructuras (Ver Tabla 4.1, Sección 4.1), se llevó a cabo usando la plataforma de Sybyl 6.8 (Sybyl). Una vez construidas, se sometieron al campo de fuerza MM de TRIPOS (Clarck *et al*, 1989) para optimizar la geometría. Una segunda optimización geométrica se hizo utilizando el método semiempírico PM3. Este último permitió calcular las cargas atómicas de Mopac, Mulliken, y Gasteiger-Marsili a todas las moléculas (Gasteiger y Marsili, 1980; Purcel y Singer, 1967; Mulliken, 1955). Posteriormente, todas las coordenadas de las moléculas se exportaron a Spartan'02 (Wavefunction) y se optimizó la geometría utilizando cálculos Hartree-Fock con funciones base 6-31G\*. Finalmente, se calculó la energía para cada molécula usando el método híbrido de funcionales de la densidad (DFT) B3LYP con funciones base 6-31G\*. Este último permitió calcular tres nuevos tipos de cargas atómicas: naturales, Mulliken y las derivadas del potencial electrostático (Reed *et al*, 1985; Stewart, 1989).

#### 3.1.2. Metodología de CoMFA

En este trabajo las moléculas se alinearon sobre los átomos de carbono del anillo de bencimidazol, asumiendo que los hidrógenos y los grupos metilo unidos a nitrógeno uno (N1) ocupan la misma posición en el sitio de unión del receptor. Se utilizó una rejilla expandida 4 Å en todas las direcciones hacia afuera de las moléculas alineadas. La rejilla se dividió en puntos de 2 Å de longitud, y sobre cada punto se evaluaron los campos estéricos y electrostáticos (descriptores). Los campos se calcularon mediante los potenciales de Lennard-Jones y de Coulomb usando un carbono con hibridación  $sp^3$  y una carga positiva ( $C^+$ ) respectivamente (Ecuaciones 1.1 y 1.2). Se empleó un filtro de 2.0 kcal/mol para eliminar aquellos puntos con poca influencia sobre la actividad biológica.

Para encontrar la correlación entre los campos estéricos y electrostáticos con la actividad biológica se empleó el método de los mínimos cuadrados parciales (*Partial Least Squares*, PLS) (Wold *et al*, 1998).

La actividad biológica, como la variable dependiente, se expresó en la forma  $pIC_{50}$  ( $pIC_{50} = -\log IC_{50}$ ; donde  $IC_{50}$  es la concentración molar que inhibe el crecimiento del 50% de la población). Los campos estéricos y electrostáticos se tomaron como variables independientes. La predictibilidad de los modelos se evaluó por validación cruzada, usando el método de dejar uno afuera (LOO). El valor de  $q^2$  para la predictibilidad se calculó como sigue:  $q^2 = (SSY - PRESS)/SSY$ , donde  $SSY$  es la variancia de la actividad biológica alrededor del valor promedio, y  $PRESS$  corresponde a la suma de los cuadrados de los errores de la predicción. Los modelos se calcularon usando el mínimo valor de componentes principales. De igual forma se calculó el valor del coeficiente de determinación  $r^2$  para cada modelo.

El resultado de este análisis fue una ecuación con cientos de coeficientes, la cual se representa gráficamente en forma de campos o contornos estéricos y electrostáticos que son relevantes para la actividad biológica.

### 3.1.3. Validación de los modelos de CoMFA

En este estudio se usaron los criterios de validación recientemente reportados por Golbraikh y Tropsha (Golbraikh y Tropsha, 2002; Golbraikh y Tropsha, 2002a). La metodología usada para la aplicación de estos criterios se resume a continuación:

Para un grupo de prueba, mínimo 5 compuestos, se graficó la actividad biológica experimental ( $y$ ) contra la actividad biológica predicha ( $\hat{y}$ ), de la misma forma se graficó ( $\hat{y}$ ) contra ( $y$ ). Seguidamente se obtuvo la estadística asociada a la regresión lineal de estas gráficas. En un modelo ideal, la recta representada por la ecuación de una recta de mínimos cuadrados (Ecuación 3.1) biseca el ángulo formado por la dirección positiva de los ejes ortogonales  $y$  y  $\hat{y}$ , la pendiente ( $m$ ) es igual a 1.0, la ordenada al origen ( $b$ ) es igual a cero y el coeficiente de correlación  $R$  es igual a 1.0.

$$y = m\hat{y} + b \quad (3.1),$$

Cuando un modelo real está cercano a un modelo ideal, éste es capaz de predecir apropiadamente y con exactitud la actividad biológica de un grupo de moléculas o grupo de prueba. Lo anterior es cierto cuando el coeficiente de determinación  $R^2$  de las rectas  $\hat{y}$  contra  $y$  y  $y$  contra  $\hat{y}$  es cercano a 1.0; los coeficientes de determinación ( $R_o^2$  y  $R_o'^2$ ) de las rectas  $\hat{y}$  contra  $y$  y  $y$  contra  $\hat{y}$  que pasan por el origen ( $y^o = k\hat{y}$  y  $\hat{y}^o = k'y$ ) son cercanas a  $R^2$

(donde  $k$  y  $k'$  son las correspondientes pendientes) y al menos una de las pendientes  $k$  o  $k'$  debe ser cercana a 1.0. Las pendientes  $k$  y  $k'$  están caracterizadas por las ecuaciones 3.2 y 3.3.

$$k = \frac{\sum y\bar{y}}{\sum \bar{y}^2} \quad (3.2)$$

$$k' = \frac{\sum y\bar{y}}{\sum y^2} \quad (3.3)$$

Parece redundante graficar  $(\bar{y})$  versus  $(y)$  y  $(y)$  versus  $(\bar{y})$ , al igual que sus rectas al origen, pero un estudio más detallado puede demostrar que estas gráficas están caracterizadas por diferente estadística. En resumen, para que un modelo tenga capacidad de predicción debe cumplir con los siguientes criterios:

- $q^2 > 0.5$
- $R^2 > 0.6$
- $((R^2 - R_0^2)/R^2) < 0.1$  o  $((R'^2 - R_0'^2)/R'^2) < 0.1$
- $0.85 \leq k \leq 1.15$  o  $0.85 \leq k' \leq 1.15$

## 3.2. QSAR-2D

### 3.2.1. Desarrollo de modelos $k$ NN-QSAR

Todas las estructuras (ver Tabla 4.1, sección 4.1) se generaron usando el constructor de moléculas del programa SYBYL 6.8 (Sybyl). Los índices de conectividad molecular (topológicos 2D y topográficos 3D (Randic, 1975; Kier y Hall, 1976; Hall y Kier, 1991)) o descriptores usados en este estudio se calcularon usando el programa MolConnZ, (MolConn-Z). En total se usaron 247 descriptores. Los valores de los descriptores se normalizaron (escalaron) de tal manera que todos los valores se encuentran entre 0 y 1. La selección de los diferentes grupos de entrenamiento y grupos de prueba se hizo usando el algoritmo de exclusión de esferas, de tal manera que cubre todo el espacio de los descriptores en todos los compuestos (Golbraikh y Tropsha, 2002).

### 3.2.2. Algoritmo general del método *k*NN-QSAR

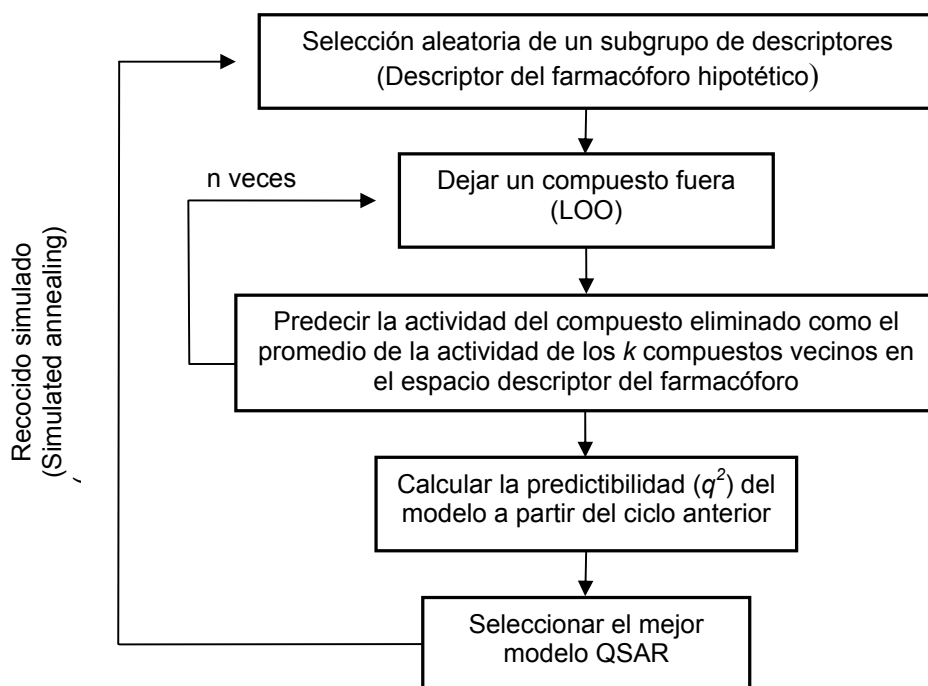
El método *k*NN-QSAR usa el principio de clasificación *el k vecino más cercano* (en inglés *kNN*, *k Nearest Neighbor*) (Sharaf *et al*, 1986) y combinado con un método de selección de variables. El método *k*NN se basa en la aproximación del análogo activo, esto implica que compuestos similares tienen un perfil de actividad biológica similar.

La actividad biológica de un compuesto se predijo como el promedio de la actividad de sus *k* (número de compuestos cercanos en el espacio descriptor del farmacóforo, *k* es un número entre 1 y 5) compuestos más similares de la base de datos. Brevemente se describe a continuación: (i) un subgrupo con un número de determinado de descriptores se seleccionó aleatoriamente; este subgrupo constituyó un descriptor del farmacóforo hipotético (**DHF**). En la selección de las variables o descriptores se utilizó el método de recocido simulado (simulated annealing) y un criterio de aceptación tipo metrópolis (Zheng y Tropsha, 2000). Los descriptores del subgrupo se modificaron hasta obtener el mejor valor posible de  $q^2$ . (ii) El DHF se validó por el método de dejar uno afuera (*Leave-One-Out*, **LOO**), donde cada compuesto se eliminó del grupo de entrenamiento y el valor de su actividad biológica se calculó como el promedio ponderado de sus *k* vecinos más cercanos en el espacio multidimensional del descriptor.

La similitud entre un par de compuestos se determinó por la distancia en el espacio multidimensional del descriptor; esta distancia es llamada distancia euclidiana. Una metodología más detallada puede verse en el trabajo de reportado por Zheng (Zheng y Tropsha, 2000). La metodología se resume en la Figura 3.1.

La robustez de los modelos QSAR generados se examinó por comparación de los valores de  $q^2$  de los datos experimentales con los valores de  $q^2$  de los mismos grupos de datos pero ahora con los valores de la actividad biológica asignados aleatoriamente.

La significatividad estadística de los modelos QSAR se evaluó usando la prueba de hipótesis de una cola (Gilbert, 1976).



**Figura 3.1.** Diagrama de flujo del método  $k$ NN-QSAR.

Para la significatividad estadística se formularon dos alternativas de hipótesis de la siguiente manera: (i)  $H_0: h = \mu$ , (ii)  $H_1: h > \mu$ , donde  $\mu$  es el valor promedio para los valores de  $q^2$  de las actividades asignadas aleatoriamente y  $h$  es el valor de  $q^2$  para la base de datos original. Así, una hipótesis nula  $H_0$  indica que el modelo QSAR desarrollado no es significativamente mejor que el modelo QSAR obtenido a partir de las actividades aleatorias.  $H_1$ , asume lo contrario. La decisión se tomó de acuerdo al siguiente procedimiento:

1. Se determinaron los valores de  $h$  ( $q^2$  de la base de datos real) y el valor promedio de  $\mu$  ( $q^2$  promedio de la base de datos con las actividades biológicas aleatorias) con su desviación estándar ( $\sigma$ ).
2. Se calculó el valor de  $Z$  correspondiente a los valores de  $q^2$  de la base de datos real:  $Z = (h - \mu) / \sigma$
3. Se comparó el valor de  $Z$  con el valor de  $Z_c$  tabulado (Tabla 3.1) a diferentes niveles de significatividad ( $\alpha$ ) (Gilbert, 1976). Así, si el valor de  $Z$  era mayor que el valor de  $Z_c$ , se podía concluir a que nivel de significatividad se aceptaba  $H_1$  o se rechazaba.

**Tabla 3.1.** Valores de  $\alpha$  y sus correspondientes  $Z_c$  para la prueba de una sola cola

$\alpha$	$Z_c$
0.10	1.28
0.05	1.64
0.01	2.33
0.001	3.10

En este estudio se tomó como criterio de aceptación para un modelo  $k$ NN-QSAR, aquellos que tuvieran  $q^2 > 0.5$  para el grupo de entrenamiento y  $R^2 > 0.6$  para el grupo de prueba.

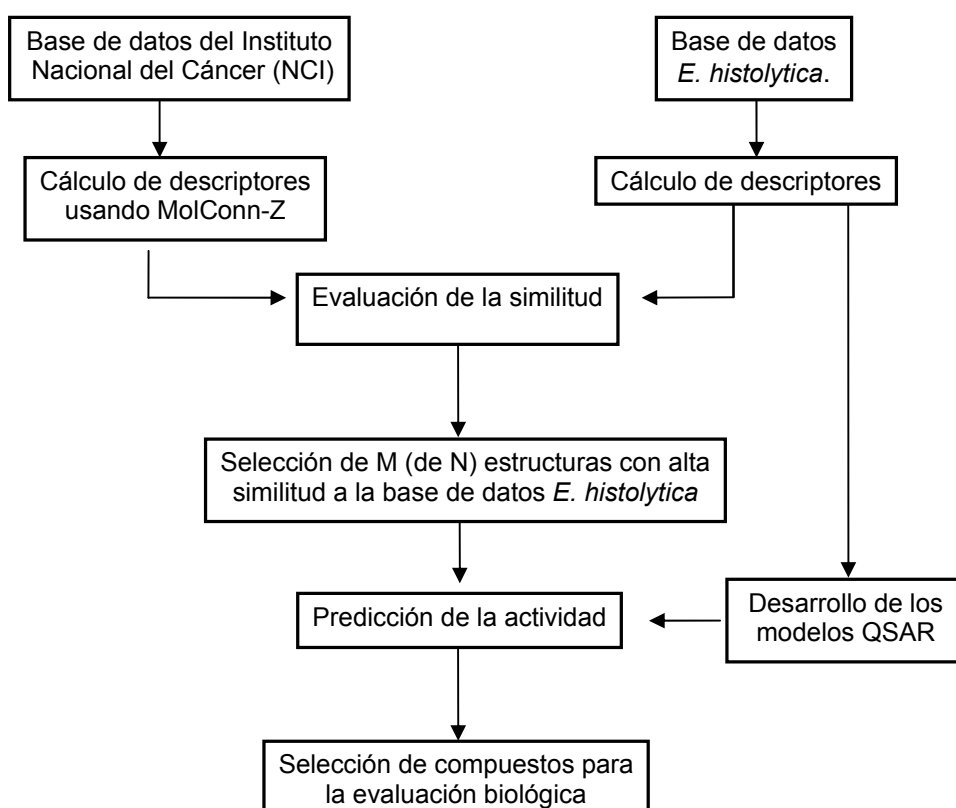
### 3.2.3. Búsqueda en la base de datos del Instituto Nacional del Cáncer (NCI) de los Estados Unidos

En esta parte del trabajo se utilizaron los modelos  $k$ NN-QSAR desarrollados y validados en la sección 3.1.1. El procedimiento para la búsqueda en bases de datos incluye varios pasos y el uso de diversos programas. Brevemente se describen a continuación.

1. Se definieron los DHF (Zheng y Tropsha, 2000) y los dominios de aplicabilidad de los modelos desarrollados como en la sección 3.2.1.
2. Se calcularon los descriptores para todos los compuestos de la base de datos del NCI. Los descriptores calculados son los mismos usados para el desarrollo de los modelos  $k$ NN-QSAR de la sección 3.1.1. Los descriptores se calcularon usando el programa MolConn-Z (MolConn-Z). Los descriptores para esta base de datos habían sido calculados previamente en el Laboratorio de modelado molecular de la Universidad de Carolina del Norte en Chapel Hill.
3. La similitud química se calculó de acuerdo a los valores de las distancias Euclidianas en el espacio multidimensional del descriptor del farmacóforo hipotético, entre los compuestos usados para el desarrollo de los modelos y cada uno de los compuestos de la base de datos del NCI.
4. Las estructuras de la base de datos se ordenaron de acuerdo a su similitud con los compuestos usados para desarrollar los modelos  $k$ NN-QSAR. Se seleccionaron un número determinado  $M$  de  $N$  compuestos de la base de datos NCI de acuerdo a valores de similitud preestablecidos.

5. La actividad biológica de los M compuestos seleccionados se predijo a partir de los modelos *k*NN-QSAR desarrollados usando un dominio de aplicabilidad.
6. Finalmente se seleccionaron los compuestos predichos por la mayoría de los modelos *k*NN-QSAR y con una alta predicción de actividad biológica. Estos compuestos constituyen compuestos líderes potenciales.

El procedimiento explicado anteriormente para la búsqueda en bases de datos usando los modelos *k*NN-QSAR se resume en la Figura 3.2.



**Figura 3.2.** Diagrama de flujo para la búsqueda de compuestos activos en la base de datos del NCI usando los modelos *k*NN-QSAR.

El mismo procedimiento que se utilizó para la búsqueda de compuestos activos en la base de datos del NCI se usó para hacer la predicción de la actividad biológica de 138 compuestos derivados de bencimidazol y benzoxazol que se han sintetizado en el Laboratorio 122 del Departamento de Farmacia, de la Facultad de Química de la UNAM. La actividad biológica contra *E. histolytica* de estos compuestos aún no se ha evaluado.



Nota: todos los programas usados para desarrollar los modelos kNN-QSAR y los utilizados para la búsqueda en la base de datos del NCI fueron desarrollados por el grupo de investigación del Laboratorio de Modelado Molecular de la Universidad de Carolina del Norte en Chapel Hill, bajo la dirección del Dr. Alexander Tropsha.

### **3.3. Purificación, inactivación y estudios de acoplamiento molecular de la triosafosfato isomerasa de *E. histolytica*.**

#### **3.3.1. Purificación de la enzima recombinante triosafosfato isomerasa de *E. histolytica*.**

La purificación de la enzima se llevó a cabo en 5 etapas, una etapa por día. A continuación se menciona brevemente el procedimiento; sin embargo, detalles de todo el proceso de purificación pueden verse en el trabajo reportado por Landa (Landa *et al*, 1997):

1. **Día 1. Revitalización de las células en glicerol.** Las células de *E. coli* tomadas de glicerol a  $-70^{\circ}\text{C}$  se sembraron en medio sólido Luria Broth con ampicilina. Las células se dejaron crecer durante 24 horas para obtener suficientes colonias. El cultivo obtenido constituye el preinóculo.
2. **Día 2. Crecimiento bacteriano.** Se inoculó un matraz con 100 mL de medio líquido Luria Broth, se agitó por dos horas a  $37^{\circ}\text{C}$  y luego se inocularon con este último 2 matraces de 1L del mismo medio y se dejó crecer con agitación durante 2.5 horas a  $37^{\circ}\text{C}$ . El medio se indujo con IPTG (Isopropil-tio- $\beta$ -D-galactopiranosido) cuando el cultivo tuvo una absorbancia entre 0.6-0.8 a 512 nm y se dejó crecer por 12 horas más a  $37^{\circ}\text{C}$ .
3. **Día 3. Tratamiento de las células.** Las células se centrifugaron por 10 minutos a 5,000 rpm a  $4^{\circ}\text{C}$ . Se decantó y se pesó el precipitado. Seguidamente, las células se resuspendieron en buffer TRIS HCl 20 mM a pH = 8.0 en DMSO. Las células suspendidas se rompieron por ultrasonificación durante dos minutos y se centrifugó por 10 minutos a 10,000 rpm a  $4^{\circ}\text{C}$ . El sobrenadante se ultracentrifugó por una hora a 45,000 rpm a  $4^{\circ}\text{C}$  y se cuantificó el volumen. El sobrenadante se precipitó en baño de hielo al llevarlo a 60% de saturación con sulfato de amonio y se dejó reposar por dos horas. Posteriormente, se centrifugó por 15 minutos a 12,000 rpm, se cuantificó el sobrenadante y se volvió a precipitar en iguales condiciones hasta el 90% de sulfato de amonio. Finalmente, se dejó en reposo durante una noche.

4. **Día 4. Separación de proteínas por cromatografía de exclusión molecular.** La fracción resultante del 60-90 % saturación se centrifugó por 15 minutos a 12000 rpm y 4°C. El precipitado de la centrifugación se resuspendió en 1 mL de TED\* 10/1/1 (pH = 7.4) y se dializó durante 6 horas contra 500 mL cambiando la solución a las tres primeras horas. La columna de Sephacryl se equilibró con 300 mL de la misma solución. La solución dializada se centrifugó por 15 min a 12,000 rpm y se cargó el FPLC con el sobrenadante. Se midió concentración de proteína a 280 nm y la actividad de las fracciones obtenidas y en ocasiones se hizo una electroforesis capilar en gel de SDS-PAGE. Las fracciones con mayor cantidad de proteína y con mayor actividad se reunieron.
5. **Día 5. Separación de proteínas por cromatografía de intercambio aniónico.** La columna Mono Q se equilibró con TED10/1/1 y TED 10/1/1 (1M de NaCl); las dos soluciones a pH 7.4. Se inyectó la columna con las fracciones ricas en proteína y se eluyó con un gradiente de 0-100 mM de NaCl. Se midió la concentración de la proteína a 280 nm y se corrió un gel de SDS-PAGE de las fracciones obtenidas. Las fracciones más activas y puras se reunieron. Finalmente, una vez que la proteína estuvo pura se dializó contra TED 100/10/1.

\*TED Corresponde a una solución amortiguadora preparada con trietanolamina, EDTA y ditioneitol. Los valores 10/1/1 y 100/10/1 corresponden a concentraciones milimolar respectivamente.

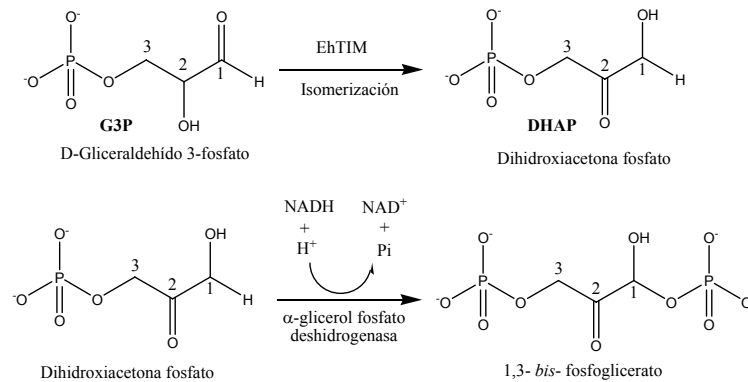
La concentración de la EhTIM se determinó por UV-VIS. Se midió la absorbancia a 280 nm. La absorptividad molar de la EhTIM usada fue de  $36,440 \text{ M}^{-1}\text{cm}^{-1}$  y su peso molecular de 27969D.

### **3.3.2. Determinación de la actividad enzimática de EhTIM**

La actividad de la EhTIM se determinó en sentido de la conversión de gliceraldehído 3-fosfato (G3P) a dihidroxiacetona fosfato (DHAP). Para el estudio de inactivación, 1  $\mu\text{g}/\text{mL}$  de EhTIM se incubó en presencia del compuesto problema a una concentración de 250  $\mu\text{M}$  por dos horas a 37°C en una solución buffer de 100 mM de trietanolamina y 10 mM de EDTA a pH 7.4 y 10% de dimetil sulfóxido (DMSO). Después de este período, una alícuota (suficiente para que la concentración de EhTIM en la celda del espectrofotómetro o de reacción fuera de 2 ng/mL) de la mezcla de incubación se adicionó a una celda espectrofotométrica de 1 mL. La celda de reacción contenía previamente 1.5 mM de G3P (El G3P se preparó a partir del dietilacetal de acuerdo a las

recomendaciones de SIGMA), 0.2 mM de NADH, 20 µg/mL de α-glicerol fosfato deshidrogenasa en una solución amortiguadora de 100 mM de trietanolamina y 10 mM de EDTA a pH 7.4. El volumen final en la celda fue de 1.0 mL. Posteriormente, en un espectrofotómetro UV-VIS se siguió la desaparición de NADH a una longitud de onda de 340 nm. En la celda de reacción o celda espectrofotométrica ocurren dos reacciones simultáneas; la transformación de **G3P** en **DHAP** y ésta última en 1,3-*bis*-fosfoglicerato (Figura 3.3).

La determinación de los parámetros cinéticos de velocidad máxima ( $V_{max}$ ) y la constante de Michaelis-Menten ( $K_m$ ) usan el mismo protocolo para la determinación de la actividad enzimática, pero en este caso se incubó sin el agente inactivante y la determinación se hizo a temperatura ambiente.



**Figura 3.3.** Reacciones de transformación del sustrato G3P.

### 3.3.3. Acoplamiento molecular (*docking*) de los compuestos inactivantes con la EhTIM

Todas las moléculas (ligandos) (Ver Tablas 6.1 y 6.6), tanto los encontrados experimentalmente como los diseñados, fueron construidos en SYBYL 6.8. Para optimizar la geometría de todos los ligandos se usó el campo de fuerza de TRIPOS (Clark *et al*, 1989).

Una vez construidos los ligandos se adicionaron las cargas de Gasteiger-Marsili (Gasteiger y Marsili, 1980). Los átomos de hidrógeno no polares se eliminaron usando el programa *AutoDock Tools* (ADT), sumando su carga parcial al átomo de carbono al cual estaba unido. Para la búsqueda del espacio conformacional accesible del ligando a la proteína, se permitió la rotación libre de todos los enlaces sencillos con excepción del enlace nitrógeno-carbono (N-CO) de amida.

La estructura tridimensional de la EhTIM se obtuvo de la base de datos del *Protein Data Bank* ([www.rcsb.org/pdb/](http://www.rcsb.org/pdb/)) bajo el código 1M6J. Todos los átomos de oxígeno, que corresponden a moléculas de agua, se removieron. Se adicionaron todos los átomos de hidrogeno polares y las cargas de Kollman (Weiner *et al*, 1984). Los volúmenes fragmentales de la proteína y los parámetros de solvatación atómica se adicionaron usando ADT.

Para hacer la evaluación energética de los diferentes modos de unión encontrados para un ligando en la proteína, el programa *AutoDock* usa el programa auxiliar *AutoGrid*. Este programa utiliza una rejilla para calcular los mapas de energía, los cuales usa posteriormente para hacer las evaluaciones energéticas entre el ligando y la proteína. En este trabajo se usó una rejilla (arreglo tridimensional de puntos) con un espaciamiento regular (0.375Å) y se situó en el posible sitio de unión de la proteína. En cada punto de la rejilla, el programa calculó el valor de la energía potencial entre un átomo de prueba y los átomos de la proteína que se encuentran a una distancia de no enlace y que están dentro de la rejilla. Los átomos de prueba que se usaron son los mismos tipos de átomos que constituyen a los ligandos. Para calcular las interacciones electrostáticas, *AutoDock* utilizó un potencial coulombico entre los átomos de la proteína y una carga de prueba e (+1.60x10<sup>-19</sup> C).

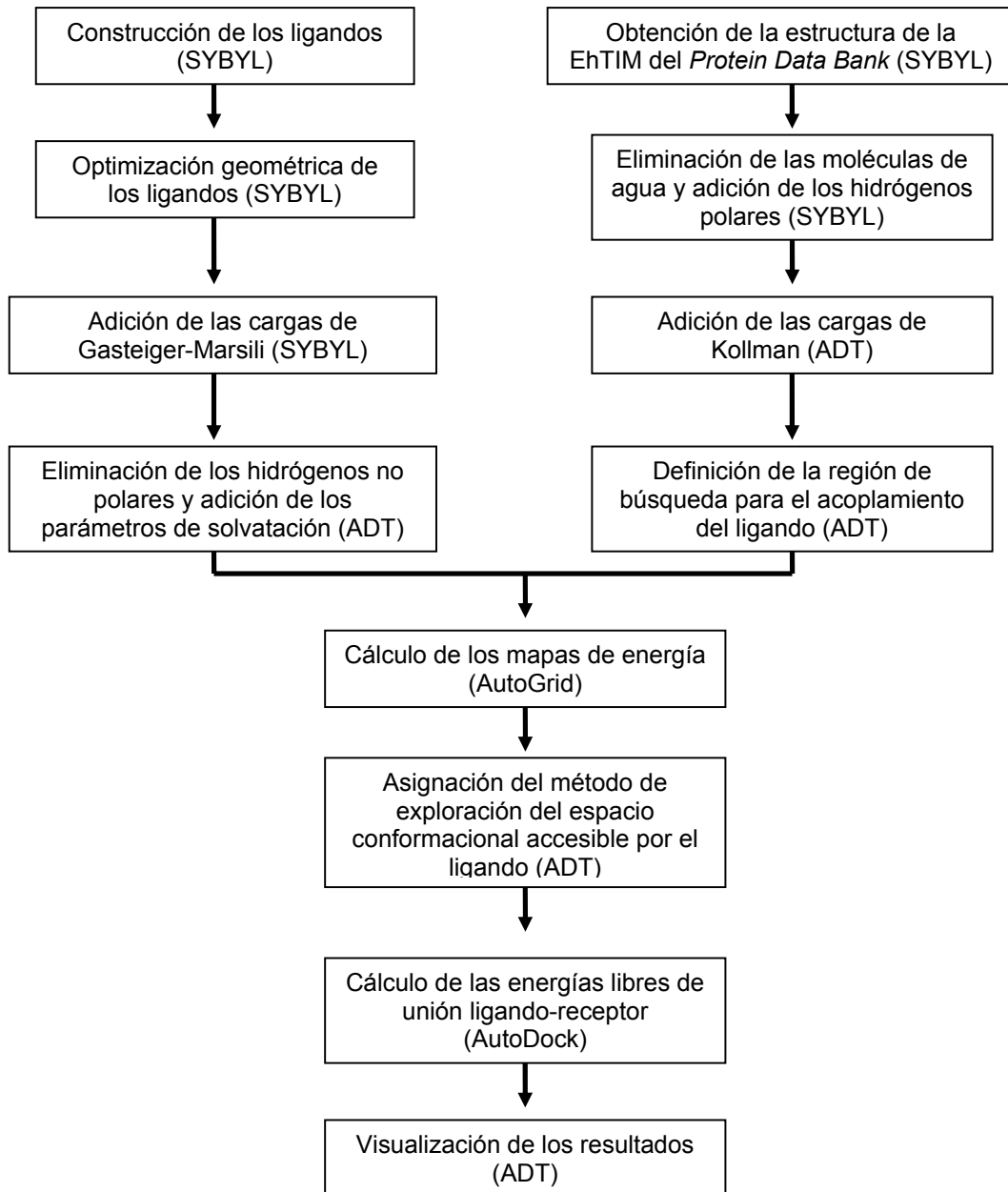
La búsqueda del espacio conformacional accesible por el ligando se realizó con el algoritmo genético implementado en *AutoDock 3.0* y los parámetros asociados a este método fueron los dados por defecto en el programa.

En cada ligando evaluado se encontraron cien soluciones con diferentes posiciones y orientaciones, las cuales fueron organizadas en grupos o *clusters*. Cada grupo contenía las soluciones para las cuales la desviación (*RMSD*, *Root Mean Square Deviations*) era menor a 1.0 Å y en cada grupo las soluciones se ordenaron en orden creciente de energía.

Posteriormente, al mejor modelo de acoplamiento ligando-EhTIM fue sometido al campo de fuerza de TRIPOS para relajar la estructura del complejo. Antes de la relajación, todos los hidrógenos fueron adicionados tanto a la proteína como al ligando (polares y no polares) y las cargas de Kollman se recalcularon sobre todos los átomos de la proteína. A los ligandos se les recalcularon las cargas de Gasteiger-Marsili sobre todos los átomos. En la relajación se utilizó una constante dieléctrica  $\epsilon = 4$  y una función dieléctrica dependiente de la distancia. Estas condiciones proveen un considerable ahorro

computacional relativo al costo-beneficio en el tratamiento de sistemas con moléculas de agua no explícitas (Christensen y Jorgensen, 1997).

El procedimiento para el estudio de acoplamiento molecular se resume en la Figura 3.4.



**Figura 3.4.** Esquema de las etapas para el estudio de acoplamiento molecular de los ligandos con la EhTIM usando AutoDock 3.0 y SYBYL 6.8.

# **RESULTADOS Y DISCUSIÓN**

# PARTE I

## Capítulo 4

# Relaciones cuantitativas estructura actividad en tres dimensiones (QSAR-3D) de compuestos derivados del bencimidazol con actividad amebicida: Un análisis comparativo de campos moleculares (CoMFA)

### 4.1. Desarrollo de los modelos CoMFA

Como se mencionó en la sección 1.3.1, la estrategia a seguir en el DiFAC depende del tipo de información experimental que se tenga a mano. Cualquier método QSAR que se desee utilizar para estudiar las relaciones estructura-actividad debe de contar con una serie de compuestos a los cuales se les conoce su actividad biológica. Los métodos QSAR son particularmente útiles cuando no se conoce la estructura del blanco o receptor asociado a un desorden biológico y ésta es la base de la estrategia usada en el presente capítulo.

Con el objetivo de estudiar los requerimientos estructurales que deben tener los derivados del bencimidazol para la actividad amebicida, se han sintetizado y reportado en la literatura algunos estudios de síntesis y evaluación biológica contra el protozoo *E. histolytica* (Valdez *et al*, 2002; Navarrete-Vázquez *et al*, 2001; Sánchez, 1999; Andrzejewska *et al*, 2002). Los compuestos reportados junto con sus valores de actividad biológica constituyen una excelente información para realizar estudios QSAR.

Existen pocos ejemplos reportados en la literatura donde se empleen métodos QSAR-3D para estudiar los requerimientos estructurales que deben tener los bencimidazoles para determinada actividad biológica (López-Rodríguez *et al*, 2001; López-Rodríguez *et al*, 2002; Vaz *et al*, 1997, Temiz-Arpaci *et al*, 2005; Bhongade y Gadad, 2004). Sin embargo, hasta el momento no se han reportado estudios QSAR-3D para correlacionar la estructura de los bencimidazoles con la actividad amebicida.

En este estudio se usaron los derivados de bencimidazol de la Tabla 4.1, incluidos los compuestos reportados por Andrzejewska (Andrzejewska *et al*, 2002). Una característica particular de algunos de estos compuestos es que presentan formas tautoméricas. Los derivados de los 1*H*-bencimidazoles, particularmente cuando tienen un hidrógeno en la posición 1, pueden presentar formas tautoméricas cuando poseen



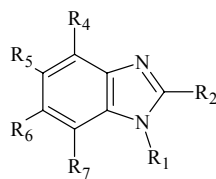
sustitución en el anillo de benceno o cuando poseen un grupo tiol unido a la posición C-2 del anillo. Los equilibrios tautoméricos para los compuestos **3**, **8**, **13**, **29**, **32**, **36**, y **40** se presentan en la Figura 4.1. Los equilibrios tautoméricos para los compuestos **6**, **7**, **9**, **10**, **17**, **21** y **23** se presentan en la Figura 4.2.

Los compuestos con formas tautoméricas son a menudo descartados en los estudios QSAR debido a que cada forma tautomérica en equilibrio representa una estructura con diferentes propiedades. Uno de los tautómeros en el equilibrio puede tener más afinidad por el receptor que el otro; de tal forma que la actividad biológica puede ser atribuida preferencialmente a una sola de las formas tautoméricas.

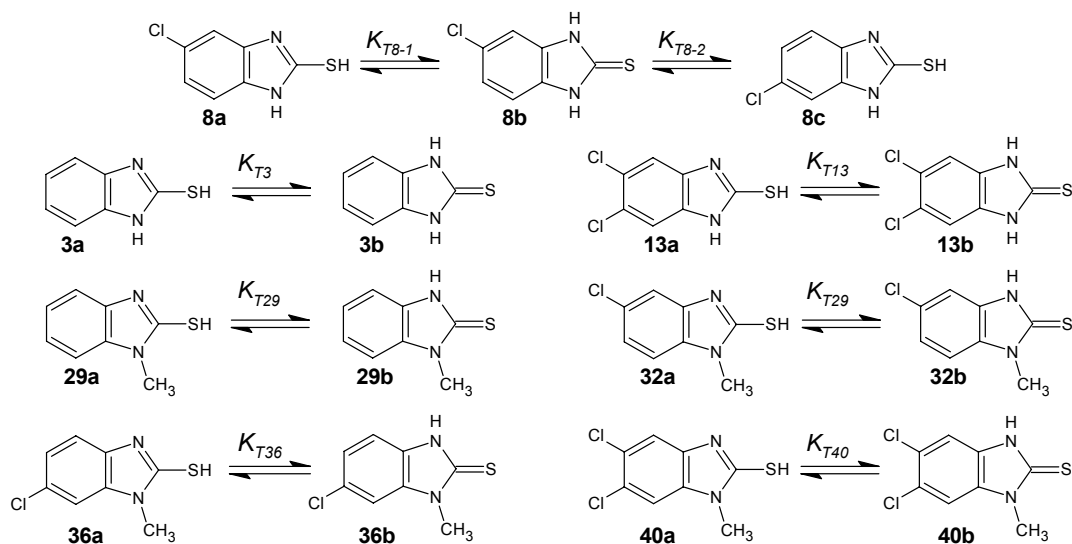
Es importante señalar que los compuestos que presentan tautómeros, Figuras 4.1 y 4.2, tienen propiedades químicas semejantes; sin embargo, cada tautómero tiene diferente complementariedad estérica y electrostática hacia un receptor biológico. Así, la elección errónea de una forma tautomérica dará como resultado modelos QSAR con una baja correlación o engañosos. La elección del tautómero más adecuado no es simple. Los equilibrios tautoméricos pueden ser influenciados por múltiples factores ambientales, tales como el disolvente, pH, temperatura y enlaces intra e intermoleculares. En el caso específico de la interacción ligando-receptor, el equilibrio puede estar desplazado en el sentido del tautómero que establezca energéticamente el complejo ligando-receptor. Es importante resaltar que el tautómero de más baja energía no es necesariamente el tautómero bioactivo (Pospisil *et al*, 2003).

Uno de los retos en este trabajo fue resolver el problema de la tautomería para obtener modelos QSAR con una mayor capacidad de predicción de las actividades biológicas. Esto quiere decir: tratar de encontrar cuál de las formas tautoméricas en los equilibrios es relevante para la actividad biológica. Para tratar de resolver esta cuestión se siguieron dos propuestas metodológicas.

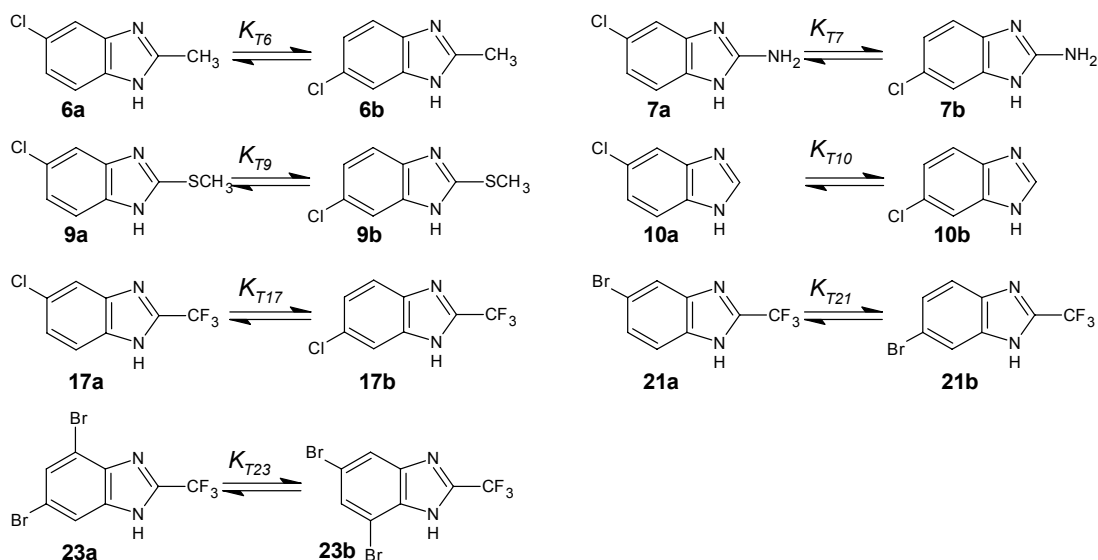
**Tabla 4.1.** Compuestos derivados del bencimidazol usados para los estudios QSAR



Compuesto	R <sub>1</sub>	R <sub>2</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	pIC <sub>50</sub>
1	H	CH <sub>3</sub>	H	H	H	H	8.15
2	H	NH <sub>2</sub>	H	H	H	H	6.94
3	H	SH	H	H	H	H	6.88
4	H	SCH <sub>3</sub>	H	H	H	H	6.41
5	H	H	H	H	H	H	7.38
6	H	CH <sub>3</sub>	H	Cl	H	H	7.08
7	H	NH <sub>2</sub>	H	Cl	H	H	6.90
8	H	SH	H	Cl	H	H	8.30
9	H	SCH <sub>3</sub>	H	Cl	H	H	6.72
10	H	H	H	Cl	H	H	7.41
11	H	CH <sub>3</sub>	H	Cl	Cl	H	7.60
12	H	NH <sub>2</sub>	H	Cl	Cl	H	7.23
13	H	SH	H	Cl	Cl	H	7.26
14	H	SCH <sub>3</sub>	H	Cl	Cl	H	6.45
15	H	H	H	Cl	Cl	H	7.02
16	H	CF <sub>3</sub>	H	H	H	H	7.16
17	H	CF <sub>3</sub>	H	Cl	H	H	7.66
18	CH <sub>3</sub>	CF <sub>3</sub>	H	H	H	H	7.40
19	CH <sub>3</sub>	CF <sub>3</sub>	H	Cl	H	H	7.34
20	CH <sub>3</sub>	CF <sub>3</sub>	H	Cl	Cl	H	7.48
21	H	CF <sub>3</sub>	H	Br	H	H	5.88
22	H	CF <sub>3</sub>	H	Br	Br	H	6.37
23	H	CF <sub>3</sub>	Br	H	Br	H	7.30
24	H	CF <sub>3</sub>	Br	Br	Br	Br	6.52
25	H	C <sub>2</sub> F <sub>5</sub>	H	Cl	Cl	H	6.68
26	H	CF <sub>3</sub>	H	NO <sub>2</sub>	NO <sub>2</sub>	H	5.87
27	H	C <sub>2</sub> F <sub>5</sub>	Br	Br	Br	Br	6.67
28	CH <sub>3</sub>	CH <sub>3</sub>	H	H	H	H	7.21
29	CH <sub>3</sub>	SH	H	H	H	H	8.22
30	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	H	H	6.71
31	CH <sub>3</sub>	CH <sub>3</sub>	H	Cl	H	H	8.30
32	CH <sub>3</sub>	SH	H	Cl	H	H	8.00
33	CH <sub>3</sub>	SCH <sub>3</sub>	H	Cl	H	H	7.43
34	CH <sub>3</sub>	CH <sub>3</sub>	H	H	Cl	H	6.69
35	CH <sub>3</sub>	NH <sub>2</sub>	H	H	Cl	H	6.61
36	CH <sub>3</sub>	SH	H	H	Cl	H	5.42
37	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	Cl	H	6.24
38	CH <sub>3</sub>	CH <sub>3</sub>	H	Cl	Cl	H	7.89
39	CH <sub>3</sub>	NH <sub>2</sub>	H	Cl	Cl	H	7.43
40	CH <sub>3</sub>	SH	H	Cl	Cl	H	6.89
41	CH <sub>3</sub>	SCH <sub>3</sub>	H	Cl	Cl	H	5.98



**Figura 4.1.** Equilibrio tautomérico de los derivados de 2-tiona bencimidazoles.



**Figura 4.2.** Equilibrio tautomérico de los derivados de los 1-*H* bencimidazoles con hidrógeno en la posición 1.

#### 4.1.1. Selección de las formas tautoméricas relevantes para la actividad biológica

Un paso importante en los estudios de CoMFA es la alineación de las moléculas. En los casos en que existen moléculas con ciertos grados de libertad, generalmente se asume que la estructura de mínima energía corresponde a la conformación bioactiva, aunque esto no es totalmente cierto (Pospisil *et al*, 2003). Esta aproximación se aplicó a las formas tautoméricas en este trabajo. En una primera aproximación se consideró a los tautómeros de baja energía como las formas tautoméricas bioactivas. En una segunda aproximación las formas tautoméricas bioactivas fueron seleccionadas por un modelo de

CoMFA que se desarrolló teniendo como grupo de entrenamiento sólo los compuestos de la Tabla 4.1 que no tienen formas tautoméricas.

#### 4.1.1.1. Estudio de CoMFA basado en el tautómero de más baja energía

Para desarrollar el modelo de CoMFA basado en esta aproximación, se calculó la energía de todas las formas tautoméricas usando el método híbrido de funcionales B3LYP con funciones base 6-31G\*, como se indicó en la sección 3.1.1. Cualquier equilibrio tautomérico tiene asociado un cambio energético ( $\Delta G$ ). Se asumió que la diferencia en los valores de energía para las formas tautoméricas involucradas en el equilibrio corresponde al valor de  $\Delta G$ . En diversos trabajos esta aproximación ha mostrado ser aceptable (Wiench *et al*, 2002; Houben *et al*, 2004; Luyten *et al*, 1998; Anandan *et al*, 2004; Duarte *et al*, 1999; Civcir, 2000; Senthilkumar y Kolandaivel, 2003; Kwiatkoski y Leszczynski, 1993). La constante de equilibrio tautomérico  $K_T$  puede derivarse de la ecuación 4.1. En la Tabla 4.2 se pueden observar los resultados para los  $\Delta G$  y las energías para las formas tautoméricas de las Figuras 4.1 y 4.2.

$$\Delta G = -RT \cdot \ln K_T \Rightarrow K_T = e^{-(\Delta G / RT)} \quad (4.1)$$

De la Tabla 4.2 puede observarse que las formas tiona **3b**, **8b**, **13b**, **29b**, **32b**, **36b**, y **40b** son más estables que sus correspondientes tautómeros en la forma tiol. El valor más bajo de  $-pK_T$  es 98.0 correspondiente al compuesto **13**, Indicando que la forma tautomérica más estable es la forma tiona para estos compuestos. Lo anterior está de acuerdo con los datos de RMN-<sup>1</sup>H, en los cuales se observa que la forma predominante es la forma tiona (Valdez, 2002). Para los derivados de los 1H-bencimidazoles con hidrógeno en posición 1 el equilibrio está desplazado ligeramente hacia las formas tautoméricas **6b**, **7b**, **9b**, **10b**, **17b**, **21b**, y **23b**. En todos los casos el valor de  $\Delta G$  es menor de 1.45 kcal/mol ( $-pK_T = 10.6$ ) indicando que estos compuestos podrían encontrarse como mezcla de tautómeros hasta casi en un 50%. El costo energético para que un compuesto de estos adopte la forma tautomérica bioactiva es muy pequeña comparada con los derivados de los 2 tiona. El costo energético para el cambio de la forma tiona a la forma tiol es de 15.62 kcal/mol para el compuesto 29 ( $-pK_T = 114.50$ ).

**Tabla 4.2.** Energías calculadas usando DFT (B3LYP) y constantes de equilibrio tautomérico para los equilibrios tautoméricos en fase gas ( $\epsilon = 1$ ) a 298.15°K

Tautómeros	Energía (a.u)	$-\Delta G$ (kcal/mol)	$-\text{p}K_T^*$
<b>3a</b>	-778.053844	15.00	109.93
<b>3b</b>	-778.076873		
<b>6a</b>	-878.785931	0.09	0.66
<b>6b</b>	-878.786070		
<b>7a</b>	-894.822511	1.28	9.37
<b>7b</b>	-894.824475		
<b>8a</b>	-1237.649110	14.30	104.78
<b>8b</b>	-1237.671060		
<b>8c</b>	-1237.649280	14.18	103.97
<b>9a</b>	-1276.967820	0.01	0.09
<b>9b</b>	-1276.967840		
<b>10a</b>	-839.461439	0.13	0.93
<b>10b</b>	-839.461634		
<b>13a</b>	-1697.238750	13.38	98.05
<b>13b</b>	-1697.259290		
<b>17a</b>	-1176.494250	0.21	1.57
<b>17b</b>	-1176.494580		
<b>21a</b>	-3290.179390	0.14	1.00
<b>21b</b>	-3290.179600		
<b>23a</b>	-5863.456620	1.45	10.60
<b>23b</b>	-5863.458840		
<b>29a</b>	-817.366243	15.62	114.50
<b>29b</b>	-817.390229		
<b>32a</b>	-1276.961710	14.77	108.26
<b>32b</b>	-1276.984390		
<b>36a</b>	-1276.961880	14.66	107.45
<b>36b</b>	-1276.984390		
<b>40a</b>	-1736.551500	14.18	103.97
<b>40b</b>	-1736.573280		

\* $\text{p}K_T = -\log K_T$

Una vez que se determinó el tautómero de más baja energía, se desarrolló un modelo de CoMFA usando todos los compuestos que no tienen formas tautoméricas de la Tabla 4.1 y los tautómeros de baja energía de la Tabla 4.2. En total se usaron 41 compuestos para desarrollar el modelo. El análisis *PLS* del modelo 1, obtenido bajo esta aproximación, tiene un coeficiente de determinación de la validación cruzada  $q^2 = 0.292$  y un coeficiente de determinación no cruzado  $r^2 = 0.634$ . Estos resultados muestran que existe una correlación muy pobre entre la estructura de los compuestos del grupo de entrenamiento y sus actividades biológicas. Es decir, las estructuras seleccionadas no tienen una buena correlación con la actividad biológica. Los resultados de este modelo se resumen en la Tabla 4.3.

#### 4.1.1.2. Estudio CoMFA basado en las formas tautoméricas mejor predichas

En esta aproximación se desarrolló primero un modelo de CoMFA usando como grupo de entrenamiento 27 compuestos de la Tabla 4.1 que no poseen formas tautoméricas. El análisis *PLS* para este grupo de entrenamiento dio como resultado el modelo 2 con valores de  $q^2 = 0.455$  y  $r^2 = 0.794$ . La exclusión de un compuesto del grupo de entrenamiento, el cual no se predijo en forma adecuada por el modelo 2, da como resultado el modelo 3 con un ligero incremento en los valores de  $q^2$  y  $r^2$  (0.490 y 0.815 respectivamente). Así, el modelo 3 se usó para predecir la actividad biológica de todas las formas tautoméricas.

**Tabla 4.3.** Modelos de CoMFA desarrollados bajos las dos aproximaciones propuestas en las secciones 4.1.1.1 y 4.1.1.2

Estadística	Estudio CoMFA basado en el tautómero de más baja energía		Estudio CoMFA basado en las formas tautoméricas mejor predichas		
	Modelo 1	Modelo 2	Modelo 3	Modelo 4	
$q^{2a}$	0.292	0.455	0.490	0.707	
$r^{2b}$	0.634	0.794	0.815	0.868	
Contribución estérica	0.604	0.654	0.660	0.620	
Contribución electrostática	0.396	0.346	0.340	0.380	
$s^c$	0.450	0.320	0.310	0.273	
$F^d$	21.389	29.613	32.325	59.339	
$N^e$	3	3	3	4	
$n^f$	41	27	26	41	

<sup>a</sup>Coefficiente de determinación de la validación cruzada  $r^2$ . <sup>b</sup>Coefficiente de determinación no cruzado  $r^2$ . <sup>c</sup>Error estándar estimado. <sup>d</sup>Fracción de la varianza explicada vs la no explicada. <sup>e</sup>Número óptimo de componentes principales. <sup>f</sup>Número de compuestos en el grupo de entrenamiento.

Los resultados de las predicciones hechas por el modelo 3 se resumen en la Tabla 4.4. Las formas tautoméricas mejor predichas corresponden a las que tienen el valor residual más bajo y están marcadas con asterisco (\*). De la Tabla 4.4 se puede observar que las formas tautoméricas **3b**, **6b**, **7b**, **21b**, **23b**, **36b**, y **40b** corresponden a las formas tautoméricas de baja energía (ver Tabla 4.2) En el caso de las formas tiona **8b**, **13b**, **29b** y **32b**, las cuales corresponden a los tautómeros de baja energía, el modelo de CoMFA 3 predice que sus tautómeros correspondientes a la forma tiol (**8a**, **13a**, **29a** y **32a**) tienen un valor residual mucho más bajo. Estos últimos tienen un valor de  $pIC_{50}$  más alto que las

tionas **3b**, **36b** y **40b**. Lo anterior muestra que las formas tiol predichas tienen un valor alto de pIC<sub>50</sub>, mientras las formas tiona poseen un valor bajo de pIC<sub>50</sub>.

Se asumió que los tautómeros mejor predichos por el modelo 3 corresponden a las formas tautoméricas bioactivas. Una vez que se asumió lo anterior, un nuevo modelo de CoMFA, el modelo 4, se desarrolló pero ahora tomando como grupo de entrenamiento a los compuestos que no tienen formas tautoméricas más los tautómeros seleccionados por el modelo 3 como tautómeros bioactivos. En total se tomaron en cuenta 41 compuestos para desarrollar el modelo 4 y poder compararlo con el modelo 1. Los resultados estadísticos para el modelo 4 se resumen en la Tabla 4.3.

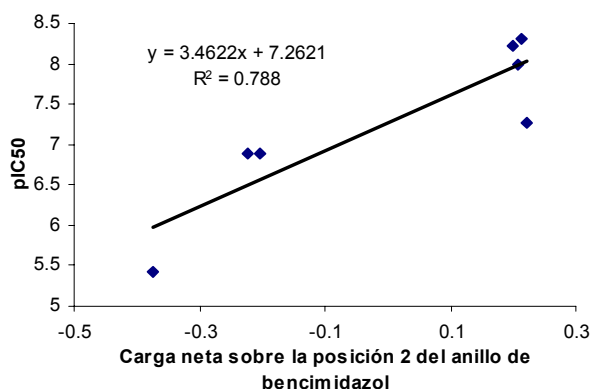
**Tabla 4.4.** Valores de pIC<sub>50</sub> experimentales y predichos hechos por el modelo de CoMFA 3 para todas las formas tautoméricas

Tautómero	pIC <sub>50</sub> Experimental	pIC <sub>50</sub> Predicho	Residual
<b>3a</b>		7.361	-0.485
<b>3b*</b>	6.876	6.786	0.090
<b>6a</b>		8.053	-0.977
<b>6b*</b>	7.076	6.978	0.098
<b>7a</b>		8.014	-1.111
<b>7b*</b>	6.903	7.046	-0.143
<b>8a*</b>		7.841	0.460
<b>8b</b>	8.301	6.771	1.530
<b>8c</b>		7.267	1.034
<b>9a*</b>		6.888	-0.171
<b>9b</b>	6.717	5.813	0.904
<b>10a*</b>		7.831	-0.422
<b>10b</b>	7.409	6.763	0.646
<b>13a*</b>		7.294	-0.034
<b>13b</b>	7.260	6.744	0.516
<b>17a*</b>		7.238	0.420
<b>17b</b>	7.658	6.144	1.514
<b>21a</b>		7.238	-1.362
<b>21b*</b>	5.876	6.116	-0.240
<b>23a</b>		6.278	1.023
<b>23b*</b>	7.301	7.097	0.204
<b>29a*</b>		7.500	0.722
<b>29b</b>	8.222	6.928	1.294
<b>32a*</b>		7.952	0.048
<b>32b</b>	8.000	7.411	0.589
<b>36a</b>		6.878	-1.455
<b>36b*</b>	5.423	5.898	-0.475
<b>40a</b>		7.440	-0.547
<b>40b*</b>	6.893	6.889	0.004

Los compuestos **3b**, **8a**, **13a**, **29a**, **32a**, **36b**, y **40b**, los cuales tienen una diferencia energética mayor en el equilibrio tautomérico (aprox. 14 kcal), presentan una interesante correlación de la carga neta del sustituyente en la posición C-2 del anillo de bencimidazol con la actividad biológica. Esto indica que hay una fuerte interacción de tipo electrostático

de los sustituyentes en la posición C-2 de los bencimidazoles. En la Figura 4.3 se puede observar que estos compuestos tienen una buena correlación con un valor de  $R^2 = 0.788$ . Se comprueba que los compuestos con una carga neta positiva sobre la posición C-2 del anillo de bencimidazol (**8a**, **13a**, **29a** y **32a**) tienen un valor alto de  $pIC_{50}$  (mayor que 7.25) y los compuestos con una carga neta negativa sobre esta posición (**3b**, **36b** y **40b**) tienen un valor bajo de  $pIC_{50}$  (menor que 6.89).

Tautómero	*Carga neta sobre la posición 2	$pIC_{50}$
<b>3T</b>	-0.224	6.880
<b>8</b>	0.213	8.300
<b>13</b>	0.221	7.260
<b>29</b>	0.199	8.220
<b>32</b>	0.208	8.000
<b>36T</b>	-0.374	5.420
<b>40T</b>	-0.204	6.890

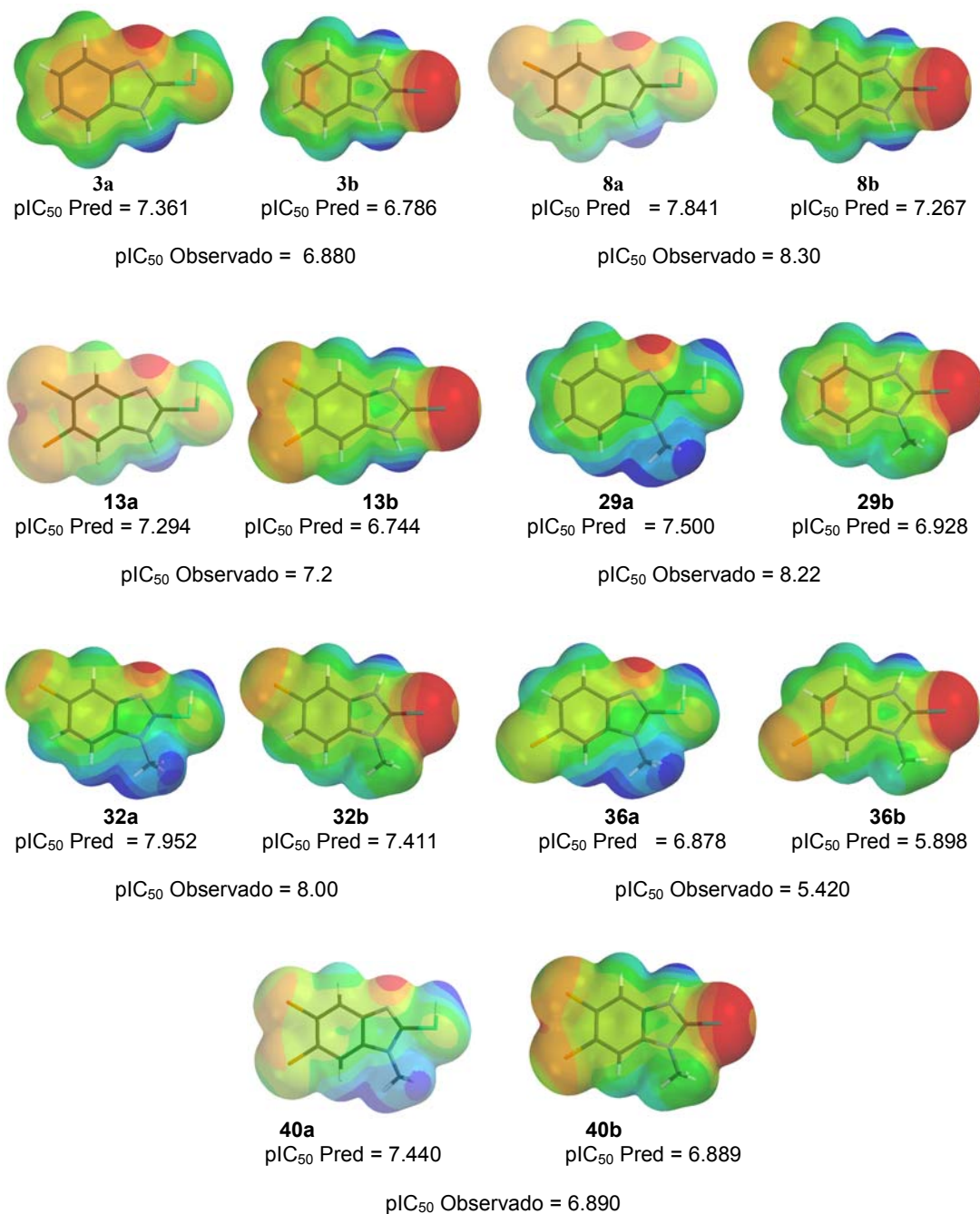


\* Carga natural obtenida de los cálculos DFT

**Figura 4.3.**  $pIC_{50}$  vs carga neta sobre la posición C-2 del anillo de bencimidazol para las tiones y tioles seleccionados por el modelo de CoMFA 3.

Para explorar la distribución electrónica de las formas tautoméricas seleccionadas por el modelo de CoMFA 3, particularmente los bencimidazoles 2-tiol y 2-tiona, se calcularon las superficies de potencial electrostático a un isovalor de densidad electrónica de 0.002. Esto se hizo a partir de los cálculos usando DFT. La Figura 4.4 muestra las superficies de potencial electrostático para las formas tautoméricas de las formas tiona y tiol. Las regiones con alta densidad electrónica se muestran en color rojo, las zonas con muy baja densidad electrónica se muestran en azul y las zonas con densidad electrónica intermedia se muestran en verde. Como es de esperarse, los bencimidazoles con el grupo tiona deben de tener alta densidad electrónica. De esta figura puede observarse que las formas tautoméricas predichas por el modelo de CoMFA 3 tienen diferentes densidades de carga en la posición C-2 del anillo de bencimidazol.





**Figura 4.4.** Superficies de potencial electrostático para las formas tautoméricas de los derivados de benzimidazol-2-tiol. Las regiones en rojo indican alta densidad electrónica. Las regiones en azul indican baja densidad electrónica. Las regiones en verde corresponden a regiones con densidad electrónica intermedia. Todas la superficies se obtuvieron a partir del cálculo de potencial electrostático obtenido por DFT (B3LYP/6-31G\*) Los valores de la actividad predicha pIC<sub>50</sub> se obtuvieron a partir del modelo de CoMFA 3 (Tabla 4.4).

Cuando el modelo de CoMFA predice una actividad alta (mayor que 7.2) la forma tautomérica más favorecida es aquella que tiene una baja densidad electrónica en la posición C-2 del anillo de benzimidazol (**8a**, **13a**, **29a**, y **32a**), es decir, la forma tiol es la favorecida. Cuando el modelo de CoMFA predice una baja actividad biológica (menor que

7.2) la forma tautomérica favorecida es aquella que tiene una alta densidad electrónica en la posición C-2 del anillo de bencimidazol (**3b**, **6b** y **40b**). De la Figura 4.4 se puede observar que en gran medida la actividad biológica puede estar determinada por la densidad electrónica en la posición C-2 del anillo de bencimidazol. En este caso, la actividad biológica se ve favorecida cuando en la posición C-2 del anillo de bencimidazol se disminuye la densidad electrónica.

#### 4.1.2. Desarrollo de modelos de CoMFA estadísticamente robustos

Una rutina normal en cualquier estudio QSAR consiste en dividir la base de datos en dos grupos de compuestos, uno de entrenamiento y uno de prueba. En este estudio la base de datos de 41 compuestos se dividió en un grupo de entrenamiento de 36 compuestos y un grupo de prueba de 5 compuestos, el cual se seleccionó aleatoriamente tratando de abarcar todo el espacio químico y biológico.

Una vez que se determinaron las posibles formas tautoméricas bioactivas, se usó el grupo de entrenamiento con 36 compuestos para desarrollar 6 modelos de CoMFA. Cabe mencionar que ahora el grupo de entrenamiento está constituido por los 23 compuestos que no tienen formas tautoméricas y los 13 tautómeros bioactivos determinados por el modelo de CoMFA 3. Los 6 modelos de CoMFA se desarrollaron usando diferentes tipos de cargas calculadas a partir del método semiempírico PM3 y usando funcionales de la densidad como se explica en la sesión 3.1. Los resultados para los 6 modelos de CoMFA desarrollados se resumen en la Tabla 4.5.

De la Tabla 4.5 se observa que todos los modelos tienen valores de  $q^2$  por encima de 0.6, con excepción del modelo obtenido a partir de las cargas de Mülliken usando PM3 (modelo 8). Todos los modelos tienen valores de  $r^2$  por encima de 8.0. Estas dos características los hacen, en principio, modelos estadísticamente robustos. El mejor modelo obtenido a partir de las cargas derivadas de cálculos DFT es el modelo obtenido usando las cargas naturales ( $q^2 = 0.720$  y  $r^2 = 0.873$ ) y el mejor modelo obtenido de los cálculos semiempíricos es el obtenido a partir de las cargas MOPAC ( $q^2 = 0.633$  y  $r^2 = 0.849$ ).

**Tabla 4.5.** Modelos de CoMFA obtenidos a diferentes niveles de teoría usando diversos tipos de cargas atómicas

	DFT B3LYP/6-31G*//HF/6-31G*			Semiempírico PM3/MMFF		
	Cargas			Cargas		
	Mulliken modelo 5	Potencial electrostático modelo 6	Naturales modelo 7	Mulliken modelo 8	MOPAC modelo 9	Gasteiger-Hückel modelo 10
$q^2$	0.686	0.633	0.720	0.595	0.633	0.666
$r^2$	0.855	0.849	0.873	0.846	0.849	0.891
$n$	36	36	36	36	36	36
$N$	3	4	4	4	4	5
$F$	63.544	43.580	53.080	42.487	43.696	48.824
$s$	0.283	0.295	0.212	0.298	0.294	0.255
Contribución estérica	0.580	0.567	0.662	0.679	0.708	0.600
Contribución electrostática	0.420	0.433	0.338	0.321	0.292	0.400
$R^{2g}$	0.954	0.943	0.951	0.922	0.963	0.982
$Ro^{2h}$	0.860	0.849	0.877	0.795	0.934	0.838
$k^i$	0.961	0.951	0.963	0.940	0.914	0.917
$(R^2-Ro^2)/R^2$	0.098	0.099	0.077	0.127	0.029	0.146
$Ro^{2j}$	0.692	0.662	0.754	0.443	0.901	0.571
$k^{k}$	1.037	1.048	1.035	1.058	1.092	1.085
$(R^2-Ro^2)/R^2$	0.274	0.298	0.277	0.519	0.064	0.417

$q^2$ ,  $r^2$ ,  $n$ ,  $N$ ,  $F$  y  $s$  tienen el mismo significado que en la Tabla 4.3. <sup>g</sup>Coefficiente de determinación  $R$  para la regresión de  $(\hat{y})$  vs  $(y)$  y  $(y)$  vs  $(\hat{y})$ . <sup>h,j</sup>Coefficiente de determinación  $R$  para la regresión de la recta que pasa por el origen. <sup>i,k</sup>Pendientes para las regresiones de las rectas que pasan por el origen.

Estadísticamente, los modelos cumplen con los criterios de aceptación para un modelo de CoMFA, es decir, tener un valor no menor de  $q^2 = 0.5$  y  $r^2 = 0.6$ .

#### 4.2. Validación de los modelos de CoMFA

Uno de los pasos más importantes en un estudio QSAR es la validación de los modelos desarrollados. En este estudio, todos los modelos se validaron usando un grupo de prueba de 5 compuestos seleccionados aleatoriamente.

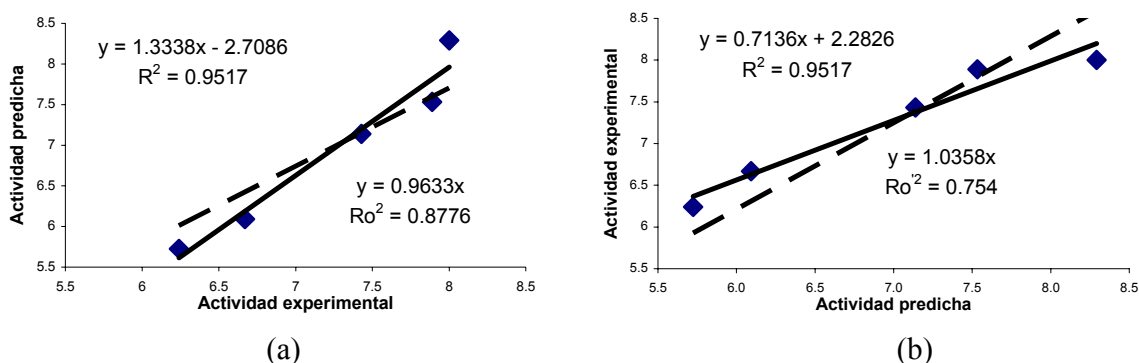
Uno de los conceptos más comúnmente usados como criterio para saber si un modelo es suficientemente robusto y tiene capacidad de predicción de las actividades biológicas, es la validación cruzada  $q^2$ . Se ha tomado el valor de  $q^2$  para definir si un modelo tiene capacidad de predicción o no. Un valor bajo de  $q^2$  indica que el modelo no es adecuado para predecir las actividades biológicas. Sin embargo, un valor alto de  $q^2$  no

necesariamente indica que el modelo tenga buena capacidad de predicción de las actividades biológicas. Por tal razón todos los modelos se validaron usando la metodología descrita en la sección 3.1.3. Los compuestos en el grupo de prueba fueron: **27**, **32a**, **33**, **37** y **38**. La predicción de la actividad biológica para este grupo de prueba, hechas por el modelo de CoMFA 7, se observan en la Tabla 4.6. El valor residual, es decir, la diferencia entre el valor experimental y el valor predicho de la actividad biológica más alta corresponde a 0.58 para el compuesto **27**.

**Tabla 4.6.** Predicciones del modelo de CoMFA 7 para el grupo de prueba

Compuesto	pIC <sub>50</sub> experimental	pIC <sub>50</sub> predicho	Residual
<b>27</b>	6.67	6.09	0.58
<b>32a</b>	8.00	8.29	-0.29
<b>33</b>	7.43	7.14	0.29
<b>37</b>	6.24	5.72	0.51
<b>38</b>	7.89	7.53	0.35

La Figura 4.5 muestra la validación para el modelo de CoMFA 7. De esta figura puede observarse que al graficar la actividad experimental contra la actividad predicha, y sus correspondientes regresiones al origen para el grupo de prueba, están todas ellas caracterizadas por una estadística diferente.



**Figura 4.5.** Gráficas de validación del modelo de CoMFA 7 (a) actividad predicha contra actividad experimental (b) actividad experimental contra actividad predicha.

La misma metodología de validación mostrada para el modelo de CoMFA 7 se aplicó a los modelos de CoMFA (6-10) y los resultados de su validación se resumen en la Tabla 4.5. Todos los modelos, con excepción de los modelos 8 y 10 para los cuales la relación  $((R^2 - R_o^2)/R^2)$  y  $((R'^2 - R_o'^2)/R'^2)$  es mayor a 0.1, cumplen los mínimos criterios de validación establecidos en la sección 3.1.3. A pesar de que el modelo de CoMFA 7 tiene el valor de  $q^2$  (0.720), su valor de  $R^2$  es más bajo que los correspondientes valores de  $R^2$

de los modelos 9 y 10. Esta falta de correlación entre los valores de  $q^2$  y  $R^2$  ya ha sido reportada (Golbraikh y Tropsha, 2002). A pesar de que el modelo 10 posee el valor más alto de  $R^2$  (0.982), este modelo no satisface los criterios mínimos de aceptación, ya que las relaciones  $((R^2-Ro^2)/R^2)$  y  $((R'^2-Ro'^2)/R'^2)$  son mayores que 0.1. Estas observaciones sugieren que es necesario aplicar los criterios de validación para saber si un modelo tiene o no una buena capacidad de predicción de la actividad biológica. Los valores altos en  $q^2$  y  $R^2$  no sugieren siempre un modelo con capacidad de predicción.

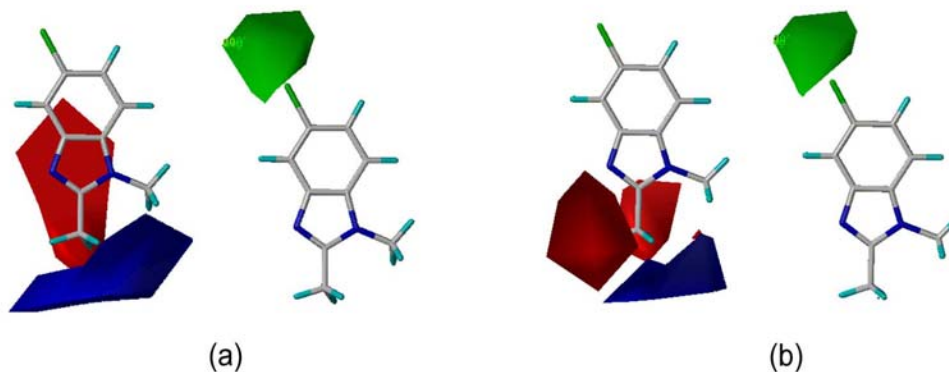
El mejor modelo obtenido a partir de DFT es el modelo 7 y el mejor modelo obtenido a partir del semiempírico PM3 es el modelo 9. Estos dos modelos son estadísticamente robustos y cumplen los criterios mínimos de validación para un grupo de prueba; cualquiera de estos dos modelos pueden ser usados para el diseño de nuevas moléculas potencialmente activas contra *E. histolytica*. Sin embargo, el modelo obtenido a partir de los cálculos de DFT resulta ser un poco impráctico en el sentido de que son computacionalmente costosos y resulta difícil la evaluación de bases de datos grandes. Así, el modelo de CoMFA 9 es una buena herramienta para diseñar nuevas moléculas activas, éste es estadísticamente robusto, cumple los criterios de validación para un grupo de prueba y resulta ser en tiempo de cómputo mucho más económico que el modelo 7.

En este trabajo se observó que hacer la evaluación de las cargas atómicas parciales usando un elevado nivel de cálculo, como lo es la teoría de funcionales de la densidad, no resultó en un mejoramiento notable para un modelo de CoMFA como en el caso reportado por Occhiato (Occhiato *et al*, 2004).

#### **4.3. Mapas de contorno de los modelos de CoMFA**

La gran ventaja que tiene la metodología de CoMFA sobre los métodos en dos dimensiones, especialmente el método de Hansch (Hansch *et al*, 1963; Hansch y Fujita, 1964), es que en CoMFA no se interpreta la ecuación de correlación entre la actividad biológica y los parámetros estéricos y electrostáticos. El resultado que se interpreta en un estudio CoMFA son los contornos estéricos y electrostáticos que tienen una fuerte influencia sobre la actividad biológica y éstos simulan las interacciones estéricas y electrostáticas en el sitio de unión del receptor. La Figura 4.6 muestra los contornos estéricos y electrostáticos para los modelos 7 y 9 obtenidos por DFT y PM3 respectivamente.

Por convención se tiene estipulado que los colores rojos y azules representan los contornos electrostáticos y los contornos verdes y amarillos representan los contornos estéricos.



**Figura 4.6.** Mapas de contorno estéricos y electrostáticos para los dos mejores modelos de CoMFA. (a) Modelo de CoMFA 7 obtenido a partir de DFT. (b) Modelo de CoMFA 9 obtenido a partir de PM3. El compuesto 31 está sobrepuesto para mejor entendimiento de los contornos.

De la Figura 4.6 puede observarse que los mapas de contorno obtenidos por métodos DFT y PM3 proveen la misma información estructural. Básicamente, el contorno verde dirigido hacia la posición C-5 del anillo de bencimidazol indica que el incremento estérico en esta posición favorece la actividad biológica y el contorno azul que se encuentra dirigido hacia la posición C-2 del anillo de bencimidazol indica que se debe disminuir la densidad electrónica en los sustituyentes en esta posición para favorecer a la actividad biológica. Esto último encuentra concordancia con las superficies de potencial electrostático de los bencimidazol 2-tiol y 2-tiona, los cuales indican que la densidad electrónica en la posición C-2 desfavorece la actividad biológica. Por tal motivo, las formas tiona activas tienen baja actividad biológica y contrariamente las formas tiol, las cuales tienen menor densidad electrónica en la posición C-2, son más activas.

Estas observaciones ayudan a explicar el comportamiento de la actividad biológica de algunos pares análogos: el 2-metilbencimidazol, **1** (8.15), es más activo que la forma 2-tiona predicha **3b** (6.88); la forma 2-tiol predicha **8a** (8.30) es más activa que el 2-metilbencimidazol **6** (7.08). Grupos electroattractores en la posición C-2 del anillo de bencimidazol desfavorecen la actividad biológica; esto puede verse al comparar el compuesto **1** con metilo en posición C-2 (8.15) y el compuesto **16** con un grupo trifluorometilo en la misma posición (7.16); el intercambio de un grupo metilo por un grupo trifluorometilo disminuye la actividad biológica casi en una unidad de  $pIC_{50}$ .

De igual manera, el contorno estérico verde en la posición C-5 del anillo de bencimidazol ayuda explicar el comportamiento de otros pares análogos y entender por qué los 5-clorobenzimidazoles son más activos que sus correspondientes análogos 5-H: así, los compuestos **9a** (6.72), **10a** (7.41), **31** (8.30) y **33** (7.43), con cloro en la posición C-5 son más activos que los compuestos con hidrógeno en esa posición **4** (6.41), **5** (7.38), **28** (7.21) y **30** (6.71), respectivamente. En el caso de los compuestos **6** y **7** las formas tautoméricas bioactivas predichas corresponden a los compuestos **6b** (7.08) y **7b** (6.90), los cuales tienen el cloro en la posición C-6 en lugar de la posición C-5; de esta forma se explica por qué son menos activos que sus análogos **11** (7.60) y **12** (7.23), respectivamente, los cuales tienen el cloro en la posición C-5.

#### 4.4. Conclusiones

Se empleó la metodología de CoMFA para establecer las relaciones cuantitativas estructura-actividad de una serie de derivados del bencimidazol. Uno de los resultados importantes en este estudio fue determinar las formas tautoméricas con un mayor impacto sobre la actividad antiamebiana. Este trabajo es el primer estudio que reporta la aplicación de un método QSAR para estudiar los requerimientos estructurales que deben tener los derivados de bencimidazol para mostrar actividad biológica contra *Entamoeba histolytica*, y el primero en usar un método QSAR para seleccionar las formas tautoméricas bioactivas.

Para determinar las formas tautoméricas más estables en fase gas se hicieron cálculos teóricos utilizando la teoría de funcionales de la densidad. Sin embargo, se encontró que no todas las formas tautoméricas energéticamente más estables conducen a modelos de CoMFA estadísticamente robustos. En un segundo intento se utilizó un modelo de CoMFA con capacidad de predicción para seleccionar las formas tautoméricas bioactivas. Los modelos obtenidos de este segundo ensayo son estadísticamente robustos, quiere decir, que las formas tautoméricas seleccionadas tienen una alta significatividad para la actividad biológica y posiblemente correspondan a las formas tautoméricas bioactivas. Una vez determinadas las formas tautoméricas con significatividad para la actividad biológica, se emplearon éstas, junto con las que no tienen tautómeros, y se desarrollaron 6 modelos de CoMFA. Para ello, se utilizaron diferentes tipos de carga de acuerdo a dos niveles de cálculo: semiempírico, con el método PM3 y funcionales de la densidad, con el método híbrido B3LYP y un conjunto de funciones base 6-31G\*. Los 6 modelos desarrollados se validaron rigurosamente de acuerdo a los

criterios de Golbraikh y Tropsha (Golbraikh y Tropsha, 2002). Con la información de la validación se determinó que el mejor modelo de CoMFA, obtenido por las funcionales de densidad, es el modelo que usa las cargas naturales; este modelo tuvo un valor de  $q^2 = 0.72$  para el grupo de entrenamiento y un valor de  $R^2 = 0.95$  para el grupo de prueba. El mejor modelo obtenido por el método semiempírico, es el que usa las cargas Mopac; este modelo tuvo un valor de  $q^2 = 0.633$  para el grupo de entrenamiento y un valor de  $R^2 = 0.96$  para el grupo de prueba. Estos dos modelos se pueden utilizar para el diseño racional de nuevas moléculas derivadas del bencimidazol con una actividad biológica potencialmente más alta que la actividad biológica de los derivados de bencimidazol de la Tabla 4.1. Por razones prácticas y de economía en tiempo de cómputo, el modelo de CoMFA de PM3 resulta ser más rápido para la predicción de la actividad biológica de nuevas moléculas derivadas de bencimidazol. Los 6 modelos de CoMFA coinciden en que dos posiciones del anillo de bencimidazol son relevantes para la actividad biológica contra *E. histolytica*; esto es, debe incrementarse el tamaño del sustituyente en la posición C-5 y disminuir la densidad electrónica de los sustituyentes de la posición C-2 del anillo de bencimidazol para favorecer la actividad biológica.



## **PARTE II**

## Capítulo 5

### Estudio de las relaciones cuantitativas estructura actividad en dos dimensiones (QSAR-2D) de compuestos derivados del bencimidazol con actividad amebicida y búsqueda de compuestos activos en bases de datos

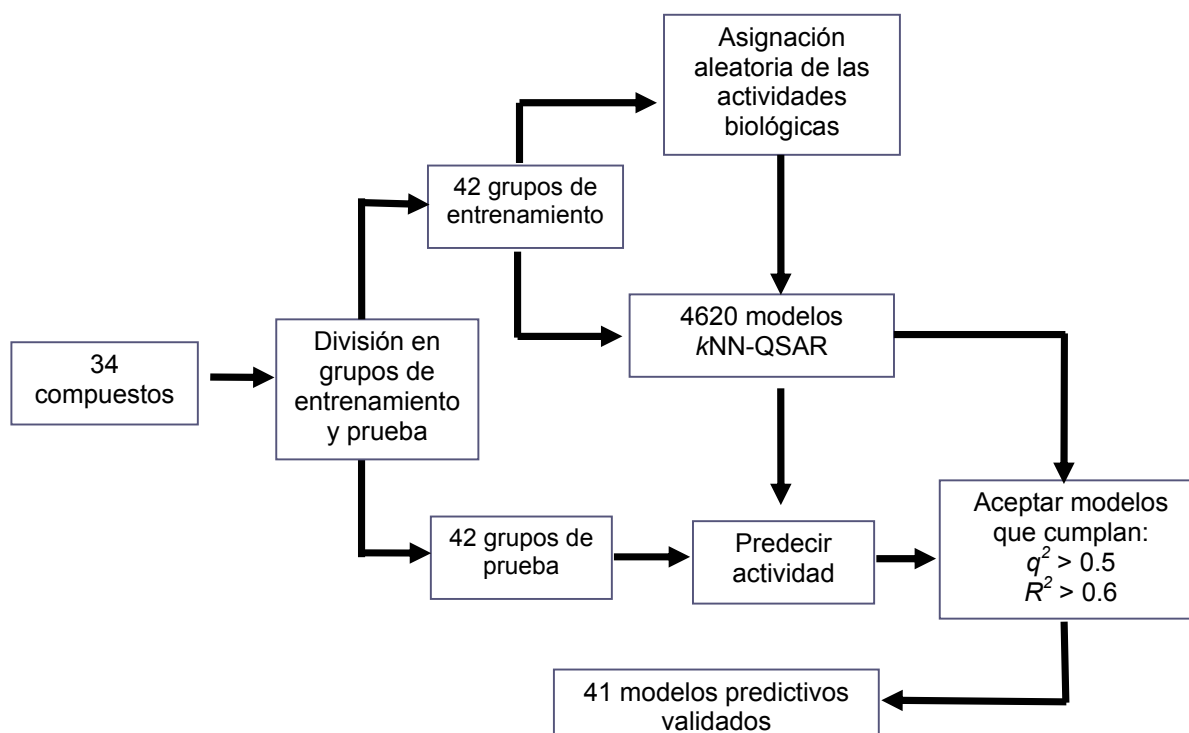
#### 5.1. Desarrollo de los modelos QSAR-2D

Usando los compuestos reportados por nuestro laboratorio (Valdez *et al*, 2002; Navarrete-Vázquez *et al*, 2001; Sánchez-Díaz, 1999) se desarrollaron modelos QSAR-2D, específicamente, usando el método *k*NN-QSAR.

Una vez que se desarrollaron y validaron los modelos, se usaron para hacer una búsqueda de compuestos activos en la base de datos del NCI y para predecir la actividad biológica de un grupo de 138 compuestos, los cuales se han sintetizado en nuestro laboratorio y a los que aún no se les ha determinado su actividad biológica contra *E. histolytica*.

##### 5.1.1. Algoritmo *k*NN-QSAR

En este estudio se usó una base de datos conformada por 34 de los 41 compuestos derivados de bencimidazol de la Tabla 4.1. La base de datos se dividió en 42 grupos de entrenamiento y 42 grupos de prueba. Los grupos de entrenamiento se emplearon para desarrollar los modelos *k*NN-QSAR de acuerdo a la metodología explicada en la sección 3.2. De esta forma se obtuvieron 4620 modelos *k*NN-QSAR, los cuales se validaron rigurosamente. La actividad biológica de los compuestos se asignó aleatoriamente para estudiar la fortaleza de los 4620 modelos obtenidos, de tal forma que sólo se aceptaron 41, los cuales cumplieron con tener un valor de  $q^2 > 0.5$  para la validación cruzada y de  $R^2 > 0.6$  para el grupo de prueba. El diagrama de flujo en la Figura 5.1 esquematiza el procedimiento.



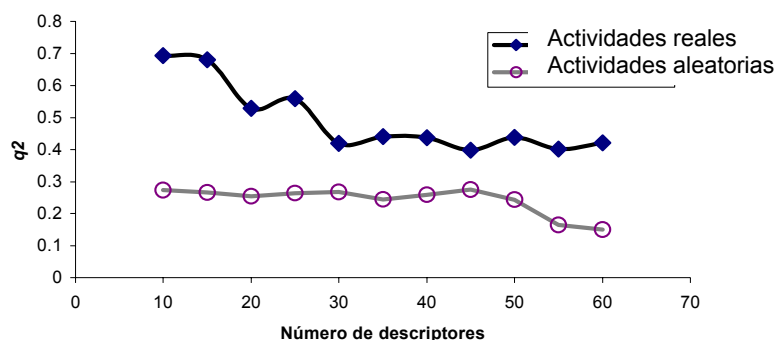
**Figura 5.1.** Diagrama de flujo para el desarrollo de los modelos *kNN-QSAR*.

### 5.1.2. Robustez de los modelos *kNN-QSAR*

La robustez de los modelos *kNN-QSAR* se estableció comparando los valores de  $q^2$  de los modelos desarrollados a partir de los datos experimentales de actividad biológica con los valores de  $q^2$  para los modelos QSAR encontrados después de asignar aleatoriamente las actividades biológicas. El valor de cada  $q^2$  corresponde al valor promedio de 10 cálculos independientes.

La Figura 5.2 muestra el gráfico de los valores de  $q^2$  versus el número de descriptores para los modelos obtenidos de los datos experimentales y los obtenidos de las actividades biológicas aleatorias. El número de descriptores (*nvar*) usados para generar los modelos fueron: 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 y 60. De esta forma puede concluirse si el modelo obtenido para la base de datos es mejor estadísticamente que el obtenido de las actividades biológicas aleatorias.

En la Gráfica 5.2 se observan diferencias entre los valores de  $q^2$  de la base de datos real, donde el rango para los valores de  $q^2$  varía entre 0.7 y 0.42, y los valores de  $q^2$  para la base de datos con las actividades aleatorias, que fluctúan entre 0.28 y 1.5. De la gráfica se puede observar que las diferencias más grandes de  $q^2$  corresponden a los modelos generados entre 10 y 25 descriptores. Esto quiere decir que los mejores modelos desarrollados son aquellos que tienen hasta 25 descriptores.



**Figura 5.2.** Gráfica de  $q^2$  para los modelos reales y los modelos con las actividades asignadas aleatoriamente vs número de descriptores (nvar).

El examen estadístico de estos resultados se describió por la prueba de hipótesis de una sola cola y se comparó con los valores tabulados de la Tabla 3.1. Para esto, se calculó el valor de  $q^2$  ( $\mu$ ) y el valor promedio de  $q^2$  ( $h$ ) de 10 cálculos independientes para las actividades biológicas asignadas al azar. Los resultados se observan en la Tabla 5.1.

**Tabla 5.1.** Prueba de hipótesis estándar de una sola cola para los modelos QSAR con diferentes números de descriptores (nvar)

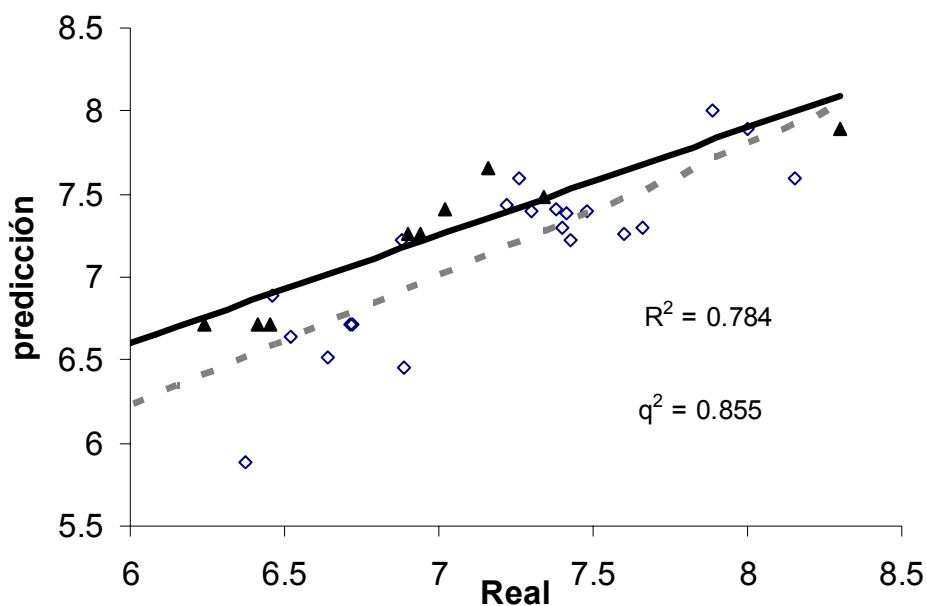
Nº de Descriptores	$h(q^2$ de los modelos reales)	$\mu$ (Prom. $q^2$ de 10 cálculos independientes)	$\sigma$	Z
10	0.693	0.274	0.039	10.897
15	0.680	0.267	0.055	7.497
20	0.529	0.255	0.064	4.262
25	0.559	0.264	0.072	4.101
30	0.419	0.268	0.050	3.056
35	0.441	0.245	0.088	2.216
40	0.437	0.259	0.067	2.653
45	0.398	0.275	0.063	1.939
50	0.438	0.243	0.069	2.823
55	0.402	0.164	0.069	3.434
60	0.421	0.150	0.058	4.681

En la mayoría de los casos el valor de Z es mayor que el valor de  $Z_c$  correspondiente a un  $\alpha = 0.001$ . Esto indica que la probabilidad de obtener un modelo QSAR, producto de la casualidad, es menor de 0.001.

### 5.1.3. Validación de los modelos *k*NN-QSAR

Es necesario asegurar que los modelos generados por el algoritmo de *k*NN pueden predecir adecuadamente la actividad biológica de un compuesto no incluido en el grupo de entrenamiento. En este trabajo se aceptaron modelos con valores mayores a 0.5 de  $q^2$  para el grupo de prueba y mayores a 0.6 de  $R^2$  para el grupo de entrenamiento. En la Tabla 5.2 se muestran los valores de  $q^2$  y  $R^2$  para los 41 modelos QSAR aceptados. Además, se observa que la mayoría de los modelos se generaron empleando entre 10 y 20 descriptores, y en la medida que disminuye el tamaño del grupo de entrenamiento, hay una disminución en el valor de  $R^2$ .

La Figura 5.3 muestra la gráfica de la actividad biológica predicha vs la actividad real, para el grupo de entrenamiento y para el grupo de prueba, utilizando el mejor de los 41 modelos aceptados.



**Figura 5.3.** Valores de la actividad real vs predicha para el grupo de entrenamiento (rombos) y grupo de prueba (triángulos) usando el modelo 3 de la Tabla 5.2.

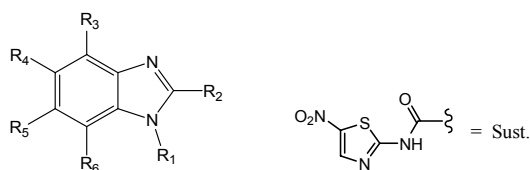
**Tabla 5.2.** 41 modelos *k*NN-QSAR desarrollados usando el algoritmo *k*NN

Modelo N°	Tamaño del grupo de entrenamiento (N° de compuestos)	Tamaño del grupo de prueba (N° de compuestos)	N° de Descriptores	$q^2$ (Grupo de entrenamiento)	$R^2$ (Grupo de prueba)
1	25	9	15	0.67	0.90
2	25	9	15	0.64	0.88
3	25	9	15	0.78	0.86
4	25	9	10	0.72	0.86
5	25	9	20	0.64	0.85
6	28	6	30	0.64	0.85
7	25	9	15	0.67	0.84
8	25	9	35	0.57	0.84
9	25	9	20	0.71	0.84
10	25	9	15	0.73	0.84
11	25	9	15	0.80	0.84
12	25	9	25	0.64	0.83
13	25	9	35	0.53	0.82
14	25	9	10	0.71	0.81
15	25	9	50	0.59	0.81
16	25	9	15	0.73	0.81
17	28	6	10	0.76	0.81
18	28	6	35	0.55	0.78
19	28	6	60	0.56	0.74
20	24	10	15	0.52	0.73
21	24	10	10	0.63	0.73
22	24	10	20	0.57	0.71
23	24	10	10	0.59	0.71
24	24	10	10	0.65	0.69
25	20	14	15	0.78	0.68
26	23	11	60	0.55	0.67
27	25	9	25	0.63	0.66
28	24	10	15	0.64	0.66
29	24	10	10	0.62	0.66
30	25	9	20	0.68	0.65
31	25	9	60	0.62	0.65
32	27	7	35	0.65	0.65
33	25	9	10	0.69	0.65
34	20	14	10	0.75	0.65
35	19	15	20	0.56	0.64
36	23	11	15	0.79	0.63
37	27	7	20	0.67	0.63
38	19	15	30	0.52	0.63
39	24	10	20	0.60	0.61
40	25	9	20	0.63	0.61
41	19	15	50	0.63	0.60

## 5.2. Predicción de la actividad amebicida para compuestos antiparasitarios sintetizados en el laboratorio L-122.

En nuestro grupo de investigación se han sintetizado una gran cantidad de derivados de bencimidazol y de benzoxazol cuya actividad biológica contra *E. histolytica* no se ha explorado. Los 41 modelos desarrollados se usaron para predecir la actividad biológica de estos derivados de bencimidazol y de benzoxazol. Los compuestos para los que se predijo la actividad biológica se muestran en la Tabla 5.3.

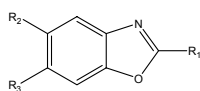
**Tabla 5.3.** Predicciones para diversos compuestos derivados de bencimidazol y benzoxazol hechas por los 41 modelos *k*NN-QSAR desarrollados



Comp	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub> pred	N° de mod
1	CH <sub>3</sub>	SCH <sub>3</sub>	H	Cl	1-piperidinilo	H	No pred.	0
2	CH <sub>3</sub>	SCH <sub>3</sub>	H	Cl	4-metilpiperidinilcarbonilo	H	No pred.	0
3	CH <sub>3</sub>	SCH <sub>3</sub>	H	1-piperidinilo	Cl	H	No pred.	0
4	CH <sub>3</sub>	SCH <sub>3</sub>	H	4-metilpiperidinilcarbonilo	Cl	H	No pred.	0
5	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	1-piperidinilo	H	No pred.	0
6	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	4-metilpiperidinilcarbonilo	H	No pred.	0
7	CH <sub>3</sub>	CF <sub>3</sub>	H	H	propiltio	H	No pred.	0
8	CH <sub>3</sub>	CF <sub>3</sub>	H	H	propilsulfinilo	H	7.34	1
9	CH <sub>3</sub>	CF <sub>3</sub>	H	propiltio	H	H	No pred.	0
10	CH <sub>3</sub>	CF <sub>3</sub>	H	propilsulfinilo	H	H	7.34	2
11	CH <sub>3</sub>	CF <sub>3</sub>	H	benzoilo	H	H	5.81	1
12	CH <sub>3</sub>	CF <sub>3</sub>	H	H	benzoilo	H	7.34	2
13	H	SCH <sub>3</sub>	H	H	CONH <sub>2</sub>	H	6.68	7
14	H	SCH <sub>3</sub>	H	H	CONHCH <sub>3</sub>	H	6.58	4
15	H	SCH <sub>3</sub>	H	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	6.28	10
16	H	SCH <sub>3</sub>	H	H	CONHC <sub>2</sub> H <sub>5</sub>	H	6.20	4
17	H	SCH <sub>3</sub>	H	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	5.84	5
18	H	SC <sub>2</sub> H <sub>4</sub> OH	Cl	H	Cl	H	6.69	9
19	H	SC <sub>2</sub> H <sub>4</sub> OH	Br	H	Br	H	6.59	7
20	H	SC <sub>2</sub> H <sub>4</sub> N(CH <sub>3</sub> ) <sub>2</sub>	Cl	H	Cl	H	5.87	24
21	H	SC <sub>2</sub> H <sub>4</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	Cl	H	Cl	H	5.88	11
22	H	SC <sub>2</sub> H <sub>4</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	Br	H	Br	H	5.87	14
23	H	SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	Cl	H	Cl	H	5.86	20
24	H	SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	Br	H	Br	H	5.86	23
25	H	SCH <sub>2</sub> CH <sub>2</sub> -(N-1-piperidinil)	Cl	H	Cl	H	5.84	2
26	H	SCH <sub>2</sub> CH <sub>2</sub> -(N-1-piperidinil)	Br	H	Br	H	5.84	2
27	H	SCH <sub>2</sub> CH <sub>2</sub> -(N-morfolinil)	Cl	H	Cl	H	5.84	1
28	H	SCH <sub>2</sub> CH <sub>2</sub> -(N-morfolinil)	Br	H	Br	H	5.84	1
29	H	SCH <sub>2</sub> CH <sub>2</sub> -(p-nitrofenil)	Cl	H	Cl	H	5.84	1
30	H	SCH <sub>2</sub> CH <sub>2</sub> -(p-nitrofenil)	Br	H	Br	H	5.84	1
31	H	CF <sub>3</sub>	H	CO <sub>2</sub> H	H	H	6.97	11
32	H	CF <sub>3</sub>	H	CO <sub>2</sub> CH <sub>3</sub>	H	H	6.8	16
33	H	CF <sub>3</sub>	H	CO <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	H	H	6.59	9
34	H	CF <sub>3</sub>	H	COCH <sub>2</sub> CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	H	No pred.	0
35	H	CF <sub>3</sub>	H	CONC <sub>5</sub> H <sub>10</sub>	H	H	7.48	1
36	H	CF <sub>3</sub>	H	propiltio	H	H	5.84	1
37	H	CF <sub>3</sub>	H	propilsulfinilo	H	H	6.96	7
38	H	CF <sub>3</sub>	H	feniltio	H	H	No pred.	0
39	H	CF <sub>3</sub>	H	fenilsulfinilo	H	H	6.74	2
40	H	CF <sub>3</sub>	H	benzoilo	H	H	6.31	3
41	H	CF <sub>3</sub>	H	CF <sub>3</sub>	H	H	5.80	18
42	H	CF <sub>3</sub>	H	2,3-diclorofenoxi	H	H	5.81	3
43	CH <sub>3</sub>	CF <sub>3</sub>	H	feniltio	H	H	No pred.	0
44	CH <sub>3</sub>	CF <sub>3</sub>	H	fenilsulfinilo	H	H	6.65	2
45	CH <sub>3</sub>	CF <sub>3</sub>	H	CF <sub>3</sub>	H	H	6.51	11
46	CH <sub>3</sub>	CF <sub>3</sub>	H	2,3-diclorofenoxi	H	H	5.81	1
47	CH <sub>3</sub>	CF <sub>3</sub>	H	H	feniltio	H	No pred.	0
48	CH <sub>3</sub>	CF <sub>3</sub>	H	H	fenilsulfinil	H	6.65	2
49	CH <sub>3</sub>	CF <sub>3</sub>	H	H	CF <sub>3</sub>	H	6.41	10
50	CH <sub>3</sub>	CF <sub>3</sub>	H	H	2,3-diclorofenoxi	H	5.81	1

Comp	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub> pred	N° de mod
51	CH <sub>3</sub>	CF <sub>3</sub>	H	CO <sub>2</sub> H	H	H	7.50	12
52	CH <sub>3</sub>	CF <sub>3</sub>	H	CO <sub>2</sub> CH <sub>3</sub>	H	H	7.25	15
53	CH <sub>3</sub>	CF <sub>3</sub>	H	C <sub>2</sub> H <sub>5</sub> CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	H	5.81	1
54	CH <sub>3</sub>	CF <sub>3</sub>	H	H	CO <sub>2</sub> H	H	7.46	12
55	CH <sub>3</sub>	CF <sub>3</sub>	H	H	CO <sub>2</sub> CH <sub>3</sub>	H	7.24	16
56	CH <sub>3</sub>	CF <sub>3</sub>	H	H	CO <sub>2</sub> CH <sub>2</sub> CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	5.81	1
57	CH <sub>3</sub>	CF <sub>3</sub>	H	CONHCH <sub>3</sub>	H	H	7.45	5
58	CH <sub>3</sub>	CF <sub>3</sub>	H	H	CONHCH <sub>3</sub>	H	7.42	5
59	CF <sub>3</sub>	CO <sub>2</sub> CH <sub>3</sub>	H	H	H	H	6.22	4
60	CF <sub>3</sub>	H	CO <sub>2</sub> CH <sub>3</sub>	H	H	H	6.31	3
61	CH <sub>3</sub>	COCH <sub>3</sub>	H	Cl	H	H	7.57	31
62	CH <sub>3</sub>	CO <sub>2</sub> H	H	Cl	H	H	7.21	16
63	CH <sub>3</sub>	CONH <sub>2</sub>	H	Cl	H	H	7.20	23
64	CH <sub>3</sub>	CONHCH <sub>3</sub>	H	Cl	H	H	7.29	10
65	CH <sub>3</sub>	CON(CH <sub>3</sub> ) <sub>2</sub>	H	Cl	H	H	7.20	14
66	CH <sub>3</sub>	CO <sub>2</sub> C <sub>2</sub> H <sub>5</sub>	H	Cl	H	H	7.26	14
67	CH <sub>3</sub>	COCH <sub>3</sub>	H	H	H	H	7.46	30
68	CH <sub>3</sub>	CO <sub>2</sub> H	H	H	H	H	6.99	14
69	CH <sub>3</sub>	CONH <sub>2</sub>	H	H	H	H	6.89	22
70	CH <sub>3</sub>	CONHCH <sub>3</sub>	H	H	H	H	7.13	11
71	CH <sub>3</sub>	CON(CH <sub>3</sub> ) <sub>2</sub>	H	H	H	H	7.21	19
72	CH <sub>3</sub>	CO <sub>2</sub> C <sub>2</sub> H <sub>5</sub>	H	H	H	H	7.24	14
73	CH <sub>3</sub>	SCH <sub>3</sub>	H	CONH <sub>2</sub>	H	H	7.25	7
74	CH <sub>3</sub>	SCH <sub>3</sub>	H	CONHCH <sub>3</sub>	H	H	6.71	1
75	CH <sub>3</sub>	SCH <sub>3</sub>	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	H	6.35	2
76	CH <sub>3</sub>	SCH <sub>3</sub>	H	CONHC <sub>2</sub> H <sub>5</sub>	H	H	No pred.	0
77	CH <sub>3</sub>	SCH <sub>3</sub>	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	H	5.84	1
78	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	CONH <sub>2</sub>	H	7.25	7
79	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	CONHCH <sub>3</sub>	H	6.48	2
80	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	6.35	2
81	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	CONHC <sub>2</sub> H <sub>5</sub>	H	No pred.	0
82	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	5.84	1
83	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	CONH <sub>2</sub>	H	7.66	2
84	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	CONHCH <sub>3</sub>	H	7.89	1
85	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	No pred.	0
86	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	CONHC <sub>2</sub> H <sub>5</sub>	H	7.89	1
87	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	No pred.	0
88	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	CONH <sub>2</sub>	H	H	7.66	2
89	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	CONHCH <sub>3</sub>	H	H	7.89	1
90	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	H	No pred.	0
91	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	CONHC <sub>2</sub> H <sub>5</sub>	H	H	7.89	1
92	CH <sub>3</sub>	NHCO <sub>2</sub> CH <sub>3</sub>	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	H	No pred.	0
93	H	CF <sub>3</sub>	H	CONH <sub>2</sub>	H	H	6.49	9
94	H	CF <sub>3</sub>	H	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	H	H	5.81	1
95	H	CF <sub>3</sub>	H	CONHC <sub>2</sub> H <sub>5</sub>	H	H	6.53	5
96	H	CF <sub>3</sub>	H	CON(CH <sub>3</sub> ) <sub>2</sub>	H	H	6.62	14
97	H	CF <sub>3</sub>	H	CONHCH <sub>3</sub>	H	H	6.90	5
98	H	CH <sub>3</sub>	H	Sust.	H	H	No pred.	0
99	CH <sub>3</sub>	CH <sub>3</sub>	H	Sust.	H	H	No pred.	0
100	CH <sub>3</sub>	CH <sub>3</sub>	H	H	Sust.	H	No pred.	0
101	H	CF <sub>3</sub>	H	Sust.	H	H	No pred.	0
102	H	SH	H	Sust.	H	H	No pred.	0
103	H	SCH <sub>3</sub>	H	Sust.	H	H	No pred.	0
104	CH <sub>3</sub>	CF <sub>3</sub>	H	Sust.	H	H	No pred.	0
105	CH <sub>3</sub>	SH	H	Sust.	H	H	No pred.	0
106	CH <sub>3</sub>	SCH <sub>3</sub>	H	Sust.	H	H	No pred.	0
107	CH <sub>3</sub>	CF <sub>3</sub>	H	H	Sust.	H	No pred.	0
108	CH <sub>3</sub>	SH	H	H	Sust.	H	No pred.	0
109	CH <sub>3</sub>	SCH <sub>3</sub>	H	H	Sust.	H	No pred.	0





Continuación Tabla 5.3 (Benzoxazoles)

Comp	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pIC <sub>50</sub> Pred	Nº de modelos
109	H	H	H	7.33	15
110	CH <sub>3</sub>	H	H	7.35	29
111	NH <sub>2</sub>	H	H	6.83	29
112	SH	H	H	6.83	30
113	CF <sub>3</sub>	H	H	6.81	22
114	SCH <sub>3</sub>	H	H	6.58	25
115	NHCO <sub>2</sub> CH <sub>3</sub>	H	H	6.74	5
116	OCO <sub>2</sub> CH <sub>3</sub>	H	H	6.57	8
117	H	Cl	H	6.96	20
118	CH <sub>3</sub>	Cl	H	7.14	26
119	CF <sub>3</sub>	Cl	H	6.93	27
120	SH	Cl	H	6.96	30
121	SCH <sub>3</sub>	Cl	H	6.59	23
122	NH <sub>2</sub>	Cl	H	7.01	30
123	NHCO <sub>2</sub> CH <sub>3</sub>	Cl	H	6.59	3
124	OCO <sub>2</sub> CH <sub>3</sub>	Cl	H	6.48	6
125	H	Cl	Cl	7.05	16
126	CH <sub>3</sub>	Cl	Cl	7.27	30
127	CF <sub>3</sub>	Cl	Cl	7.00	20
128	SH	Cl	Cl	6.93	28
129	SCH <sub>3</sub>	Cl	Cl	6.59	16
130	NH <sub>2</sub>	Cl	Cl	7.12	25
131	NHCO <sub>2</sub> CH <sub>3</sub>	Cl	Cl	6.89	1
133	H	H	Cl	7.13	15
134	CH <sub>3</sub>	H	Cl	7.14	26
135	CF <sub>3</sub>	H	Cl	6.84	22
136	SH	H	Cl	6.92	27
137	SCH <sub>3</sub>	H	Cl	6.63	19
138	NHCOCH <sub>3</sub>	H	Cl	6.17	2

Se debe tener en cuenta que la actividad se predice como el promedio de las actividades de los compuestos más cercanos en el espacio descriptor del farmacóforo. Es decir, si un modelo no puede predecir la actividad de un compuesto, es porque éste no está dentro del dominio de aplicación del modelo. De ninguna manera debe interpretarse que el compuesto, para el cual su actividad biológica no pudo ser predicha, no tiene actividad biológica; es sólo que el modelo es incapaz de predecirla. Esto se debe básicamente a que el compuesto no presentó una similitud clara con respecto a los compuestos usados para desarrollar los modelos (Tabla 4.1).

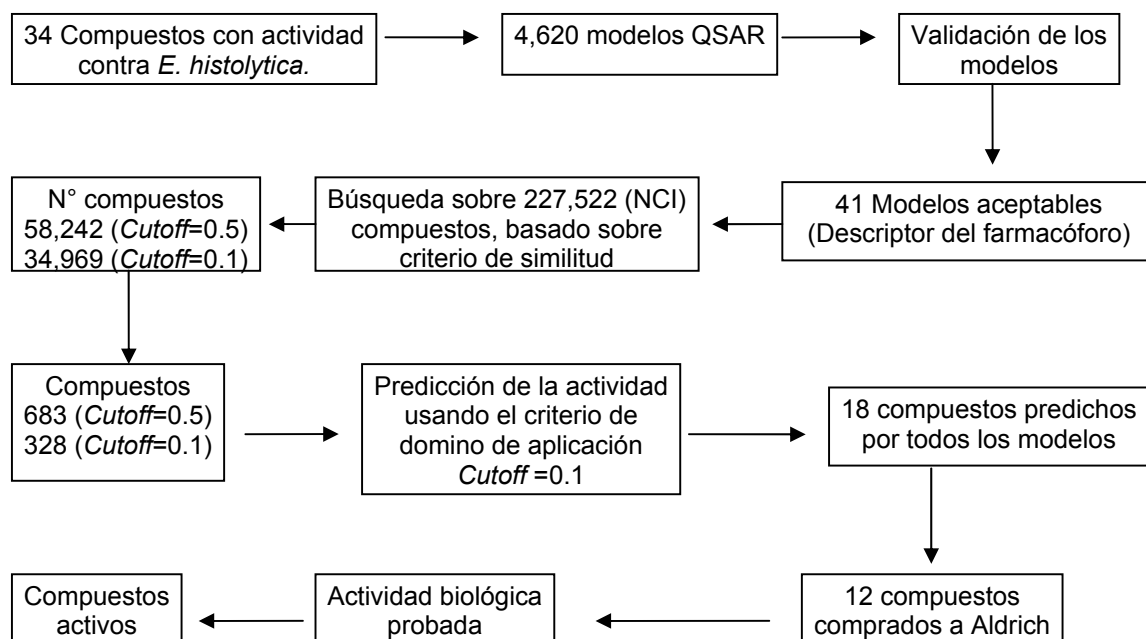
Entre mayor es el número de modelos que predicen la actividad biológica de un compuesto, mayor es la probabilidad de que la predicción sea la adecuada. En este caso

sólo se aceptaron las predicciones para los compuestos predichos por al menos el 75% (30 modelos) de los modelos y que tuvieron una desviación estándar menor a 0.3.

De la Tabla 5.3 se observa que muchos de los compuestos no tienen predicción para su actividad biológica y en otros casos la actividad se predice por menos del 75% de los modelos. El único compuesto que cumple los criterios de aceptación es el compuesto **122**; para el cual su  $pIC_{50}$  calculado es 7.01. El  $pIC_{50}$  para este compuesto se predice por 30 modelos con una desviación estándar de 0.23.

### 5.3. Búsqueda de compuestos activos en la base de datos del Instituto Nacional del Cáncer (NCI)

Para la búsqueda de compuestos activos en la base de datos NCI se usaron los 41 modelos desarrollados. En este estudio se usó un umbral (*Cutoff*) de 0.1 de distancia euclidiana como valor de similitud para la selección de las moléculas provenientes de la base de datos NCI. La Figura 5.4 resume la estrategia usada en la búsqueda.

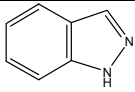
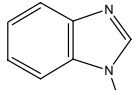
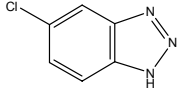
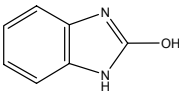
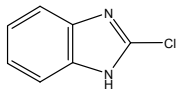
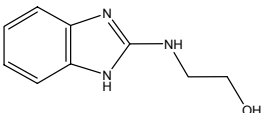
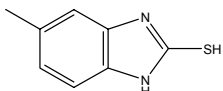
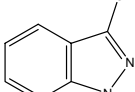
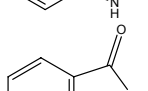
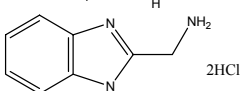
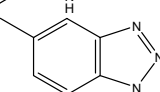
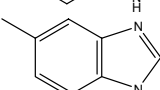


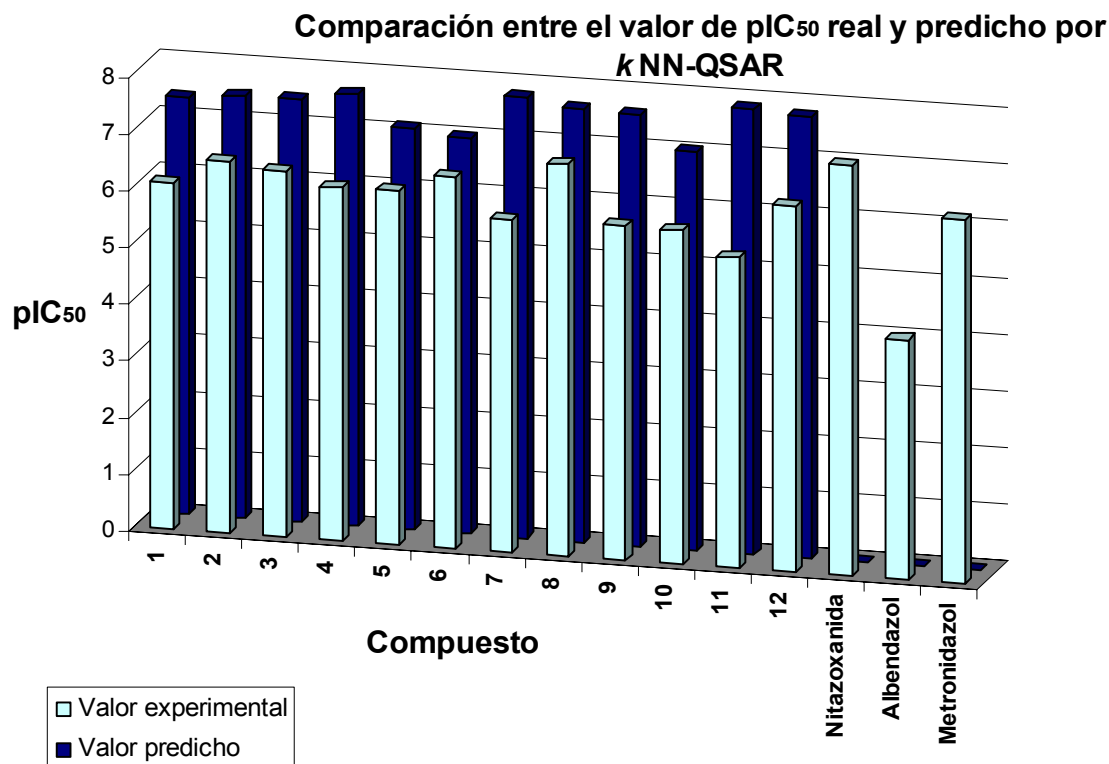
**Figura 5.4.** Estrategia de búsqueda en la base de datos del NCI usando los 41 modelos kNN-QSAR desarrollados.

De acuerdo a los resultados obtenidos de la base de datos del NCI, algunos compuestos con potencial actividad antiparasitaria contra *E. histolytica* son: benzotriazoles, tetraindazoles, azaindazoles, indazoles, indazolinonas y benzotriazoles, entre otros.

Como se indica en la Figura 5.4, 12 compuestos comercialmente disponibles (Aldrich) se compraron y se determinó su actividad biológica contra *E. histolytica* (Tabla 5.4). La actividad biológica de los compuestos se determinó en el Centro Médico Nacional Siglo XXI, empleando el mismo protocolo de ensayo que se usó para determinar la actividad biológica de los compuestos del grupo de entrenamiento en la Tabla 4.1.

**Tabla 5.4.** Compuestos obtenidos de la búsqueda en la base de datos del NCI

Estructura	Compuesto	% Modelos	pIC <sub>50</sub> exp	pIC <sub>50</sub> pred	Residual
	<b>1</b>	100	6.13	7.37-7.38	1.25
	<b>2</b>	87	6.56	7.46-8.23	1.28
	<b>3</b>	85	6.47	7.06-7.89	1.01
	<b>4</b>	78	6.24	7.11-8.16	1.39
	<b>5</b>	100	6.26	7.08-7.12	0.84
	<b>6</b>	85	6.58	7.00-7.02	0.43
	<b>7</b>	86	5.87	7.32-8.29	1.93
	<b>8</b>	97	6.93	7.20-8.08	0.71
	<b>9</b>	90	5.91	7.16-8.09	1.71
	<b>10</b>	90	5.9	6.63-7.44	1.13
	<b>11</b>	82	5.49	7.53-8.22	2.38
	<b>12</b>	95	6.48	7.39-8.17	1.30
	Nitazoxanida		7.26		
	Albendazol		4.25		
	Metronidazol		6.45		



**Figura 5.5.** Comparación entre los valores de pIC<sub>50</sub> de la actividad biológica contra *E. histolytica* para las predicciones hechas por los modelos *k*NN-QSAR de los compuestos obtenidos de la base de datos NCI. Comparación con los medicamentos comerciales como referencia.

De la Figura 5.5 se observa que todos los valores de pIC<sub>50</sub> que predijeron los modelos *k*NN-QSAR están por encima de los valores experimentales. En todos los casos, los compuestos tienen mayor actividad que el albendazol. Comparado con el metronidazol, los compuestos **7**, **9**, **10** y **11** tienen una actividad biológica menor. Los compuestos **1**, **4** y **5** tienen una actividad biológica casi equiparable y los compuestos **2**, **3**, **6**, **8** y **12** fueron más activos que el metronidazol. Hay que resaltar que entre los compuestos más activos que el metronidazol se encuentra un derivado de benzotriazol (compuesto **2**) y un derivado de indazol (compuesto **8**). Al comparar la actividad biológica de los compuestos con la actividad biológica de la nitazoxanida, se observa que ninguno fue más activo que ésta.

#### 5.4. Conclusiones

Se desarrollaron 41 modelos *k*NN-QSAR en dos dimensiones (QSAR-2D) usando el método *el k vecino más cercano*. Todos los modelos desarrollados tuvieron valores de

$q^2$ , para el grupo de entrenamiento, mayores a 0.5 y valores de  $R^2$ , para el grupo de prueba, mayores a 0.6.

Usando los 41 modelos  $k$ NN-QSAR desarrollados se logró hacer una búsqueda virtual de compuestos activos contra *E. histolytica* usando la base de datos del Instituto Nacional del Cáncer (NCI) de los Estados Unidos.

De la base de datos NCI se logró identificar 12 moléculas activas contra *E. histolytica*. Algunas de las moléculas activas son derivados de bencimidazol, sin embargo, se encontraron moléculas con núcleos diferentes al bencimidazol y que tienen actividad biológica. Hasta el momento no existen reportes previos sobre la actividad contra *E. histolytica* de estas moléculas. Las nuevas moléculas activas tienen núcleos de indazol, indazolona y benzotriazol. Esto es importante ya que abre nuevas líneas de investigación con el objeto de estudiar los requerimientos estructurales que deben tener los derivados de indazol y benzotriazol para la actividad contra *E. histolytica*.

## **PARTE III**

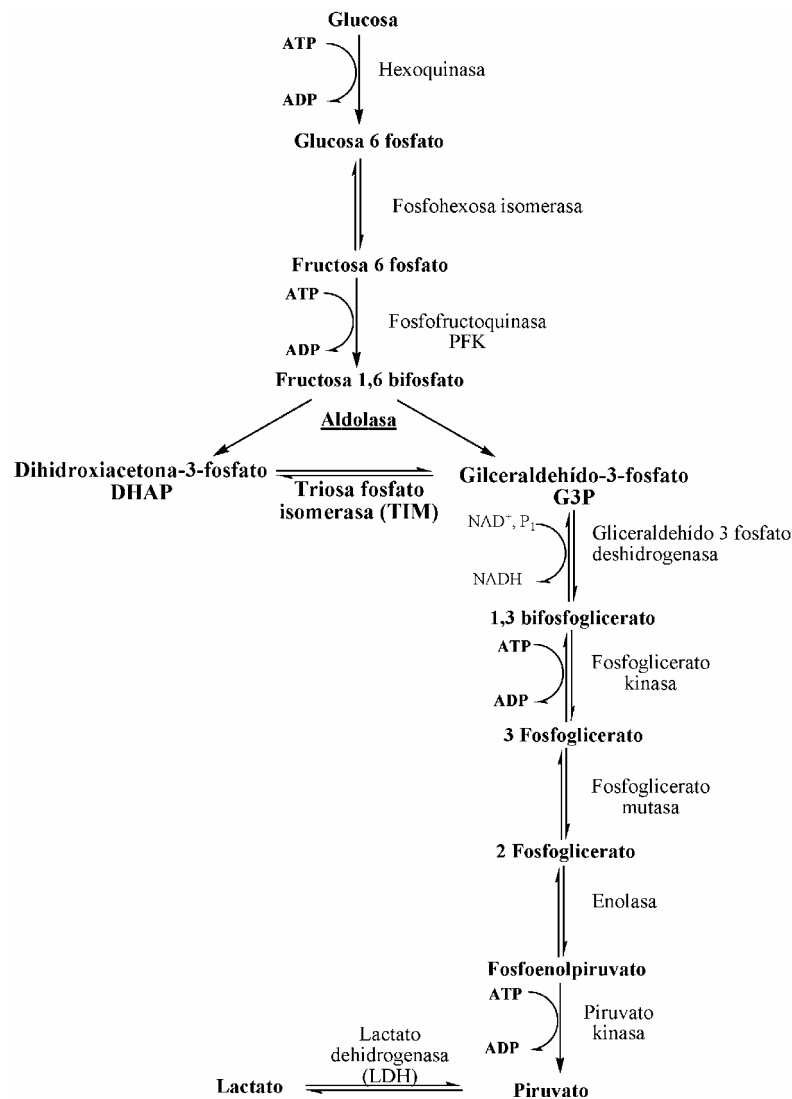
## Capítulo 6

### Estudios de inactivación, acoplamiento molecular y diseño de nuevos agentes inactivantes de la triosafosfato isomerasa de *E. histolytica* (EhTIM)

#### 6.1 La triosafosfato isomerasa (TIM) como blanco para el diseño de fármacos

La glucólisis o glicólisis consiste en la degradación de la glucosa hasta piruvato. Generalmente ocurre en el citoplasma y puede realizarse tanto en microorganismos anaerobios como aerobios. En la glicólisis se distinguen dos fases: una fase de inversión de energía, con gasto de ATP y una fase con generación de energía en forma de ATP. Se ha propuesto que las enzimas involucradas en la glicólisis pueden ser un buen blanco para el diseño de fármacos (Lakhdar-Ghazal *et al*, 2002). La síntesis de ATP en el parásito *E. histolytica* solamente se lleva a cabo por vía glicolítica, por tal razón, si se logra bloquear la producción de energía en el parásito esto conduciría a su muerte (Saavedra *et al*, 2005).

La triosafosfato isomerasa (TIM) es una enzima que juega un papel importante en la glicólisis y está involucrada en la última reacción de su primera fase o de inversión de energía. Como se muestra en la Figura 6.1, la triosafosfato isomerasa cataliza la interconversión de gliceraldehído-3-fosfato (**G3P**) y dihidroxiacetona-3-fosfato (**DHAP**). Sólo el G3P puede continuar la ruta de la glicólisis. La deficiencia de la TIM está asociado con la acumulación de DHAP; por tanto el DHAP debe convertirse, por acción de la TIM, a G3P para continuar la glicólisis. La deficiencia o baja concentración de TIM en la célula está asociada a una acumulación de DHAP y éste se degrada espontáneamente a metilglioxal (Ahmed *et al*, 2003). El aldehído metilglioxal causa la muerte de todo tipo de células (Maeta *et al*, 2005) y en el ser humano la deficiencia de TIM está asociada a una severa anemia y a un síndrome clínico de degeneración neuromuscular progresivo (Ahmed *et al*, 2003).



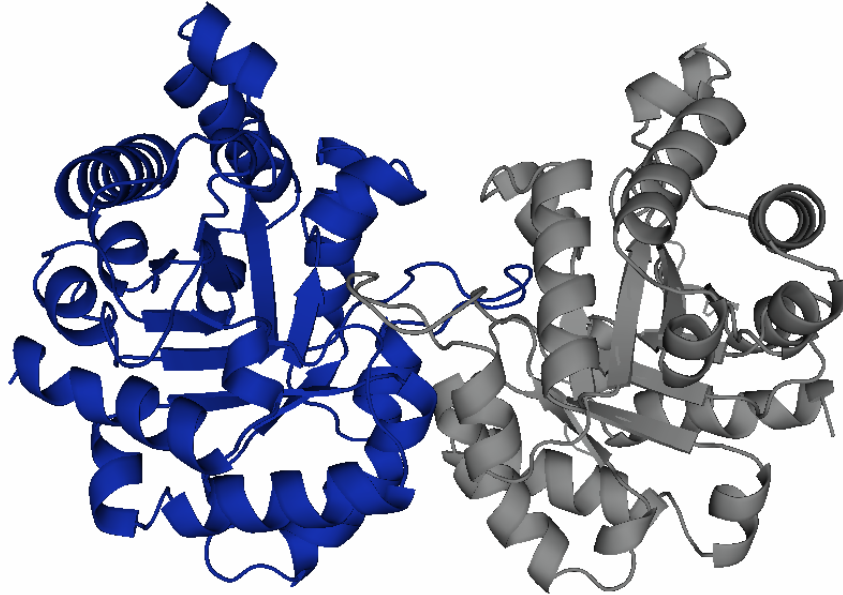
**Figura 6.1.** Diagrama de la glicólisis. Punto de intervención de la TIM en la interconversión de gliceraldehído-3-fosfato (G3P) y dihidroxiacetona-3-fosfato (DHAP).

## 6.2. Estructura general de la TIM

La TIM es una enzima glicolítica ampliamente estudiada. Esta enzima es un homodímero con estructura de barril ( $\alpha$ - $\beta$ )<sub>8</sub> (Rodríguez-Romero *et al*, 2002). Se ha resuelto la estructura tridimensional de la TIM de 12 diferentes especies con una buena resolución (Monfort *et al*, 2002), incluyendo la del humano (Mande *et al*, 1996). Excepto las TIM de *Pyrococcus woesei* y *Methanothermus fervidus* que son tetrámeros, todas las otras TIM conocidas son homodímeros (Figura 6.2) (Kohlhoff *et al*, 1996). Cada subunidad del homodímero de la TIM se compone de 8 láminas  $\beta$ -plegadas y 8  $\alpha$ -hélices intercaladas; las láminas y las  $\alpha$  hélices están unidas por 8 asas o *loops*. El asa 3 mantiene la estructura dimérica de la enzima y forma gran parte de los residuos de



contacto con el asa 1 de la otra subunidad (Figura 6.2). Es importante tener en cuenta que cada monómero tiene un sitio activo y la enzima sólo es activa en la forma dimérica (Rodríguez-Romero *et al*, 2002).

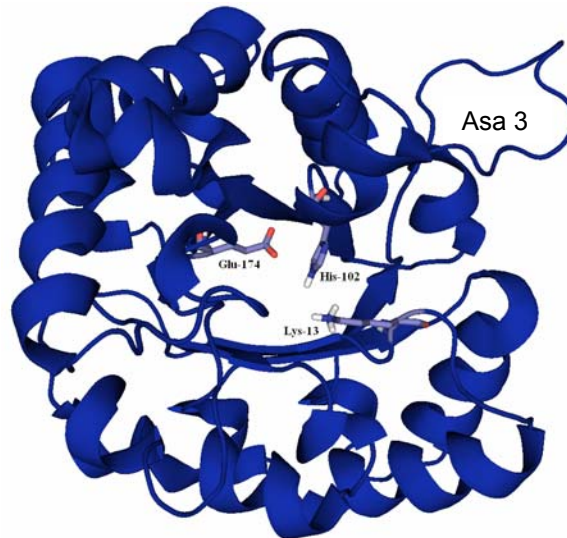


**Figura 6.2.** Estructura dimérica de la TIM, se destaca con diferente color cada subunidad o monómero

La interfaz formada por la unión de las dos subunidades de monómeros es relativamente grande,  $1400 \text{ \AA}^2$ . La TIM del humano (HsTIM) comparte la misma estructura cuaternaria, sin embargo, la HsTIM posee un residuo de más en el asa 3 respecto al asa 3 de la EhTIM. Además, en el asa 3, la HsTIM posee una metionina en lugar de una cisteína, como es el caso del resto de TIMs conocidas provenientes de organismos patógenos. Esta diferencia ha hecho posible inactivar selectivamente la TIM de *Entamoeba histolytica* (EhTIM), la TIM de *Trypanosoma cruzi* (TcTIM) y la TIM de *Trypanosoma brucei* (TbTIM) por derivatización de la cisteína con metil metano tiosulfonato (MMTS) (Rodríguez-Romero *et al*, 2002; Pérez-Monfort *et al*, 1999).

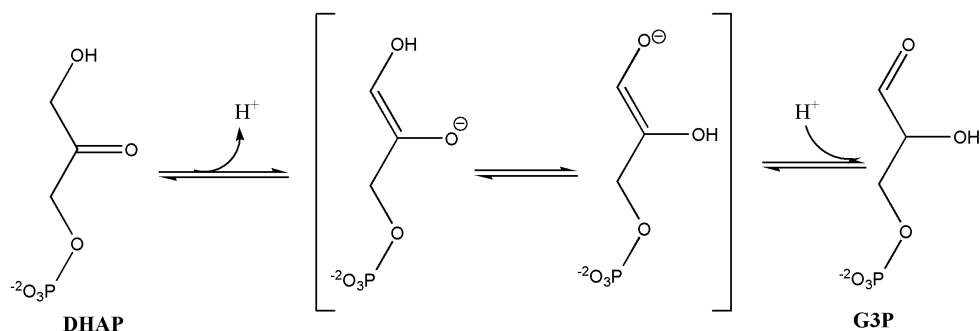
## 6.2. Catálisis de la TIM

El sitio catalítico de la TIM es altamente conservado en todas las especies incluyendo la del humano. Se ha propuesto que tres residuos están implicados directamente sobre la interconversión del G3P y DHAP. Los residuos del sitio catalítico son: histidina, ácido glutámico y una lisina. Como se muestra en la Figura 6.3, en el caso particular de la EhTIM, los residuos involucrados son: lisina 13, histidina 102 y ácido glutámico 174.



**Figura 6.3.** Monómero o subunidad de la EhtTIM. Residuos involucrados en la catálisis: Lys-13, His-102 y Glu-174.

El mecanismo catalítico se ha estudiado por diferentes métodos: mutación de residuos catalíticos (Schliebs *et al*, 1996), cristalografía de rayos X (Joseph-McCarthy *et al*, 1994; Kursula y Wierenga, 2003; Jogl *et al*, 2003), por resonancia magnética nuclear (Harris *et al*, 1998) y por cálculos teóricos (Guallar *et al*, 2004; Cui y Karplus, 2001). Se ha propuesto que la reacción catalítica transcurre vía un intermediario enediol o enediolato. Una de las vías propuestas es la siguiente: en un primer paso un protón se extrae de C-1 por la cadena lateral del ácido glutámico 174. Posteriormente, una carga negativa se desarrolla sobre el oxígeno 2, el cual se estabiliza por la interacción con el hidrógeno unido al nitrógeno de la histidina 102. En el segundo paso la transferencia de un protón es asistida por la histidina 102 del oxígeno 1 al oxígeno 2. En el tercer y último paso, el ácido glutámico 174 adiciona un protón al carbono 2 (Figura 6.4) (Wierenga *et al*, 1992).



**Figura 6.4.** Mecanismo propuesto para la interconversión de DHAP a G3P vía intermediario enediol.

### 6.3. Interfaz de la TIM como blanco para el diseño de fármacos

En vista de que el sitio catalítico de la TIM es altamente conservado en todas las especies, incluyendo la del humano, es necesario buscar otro sitio de la proteína menos conservado para el diseño de fármacos. Como se mencionó anteriormente, la TIM es un homodímero y sólo en la forma dimérica es activa (Waley, 1973).

Muchas de las proteínas en los seres vivos son oligoméricas, es decir, son asociaciones de subunidades proteicas. Esta asociación o disociación en las proteínas permite regular su función, y el buen funcionamiento puede ser provocado o detenido a través de la comunicación alostérica entre las subunidades (Monfort *et al*, 2002). Los residuos que hacen parte de la interfaz de unión de los monómeros juegan un papel importante en la estabilidad y función de los oligómeros. Estas características se pueden aprovechar para el diseño de moléculas que puedan interactuar con los residuos de la interfaz de los oligómeros e inhibir su función catalítica.

Existen varios ejemplos en los cuales se han utilizado pequeñas moléculas que actúan a nivel de la interfaz de la TIM para inhibir su función. Se ha reportado, en el caso de la TIM de *Plasmodium falciparum* (PfTIM), que pequeños péptidos que tienen la misma secuencia o mimetizan una parte de la interfaz de la TIM son capaces de inactivarla (Singh *et al*, 2001).

Se ha reportado, además, un estudio donde encontraron moléculas no peptídicas que inhiben la PfTIM. Las moléculas se encontraron por medio de la evaluación *in silico* de la base de datos del Instituto Nacional del Cáncer (NCI) de los Estados Unidos. La evaluación de esta base de datos se hizo sobre el sitio catalítico (Joubert *et al*, 2001). En estudios recientes encontraron, por búsqueda al azar, que derivados de benzotiazol inactivan la TcTIM, la estructura tridimensional de uno de los complejos de TcTIM-benzotiazol fue resuelta por difracción de rayos X. En este estudio se observó que el derivado de benzotiazol se encuentra cerca de una región hidrofóbica de la interfaz (Téllez-Valencia *et al*, 2002; Téllez-Valencia *et al*, 2004; Olivares-Illana *et al*, 2006). En otros estudios lograron resolver la estructura cristalográfica de la TcTIM a partir de cristales que incorporan hexano y se observó que dos moléculas de hexano se encuentran en una región de la interfaz muy cerca de la posición que ocupó el derivado de benzotiazol en los estudios anteriores (Gao *et al*, 1999). Algunos estudios de acoplamiento molecular por computadora se han hecho para intentar explicar los modos de unión de estos derivados de benzotiazol en la Interfaz de la TcTIM y sobre la interfaz

de la HsTIM para determinar de manera teórica la selectividad de estos compuestos (Espinoza-Fonseca y Trujillo-Ferrara, 2004; Espinoza-Fonseca y Trujillo-Ferrara, 2005; Espinoza-Fonseca y Trujillo-Ferrara, 2006).

#### 6.4. Purificación de la EhTIM

El proceso de purificación de la EhTIM en esta tesis se llevó a cabo en varias ocasiones con el objeto de contar siempre con proteína nueva. La concentración final de la EhTIM varió en cada purificación. Lo importante fue ajustar la concentración a 1 µg/mL en el momento de incubar con el agente inactivante.

Una vez obtenida la EhTIM se determinaron sus parámetros cinéticos: velocidad máxima ( $V_{max}$ ) y constante de Michaelis ( $K_m$ ). Estos parámetros se determinaron analizando la variación de la actividad enzimática de la EhTIM a concentración constante de 2 ng/mL y variando la concentración de sustrato **G3P** entre 0.1 mM y 3 mM.

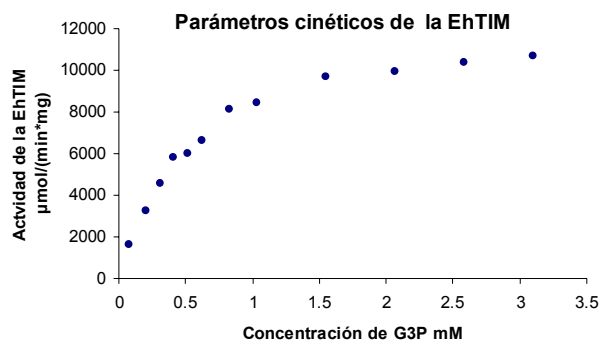


Figura 6.5. Determinación de los parámetros cinéticos  $K_m$  y  $V_{max}$  de la EhTIM.

El ajuste de la gráfica (Figura 6.5) es una hipérbola según la ecuación  $Y = P1 * X / (P2 + X)$  donde  $P1$  corresponde a la  $V_{max}$ ,  $P2$  a  $K_m$  y  $X$  la concentración inicial de sustrato ( $[So]$ ), según la ecuación de Michaelis-Menten (Ecuación 6.1):

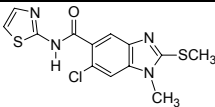
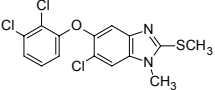
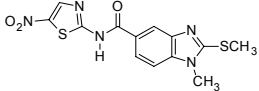
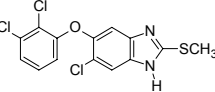
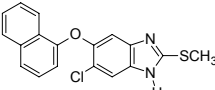
$$V_o = V_{max} \frac{[So]}{K_m + [So]} \quad (6.1)$$

Los valores obtenidos para la constante de Michaelis-Menten y la velocidad máxima son 0.53 mM y 12,663 µmol/(min\*mg) respectivamente. Los valores de  $K_m$  y  $V_{max}$  son cercanos a los reportados por Landa (Landa *et al*, 1997). Estos valores indican que la enzima se encuentra en óptimas condiciones de actividad.

## 6.5. Evaluación de compuestos orgánicos como agentes inactivantes de la EhTIM

Inicialmente, se evaluó el efecto sobre la actividad de la EhTIM con 97 compuestos, a una concentración de 250  $\mu\text{M}$ . La mayoría de los compuestos son derivados del bencimidazol y algunos intermediarios de sus síntesis. En la Tabla 6.1 se presentan sólo los cinco compuestos que mostraron efecto sobre la actividad de la EhTIM. Adicional a esto, se evaluó la actividad biológica *in vitro* de los cinco compuestos contra la *E. histolytica*. Los compuestos presentaron una alta actividad biológica. Los valores de actividad biológica para estos compuestos en concentración micromolar varían entre 0.0257 y 1.38 ( $\text{pIC}_{50}$  entre 7.59 y 5.86 respectivamente); esto indica que el mecanismo de acción de los compuestos no necesariamente es la inactivación de la EhTIM. Los compuestos pueden estar interactuando con otros blancos o receptores, de los cuales no existe evidencia alguna hasta el momento. Sin embargo, el efecto de inactivación de la EhTIM por los cinco compuestos brinda una valiosa información que permite estudiar, de manera computacional, la forma en que éstos interactúan con la enzima. A partir de los modelos de interacción ligando-receptor es posible diseñar nuevas moléculas más activas contra la EhTIM.

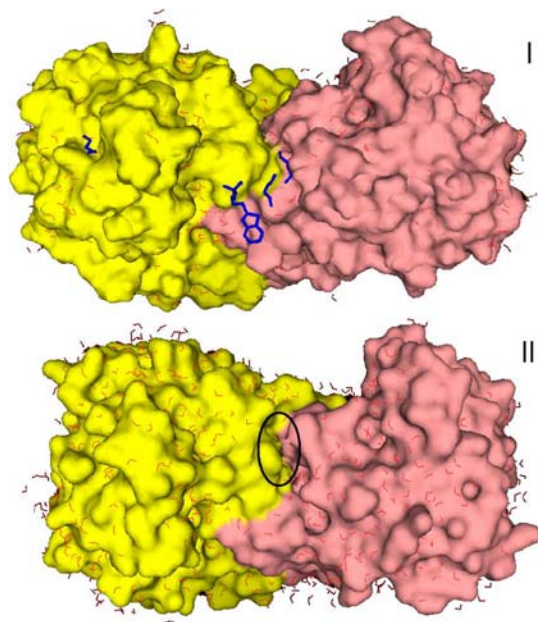
**Tabla 6.1.** Porcentajes de actividad residual de la EhTIM por acción de los compuestos 1-5 probados a 250  $\mu\text{M}$ .  $\text{pIC}_{50}$  representa la actividad biológica *in vitro* sobre el parásito. ClogP representa el coeficiente de partición octanol/agua calculado.

Compuesto	Estructura	% actividad residual de la EhTIM	$\text{pIC}_{50}$	ClogP
1		47	7.21	2.45
2		46	5.86	6.40
3		42	7.59	2.27
4		55	6.38	6.44
5		34	7.58	6.53

De la Tabla 6.1 pueden observarse los valores del coeficiente de partición octanol agua calculado (ClogP). Los valores de ClogP para estos compuestos varían entre 2.27 y 6.53; esto indica que son moléculas de relativa baja polaridad.

Como se mencionó anteriormente, en estudios recientes se ha logrado obtener dos estructuras tridimensionales de la TcTIM: una con tres moléculas de hexano y la otra con un derivado de benzotiazol. Estas moléculas se encuentran cerca de algunos residuos de la interfaz de la enzima, precisamente en una región donde hay ausencia de moléculas de agua. En los trabajos publicados por Mattos y Ringe (Mattos y Ringe, 1996; Ringe, 1995) acerca de los sitios de unión, han concluido que las regiones hidrofóbicas representan un sitio de unión potencial para moléculas orgánicas.

Al comparar la estructura cristalográfica de la EhTIM con las estructuras superpuestas de la TcTIM (1SUX y 1CI1) se observa que en la zona que se encuentran las moléculas de hexano y el derivado de benzotiazol es una zona con ausencia de moléculas de agua. En la cavidad formada por algunos residuos de la interfaz de la EhTIM, alrededor de las histidinas 110 de la cadena A y la cadena B, hay ausencia de moléculas de agua. Esto implica una región hidrofóbica en la EhTIM que puede ser un sitio de unión potencial para moléculas orgánicas (Figura 6.6).



**Figura 6.6.** Superficies de la interfaz de la TcTIM (I) (1CI1, 1SUX). Se observan dos moléculas de hexano y una de benzotiazol (color azul) cerca de los residuos de la interfaz. EhTIM (II) (1M6J). En el óvalo se observa una región cerca de los residuos de la interfaz con ausencia de moléculas de agua.

Estos antecedentes y observaciones sugieren que un posible sitio de unión para los derivados del bencimidazol podría ser la cavidad expuesta al solvente y formada por

los residuos de la interfaz de la EhTIM. Éstos podrían unirse a la interfaz del dímero alterando el equilibrio asociación-disociación y afectando su actividad.

### **6.7. Exploración de los posibles sitios de unión de los derivados de bencimidazol en la EhTIM usando el acoplamiento molecular *in silico***

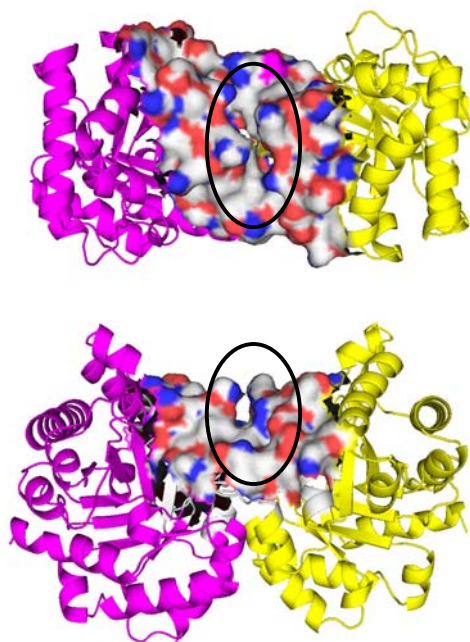
En vista de que no se conoce el sitio de unión de los ligandos, inicialmente se decidió hacer primero un estudio de acoplamiento molecular ciego (*blind docking*).

Como se mencionó anteriormente, el programa AutoDock 3.0.5 utiliza una rejilla en la región del sitio de unión para calcular las interacciones estéricas y electrostáticas con los átomos de prueba que se encuentran presentes en los ligandos. El tamaño de la rejilla puede seleccionarse de acuerdo a las necesidades que se requieran. Normalmente, las distancias entre los puntos de la rejilla son de 0.375 Å y el valor máximo posible para cada una de las aristas de la rejilla es de 126 puntos, es decir, la rejilla más grande es un cubo de 46.875 Å por arista. Este último se seleccionó para realizar una primera exploración con los 5 ligandos de la Tabla 6.1.

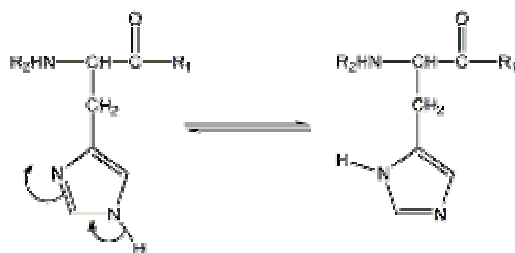
Se seleccionó una de las dos cadenas para llevar a cabo la primera exploración de acoplamiento molecular ciego. Para esto, una de las dos cadenas del homodímero se incluyó dentro de una rejilla de 46.875 Å de arista; el tamaño de la rejilla fue lo suficientemente grande como para abarcar la cadena completa, más los residuos de contacto de la otra cadena en la interfaz del homodímero. El resultado de esta primera exploración con los 5 ligandos sugirió que el sitio de unión más probable era la interfaz del homodímero superando incluso al sitio catalítico.

El siguiente paso entonces fue reducir el tamaño de la rejilla centrándola en la interfaz del homodímero. El tamaño final de la rejilla fue de 60 x 40 x 70 puntos (22.5 Å x 15.0 Å x 26.2 Å). El tamaño de la rejilla fue suficiente como para cubrir el tamaño de la cavidad formada en la interfaz del dímero de la EhTIM (Figura 6.7).

En la interfaz de la EhTIM se encuentran dos histidinas (His 110), una por monómero. Estos residuos se encuentran expuestos al solvente. La cadena lateral de la histidina está compuesta por un grupo imidazol. El cual es un anillo aromático heterocíclico de cinco átomos (Figura 6.8).



**Figura 6.7.** Cavity formada por los residuos de la Interfaz de la EhTIM. Posible sitio de unión de los derivados del bencimidazol como resultado de la primera exploración por acoplamiento molecular ciego.



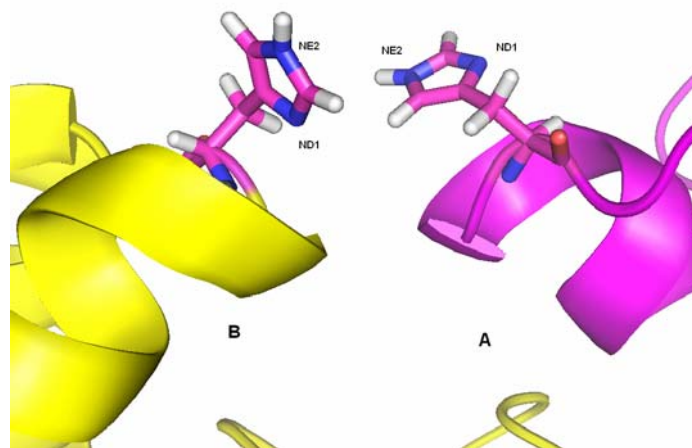
**Figura 6.8.** Tautomería de la histidina. Desplazamiento prototrópico entre los nitrógenos.

Usando el módulo pKa del programa ACD/Labs fue posible calcular los valores de pKa para las dos formas tautoméricas y las posibles protonaciones de la histidina libre como una aproximación a la histidina unida a proteína. El resultado muestra que a pH =7.4, valor de pH al que se realizaron los ensayos de evaluación de inactivación, la histidina no se encuentra protonada. Se calculó la diferencia de energía para los dos tautómeros de la histidina con el método semiempírico AM1 usando Spartan'02. La diferencia de energía entre las formas tautoméricas es de 1.613 kcal/mol, lo que indica que a una temperatura de 37 °C (temperatura del ensayo) hay un equilibrio prototrópico en el cual están presentes las dos especies tautoméricas.

Las histidinas 110 de las dos cadenas del homodímero que se encuentran en la interfaz, pueden ejercer una fuerte influencia sobre la afinidad de un compuesto orgánico



que se acople a la interfaz. Como se observa en la Figura 6.9, las histidinas no protonadas en la interfaz de la EhTIM pueden existir en cuatro formas tautoméricas o cuatro combinaciones de tautómeros, dependiendo de si el hidrógeno está unido a los nitrógenos ND1 o NE2: ND1-ND1, ND1-NE2, NE2-ND1 y NE2-NE2. El primer hidrógeno en cada combinación corresponde al hidrógeno de la histidina de la cadena B y el segundo al hidrógeno de la cadena A.



**Figura 6.9.** Posibles formas tautoméricas de las histidinas 110 de las cadenas A y B en la interfaz de la EhTIM.

Para saber si hay una influencia importante de las formas tautoméricas sobre el acoplamiento molecular de los derivados de bencimidazol de la Tabla 6.1, se generaron los mapas de energía para cada una de las combinaciones de las formas tautoméricas usando una rejilla del tamaño de la cavidad de la interfaz. El resultado de las energías de acoplamiento para cada uno de los acoplamientos de las cuatro combinaciones se presenta en la Tabla 6.2.

**Tabla 6.2.** Energías de acoplamiento molecular de los derivados de bencimidazol con las diferentes combinaciones tautoméricas de las histidinas 110 en la interfaz de la EhTIM. Valores obtenidos de la evaluación *in silico* con AutoDock 3.0.5

Compuesto	Energías libres de unión $-\Delta G_{\text{unión}}$ (kcal/mol)				% actividad residual de la EhTIM
	ND1-ND1	ND1-NE2	NE2-ND1	NE2-NE2	
1	9.03	8.96	8.98	8.92	47
2	8.64	8.60	8.63	8.57	46
3	9.00	9.06	9.01	8.95	42
4	8.38	8.31	8.37	8.29	55
5	10.15	10.08	10.15	10.07	34

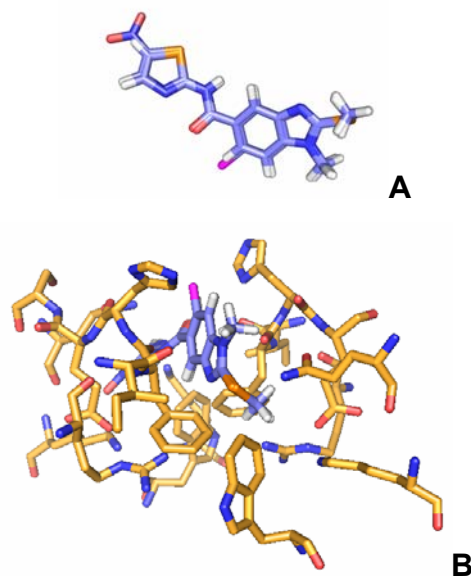
Al comparar las energías libres de unión para cada compuesto entre las diferentes combinaciones de los tautómeros de histidinas 110, se puede observar que no existe una variación notable. La diferencia más grande para el compuesto **1** es de 0.11 kcal/mol, para el compuesto **2** es de 0.07 kcal/mol, para el compuesto **3** es de 0.11 kcal/mol, para el compuesto **4** es de 0.09 kcal/mol, y para el compuesto **5** es de 0.08 kcal/mol.

La estructura con la combinación tautomérica ND1-ND1, la cual presenta una energía de unión ligeramente más favorable, fue seleccionada para el análisis y para el diseño de nuevas moléculas. Las demás combinaciones se descartaron para posteriores análisis.

Los 5 modelos de acoplamiento de ND1-ND1 se sometieron a una optimización geométrica usando el campo de fuerza de TRIPOS en las condiciones mencionadas en la sección 3.3.3. Después de la relajación por efecto del campo de fuerza sobre toda la proteína, el movimiento de la proteína y del ligando fue pequeño. Básicamente, los modelos de acoplamiento se pueden dividir en dos grupos: El primero comprende a los compuestos **1** y **3**, y el segundo a los compuestos **2**, **4** y **5**.

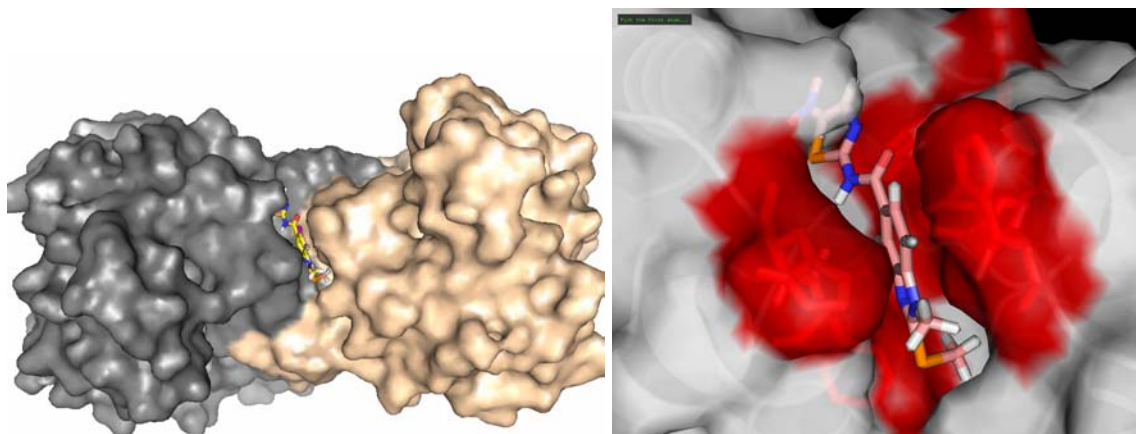
Dentro del primer grupo, los compuestos **1** y **3**, difieren poco en su estructura. Como puede observarse en la Tabla 6.1, el compuesto **3** posee un grupo metilo en la posición 1 del anillo de bencimidazol y un grupo nitro en el anillo de tiazol respecto al compuesto **1** donde las mismas posiciones están ocupadas por átomos de hidrógeno. Además, el compuesto **3** carece del átomo de cloro que posee el compuesto **1** en la posición C-6 del anillo de bencimidazol. La presencia de grupo nitro y del metilo en el compuesto **3**, parece no modificar drásticamente el efecto sobre la EhTIM y la energía de acoplamiento. El porcentaje de inactivación sólo mejora en un 5% y la energía de unión calculada por AutoDock mejora en 0.03 kcal/mol. Según el modelo encontrado por AutoDock y después de la relajación, los compuestos **1** y **3** ocupan la misma posición en la interfaz de la EhTIM (Figura 6.10A).

Los residuos que se encuentran a una distancia de menos de 5 Å de los compuestos **1** y **3** en la interfaz son: cadena A (Trp75, Arg105, Ile108, Phe109, His110, Glu111, Gln115 y Lys119) y cadena B (Trp75, Lys77, Arg105, Ile108, Phe109, His110, Glu111, Ser112, Gln115 y Lys119) (Figura 6.10B).



**Figura 6.10.** Modelo de acoplamiento para los compuestos **1** y **3**. (A) Superposición de los compuestos **1** y **3** en el modelo acoplamiento. (B) Residuos de contacto con los compuestos **1** y **3** en la interfaz de la EhTIM.

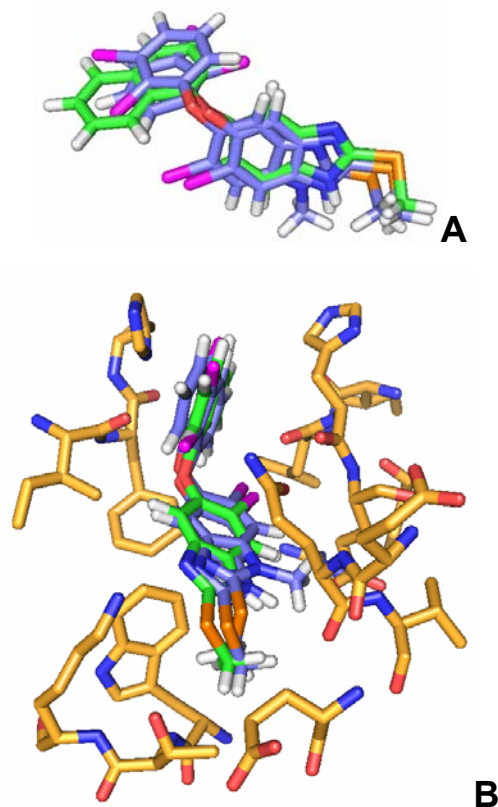
En la Figura 6.11 se puede observar cómo los compuestos **1** y **3** se acoplan perfectamente en el centro de la interfaz de la EhTIM. El anillo de benzimidazol está ocupando una región en la cual la estructura cristalográfica nativa de la EhTIM tiene ausencia de moléculas de agua.



**Figura 6.11.** Acoplamiento de los compuestos **1** y **3** en la superficie de la interfaz de la EhTIM (Figura izquierda). Acercamiento del acoplamiento del compuesto **3** (Figura derecha).

Los compuestos del segundo grupo, compuestos **2**, **4** y **5**, tienen un modelo de acoplamiento muy similar. Los compuestos **2** y **4** difieren en su estructura sólo en un grupo metilo en la posición 1. El compuesto **4** es un medicamento comercial conocido como triclabendazol y se usa normalmente para el tratamiento de la fasciolosis animal y humana (Keiser *et al*, 2005; El-Morshedy *et al*, 1999). Al comparar las energías de unión

calculadas por AutoDock, y el porcentaje de inactivación de los compuestos **2** y **4**, se puede notar que el grupo metilo de la posición C-2 en el compuesto **2** mejora la inactivación en un 9% y la energía de unión se favorece en -0.26 kcal/mol. El compuesto **5** se diseñó inicialmente como un isómero del triclabendazol (compuesto **4**), por reemplazamiento de los átomos de cloro en el anillo de benceno por un anillo de benceno fusionado para dar un anillo de naftaleno. El compuesto **5** ha resultado efectivo en ensayos *in vivo* e *in vitro* contra la *Fasciola hepatica* (Hernández-Campos *et al*, 2002; Vera-Montenegro *et al*, 2003). En este trabajo, el compuesto **5** mejoró el porcentaje de inactivación y la energía de unión a la EhTIM en un 21% y 1.77 kcal/mol, respectivamente, con respecto al compuesto **4**. En la Figura 6.12A se puede observar que el anillo de bencimidazol en los compuestos **2**, **4** y **5** ocupan casi la misma posición en el modelo de acoplamiento.

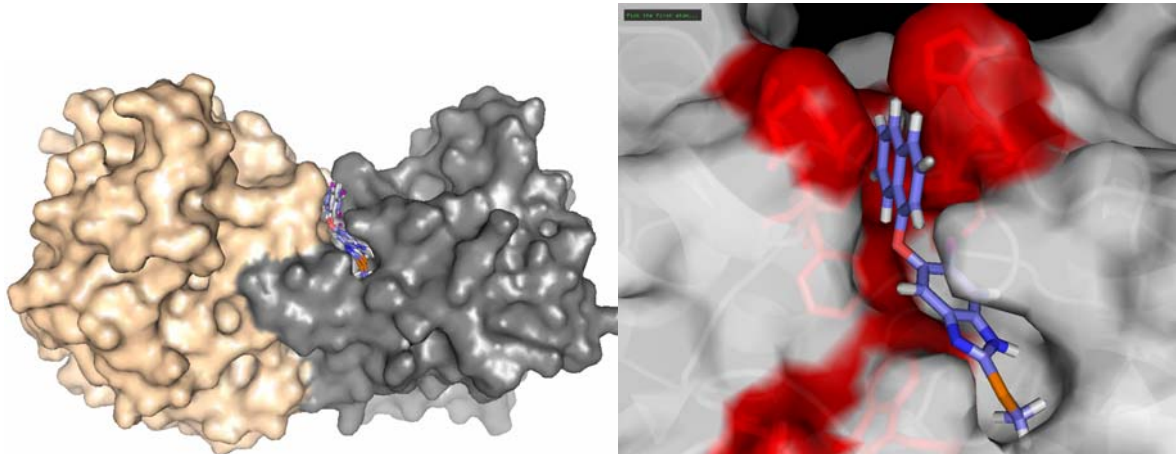


**Figura 6.12.** Modelo de acoplamiento para los compuestos **2**, **4** y **5**. (A) Superposición de los compuestos **2**, **4** y **5** en el modelo acoplamiento. (B) Residuos de contacto con los compuestos **2**, **4** y **5** en la interfaz de la EhTIM.

Los residuos que se encuentran a menos de 5 Å de los compuestos **2**, **4** y **5** en la interfaz son: cadena A (Ile108, Phe109 y His110) y cadena B (Thr76, Trp75, Lys77,

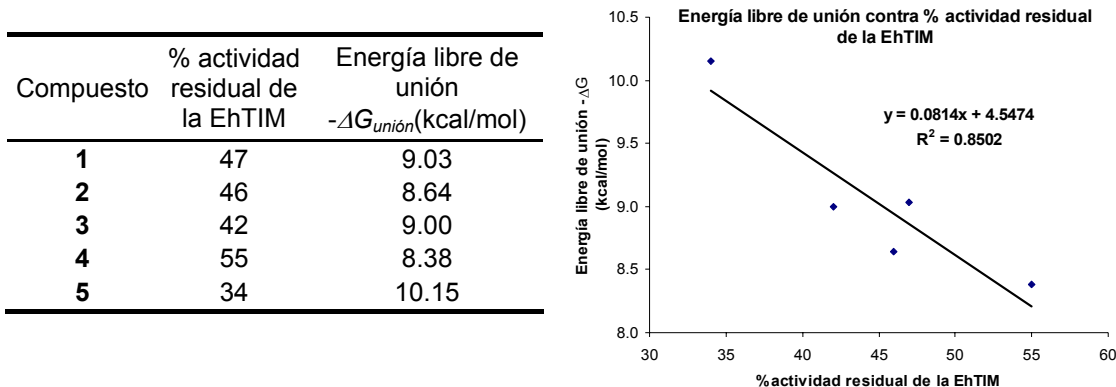
Arg105, Ile108, Phe109, His110, Glu111, Glu114, Gln115, Val116, Glu118 y Lys119 (Figura 6.12B).

En la Figura 6.13 puede observarse que los compuestos **2**, **4** y **5** se acoplan perfectamente en la interfaz del dímero de la EhTIM y están en contacto con residuos de las cadenas A y B.



**Figura 6.13.** Acoplamiento de los compuestos **2**, **4** y **5** en la superficie de la interfaz de la EhTIM (Figura izquierda). Acercamiento del acoplamiento sólo del compuesto **5** (Figura derecha).

Al analizar las energías de unión entre los ligandos y el porcentaje de inactivación de la EhTIM, existe una interesante correlación que muestra que entre menor es la actividad residual de la EhTIM por efecto del agente inactivante, más negativa es la energía libre de unión calculada por AutoDock, es decir, mayor es la afinidad de los ligandos por la interfaz de la EhTIM. El coeficiente de determinación encontrado para esta correlación es de 0.85 (Figura 6.14).



**Figura 6.14.** Correlación entre el porcentaje de inactivación de los compuestos **1-5** y su energía libre de unión calculada por AutoDock.

AutoDock calcula la energía libre de unión ( $\Delta G_{unión}$ ) como la suma de la energía intermolecular final más la energía libre de torsión del ligando; regularmente esta última energía tiene un valor positivo y desfavorece el  $\Delta G_{unión}$ . El compuesto **5**, el cual mostró una mayor  $\Delta G_{unión}$ , fue el que resultó con una mejor energía intermolecular y una menor energía de torsión. En orden descendente de energía intermolecular están los compuestos **1**, **3**, **2** y **4**. Las interacciones que aportaron más al  $\Delta G_{unión}$  fueron las interacciones de van der Waals. Es sabido que estas dependen mucho del tamaño y de la forma de la molécula. Así, para las moléculas de mayor longitud se dieron las mejores afinidades.

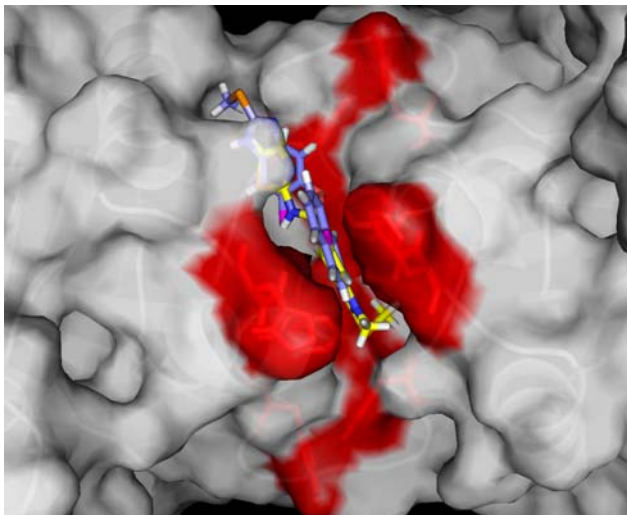
Algunos de los residuos que forman la interfaz del dímero de la EhTIM son residuos aromáticos. Estos residuos aromáticos forman un *cluster* aromático, el cual estabiliza la unión entre los monómeros. El *cluster* está formado por la interacción entre el residuo Phe109 de una cadena con los residuos Tyr81 y Trp75 de la otra cadena y viceversa. Lo anterior ha sido reportado para el caso de la TcTIM (Espinoza-Fonseca y Trujillo-Ferrara, 2004; Espinoza-Fonseca y Trujillo-Ferrara, 2005; Espinoza-Fonseca y Trujillo-Ferrara, 2006) donde el cluster está formado por el residuo Phe75 de una cadena con los residuos Tyr102 y Tyr103 de la otra cadena. Los autores reportaron los estudios de *docking* de una serie de derivados de benzotiazol. Estos derivados se acoplan en la interfaz de la TcTIM y los autores sugieren que el acoplamiento se estabiliza principalmente por interacciones de tipo  $\pi-\pi$ , donde los centroides de los anillos aromáticos de los residuos de aminoácido están a una distancia de entre 3-4 Å al anillo aromático del benzotiazol.

Como se mencionó anteriormente, el acoplamiento molecular de todos los ligandos derivados del bencimidazol se acoplan en la interfaz del dímero de la EhTIM y en todos los casos los ligandos se encuentran a menos de 5 Å de distancia de los residuos que forman los dos *clusters* aromáticos. En los compuestos **2**, **4** y **5**, el anillo de bencimidazol se encuentra perpendicular al anillo de indol del residuo Trp75. Aunque AutoDock es basado en un campo de fuerza y no simula las interacciones de tipo  $\pi-\pi$ , mediante un análisis visual es posible suponer que estos tres ligandos podrían estar estabilizados por interacciones de tipo apilamiento T (*T-stacking*). Los tres ligandos poseen un modo de unión casi idéntico y sus complejos están estabilizados por el mismo tipo de interacciones. En el caso del complejo formado con el ligando **5** la energía de unión es la más favorable

y su capacidad de inactivación es mayor que los demás compuestos. Lo anterior se puede explicar debido a que el compuesto **5** posee un sistema aromático más grande que los demás compuestos, y es bien sabido que las interacciones tipo *stacking* dependen de la cantidad de electrones  $\pi$  (McGaughey *et al*, 1998; Meyer *et al*, 2003).

En los ligandos **1** y **3**, el anillo de tiazol ocupa el mismo lugar que el anillo de bencimidazol de los ligandos **2**, **4** y **5**; el anillo de tiazol se encuentra perpendicular al residuo Trp75, el cual podría estabilizarse por una interacción *T-stacking*. Otra contribución a la estabilización de estos dos complejos está dada por la interacción de tipo electrostática entre el nitrógeno 3 del anillo del bencimidazol, el cual tiene una alta densidad electrónica, con un hidrógeno aromático del residuo Phe109. De esta manera, los residuos Phe109 y Trp75 podrían jugar un papel importante en la estabilización de los complejos EhTIM-ligando. Esta información es útil para el diseño de nuevos agentes inactivantes de la EhTIM.

La Figura 6.15 muestra el acoplamiento de los compuesto **1** y **5** superpuestos en la interfaz de la EhTIM. En esta figura se puede observar una superficie roja translúcida en la interfaz debido a los residuos aromáticos que están expuestos al disolvente y que interaccionan con los ligandos **1-5**.

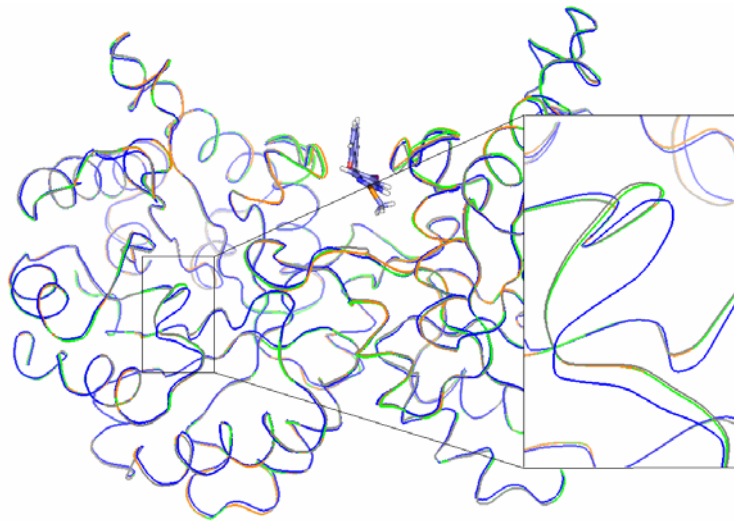


**Figura 6.15.** Residuos aromáticos en la interfaz de la EhTIM (color rojo).

Al alinear sobre los carbonos alfa de la cadena peptídica de los complejos de EhTIM-compuestos **1-5** con la EhTIM sola, sometidos todos al campo de fuerza, se observa que no hay una variación notable en la conformación de los residuos, ni en la cadena peptídica con excepción del complejo formado por EhTIM-compuesto **5**. El

compuesto **5** mostró ser el compuesto con una mayor capacidad de inactivación (34 % de actividad residual) y una mejor energía de unión calculada por AutoDock (-10.15 kcal/mol).

Se observó que en el complejo EhTIM-compuesto **5** hay cambios más drásticos en algunos residuos involucrados en el sitio catalítico respecto de los demás complejos y de la EhTIM sola (Figura 6.16).



**Figura 6.16.** Superposición de los complejos EhTIM-ligando con la estructura de la EhTIM nativa. Cambios conformacionales observados en el asa 8 y que está involucrada en el sitio catalítico. Todas las estructuras se optimizaron con el campo de fuerza MM de Tripos.

Algunos de los residuos que hacen parte del asa 8 y que cambian más drásticamente su conformación al unir al compuesto **5** son: valina 240 (Val240), glicina 241 (Gly241) y glicina 242 (Gly242). Los cambios se observaron en ambos monómeros de la EhTIM (ver recuadro Figura 6.16).

Se ha reportado que el hidrógeno del grupo amino del enlace peptídico de la asparagina 233, en el caso de la TIM del parásito *Plasmodium falciparum* (agente causante de la malaria), y el hidrógeno del grupo amino de la glicina 233 en la TIM del tripanosoma son importantes para la estabilización del complejo entre la TIM y su ligando natural, la DHAP. Estos átomos forman un enlace de hidrógeno con uno de los oxígenos del grupo fosfato de la DHAP (Eaazhisai *et al*, 2004). La estabilización por el enlace de hidrógeno mantiene unido el grupo fosfato al ligando natural y evita que se elimine con la consecuente formación de metil glioxal.



El residuo correspondiente a la Asp233 y Gly233 en la EhTIM es la Gly242. Este residuo en la EhTIM estaría cumpliendo la misma función que Asp233 y Gly233, estabilizar el complejo EhTIM-DHAP. Como se mencionó anteriormente, el residuo Gly242 hace parte del asa 8 el cual sufre cambios conformacionales en el complejo EhTIM-compuesto **5**. Algunos cambios conformacionales más detallados en los residuos Gly241 y Gly242 promovidos por la unión del compuesto **5** se muestran a continuación:

En la cadena B, el hidrógeno del grupo amino de la Gly242 varía en un ángulo aproximado de 21° con respecto al mismo hidrógeno de la estructura de la EhTIM sin compuesto. En la cadena A, este hidrógeno se desvía aproximadamente unos 41°. Es bien conocido que la fuerza de un enlace de hidrógeno depende principalmente del ángulo de enlace; por tanto, la variación en estos ángulos de enlace puede disminuir la fuerza del enlace de hidrogeno del oxígeno del grupo fosfato de la DHAP con el hidrógeno de la Gly242. Esto puede provocar la desestabilización del complejo EhTIM-DHAP causando una baja en su actividad enzimática.

Las distancias entre los átomos de la cadena peptídica de los residuos Gly241 y Gly242 (asa 8) del complejo EhTIM-compuesto **5** y la proteína sin ligando o nativa se muestran en la Tabla 6.3.

**Tabla 6.3.** Distancias entre los átomos de carbono alfa (C), oxígeno (O) y nitrógeno del enlace peptídico (N) de los residuos Gly 241 y Gly 242 del complejo EhTIM-compuesto **5** con los correspondientes átomos de la EhTIM sin compuesto

Distancias entre los átomos C, O y N del complejo EhTIM-compuesto <b>5</b> y la EhTIM sin compuesto (Ángstrom)						
Residuo	Cadena A			Cadena B		
	C (alfa)	O (carbonilo)	N (enlace peptídico)	C (alfa)	O (carbonilo)	N (enlace peptídico)
Gly241	2.058	2.782	1.546	0.920	1.176	0.666
Gly242	1.267	1.279	1.335	1.238	1.190	1.161

Cuando se generan las superficies de van der Waals, tanto para el complejo EhTIM-compuesto **5** como para la proteína sola, se observa que el espacio del sitio catalítico pierde volumen, de tal forma que podría pensarse, además, que se bloquea la entrada del ligando natural, impidiendo de esta forma la reacción del ligando natural.

Si bien el efecto de inactivación del compuesto **5** podría explicarse por los cambios conformacionales que provocan en el sitio catalítico, particularmente, en el asa 8, los

efectos de inactivación de los demás compuestos no pueden explicarse por medio de la optimización geométrica usando mecánica molecular (campo de fuerza TRIPOS) en las condiciones mencionadas.

### **6.8. Estudio de acoplamiento molecular de los derivados de bencimidazol 1-5 sobre la TIM de humano (HsTIM) y selectividad de los compuestos por la EhTIM**

Para saber si los compuestos que inhiben la actividad de la EhTIM podrían ser selectivos a esta y no afectan a la TIM de humano (HsTIM), se llevó a cabo un estudio de acoplamiento molecular sobre la HsTIM con los mismos ligandos. La estructura cristalográfica de la HsTIM se puede obtener de la base de datos del *protein data bank* bajo el ID 1HTI (Mande, 1994). El procedimiento que se usó en este estudio fue igual al utilizado para EhTIM. Los resultados de las energías libres de unión se presentan a manera de resumen en la Tabla 6.4. Según los resultados, las energías libres de unión de los ligandos sobre la HsTIM son más negativas que las de la EhTIM. Todos los ligandos tienen una preferencia por la TIM de humano más que por la TIM de *E. histolytica*, con excepción del compuesto **5**, el cual presenta una ligera preferencia por la EhTIM. Esta condición es indeseable ya que los compuestos están mostrando que podrían tener selectividad por la enzima del humano más que la TIM del parásito de *E. histolytica*. Es importante notar que la resolución de la estructura cristalográfica de la HsTIM (1HTI) usada en el estudio de acoplamiento es de 2.8 Å.

El dímero de la HsTIM, al igual que las demás TIM, se estabiliza por *clusters* aromáticos. El *cluster* de la HsTIM está formado por el residuo Phe102 de un monómero con los residuos Phe74 y Tyr67 del otro monómero y viceversa.

Al comparar los residuos que forman la cavidad de la interfaz de la EhTIM con la HsTIM, se observa que hay un 25% de identidad. De 8 residuos que forman la cavidad, en cada monómero, existe identidad sólo en dos residuos, fenilalanina y lisina. El triptófano y la tirosina en la cavidad de la interfaz de la EhTIM se reemplazan por una tirosina y una fenilalanina en la HsTIM, respectivamente. Esto quiere decir que las dos enzimas tienen requerimientos similares respecto a un ligando que se una a su interfaz. Sin embargo, las diferencias entre una y otra hace posible diseñar moléculas con cierta selectividad por una de ellas.

**Tabla 6.4.** Comparación de las energías libres de unión entre la TIM de humano y la de *E. histolytica*

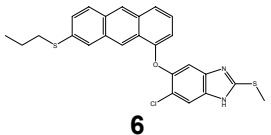
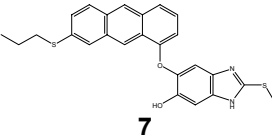
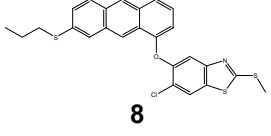
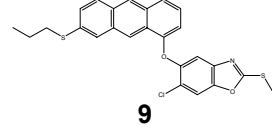
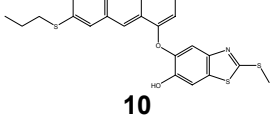
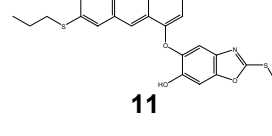
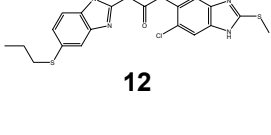
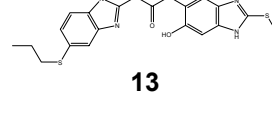
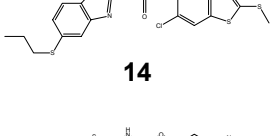
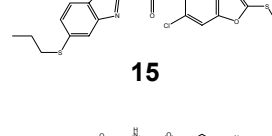
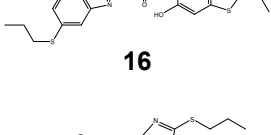
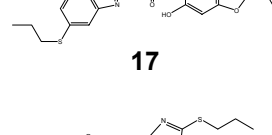
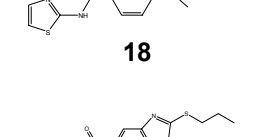
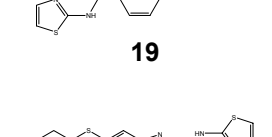
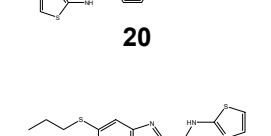
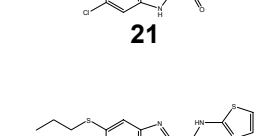
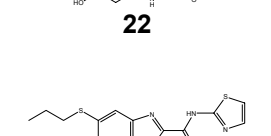
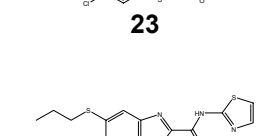
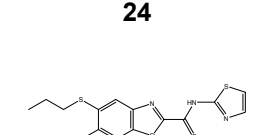
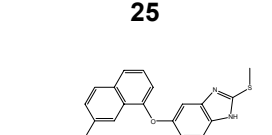
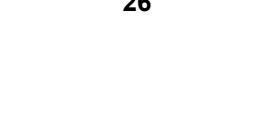

Compuesto	Estructura	% actividad residual de la EhTIM	$(-\Delta G_{\text{unión}})$ EhTIM	$(-\Delta G_{\text{unión}})$ HsTIM
1		53	9.03	10.18
2		54	8.64	9.80
3		58	9.51	9.93
4		45	8.38	9.50
5		66	10.15	9.95

### 6.9. Diseño de nuevas moléculas con alta afinidad por la interfaz de la EhTIM como posibles agentes inactivantes.

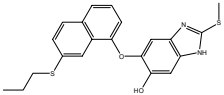
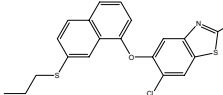
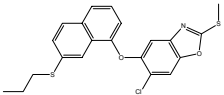
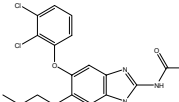
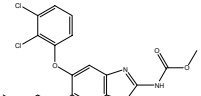
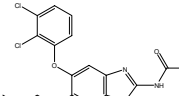
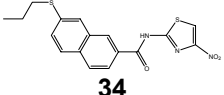
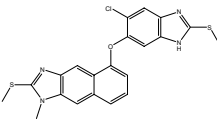
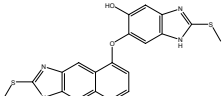
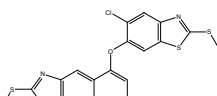
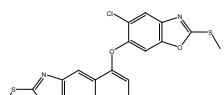
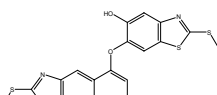
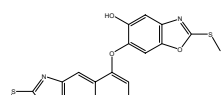
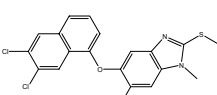
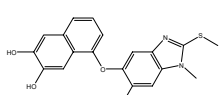
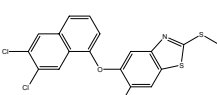
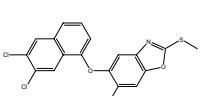
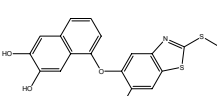
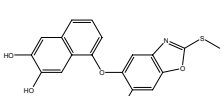
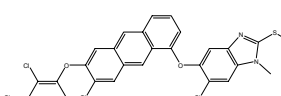
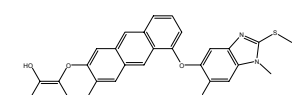
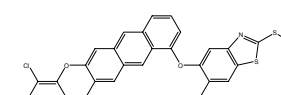
Teniendo en cuenta los requerimientos antes mencionados y el hecho de que la interfaz de la EhTIM tiene una cavidad de una longitud de casi el doble que la longitud de los compuestos probados, se diseñaron 83 moléculas. Los nuevos ligandos se diseñaron de varias formas: por visualización directa de su ambiente en la interfaz de la proteína, por superposición de diferentes modos de unión de un mismo ligando y de diferentes ligandos, por sustitución de las moléculas de agua cristalográficas y usando los conceptos de reemplazamiento isostérico e hibridación molecular (Silverman, 2004; Wermuth, 2003a).

El estudio de acoplamiento con la interfaz de la EhTIM de las moléculas diseñadas se realizó siguiendo el mismo procedimiento usado para los compuestos **1-5**. La Tabla 6.5 muestra las estructuras de los compuestos diseñados, los resultados de las energías libres de unión calculados por AutoDock y sus correspondientes valores de ClogP.

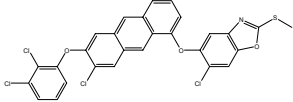
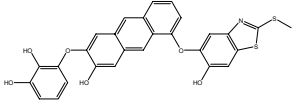
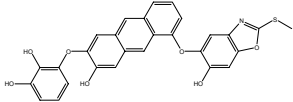
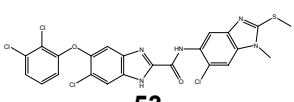
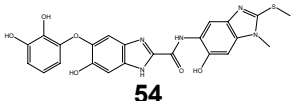
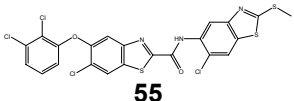
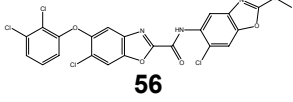
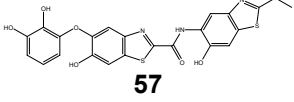
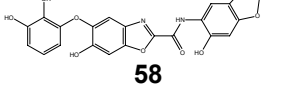
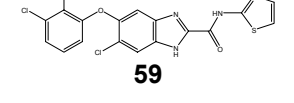
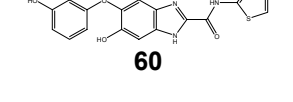
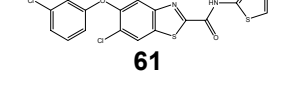
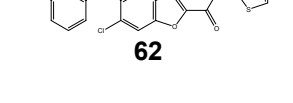
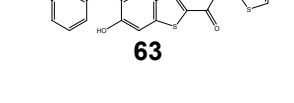
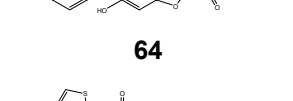
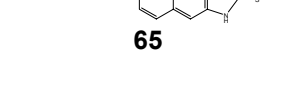
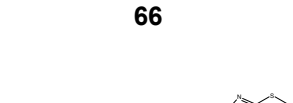
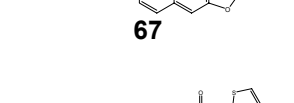
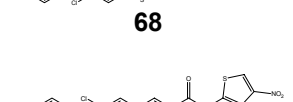

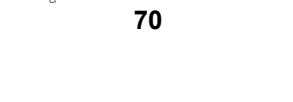
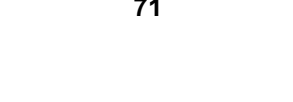
**Tabla 6.5.** Ligandos diseñados como posibles moléculas inactivantes de la EhTIM

Ligandos diseñados					
Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP	Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP
	9.62	9.33		12.45	7.19
	12.23	9.80		9.34	9.16
	10.27	7.72		10.11	7.04
	9.81	6.17		11.12	4.02
	8.66	7.15		8.31	5.84
	10.13	5.31		8.52	3.96
	9.81	3.62		9.62	4.19
	9.23	3.65		9.71	3.84
	9.43	2.14		9.64	4.89
	8.52	4.27		9.21	4.26
	8.24	2.93		8.99	8.15

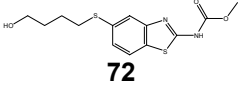
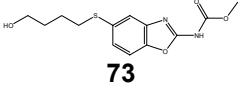
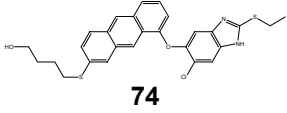
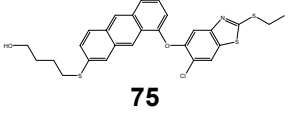
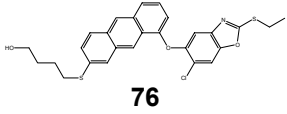
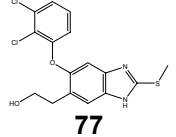
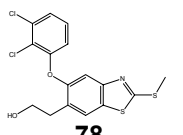
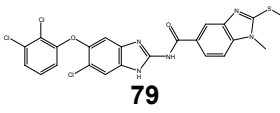
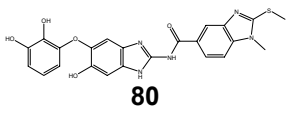
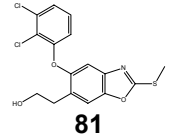
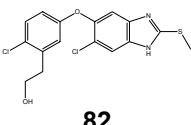
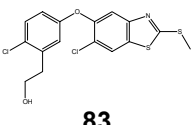
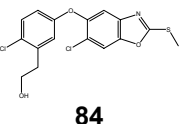
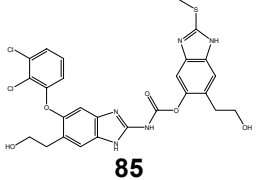
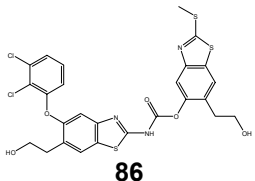
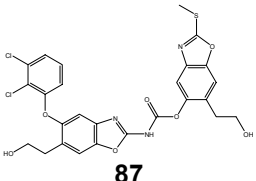
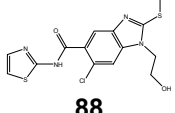
Ligandos diseñados

Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP	Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP
 <b>28</b>	ND	6.01	 <b>29</b>	8.51	8.62
 <b>30</b>	8.57	7.99	 <b>31</b>	8.69	6.23
 <b>32</b>	8.42	6.81	 <b>33</b>	7.32	6.08
 <b>34</b>	9.64	4.63	 <b>35</b>	12.45	5.67
 <b>36</b>	12.15	4.98	 <b>37</b>	12.25	8.09
 <b>38</b>	11.99	6.78	 <b>39</b>	11.84	7.45
 <b>40</b>	11.73	6.03	 <b>41</b>	10.65	7.81
 <b>42</b>	10.48	3.15	 <b>43</b>	10.17	8.31
 <b>44</b>	10.79	7.68	 <b>45</b>	10.41	3.66
 <b>46</b>	10.25	4.36	 <b>47</b>	13.99	11.33
 <b>48</b>	13.46	6.66	 <b>49</b>	13.67	11.83

Ligandos diseñados

Ligando	$(-\Delta G_{\text{unión}})$ (kcal/mol)	ClogP	Ligando	$(-\Delta G_{\text{unión}})$ (kcal/mol)	ClogP
 <b>50</b>	8.99	11.20	 <b>51</b>	9.10	5.77
 <b>52</b>	12.88	6.46	 <b>53</b>	12.18	7.85
 <b>54</b>	11.28	3.62	 <b>55</b>	11.85	8.48
 <b>56</b>	9.95	7.23	 <b>57</b>	9.99	3.82
 <b>58</b>	10.17	4.32	 <b>59</b>	9.81	6.64
 <b>60</b>	9.17	0.97	 <b>61</b>	9.61	6.08
 <b>62</b>	8.20	5.46	 <b>63</b>	9.67	1.43
 <b>64</b>	8.25	2.19	 <b>65</b>	10.50	3.57
 <b>66</b>	10.05	4.08	 <b>67</b>	10.31	3.41
 <b>68</b>	10.21	8.37	 <b>69</b>	8.27	6.53
 <b>70</b>	9.90	6.63	 <b>71</b>	7.96	2.00

Ligandos diseñados

Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP	Ligando	(-) $\Delta G_{\text{unión}}$ (kcal/mol)	ClogP
 <b>72</b>	8.72	2.58	 <b>73</b>	7.48	1.85
 <b>74</b>	9.65	8.40	 <b>75</b>	10.10	8.87
 <b>76</b>	9.10	8.23	 <b>77</b>	10.24	5.06
 <b>78</b>	ND	5.59	 <b>79</b>	12.45	7.32
 <b>80</b>	11.99	3.03	 <b>81</b>	ND	4.91
 <b>82</b>	8.74	5.26	 <b>83</b>	8.59	5.73
 <b>84</b>	8.48	5.10	 <b>85</b>	10.59	5.32
 <b>86</b>	10.03	6.35	 <b>87</b>	7.05	5.01
 <b>88</b>	9.08	1.62			

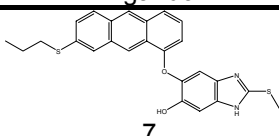
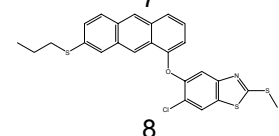
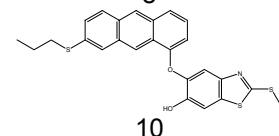
De los 83 ligandos diseñados, 33 mostraron energías libres de unión más negativas que -10 kcal/mol. La energía libre de unión de estos 33 compuestos con la HsTIM se evaluó del mismo modo. Lo anterior se hizo con el objetivo de determinar la

posible selectividad de los ligandos por la enzima del parásito. La selectividad de los compuestos por la EhTIM puede determinarse por la diferencia entre las energías libres de unión de los compuestos por la EhTIM y la de HsTIM ( $\Delta\Delta G_{unión} = \Delta G_{unión}^{EhTIM} - \Delta G_{unión}^{HsTIM}$ ). La energía de la EhTIM debe ser más negativa que la de la HsTIM, es decir, entre más negativa sea la diferencia podría haber más selectividad de los compuestos por la TIM del parásito que por la TIM del humano. De estos 33 ligandos, 20 mostraron selectividad sobre la EhTIM. De igual forma, los 33 ligandos se sometieron a la regla de Lipinski o de los cinco (Lipinski *et al*, 1997; Lipinski *et al*, 2001). Esta regla es empírica y comúnmente se usa para predecir si un compuesto puede tener buenas propiedades farmacocinéticas. La regla dice que un compuesto tendrá buenas propiedades biofarmacéuticas si no viola más de una de las siguientes reglas:

- ✚ Peso molecular menor a 500 g/mol
- ✚ ClogP menor a 5
- ✚ No poseer más de cinco donadores de enlace de hidrógeno
- ✚ No poseer más de 10 aceptores de enlace de hidrógeno

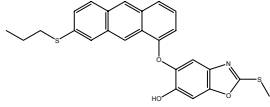
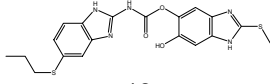
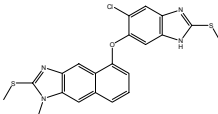
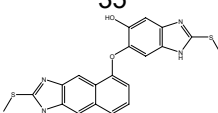
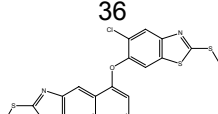
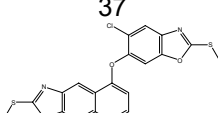
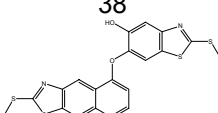
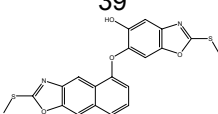
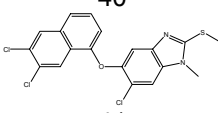
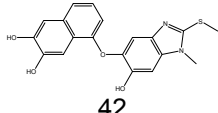
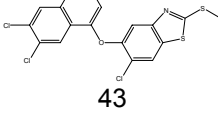
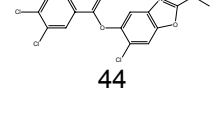
En la Tabla 6.6 se pueden observar los resultados para el acoplamiento molecular de los ligandos seleccionados con la HsTIM y el número de violaciones a la regla de Lipinski.

**Tabla 6.6.** Comparación de las energías libres de unión para los ligandos seleccionados y violaciones a la regla de Lipinski

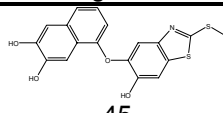
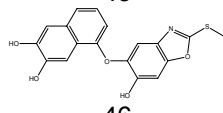
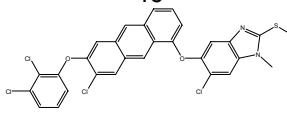
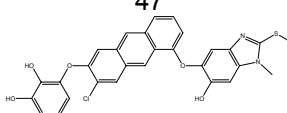
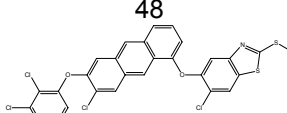
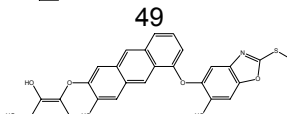
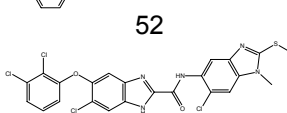
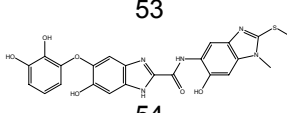
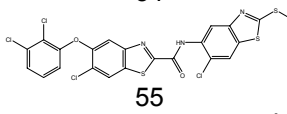
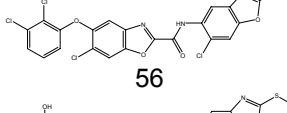
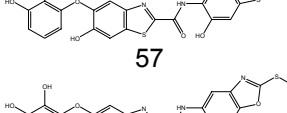
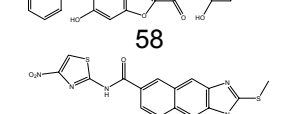
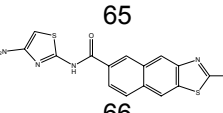

Regla de Lipinski o de los cinco								
Ligando	$(-)\Delta G_{unión}^{EhTIM}$	$(-)\Delta G_{unión}^{HsTIM}$	$\Delta\Delta G_{unión}$	ClogP	Peso mol (g/mol)	N° de donadores de hidrógeno	N° de aceptores de hidrógeno	N° de faltas a la regla
				< 5	< 500	< 5	< 10	< 1
 7	12.45	11.58	-0.87	7.19	446.58	2	3	1
 8	12.23	11.91	-0.32	9.8	482.08	0	2	1
 10	10.27	11.52	1.25	7.72	463.63	1	3	1

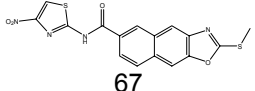
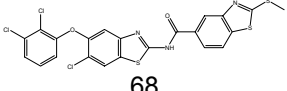
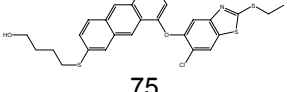
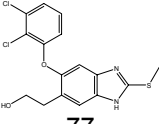


## Regla de Lipinski o de los cinco

Ligando	$(-)\Delta G_{unión}$ EhTIM	$(-)\Delta G_{unión}$ HsTIM	$\Delta\Delta G_{unión}$	ClogP	Peso mol (g/mol)	N° de donadores de hidrógeno	N° de aceptores de hidrógeno	N° de faltas a la regla
				< 5	< 500	< 5	< 10	< 1
 11	10.11	11.49	1.38	7.04	447.57	1	4	1
 13	11.12	11.22	0.10	4.02	429.52	4	6	0
 35	12.45	10.37	-2.08	5.76	440.97	1	3	1
 36	12.15	11.38	-0.77	4.98	422.52	2	4	0
 37	12.25	11.50	-0.75	8.09	461.04	0	3	1
 38	11.99	10.89	-1.10	6.78	428.91	0	5	1
 39	11.84	9.87	-1.97	7.45	442.60	1	4	1
 40	11.73	10.85	-0.88	6.03	410.47	1	6	1
 41	10.65	10.92	0.27	7.81	423.74	0	2	1
 42	10.48	10.83	0.35	3.15	368.41	3	5	0
 43	10.17	11.57	1.40	8.31	426.77	0	2	1
 44	10.79	10.20	-0.59	7.68	410.70	0	3	1

## Regla de Lipinski o de los cinco

Ligando	$(-)\Delta G_{unión}$ EhTIM	$(-)\Delta G_{unión}$ HsTIM	$\Delta\Delta G_{unión}$	ClogP	Peso mol (g/mol)	N° de donadores de hidrógeno	N° de aceptores de hidrógeno	N° de faltas a la regla
				< 5	< 500	< 5	< 10	< 1
 45	10.41	9.94	-0.47	3.66	371.43	3	5	0
 46	10.25	10.16	-0.09	4.36	355.36	3	6	0
 47	13.99	11.40	-2.59	11.33	600.34	0	3	1
 48	13.46	12.47	-0.99	6.66	545.01	3	6	2
 49	13.67	12.05	-1.62	11.83	603.37	0	3	2
 52	12.88	11.64	-1.24	6.46	513.52	4	8	2
 53	12.18	11.50	-0.68	7.85	567.27	2	5	2
 54	11.28	11.39	0.11	3.62	493.49	5	9	0
 55	11.85	11.71	-0.14	8.48	587.35	1	5	2
 56	9.95	12.71	2.76	7.23	555.22	1	7	2
 57	9.99	11.78	1.79	3.82	513.57	5	9	1
 58	10.17	10.41	0.24	4.32	481.43	5	11	1
 65	10.50	10.63	0.13	3.57	385.42	2	6	0
 66	10.05	9.87	-0.18	4.08	402.47	1	6	0

Regla de Lipinski o de los cinco								
Ligando	$(-)\Delta G_{\text{unión}}^{\text{EhTIM}}$	$(-)\Delta G_{\text{unión}}^{\text{HsTIM}}$	$\Delta\Delta G_{\text{unión}}$	ClogP	Peso mol (g/mol)	N° de donadores de hidrógeno	N° de aceptores de hidrógeno	N° de faltas a la regla
	< 5	< 500	< 5	< 10	< 1			
 67	<b>10.31</b>	10.13	<b>-0.18</b>	3.41	386.40	1	7	0
 68	10.21	10.46	0.25	8.37	552.90	1	5	2
 75	10.10	11.91	1.81	8.87	526.13	1	3	2
 77	<b>10.24</b>	9.43	<b>-0.81</b>	5.59	369.27	2	3	1

De los 20 ligandos que mostraron selectividad por la EhTIM, 5 violan más de una regla (**48**, **49**, **52**, **53** y **55**), sin embargo, dentro de estos 5 ligandos, los complejos que forman con la EhTIM **48** y **49** poseen las mejores energías de acoplamiento, lo cual sugiere que no deben descartarse.

De acuerdo a los requerimientos estructurales que deben tener los ligandos para que se unan a la interfaz de la EhTIM, éstos deben tener uno o varios sistemas aromáticos. El problema inherente a grandes sistemas aromáticos es que a menudo los compuestos con estos sistemas, como el caso de **48** y **49**, podrían resultar agentes intercalantes del ADN, lo cual es indeseable. Otro problema asociado a esto es la solubilidad, lo anterior es un factor que se debe de tomar en cuenta, principalmente porque los ensayos de inactivación se hacen en medio acuoso. De acuerdo a los resultados obtenidos en el diseño de los nuevos ligandos, los compuestos cuyas energías libres de unión aparecen en negrita en la Tabla 6.6 son posibles candidatos a ser sintetizados.

## 6.10. Conclusiones

Experimentalmente se encontraron 5 derivados de bencimidazol que tienen un efecto de inactivación sobre la trifosfato isomerasa de *Entamoeba histolytica*. A pesar de que el efecto de los derivados de bencimidazol sobre la actividad catalítica de la EhTIM fue moderado, los cinco compuestos mostraron un efecto inhibitorio excelente sobre el crecimiento *in vitro* del parásito *E. histolytica*.

De acuerdo a los antecedentes que existen para la TcTIM y a los resultados del estudio de acoplamiento molecular ligando-receptor usando AutoDock, los 5 derivados de bencimidazol interactúan con los residuos de la interfaz de la EhTIM. Las energías de unión para los 5 derivados de bencimidazol se pudieron calcular y mostraron tener una interesante correlación con los valores del % de actividad residual de la EhTIM, el coeficiente de correlación ( $R$ ) fue de 0.92.

Según el modelo de acoplamiento o de unión predicho por AutoDock, los derivados de bencimidazol se pueden agrupar en dos grupos: un primer grupo para los compuestos con un anillo de tiazol unido a la posición C-5 del anillo de bencimidazol a través de un grupo amida (compuestos **1** y **3**) y un segundo grupo para los derivados de bencimidazol con un anillo de benceno y naftaleno unido a la posición C-5 del anillo de bencimidazol a través de un grupo éter (compuestos **2**, **4** y **5**).

El compuesto **5** resultó ser el compuesto con mayor efecto de inactivación, es de notar que este compuesto tiene un sistema aromático más grande que los demás. El *cluster* aromático en la interfaz de la EhTIM, formado por los residuos Phe109, His110 y Trp75, puede desempeñar un papel importante en el acoplamiento de ligandos aromáticos. En todos los casos, el Trp75 se encuentra en posición perpendicular a un sistema aromático en los derivados de bencimidazol. Un análisis comparativo más minucioso de las estructuras de los complejos ligando-receptor mostró que para el caso del compuesto **5** su efecto de inactivación podría deberse a cambios conformacionales provocados en los residuos Val240, Gly241 y principalmente Gly242; este último es importante en la estabilización del complejo enzima-sustrato.

El estudio de acoplamiento molecular para los 5 derivados de bencimidazol con la HsTIM mostró que sólo el compuesto **5** podría ser ligeramente selectivo a la enzima del parásito, más que a la TIM del humano.

Partiendo de los modelos de acoplamiento molecular de los 5 derivados de bencimidazol, se lograron diseñar 33 nuevas moléculas con una mejor afinidad por la interfaz de la EhTIM. Las moléculas diseñadas con las energías de acoplamiento más negativas se analizaron en función de la regla de Lipinski o de los cinco y sólo algunas de ellas violan más de una regla.

El estudio de acoplamiento molecular con la HsTIM de los compuestos diseñados muestra que *in silico* una parte de ellos tienen selectividad por la EhTIM y estas moléculas

se convierten en agentes inactivantes potenciales de la EhTIM y como posibles candidatos a ser sintetizados y probados contra la enzima.

# Capítulo 7

## Conclusiones generales

- En este trabajo se aplicaron métodos del diseño de fármacos asistido por computadora para buscar y diseñar moléculas con interés antiamebiano. Los métodos usados fueron en base a la estructura del receptor y en el ligando.
- El método *k*NN-QSAR (método en base al ligando) demostró ser una buena herramienta en la búsqueda de nuevas moléculas activas contra *E. histolytica*. Este método permitió encontrar 12 compuestos, los cuales presentaron actividad biológica. Los compuestos encontrados tienen núcleos del bencimidazol, benzotriazol e indazol. Estos dos últimos son de gran importancia, puesto que hasta el momento no hay reportes sobre la actividad biológica contra *E. histolytica* de este tipo de compuestos.
- Otro de los métodos en base al ligando que se utilizó fue el análisis comparativo de campos moleculares (CoMFA). Éste permitió, no sólo desarrollar modelos QSAR para estudiar los requerimientos estructurales que deben de tener los derivados del bencimidazol para tener actividad biológica contra *E. histolytica*, sino que también, fue una ayuda importante para encontrar cuál de las formas tautoméricas de los derivados del bencimidazol son significativas para la actividad biológica. En vista de los resultados obtenidos con CoMFA se puede concluir que los métodos QSAR, en particular CoMFA, es una alternativa para la selección del tautómero bioactivo cuando en el grupo de entrenamiento se presentan compuestos que pueden tener formas tautoméricas.
- Se observó que en los métodos QSAR es necesario hacer la validación de los modelos que se desarrollan. La validación interna y su parámetro de calidad  $q^2$  no es determinante para asegurar que el modelo obtenido tiene una buena predictibilidad. De tal forma que es necesario usar otros criterios, como los reportados por Golbraikh y Tropsha, que aseguren la predictibilidad del modelo

obtenido. Por lo tanto siempre es necesario usar un grupo de prueba para validar un modelo QSAR.

- En los modelos de CoMFA se demostró que no siempre es posible mejorar su estadística al usar métodos de cálculo de mecánica cuántica más avanzados, como lo es el método híbrido B3LYP de la teoría de funcionales de la densidad (DFT); en muchos casos, un cálculo a nivel semiempírico es suficiente para obtener un buen modelo, y además, ofrece la ventaja que se requiere menor tiempo de cálculo.
- El acoplamiento molecular o *docking*, el cual es un método en base a la estructura, mostró ser una excelente herramienta para estudiar los posibles modos de unión entre un ligando y un receptor. La correlación encontrada entre las energías libres de unión ( $\Delta G_{unión}$ ) y las actividades residuales de la triosafosfato isomerasa, ocasionadas por la interacción de los agentes inactivantes, sugieren que los modelos de unión encontrado por el programa AutoDock son representativos del complejo ligando-receptor.
- El estudio del modelo de acoplamiento ligando-receptor permite por visualización directa sugerir y diseñar nuevas moléculas con una mejor afinidad por el receptor. Las moléculas así diseñadas son un prototipo que tienen una mayor probabilidad de éxito de ser más activas que las moléculas a partir de las cuales fueron diseñadas.

## Referencias

- Ahmed, N.; Battah, S.; Karachalias, N.; Babaei-Jadidi, R.; Horanyi, M.; Baroti, K.; Hollan, S.; Thornalley, P. J. Increased formation of methylglyoxal and protein glycation, oxidation and nitrosation in triosephosphate isomerase deficiency. *Biochim. Biophys. Acta.* **2003**, 2, 121-32.
- Anandan, K.; Kolandaivel, P.; Kumaresan, R. Ab initio and DFT studies on tautomerism of indazole in gaseous and aqueous phases. *J. Mol. Strut.: THEOCHEM.* **2004**, 686, 83-89.
- Andrzejewska, M.; Yépez-Mulia, L.; Cedillo-Rivera, R.; Tapia, A.; Vilpo, L.; Kazimierczuk, J. Synthesis, antiprotozoal and anticancer activity of substituted 2-trifluoromethyl- and 2-pentafluoroethylbenzimidazoles. *Eur. J. Med. Chem.* **2002**, 12, 973-978.
- Baldwin, J. J.; Ponticello, G. S.; Anderson, P. S.; Christy, M. E.; Murcko, M. A.; Randall, W. C.; Schwam, H.; Sugrue, M. F.; Gautheron, P.; Grove, j.; Mallorga, P.; Viader, M-P.; McKeever, B. M.; Navia, M. A. Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. *J. Med. Chem.* **1989**, 32, 2510-2513.
- Bansal, D.; Malla, N.; Mahajan, R. C. Drug resistance in amoebiasis. *Indian J. Med. Res.* **2006**, 123, 115-118.
- Bendesky A.; Menendez, D.; Otrrosky-Wegman, P. Is metronidazole carcinogenic? *Mutat. Res.* **2002**, 511, 133-144.
- Bhongade, B. A.; Gadad, A. K. 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) inhibitors: a strategic design in novel anticancer agents. *Bioorg. Med. Chem.* **2004**, 12, 2797-2805.
- Böhm, H. J. The computer program LUDI: a new simple method for de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **1992**, 6, 61-78.
- Bucknham, J.; Glen, R. C.; Hill, A. P.; Hyde, R. M.; Martin, G. R.; Robertson, A. D.; Salmon, J. A.; Woollard, P. M. Computer-aided design and synthesis of 5-substituted tryptamines and their pharmacology at the 5-HT<sub>1D</sub> receptor: discovery of compounds with potential anti-migraine properties. *J. Med. Chem.* **1995**, 38, 3566-3580.
- Byington, C. L.; Dunbrack, R. L.; Whitby, F. G.; Cohen, F. E.; Agabian, N. *Entamoeba histolytica*: computer-assisted modeling of phosphofructokinase for prediction of broad-spectrum antiparasitic agents. *Exp. Parasitol.* **1997**, 87, 194-202.
- Cardozo, M. G.; Hawai, T.; Limura, Y.; Sugimoto, H.; Yamanishi, Y.; Hopifingfer, A.J. Conformational analyses and molecular-shape comparisons of a series of



- indanone-benzylpiperidine inhibitors of acetylcholinesterase. *J. Med. Chem.* **1992**, *35*, 590-601.
- (a) Cardozo, M. G.; Limura, Y.; Sugimoto, H.; Yamanishi, Y.; Hopfinger, A.J. Quantitative structure-activity relationship. QSAR, analyses of the substituted indanone and benzylpiperidine rings of a series of indanone-benzylpiperidine inhibitors of acetylcholinesterase. *J. Med. Chem.* **1992**, *35*, 584-589.
  - Cedillo-Rivera, R.; Muñoz, O. In-vitro susceptibility of *Giardia lamblia* to albendazole, mebendazole and other chemotherapeutic agents. *J. Med. Microbiol.* **1992**, *37*, 221-224.
  - Chavez, B.; Cedillo-Rivera, R.; Martinez-Palomo, A. *Giardia lamblia*: ultrastructural study of the in vitro effect of benzimidazoles. *J. Protozool.* **1992**, *39*, 510-515.
  - Christensen, I. T.; Jorgensen, F. S. Molecular Mechanics calculation of proteins. Comparison of different energy minimization strategies. *J. Biomol. Struct. Dyn.* **1997**, *15*, 473-487.
  - Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 198-210.
  - Civcir, P.Ü. A theoretical study of tautomerism of cytosine, thymine, uracil and their 1-methyl analogues in the gas and aqueous phases using AM1 and PM3. *J. Mol. Struct.: THEOCHEM.* **2000**, *532*, 157-169.
  - Claramun, R.; Avendaño, C. Optimización de un prototipo: Correlaciones estructura química-actividad biológica cuantitativas. En *Introducción a la Química Farmacéutica*. Avendaño, C., Mc Graw Hill, 2da Ed., **2001**; pp 85-113.
  - Clark, M.; Cramer R. D.; Opdenbosch, N. V. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982-1012.
  - Cohen, C. N.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. Molecular modeling software and methods for medicinal chemistry. *J. Med. Chem.* **1990**, *3*, 883-894.
  - Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
  - Dorsey, B. D.; Levin, R. B.; McDaniel, S.L.; Vacca, J. P.; Guare, J. P.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Quintero, J. C.; Lin, J. H.; Chen, I. W.; Holloway, M. K.; Fitzgerald, P. M. D.; Axel, M. G.; Ostovic, D.; Anderson, P. S.; UHF, J. R. L-735,524: The design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.* **1994**, *37*, 3443-3451.
  - Duarte, H. A.; Carvalho, S.; Paniago, E. B.; Simas, A. M. Importance of tautomers in the chemical behavior of tetracyclines. *J. Pharm. Sci.* **1999**, *88*, 111-120.

- Duncia, J. V.; Chiu, A. T.; Carini, D. J.; Gregory, G. B.; Johnson, A. L.; Price, W. A.; Wells, G. J.; Wong, P.C.; Calabrese, J. C.; Timmermans, P. B. M. W.M. The discovery of potent nonpeptide angiotensin II receptor antagonists: a new class of potent antihypertensives. *J. Med. Chem.* **1990**, *33*, 1312-1329.
- Eaazhisai, K. Balaram, H.; Balaram, P; Murthy, M. R. N. Structure of unliganded and inhibitor complexes of W168F, a loop 6 hinge mutant of *Plasmodium falciparum* triosephosphate isomerase: observation of an intermediate position of loop 6. *J. Mol. Biol.* **2004**, *343*, 671-684.
- El-Morshedy, H.; Farghaly, A.; Sharaf, S.; Abou-Basha, L.; Barakat, R. Triclabendazole in the treatment of human fascioliasis: a community-based study. *Eastern Mediterranean Health Journal.* **1999**, *5*, 888-894.
- Espinoza-Fonseca, M.; Trujillo-Ferrara, J. Exploring the possible binding sites at the interface of triosephosphate isomerase dimer as a potential target for anti-trypansomal drug design. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3151-3154.
- Espinoza-Fonseca, M.; Trujillo-Ferrara, J. Structural considerations for the rational design of selective anti-trypansomal agents: The role of the aromatic clusters at the interface of triosephosphate isomerase dimer. *Biochem. Biophys. Res. Commun.* **2005**, *328*, 922-928.
- Espinoza-Fonseca, M.; Trujillo-Ferrara, J. Toward a rational design of selective multi-trypansomatid inhibitors: A computational docking study. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 6288-6292.
- Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK4.0: Search strategies for automated molecular docking of flexible molecules databases. *J. Comput-Aided Mol. Des.* **2001**, *15*, 411-482.
- Free, S. M.; Wilson, J. W. A Mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395-399.
- Fujita, T. The extrathermodynamic approach to drug design. En *Comprehensive in Medicinal Chemistry* (C. A. Ramsden, ed.); Pergamon Press: New York, NY; **1990**, Vol 4, pp. 497-560.
- Gao, X.G.; Maldonado, E.; Perez-Montfort, R.; Garza-Ramos, G.; de Gomez-Puyou, M.T.; Gomez-Puyou, A.; Rodriguez-Romero, A. Crystal structure of triosephosphate isomerase from *Trypanosoma cruzi* in hexane. *Proc. Natl. Acad. Sci. USA.* **1999**, *96*, 10062-10067.
- Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron.* **1980**, *36*, 3219-3228.
- Gilbert, N. *Statistics*, W. B. Saunders, Co., Philadelphia, PA, 1976.

- Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the test set and training set selection. *J. Comput.-Aid. Mol. Des.* **2002**, *16*, 367-369.
- (a) Golbraikh, A.; Tropsha, A. Beware of  $q^2$ . *J. Mol. Graphics Model.* **2002**, *20*, 269-276.
- Gómez-Puyou, A.; Saavedra-Lira, E.; Becker, I.; Zubillaga, R. A.; Rojo-Domínguez, A.; Pérez-Monfort, R. Using evolutionary changes to achieve species-specific inhibition of enzyme action-studies with triosephosphate isomerase. *Chem & Biol.* **1995**, *2*, 847-855.
- Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biological important macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.
- Greco, G.; Novellino, E.; Martin, Y. C., Approaches to three-dimensional quantitative structure–activity relationships. En *Reviews in Computational Chemistry*. Lipkowitz, K.; Boyd, D., Eds; VCH Publishers, New York, **1997**, Vol. 11, pp. 183-240.
- Greer, J; Erickson, J. W.; Baldwin, J. L; Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37*, 1035-1054.
- Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. Computational modeling of the catalytic reaction in triosephosphate isomerase. *J. Mol. Biol.* **2004**, *337*, 227-239.
- Guillet, V. J.; Myatt, G.; Zsoldos, Z.; Johnson, A. P. SPROUT, HIPPO and CAESA: tools for de novo structure generation and estimation of synthetic accessibility. *Perspect. Drug Discov. Des.* **1995**, *3*, 34-50.
- Hall, L. H.; Kier, L. B. The molecular connectivity Chi indexes and kappa shape indexes in structure-property modeling. En *Reviews in Computational Chemistry II*; Lipkowitz, K. B.; Boyd, D. B. Eds.; VCH: Cambridge, U.K., **1991**, pp 367-422.
- Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins.* **2002**, *47*, 409-43.
- Hansch, C.; Fujita, T. r-s-p analysis: method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *85*, 1616-1626.
- Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391-3400.
- Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, E.; Streich, M The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J. Am. Chem. Soc.* **1963**, *85*, 2817-2824.

- Harris, T.; Cole, R.; Comer, F.; Mildvan, A. Proton transfer in the mechanism of triosephosphate isomerase. *Biochem.* **1998**, *37*, 16828-16838.
- Hernandez-Campos, A; Ibarra-Velarde, F.; Vera-Montenegro, Y.; Rivera-Fernandez, N.; Castillo R. Synthesis and fasciolicidal activity of 5-chloro-2-methylthio-6-(1-naphthyloxy)-1*H*-benzimidazole. *Chem Pharm Bull.* **2002**, *50*, 649-52.
- Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discov. Today.* **2004**, *9*, 659-669.
- Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H. A novel approach for the development of selective Cdk4 inhibitors: library design base on locations of Cdk4 specific amino acid residues. *J. Med. Chem.* **2001**, *44*, 4628-4640.
- Hopfinger, A. J. Computer-assisted drug design. *J. Med. Chem.* **1985**, *28*, 1133-1139.
- Houben, L.; Ramaekers, R.; Adamowicz, L.; Maes, G. Tautomerism and non-planarity of the amino group in 4(7)-aminobenzimidazole: A theoretical and Matrix-isolation FT-IR study. *Internet Electron. J. Mol. Des.* **2004**, *3*, 163-181.
- Instituto Nacional de Estadística, Geografía e Informatica/Dirección General de información y Evaluación del Desempeño, Secretaria de Salud de México, **2001**.
- Irwin, J. J.; Shoichet, B. ZINC – Afree database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
- Ji, H.; Zhang, W.; Zhang, M.; Kudo, M.; Aoyama, Y.; Yoshida, Y.; Sheng, C.; Song, Y.; Yang, S.; Zhou, Y.; Lü, J.; Zhu, J. Structure-based *de novo* design, synthesis and biological evaluation of non-azole inhibitors specific for lanosterol 14a-demethylase of fungi. *J. Med. Chem.* **2003**, *46*, 474-485.
- Jogl, G.; Rozovsky, S.; McDermott, A.; Tong L. Optimal alignment for enzymatic proton transfer: Structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. *PNAS.* **2003**, *100*, 50–55.
- Johnson, P. J. Metronidazole and drug resistance. *Parasitol. Today.* **1993**, *9*, 183-186.
- Joseph-McCarthy, D.; Rost, L.; Komives, E.; Petsko, G. Crystal structure of the mutant yeast triosephosphate isomerase in which the catalytic base glutamic acid 165 is changed to aspartic acid. *Biochem.* **1994**, *33*, 2824-2829.
- Joubert, F.; Neitz, A. W. H.; Louw, A. Structure-based inhibitor screening: a family of sulfonated dye inhibitors for malaria parasite triosephosphate isomerase. *Protein: Struct. Funct. Genet.* **2001**, *45*, 136-143.

- Kaldor, S. W.; Cáliz, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40*, 3979-3985.
- Karplus, M.; Cui, Q. Triosephosphate isomerase: A theoretical comparison of alternative pathways. *J. Am. Chem. Soc.* **2001**, *123*, 2284-2290.
- Kassai, T.; Cordero del Campillo, M.; Euzeby, J.; Gaafar, S.; Hiepe, Th.; Himonas, C. A. standarized nomenclature of animal parasitic diseases (SNOAPAD). *Vet. Parasitol.* **1988**, *29*, 299-326.
- Kawkami, Y.; Inoue, A; Kawwai, T.; Wakita, M.; Sugimoto, H.H; Hopfinger, A. J. The rationale for E2020 as a potent acetylcholinesterase inhibitor. *Bioorg. Med. Chem.* **1996**, *4*, 1429-1446.
- Keenan, R. M.; Weinstock, J.; Finkelstein, J.A.; Franz, R. G.; Gaitanopoulos, D. E.; Girard, G. R.; Hill, D. T.; Morgan, T: M; Samanen, J. M.; Peishoff, C. E.; Tucker, L. M.; Aiyar, N.; Griffin, E.; Ohlstein, E. H.; Stack, E. J.; Weidley, E. F.; Edwards, R. M. Potent nonpeptide angiotensin II receptor antagonist. 2. 1-(carboxybenzyl)imidazole-5-acrylic acids. *J. Med. Chem.* **1993**, *36*, 1880-1892.
- Keiser, J.; Engels, D.; Büscher, G.; Utzinger, J. Triclabendazole for the treatment of fascioliasis and paragonimiasis. *Exp. Opin. Invest. Drugs*, **2005**, *14*, 1513-1526.
- Kier, L.; Ball, L. H. Molecular connectivity in chemistry and drug research; Academic Press: New York, **1976**.
- Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *110*, 4130-4146.
- Knight, R. The chemotherapy of amoebiasis. *J. Antimicrob Chemother.* **1980**, *6*, 577-593.
- Koga, H.; Itoh, A.; Murayama, S.; Suzue, S.; Irikura, T. Structure-activity relationships of antibacterial 6,7- and 7,8-disubstituted 1-alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic acids. *J. Med. Chem.* **1980**, *23*, 1358-1363.
- Kohlhoff, M.; Dahm, A.; Hensel, R. Tetrameric triosephosphate isomerase from hyperthermophilic archaea. *FEBS Lett.* **1996**, *383*, 245-250.
- Kotani, T.; Higashiura, K., Comparative Molecular Active Site Analysis (CoMASA). 1. An Approach to Rapid Evaluation of 3D QSAR. *J. Med. Chem.* **2004**, *47*, 2732-2742.

- Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Curr. Comput-Aided Drug Des.* **2005**, *1*, 93-102.
- Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269-288.
- Kursula, I.; Wierenga, K. Crystal structure of triosephosphate isomerase complexed with 2-phosphoglycolate at 0.83-Å resolution. *J. Biol. Chem.* **2003**, *278*, 9544-9551.
- Kwiatkoski, M. J.; Leszczynski, J. Temperature-dependent thermodynamic contributions to the relative tautomeric stabilities of nucleic acid bases. *Chem. Phys. Lett.* **1993**, *204*, 430-434.
- Lakhdar-Ghazal, F.; Blonski, C.; Willson, M.; Michels P.; Perie, J. Glycolysis and proteases as targets for the design of new anti-trypanosome drugs. *Curr. Top. Med. Chem.* **2002**, *2*, 439-456.
- Landa, A.; Rojo-Dominguez, A.; Jiménez, L.; Fernandez-Velasco, D. A. Sequencing, expresion and properties of triosephosphate isomerase from *Entamoeba histolytica*. *Eur. J. Biochem.* **1997**, *247*, 348-355.
- Li, A. P. Screening for human ADME/Tox drug properties in drug discovery. *DDT.* **2001**, *6*, 357-366.
- Li, A. P., Early ADME/Tox studies and *in silico* screening. *DDT.* **2002**, *7*, 25-27.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **1997**, *23*, 3-25.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **2001**, *46*, 3-26.
- López-Rodríguez, M.; Murcia, M.; Benhamú, B.; Viso, A.; Campillo, M.; Pardo, L. 3-D-QSAR/CoMFA and recognition models of benzimidazole derivatives at the 5-HT(4) receptor. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2807-2811.
- López-Rodríguez, M.; Murcia, M.; Benhamú, B.; Viso, A.; Campillo, M.; Pardo, L. Benzimidazole Derivatives. 3. 3D-QSAR/CoMFA Model and computational simulation for the recognition of 5-HT4 receptor antagonists. *J. Med Chem.* **2002**, *45*, 4806-4815.
- Luyten, I.; Pankieewicz, K. W.; Watanabe, K. A.; Chattopadhyaya, J. Determination of the tautomeric equilibrium of Y-uridine in the basic solution. *J. Org. Chem.* **1998**, *63*, 1033-1040.
- Maeta, K.; Mori, K.; Takatsume, Y.; Izawa, S.; Inoue, Y. Diagnosis of cell death induced by methylglyoxal, a metabolite derived from glycolysis, in *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.* **2005**, *243*, 87-92.

- Maldonado, G.; Garza-Ramos, G. de Gomez- Puyou, M., Gomez- Puyou, A.; Rodríguez-Romero, A. Crystal structure of triosephosphate isomerase from *Trypanosoma cruzi* in hexane. *Proc. Natl. Acad. Sci. USA*. **1999**, *96*, 10062-10067.
- Mande, S. C.; Mainfroid, V.; Kalk, K. H.; Goraj, K.; Martial, J. A.; Hol, W. G. Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Sci*. **1994**, *3*, 810-821.
- Mande, S. C.; Mainfroid, V.; Kalk, K. H.; Goraj, K.; Martial, J. A.; Hol, W. G. Crystal structure of recombinant triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase related human genetic disorders and comparison with the trypanosomal enzyme. *Protein. Sci*. **1996**, *3*, 114-120.
- Martin, Y. C., A practitioner's perspective of the role of quantitative structure-activity analysis in medicinal chemistry. *J. Med. Chem*. **1981**, *24*, 229-237.
- Mattos, C.; Ringe, D. Locating and characterizing binding sites, *Nat. Biotechnol*. **1996**, *14*, 595-599.
- McGaughey, G.; Gagné, M.; Rappé, A.  $\pi$ -Stacking interactions, *J. Biol. Chem*. **1998**, *273*, 15458-15463.
- Medina-Franco, J. L.; Lopez-Vallejo, F.; Castillo R. Diseño de fármacos asistido por computadora, *Edu. Quím*. **2007**, *17*, 452-457
- Medina-Franco, J.; Golbraikh, A.; Oloff, S.; Castillo, R.; Tropsha, A. Quantitative structure-activity relationship analysis of pyridinone HIV-1 reverse transcriptase inhibitors using the k nearest neighbor method and QSAR-based database mining, *J Comput Aided Mol Des*. **2005**, *19*, 229-242.
- Menendez, D.; Rojas, E.; Herrera, L. A.; López, M. C.; Sordo, M.; Elizondo, G.; Otrosky-Wegman, P. DNA breakeage due to metronidazole treatment. *Mutat. Res*. **2001**, *478*, 153-158.
- Meyer, E.; Castellano, R.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition, *Angew. Chem. Int. Ed*. 2003, *42*, 1210-1250.
- Molconn-Z 4.05, Hall Associates Consulting: Quincy, MA.
- Monfort, R.; Tuena de Gómez-Puyou, M.; Gómez-Puyou, A. The interfaces of oligomeric proteins as targets for drug design against enzymes from parasites. *Curr. Top. Med. Chem*. **2002**, *2*, 457-470.
- Mülliken, R. S. Electronic population analysis on LCAO MO molecular wave function, *J. Chem. Phys*. **1955**, *23*, 1833-1840.
- Navarrete-Vazquez, G.; Cedillo, R.; Campos-Hernández, A.; Yépez, L.; Hernández-Luis, F.; Valdez, J.; Morales, R.; Hernández, M.; Castillo, R. Synthesis and

- antiparasitic activity of 2-(trifluoromethyl)-benzimidazole derivatives. *Bioorg. Med. Chem.Lett.* **2001**, *11*, 187-190.
- Navarrete-Vazquez, G.; Rojano-Vilchis, M.; Yépez, L.; Victor, M.; Gerena, L.; Campos-Hernandez, A.; Hernandez-Luis, F.; Castillo, R. Synthesis and antiprotozoal activity of some 2-(trifluoromethyl)-1H-benzimidazole bioisosteres. *Eur. J. Med Chem.* **2006**, *41*, 135-141.
  - Navarrete-Vazquez, G.; Yépez, L.; Campos-Hernández, A.; Amparo, T.; Hernández-Luis, F.; Cedillo, R.; Gonzales, J.; Martínez-Fernández, A.; Martínez-Grueiro, M.; Castillo, R. Synthesis and antiparasitic activity of albendazole and mebendazole analogues. *Bioorg. Med. Chem.* **2003**, *11*, 4615-4619.
  - Norel, R.; Lin, S. L.; Wolfson, H.; Nussinov, R. Shape complementarity at protein-protein interfaces. *Biopolymers*, **1994**, *34*, 933-940.
  - Norel, R.; Petrey, D.; Wolfson, H.; Nussinov, R. Examination of shape complementarity in docking of unbound proteins. *Proteins*. **1999**, *35*, 403-419.
  - Occhiato, E. G.; Ferrali, A.; Menchi, G.; Guarna, A.; Danza, G.; Comerci, A.; Mancina R.; Serio, M.; Gianni Garotta, A.; Cavalli, M de Vivo; Recanatini, M., Synthesis, biological activity, and three-dimensional quantitative structure-activity relationship model for a series of benzo(c)quinolizin-3-ones, nonsteroidal inhibitors of human steroid 5-reductase 1. *J. Med. Chem.* **2004**, *47*, 3546-3560.
  - Olivares-Illana, V.; Pérez-Monfort, R.; López-Calahorra, F.; Costas, M.; Rodríguez-Romero, A; Tuena de Gómez-Puyou, M.; Gómez-Puyou, A. Structural differences in triosephosphate isomerase from different species and discovery of a multitypanosomatid inhibitor. *Biochem.* **2006**, *45*, 2556 -2560.
  - Ooms, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.* **2000**, *7*, 141-158
  - Oprea, T. L.; Waller, C. L., Theoretical and practical aspects of three-dimensional quantitative structure-activity relationships. En *Reviews in Computational Chemistry*. Lipkowitz, K.; Boyd, D., Eds; VCH Publishers, New York, **1997**, Vol. 11, pp. 127-182.
  - Pensak, D. A. Molecular modelling: scientific and technological boundaries. *Pure & Appl. Chem.* **1989**, *62*, 601-603.
  - Pérez-Monfort, R; Garza-Ramos, G.; Hernández-Alcántara, G.; Reyes-Vivas, H.; Gao, X. G.; Maldonado, E. Derivatization of the interface cysteine of triosephosphate isomerase from *Trypanosoma brucei* and *Trypanosoma cruzi* as a probe of the interrelationship between the catalytic sites and the dimer interface. *Biochem.* **1999**, *38*, 4114-4120.
  - Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in computer-aided drug design. *J. Recept. Signal Transduct.* **2003**, *23*, 361-371.



- Purcel, W.; Singer, P. A brief review and table of semiempirical parameters used in the Hückel molecular orbital method. *J. Chem. Eng. Data.* **1967**, *12*, 235-246.
- Randic, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735-746.
- Reynoldson, J. A.; Thompson, R. C.; Horton, R. J. Albendazole as a future anti-giardial agent. *Parasitol. Today.* **1992**, *8*, 412-414.
- Ringe, D. What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **1995**, *5*, 825-829.
- Rodríguez-Romero, A.; Hernández-Santoyo, A.; Yauner, L. del P.; Kornhauser, A. Fernández-Valasco. Structure and inactivation of triosephosphate isomerase from *Entamoeba histolytica*. *J. Mol. Biol.* **2002**, *322*, 669-675.
- Rogers-Evans, M.; Alanine, A. I.; Bleicher, K. H.; Kube, D.; Schneider, G. Identification of novel cannabinoid receptor ligands via evolutionary de novo design and rapid parallel synthesis, *QSAR Comb. Sci.* **2004**, *23*, 426-430.
- Saavedra, E.; Encalada, R.; Pineda, E.; Jasso-Chávez, R.; Moreno-Sánchez, R. Glycolysis in *Entamoeba histolytica*: Biochemical characterization of recombinant glycolytic enzymes and flux control analysis. *FEBS Journal.* **2005**, *272*, 1767-1783.
- Saibil, H. R. Macromolecular structure determination by cryo-electron microscopy. *Acta Crysta.* **2000**, *D56*, 1215-1222.
- Sánchez-Díaz, R. Síntesis de derivados del 1-metilbencimidazol con actividad antihelmíntica. Tesis de Maestría. Facultad de Química, UNAM, México D.F. 1999.
- Schliebs, W.; Thanki, N.; Eritja, R.; Wierenga, R. Active site properties of monomeric triosephosphate isomerase (monoTIM) as deduced from mutational and structural studies. *Protein Sci.* **1996**, *5*, 229-239.
- Scheneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649-663.
- Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking. *Curr. Med. Chem.* **2004**, *11*, 91-107.
- Sears, S. D.; O'Jare, J. In vitro susceptibility of *Trichomonas vaginalis* to 50 antimicrobial agents. *Antimicrob. Agents. Chemother.* **1988**, *32*, 144-146.
- Senthilkumar, L.; Kolandaivel, P. Post Hartree-Fock and density functional theory studies on tautomerism of 6-thioxantine in gas phase and in solution. *J. Mol. Struct.: THEOCHEM.* **2003**, *638*, 69-78.

- Sharaf, m. A.; Illman, D. L.; Kowalski, b. R. Chemometrics; John Wiley & Sons, New York, **1986**, pp 234-239.
- Shen, M.; Béguin, C.; Golbraikh, A.; Stables J.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **2004**, *47*, 2356-2364.
- Shoichet, B. K.; McGovern, S.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Current Opin. Chem. Biol.* **2002**, *6*, 439-446.
- Silverman, R. The organic chemistry of drug design and drug action. Elsevier Science. **2004**, pp.29-33.
- Singh, K.; Maithal, K.; Balaram, H.; Balaram, P. Synthetic peptides as inactivators of multimeric enzymes: inhibition of *Plasmodium falciparum* triosephosphate isomerase by interface peptides. *FEBS Lett.* **2001**, *501*, 19-23.
- Sotriffer, C.; Klebe, G.; Stahl, M.; Böhm, H-J. Docking and scoring function/virtual screening. En Burger's Medicinal Chemistry and Drug Discovery, New Jersey, 2003, Volumen 1, 281-331.
- Stewart, J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209-220.
- Sybyl 6.8; Tripos Associates, Inc., St. Louis, MO, USA.
- Taylor, J. S.; Burnett, R. M. DARWIN: A program for docking flexible molecules. *Proteins* **2000**, *41*, 173-191.
- Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comp-Aided Mol. Design.* **2002**, *16*, 151-166.
- Tellez-Valencia, A.; Ávila-Rios, S.; Pérez-Monfort, R.; Rodríguez-Romero, A; Tuena de Gómez-Puyou, M.; López-Calahorra, F.; Gómez-Puyou, A. Highly specific inactivation of triosephosphate isomerase from *Trypanosoma cruzi*. *Biochem. Biophys. Res. Commun.* **2002**, *295*, 958-963.
- Tellez-Valencia, A.; Olivares-Illana, V.; Hernández-Santoyo, A. Pérez-Monfort, R.; Costas, M.; Rodríguez-Romero, A; López-Calahorra, F.; Tuena de Gómez-Puyou, M.; Gómez-Puyou, A. Inactivation of triosephosphate isomerase from *Trypanosoma cruzi* by an agent that perturbs its interface. *J. Mol. Biol.* **2004**, *341*, 1355-1365.
- Temiz-Arpaci, O.; Tekiner-Gulbas, B.; Yildiz, I.; Aki-Sener, E.; Yalcin, I. 3D-QSAR analysis on benzazole derivatives as eukaryotic topoisomerase II inhibitors by using comparative molecular field analysis method. *Bioorg. Med. Chem.* **2005**, *13*, 6354-6359.
- Tripos, Inc., St. Louis, MO.

- Turner, J. V.; Glass, B. D.; Agatonovic-Kustrin, S. Prediction of drug bioavailability based on molecular structure. *Anal. Chim. Acta.* **2003**, *485*, 89-102.
- Upcroft, J A.; Upcroft, P. Drug resistance and giardia. *Parasitol. Today*, **1993**, *9*, 187-190.
- Upcroft, P.; Upcroft, J A. Drug targets and mechanisms of resistance in the anaerobic protozoa. *Clin. Microbiol. Rev.* **2001**, *14*, 15-164.
- Valdez, J.; Cedillo, R.; Campos-Hernández, A.; Yépez, L.; Hernández-Luis, F.; Navarrete-Vazquez, G.; Hernández, M.; Castillo, R. Synthesis and antiparasitic activity of 1H-benzimidazole derivatives. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2221-2224.
- Vaz, R. J.; Maynard, G. D.; Kudlacz, E. M.; Bratton, L. D.; Kane, J. M.; Shatzer, S. A.; Knippenberg, R. W. Use of CoMFA in validating the conformation used in designing 4-(1H-benzimidazole-2-carbonyl)piperidines with H1/NK1 receptor antagonist activity. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2825-2830.
- Vera-Montenegro, Y.; Ibarra Velarde, F.; Quiroz Romero, H.; Hernández Campos, A.; Castillo, R. Field trial on the efficacy of an experimental fasciolicide compared with some commercial compounds in naturally infected cattle. *Parasitol. Res.* **2003**, *91*, 1-4.
- Veselovsky, A. V.; Ivanov, A. S. Strategy of computer-aided drug design current drug targets. *Infectious Disorders*, **2003**, *3*, 33-40.
- Von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B. Van Phan, T.; Smythe, M. L.; White, H. F. Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hothman, V. J.; Cameron, J. M.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, **1993**, *363*, 418-423.
- Voolmann, T.; Boreham, P. F. L. Metronidazole resistant *Trichomonas vaginalis* in brisbane. *Med. J. Aust.* **1993**, *159*, 490-497.
- Waley, Stephen. Refolding of triosephosphate isomerase. *Biochem. J.* **1973**, *135*, 165-172.
- Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Curr. Opin. Drug. Discov. Devel.* **2002**, *5*, 407-413.
- Waterbeemd, H.; Gifford, E., ADMET in silico modeling: towards prediction paradise. *Nat. Rev. Drug Disc.* **2003**, *2*, 192-204.
- Wavefunction, Inc., Irvine, CA.
- Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J. A new force field for molecular mechanical simulation of nucleic acids and proteins *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

- Weinstock, J.; Keenan, R. M.; Samanen, J.; Hempel, J.; Finkelstein, J.A.; Franz, R. G.; Gaitanopoulos, D. E.; Girard, G. R.; Gleason, J. G.; Hill, D. T.; Morgan, T. M.; Peishoff, C. E.; Aiyar, N.; Brooks, D. P.; Fredrickson, T. A.; Ohlstein, E. H.; Ruffolo, Jr. R. R.; Stack, E. J.; Sulpizio, A. C.; Weidley, E. F.; Edwards, R. M. 1-(Carboxybenzyl)imidazole-5-acrylic acids: potent and selective angiotensin II receptor antagonist. *J. Med. Chem.* **1991**, *34*, 1514-1517.
- Wermuth, C. G. Strategies in the search for new lead compounds or original working hypotheses. En *The Practice of Medicinal Chemistry*, Elsevier Science, **2003**, pp 65-83.
- (a) Wermuth, C. G. Identical and non-identical twin drugs. En *The Practice of Medicinal Chemistry*, Elsevier Science, **2003**, pp 251-273.
- Wieman, H.; Tondel, K.; Anderssen, E.; Drablos, F. Homology-based modelling of targets for rational drug design. *Mini-Rev. Med. Chem.* **2004**, *4*, 793-804.
- Wiench, J. W.; Koprowski, M.; Stefaniak, L.; Webb, G. A. Composition of tautomeric mixtures studied by nitrogen chemical shifts and *ab initio* molecular orbital calculations. *Polish J. Chem.* **2002**, *76*, 525-536.
- Wierenga, R. K.; Borchert, T. V.; Noble, M. E. M. Crystallographic binding studies with triosephosphate isomerase: conformational changes induced by substrate and substrate-analogues. *FEBS.* **1992**, *307*, 34-39.
- Wold, S.; Sjostrom, M.; Erikson, L. Partial least squares projection to latent structures (PLS) in chemistry. En *Encyclopedia of Computational Chemistry*. Schleyer, P. V. R., Ed; John Wiley and sons, West Sussex, UK, **1998**; Vol. 3, pp 2006-2021.
- Zavala J. T. Sánchez, V. T., Monografía, Laboratorio de Parasitología, Dpto. de Microbiología y Parasitología, Facultad de Medicina; UNAM, Unidad de Medicina Familiar No. 28, IMSS. **2001**.
- Zheng, W.; Tropsha, A., Novel variable selection quantitative structure-property relationship approach based on the *k*-nearest neighbor principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185-194.

# Molecular modeling of some 1*H*-benzimidazole derivatives with biological activity against *Entamoeba histolytica*: A comparative molecular field analysis study<sup>☆</sup>

Fabian López-Vallejo,<sup>a</sup> José Luis Medina-Franco,<sup>b</sup> Alicia Hernández-Campos,<sup>a</sup> Sergio Rodríguez-Morales,<sup>a</sup> Lilián Yépez,<sup>c</sup> Roberto Cedillo<sup>c</sup> and Rafael Castillo<sup>a,\*</sup>

<sup>a</sup>Departamento de Farmacia, Facultad de Química, UNAM, CU, D.F. 04510, Mexico

<sup>b</sup>BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA

<sup>c</sup>Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, IMSS. México, DF 06720, Mexico

Received 15 July 2006; revised 23 September 2006; accepted 11 October 2006

Available online 18 October 2006

**Abstract**—Comparative molecular field analysis (CoMFA) was performed on a set of 1*H*-benzimidazole derivatives. Molecular modeling and 3D-QSAR were employed to determine the tautomeric form that would probably fit a target receptor in *Entamoeba histolytica*. CoMFA results suggest that the antiameobic activity is favored with steric bulk at position 5 of the benzimidazole ring and low electron density on the group at position 2. To the best of our knowledge this is the first 3D-QSAR study performed for benzimidazoles as antiameobic agents. The CoMFA models derived will be very valuable to design new and more potent compounds against *E. histolytica*.

© 2006 Elsevier Ltd. All rights reserved.

## 1. Introduction

Parasitic infections such as amoebiasis and other protozooses are still major threats against public health, especially in developing countries. The infections caused by *Entamoeba histolytica* mainly affect the infant population with high morbidity and mortality indices due to severe diarrhea and dehydration.<sup>1</sup> Although drug therapy exists for the control of these infections, such as the nitroheterocycles, they produce severe side effects and in some cases resistance by the parasites have also been reported.<sup>2</sup> Consequently, the search for new therapeutic agents is of utmost importance.

We have recently reported that some 1*H*-benzimidazole derivatives with hydrogen at position 1, non 2-carbamates, were very active in vitro against *E. histolytica*.<sup>3–6</sup> Since there are no reports on the target receptor for these types of compounds and in order to obtain more informa-

tion regarding the structural requirements of benzimidazole derivatives to be active against *E. histolytica*, computational studies were carried out.

Three-dimensional quantitative structure–activity relationship (3D-QSAR) methods, such as comparative molecular field analysis (CoMFA),<sup>7</sup> have been applied in many cases for the design of new active molecules; especially, when the binding site and mechanism of action are known.<sup>8</sup> Additionally, it is also useful when the binding site is unknown.<sup>9</sup> Although CoMFA studies have been conducted for benzimidazole derivatives,<sup>10–14</sup> there are no computational analyses that study the activity against *E. histolytica*.

In order to rationalize the structure–activity relationships of a set of 41 1*H*-benzimidazole derivatives (cf., Table 1), in this paper 3D-QSAR studies using CoMFA were carried out. A particular challenge for the present CoMFA study is to deal with the tautomeric forms of compounds **3**, **8**, **13**, **29**, **32**, **36**, and **40**; and the tautomeric forms of compounds **6**, **7**, **9**, **10**, **17**, **21**, and **23** shown in Figures 1 and 2, respectively. The challenge is to determine which tautomer is most likely to be the active one when the QSAR studies are carried out. Each

**Keywords:** Amoebiasis; CoMFA; QSAR validation; Tautomers.

<sup>☆</sup> Taken in part from the Ph.D. dissertation of Fabian López-Vallejo.

\* Corresponding author. Tel./fax: +52 5622 5329; e-mail: rafaelc@servidor.unam.mx