

**UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS
FACULTAD DE QUÍMICA-FARMACIA
DEPARTAMENTO DE FARMACIA**



**“Métodos de Correlación Estructura-Actividad
Multiobjetivos Aplicados al Desarrollo Racional de
Fármacos”**

Tesis presentada en opción al grado científico de Doctor en Ciencias Farmacéuticas

Prof. Asist., Lic. Maykel Cruz Monteagudo

Santa Clara, Cuba

2009

**UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS
FACULTAD DE QUÍMICA-FARMACIA
DEPARTAMENTO DE FARMACIA**



**“Métodos de Correlación Estructura-Actividad
Multiobjetivos Aplicados al Desarrollo Racional de
Fármacos”**

Tesis presentada en opción al grado científico de Doctor en Ciencias Farmacéuticas

Autor: Prof. Asist., Lic. Maykel Cruz Monteagudo

Tutores: Prof. Tit., Ing. Elena Rosa Domínguez, Dr. C.

Prof. Tit., Lic. Maria Fernanda Martins Borges, Dr. C.

Santa Clara, Cuba

2009

SÍNTESIS

En la actualidad se ha comprobado que un pobre perfil farmacocinético y/o toxicológico del candidato constituye la principal causa de fallas en el proceso de descubrimiento de fármacos. Esto sugiere que la optimización de un compuesto líder basándose solamente en su actividad farmacológica ha sido sobreestimada. De aquí la necesidad de considerar simultáneamente las propiedades farmacológicas, toxicológicas y farmacocinéticas del candidato, lo que ha conducido a una reconsideración del paradigma secuencial del proceso de desarrollo de nuevos fármacos en favor de un enfoque multiobjetivos. En este trabajo, la capacidad de considerar simultáneamente múltiples propiedades se logra mediante la modificación de los diferentes elementos involucrados en la relación estructura-propiedad. Inicialmente se introduce una familia de descriptores moleculares basados en cadenas de Markov capaces de codificar de manera diferente una misma molécula en dependencia del sistema con que interactúe. Estos descriptores moleculares son empleados en el estudio simultáneo del perfil toxicológico (teniendo en cuenta múltiples efectos adversos asociados a varios sistemas de órganos) de un conjunto de fármacos. También se desarrolla un modelo unificado del proceso de partición capaz de clasificar un conjunto de compuestos estructuralmente diversos de acuerdo a su comportamiento en múltiples sistemas de partición bifásicos de relevancia biológica y ambiental. Posteriormente, son las propiedades las que son transformadas de manera que sea posible su estudio de forma simultánea. Para esto se propone un método de optimización multiobjetivos (OMO) basado en las funciones de deseabilidad de Derringer el cual permite conducir estudios globales de relación estructura-actividad al

ser capaz de considerar simultáneamente el perfil farmacológico, farmacocinético y toxicológico de un conjunto de candidatos. Los resultados del método de OMO son empleados como base para la aplicación de un nuevo método de ordenamiento también basado en funciones de deseabilidad. El ordenamiento se basa solamente en parámetros estimados a partir de la estructura molecular (grado de similitud estructural con el candidato óptimo determinado mediante el proceso de OMO) por lo que permite su aplicación a candidatos provenientes de bibliotecas combinatorias de los cuales solo se conoce su estructura. La estrategia de OMO fue aplicada en la obtención de un candidato a fármaco con un perfil analgésico, antiinflamatorio y ulcerogénico simultáneamente mejorado. Por su parte la combinación del proceso de OMO y ordenamiento se empleó en el desarrollo de un esquema de ordenamiento para el filtrado de candidatos con la mejor combinación posible de sus propiedades antibacterianas y citotóxicas. Todos estos resultados sugieren que los métodos multiobjetivos propuestos pueden aportar al proceso de desarrollo de fármacos un grado de racionalidad superior al alcanzado mediante los métodos actuales de correlación estructura-actividad contribuyendo al necesario cambio de paradigma del proceso.

A la memoria de Maykel Perez González.

A Silvana, por su paciencia.

Hay personas que no puedo dejar de mencionar...

A Humbe, porque gran parte de este trabajo es de el y la otra no hubiera podido hacerla sin lo que aprendí con el.

A mi familia, por estar siempre pendiente de mí.

Al fiero, Nereyda, Eduardo, Denia y los chiquillos por ser mi otra familia.

Evidentemente, no puedo olvidar a Silvana por estar siempre ahí y por ser como es.

Al profe Grau y a Leysi, por sus acertados comentarios y su dedicación en la revisión de este trabajo.

A Fernanda, la "doutorinha", por esa sonrisa eterna y por ese ánimo incansable.

A Elena por ser el puente entre la ingeniería y la farmacia... y sobre todo por su paciencia con mis despistes.

A mi gente del "grupo de diseño", a Aliuska, el chiqui, jabao y Belkis, Migue, el guille y Giselle, a Lianesilla... a los que están cerca, a los que están menos cerca, a los que ya no están...

A mi gente del otro lado del Atlántico, Cristian, Marta, Luis el gallego, Natalia, Angela y Eugenio...

A la gente del CEI, a Morell, Oscariño, Mario Pupo... a la gente del CEQA, Vladimir, Belkis, Marlene, Surey, Guille, Alina...

Y a los que por mi mala memoria olvidé mencionar pero que también me hicieron más fácil el camino hasta aquí.

A todos ustedes...

GRACIAS

TABLA DE CONTENIDOS

GLOSARIO

INTRODUCCIÓN.....	1
-------------------	---

1 QUIMIOINFORMÁTICA, QUIMIOMETRÍA Y DESARROLLO RACIONAL DE

FÁRMACOS.....	8
---------------	---

1.1 INTRODUCCIÓN A LA INVESTIGACIÓN TERAPÉUTICA MODERNA.....	8
--	---

1.1.1 Elementos biofarmacéuticos de la investigación terapéutica.....	8
---	---

1.1.2 Generalidades del proceso de descubrimiento de fármacos	12
---	----

1.1.3 Métodos computacionales	16
-------------------------------------	----

1.2 TEMAS ABIERTOS EN QUIMIOINFORMÁTICA.....	21
--	----

1.2.1 Representación y codificación de la estructura molecular.....	21
---	----

1.2.2 Determinación de la similitud molecular.....	27
--	----

1.2.3 Quimiometría y modelación de la relación estructura-actividad.....	29
--	----

2 MATERIALES Y MÉTODOS.....	44
------------------------------------	-----------

2.1 ESTUDIOS DE CORRELACIÓN ESTRUCTURA-PROPIEDAD MULTISISTEMAS BASADOS EN EL USO DE ÍNDICES MARKOVIANOS SISTEMA-DEPENDIENTES.....	44
--	----

2.1.1 Cálculo de los Índices Markovianos Sistema-Dependientes.....	44
--	----

2.1.2 Conformación de la base de datos	44
--	----

2.1.2.1 Efectos adversos de fármacos sobre múltiples sistemas de órganos	44
---	----

2.1.2.2 Coeficientes de partición de fármacos, compuestos químicos y contaminantes ambientales en múltiples sistemas de partición bifásicos	44
2.1.3 Análisis estadístico	46
2.1.4 Ensayos biológicos y animales de laboratorio	48
2.2.1 Optimización multiobjetivos basada en la estimación de la deseabilidad de varias respuestas interrelacionadas	50
2.2.2 Ordenamiento de soluciones basado en funciones de deseabilidad	55
2.2.2.1 Validación del método de ordenamiento y estimación del índice de calidad del ordenamiento	57
2.2.3 Conformación de la base de datos	59
2.2.3.1 Perfil antiinflamatorio, analgésico y ulcerogénico de quinazolinonas	59
2.2.3.2 Perfil antibacteriano y citotóxico de fluoroquinolonas.....	60
2.2.4 Especificaciones de las funciones de deseabilidad	60
2.2.5 Métodos computacionales y análisis estadístico	62
2.2.6 Análisis comparativo de distintos enfoques de tamizaje virtual	63
3 RESULTADOS Y DISCUSIÓN.....	65
3.1 ESTUDIOS DE CORRELACIÓN ESTRUCTURA-PROPIEDAD MULTISISTEMAS BASADOS EN EL USO DE ÍNDICES MARKOVIANOS SISTEMA-DEPENDIENTES.....	66
3.1.1 Índices Markovianos Sistema-Dependientes	66
3.1.2 Predicción de múltiples efectos adversos de medicamentos.....	73
3.1.3 Predicción del perfil de distribución de fármacos y compuestos químicos en sistemas de partición bifásicos de relevancia biológica.....	77

3.1.4 Predicción del perfil de distribución de fármacos, compuestos químicos y contaminantes ambientales en sistemas de partición bifásicos de relevancia biológica y ambiental	80
3.2 EMPLEO DE MÉTODOS DE OPTIMIZACIÓN MULTIOBJETIVOS Y ORDENAMIENTO BASADOS EN FUNCIONES DE DESEABILIDAD EN ESTUDIOS GLOBALES DE CORRELACIÓN ESTRUCTURA-ACTIVIDAD.....	84
3.2.1 Optimización Simultánea del perfil analgésico, antiinflamatorio y ulcerogénico de una familia de quinazolinonas	84
3.2.2 Optimización simultánea del perfil antibacteriano y citotóxico de fluoroquinolonas	91
CONCLUSIONES.....	99
RECOMENDACIONES.....	100
REFERENCIAS	
BIBLIOGRÁFICAS.....	102
ANEXOS	

GLOSARIO

Aa	Actividad antiinflamatoria
ABC	Área bajo la curva
ACF	Familia de descriptores moleculares basados en fragmentos centrados en átomos, <u>Atom Centred Fragments</u>
ACP	Análisis de Componentes Principales
ADG	Análisis Discriminante General
ADL	Análisis Discriminante Lineal
ADME	Absorción, Distribución, Metabolismo y Excreción
AG	Algoritmo Genético
AINES	Antiinflamatorios no Esteroides
An	Actividad analgésica
C ₀	Concentración Inicial
CECMED	Centro para el Control Estatal de la Calidad de los Medicamentos
CENPALAB	Centro Nacional de Animales de Laboratorio
CI	Aclaramiento plasmático
CM	Cadenas de Harkov
<u>CoMFA</u>	<u>Comparative Molecular Field Analysis</u>
COX	Ciclooxigenasa
CTP	Con tendencia a la partición
DDNF	Descubrimiento y desarrollo de nuevos fármacos
DM	Descriptor Molecular
F	Biodisponibilidad
fl	Fracción de fármaco libre
<u>FDA</u>	<u>Food and Drugs Administration</u>
<u>HTS</u>	<u>High-Throughput Screening</u>
IC ₅₀	Concentración inhibitoria media
IM	Índice Molecular
IT	Índice Topológico
Log P	Coefficiente de partición del sistema bifásico (expresado como su logaritmo natural)
<u>LOO-CV</u>	Procedimiento de validación cruzada dejando un caso fuera, <u>leave-one-out cross validation</u>

GLOSARIO

<u>MARCH-INSIDE</u>	<u>MAR</u> kovian <u>CH</u> emical <u>IN</u> <u>SI</u> llico <u>DE</u> sign
MCO	Mínimos Cuadrados Ordinarios
MCP	Mínimos Cuadrados Parciales
MIC	Mínima Concentración Inhibitoria
MLG	Modelos Lineales Generales
<u>MOOP-DESIRE</u>	<u>M</u> ulti- <u>O</u> bjective <u>O</u> ptimization based on the <u>D</u> esirability <u>E</u> stimation of <u>S</u> everal <u>I</u> nterrelated <u>R</u> esponses
MP	Modelo de predicción
MRG	Modelos de Regresión Generales
MSC	Máximo sub-grafo común
<u>OLAP</u>	<u>O</u> n <u>L</u> ine <u>A</u> nalytical <u>P</u> rocessing
OMO	Optimización multiobjetivos
PR	Punto de ruptura
<u>QSAR</u>	<u>Q</u> uantitative <u>S</u> tructure- <u>A</u> ctivity <u>R</u> elationships
<u>QSPR</u>	<u>Q</u> uantitative <u>S</u> tructure- <u>P</u> roperty <u>R</u> elationships
RAPG	Receptores acoplados a proteínas G
REA	Relación Estructura-Actividad
REP	Relación Estructura-Propiedad
RL	Regresión Logística
RLM	Regresión Lineal Múltiple
RNAs	Redes Neuronales Artificiales
SPB	Sistema de partición bifásica
SPFO	Suma Ponderada de Funciones Objetivos
STP	Sin tendencia a la partición
<u>SVM</u>	<u>S</u> upport <u>V</u> ector <u>M</u> achines
$t_{1/2}$	Tiempo de vida media
Tr	Operador Traza de una matriz
U	Índice ulcerogénico
Vd	Volumen de distribución
Vdl	Volumen de distribución de la fracción libre

INTRODUCCIÓN

INTRODUCCIÓN

El descubrimiento y desarrollo de nuevos medicamentos es un proceso en extremo complejo y costoso en términos de tiempo y dinero. Durante la década de los 90 el costo promedio del desarrollo de un nuevo fármaco, desde la identificación del blanco hasta su aprobación, estaba alrededor de los 15 años y US\$ 900 millones (1). En las dos décadas pasadas sin embargo, los avances tecnológicos, junto con la progresiva reconsideración del proceso en si, ha conducido a una gran revolución del proceso de encontrar un nuevo fármaco.

Hasta los años 80, el paso de generación de hits o candidatos potenciales (moléculas que muestran una determinada actividad química pero que no necesariamente cumple con los requerimientos de eficiencia de un lead o compuesto líder) constituía el principal paso limitante del proceso de descubrimiento y desarrollo de nuevos fármacos (DDNF) debido al costo de la síntesis y evaluación de nuevas moléculas (2, 3). Durante esta etapa las esperanzas de resolver el problema del DDNF fueron puestas en el desarrollo de las tecnologías de alto rendimiento (4, 5) y la química combinatoria (6, 7), a través de una paralelización masiva del proceso. Se esperaba que la sinergia entre estas dos tecnologías facilitara el ciclo de descubrimiento de fármacos a través de la evaluación sistemática de grandes cantidades de compuestos (3, 8).

En la práctica se evidenció que si no eran utilizadas cuidadosamente, el uso indiscriminado de estas técnicas podría conducir a un aumento dramático del número de moléculas o candidatos, de manera que el descubrimiento de un nuevo fármaco sería como hallar una aguja en un pajar. Mientras que el número de hits

identificados pudo ser incrementado sustancialmente, se observó que no existía una correspondencia con el crecimiento del número de fármacos que entraban al mercado, dejando esto claro que el verdadero paso limitante del descubrimiento de fármacos no era la generación de hits, sino los pasos de identificación y optimización del compuesto líder (2, 9). Como resultado, este tipo de solución a gran escala ha sido abandonada progresivamente en los últimos años, favoreciéndose una racionalización del proceso, en la que los métodos computacionales han ganado una importancia creciente (9).

Desarrollar un medicamento exitoso es el resultado del descubrimiento del mejor compromiso entre numerosos objetivos que muy a menudo compiten entre si. El fracaso de un candidato a fármaco con una potencia adecuada durante el proceso de desarrollo es principalmente debido a una pobre biodisponibilidad, y/o toxicidad (10). De forma simplificada, el fármaco ideal debería tener la mayor eficacia terapéutica y biodisponibilidad y la mínima toxicidad posible; lo que evidencia la naturaleza multiobjetiva del proceso de descubrimiento de fármacos.

Lo anterior sugiere que en la fase de optimización del líder, la capacidad de mejorar el perfil terapéutico del candidato seleccionado basándose solamente en su actividad farmacológica se ha sobreestimado, lo que refuerza la necesidad de considerar las propiedades toxicológicas y farmacocinéticas del candidato paralelamente a sus propiedades farmacológicas en etapas anteriores a la optimización, durante la fase de identificación del líder (2, 9).

Todo lo anterior ha llevado tanto a la academia como a la industria farmacéutica a una reconsideración del paradigma secuencial del proceso de descubrimiento de

fármacos en favor de un enfoque multiobjetivos del proceso del mismo. Este cambio de paradigma marca un avance significativo hacia su racionalización (1, 2). En este sentido, los métodos de optimización multiobjetivos (OMO) introducen una nueva filosofía para la optimización basada en la obtención de compromisos entre varios objetivos. Por medio de los métodos OMO es posible encontrar una solución óptima global mediante la optimización simultánea de múltiples propiedades interrelacionadas (11).

Específicamente, el paradigma de las relaciones cuantitativas estructura-actividad (identificado por sus siglas en inglés, QSAR) por mucho tiempo ha sido de interés para el descubrimiento de fármacos (12, 13). Sin embargo, a pesar de los avances logrados en el campo de la quimioinformática, la naturaleza multiobjetiva del proceso todavía es ignorada por estos métodos al optimizar secuencialmente cada propiedad involucrada en el perfil terapéutico global de un candidato (14-29).

En este sentido, una forma de manejar múltiples criterios mediante un único enfoque lo constituye la redefinición de descriptores moleculares previamente definidos (30). En este sentido, los descriptores moleculares basados en la teoría de las cadenas de Markov (CM) (31), a diferencia de otros descriptores moleculares también de naturaleza topológica pero basados en representaciones algebraicas que tienden a ser más abstractas (32, 33) ofrecen una interpretación físicamente coherente de los fenómenos abordados mediante su empleo (34, 35). El trabajo de González-Díaz y colaboradores (20, 36-74) es un ejemplo de lo anteriormente expuesto.

A su vez, a pesar de la disponibilidad de numerosos objetivos de optimización, las técnicas de optimización multiobjetivos han sido sólo recientemente aplicadas al

área de los estudios QSAR. Realmente, muy poca información existe acerca de la aplicación de estos métodos a estudios QSAR (75-77). Hasta hoy es realmente escaso el reporte de estudios QSAR que consideren la optimización simultánea de múltiples objetivos directamente relacionados con el perfil terapéutico global del fármaco como la eficacia terapéutica, biodisponibilidad y/o toxicidad del mismo.

Por todo lo anterior se plantea el siguiente problema científico:

Los métodos actuales de correlación estructura-actividad empleados en el desarrollo de nuevos fármacos no consideran la naturaleza multiobjetiva de este proceso al optimizar secuencialmente cada propiedad involucrada en el perfil terapéutico de un fármaco.

Para dar respuesta al problema científico planteado, se propone el siguiente objetivo general:

Objetivo General

Establecer una metodología de correlación estructura-actividad multiobjetivos, que permita predecir el perfil de biodistribución o toxicológico de un fármaco en múltiples sistemas biológicos, que considere simultáneamente y de modo integrador múltiples objetivos involucrados en el perfil terapéutico de un fármaco; y que aporte al desarrollo de fármacos un grado de racionalidad superior al alcanzado mediante los métodos actuales de correlación estructura-actividad.

Objetivos Específicos

- ✓ Introducir una familia de descriptores moleculares basados en cadenas de Markov que codifiquen de manera diferente una misma molécula en dependencia del sistema con que interactúe, y emplear estos descriptores en la obtención de modelos de correlación estructura-actividad que permitan predecir el perfil de biodistribución o toxicológico de un fármaco en múltiples sistemas biológicos.
- ✓ Obtener y validar una metodología de optimización multiobjetivos basada en funciones de deseabilidad que permita determinar los niveles de las variables predictivas necesarios para obtener un candidato a fármaco con una combinación óptima de propiedades involucradas en el perfil terapéutico global del mismo.
- ✓ Obtener y validar una metodología que permita ordenar candidatos a fármaco por su similitud respecto al candidato óptimo derivado del proceso de optimización multiobjetivos.

Los resultados derivados de la ejecución de los objetivos permitirán evaluar la veracidad de la siguiente hipótesis:

Un método capaz de considerar de manera simultánea múltiples objetivos involucrados en el perfil terapéutico de un fármaco puede aportar al proceso desarrollo de nuevos fármacos un grado de racionalidad superior al alcanzado mediante el empleo secuencial de los métodos actuales de correlación estructura-actividad.

La *novedad científica* de este trabajo se fundamenta en la obtención de una metodología de correlación estructura-actividad global capaz de considerar simultáneamente múltiples objetivos involucrados en el perfil terapéutico de un fármaco a través del empleo de: i) una familia de descriptores moleculares basados en cadenas de Markov capaces de codificar de manera diferente una misma molécula en dependencia del sistema con que interactúe; ii) una metodología de optimización multiobjetivos y ordenamiento basada en las funciones de deseabilidad de Derringer.

El *aporte teórico* de este trabajo se centra en: i) la introducción de una familia de descriptores moleculares basados en cadenas de Markov que codifiquen de manera diferente una misma molécula en dependencia del sistema con que interactúe; ii) la obtención de una metodología de optimización multiobjetivos y ordenamiento basada en las funciones de deseabilidad de Derringer; iii) la propuesta de criterios cuantitativos de la calidad del proceso de optimización multiobjetivos, de un método de validación del proceso de ordenamiento y de una medida cuantitativa de la calidad del mismo.

El *aporte práctico* del trabajo consiste en el empleo de modelos de correlación estructura-actividad multisistemas basados en cadenas de Markov, así como de una metodología de optimización multiobjetivos y ordenamiento basada en funciones de deseabilidad en el proceso de descubrimiento y desarrollo de nuevos candidatos a fármaco con perfiles farmacológico, farmacocinético y/o toxicológico simultáneamente mejorados. La metodología esta dirigida a conferir racionalidad al proceso de descubrimiento y desarrollo de nuevos fármacos.

El *aporte metodológico* reside en: i) el uso de la teoría de las cadenas de Markov como fuente de descriptores moleculares capaces de describir fenómenos biológicos de interés farmacéutico y químico-médico; y ii) en la integración de métodos y teorías ya establecidas (estudios de correlación estructura-actividad, las funciones de deseabilidad de Derringer y métodos de optimización) que resultan en una metodología de optimización multiobjetivos aplicable a estudios globales de correlación estructura-actividad.

La presente tesis se encuentra estructurada de la siguiente manera: la síntesis, el índice que se emplea en la tesis y un glosario de términos y abreviaturas que pueden ayudar a la lectura del documento. A la presente introducción le siguen los tres capítulos que forman el cuerpo de la tesis. El capítulo I está dedicado al empleo de la quimiinformática y la quimiometría en el proceso de descubrimiento y desarrollo racional de nuevos fármacos, el segundo capítulo se dedica a los materiales y métodos empleados en el trabajo y finalmente en el tercer capítulo se exponen y comentan los principales resultados del trabajo de tesis. A ello le siguen las conclusiones, las recomendaciones, las referencias bibliográficas y los anexos. Esta última sección incluye una serie de tablas que ofrecen detalles de los resultados presentados en esta tesis y ayudan a comprobar la reproducibilidad de los experimentos realizados, así como la producción científica del autor, subdividida en: publicaciones originales que constituyen el hilo conductor de los resultados y que se refieren con números romanos (**I** a **VII**); publicaciones relacionadas con la tesis o el autor, igualmente numerados románicamente (**VIII** a **XVI**); y la participación en eventos científicos con una reseña de los principales premios alcanzados.

1. QUIMIOINFORMÁTICA, QUIMIOMETRÍA Y DESARROLLO RACIONAL DE FÁRMACOS

1 QUIMIOINFORMÁTICA, QUIMIOMETRÍA Y DESARROLLO RACIONAL DE FÁRMACOS

1.1 Introducción a la Investigación Terapéutica Moderna

Este capítulo ofrece una perspectiva general de la investigación terapéutica así como los principios y bases teóricas de las herramientas quimioinformáticas empleadas en el desarrollo racional de fármacos, dirigida a no especialistas en las respectivas áreas.

1.1.1 Elementos biofarmacéuticos de la investigación terapéutica

Relaciones dosis-efecto y dosis-respuesta. Se denomina *efecto* a la manifestación de la acción de un fármaco que modifica algún mecanismo biológico o función fisiológica. Este cambio debido a la interacción a nivel molecular entre el xenobiótico y los constituyentes biológicos puede no ser evidente (efecto subclínico) o manifiesto (efecto clínico), siendo posible en ocasiones medirlo de forma objetiva. El efecto va indefectiblemente ligado a dos variables: dosis y tiempo, aunque con frecuencia se considera solo el binomio dosis-efecto. Esta relación puede ser de dos tipos:

- *Cuántica*: Responde a la ley del todo o nada; ante una dosis el individuo presenta el máximo efecto posible o no experimenta nada.
- *Gradual*: El efecto es función de la dosis; la representación gráfica de esta relación se aproxima (porque generalmente es una nube de puntos) a una hipérbola que muestra como al aumentar la dosis se incrementa el efecto, hasta llegar a una dosis máxima en que se alcanza un efecto máximo, y la curva se

hace asintótica. En ocasiones la correlación fluctúa y gráficamente es una sigmoide. Cuando se representa el efecto frente al logaritmo de la dosis se obtiene una recta.

De estas curvas es posible extrapolar la respuesta que producirá una dosis. Al representar gráficamente la relación dosis-efecto de un fármaco, encontramos las siguientes áreas de correspondencia: zona sin efecto aparente favorable o desfavorable (A); zona de aplicación terapéutica (B); zona de efectos tóxicos transitorios o reversibles (C); y zona de efectos tóxicos irreversibles (D).

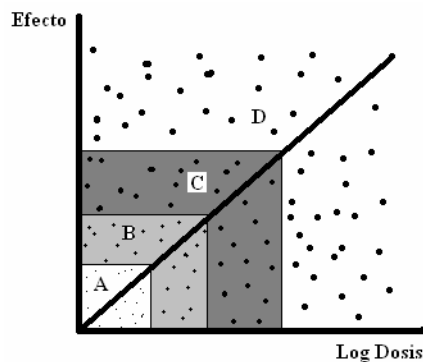


Figura 1.1 Zonas de relación dosis-efecto. (Adaptado de: Repetto, M. Toxicología Fundamental. Científico Médica. Barcelona. 1988).

Farmacocinética. De manera general, la farmacocinética está relacionada con el comportamiento de un fármaco en el organismo, tales como su capacidad de pasar al torrente circulatorio y alcanzar el blanco, y ser posteriormente destruido y eliminado por el organismo. Las principales propiedades farmacocinéticas se resumen en el acrónimo ADME, incluyendo los procesos de Absorción, Distribución, Metabolismo y Excreción (3, 78).

La mayoría de los fármacos se administran oralmente por razones de conveniencia y comodidad. Típicamente, un fármaco se disuelve en el tracto gastrointestinal, se absorbe a través de la pared intestinal y, a continuación, pasa a través del hígado a la circulación sanguínea. El porcentaje de la dosis que alcanza la circulación determina la biodisponibilidad (F). A partir de aquí el fármaco se distribuirá a los diferentes tejidos y órganos en el cuerpo. El alcance de la distribución dependerá de la estructura y las propiedades fisicoquímicas de los compuestos. Algunos fármacos logran llegar al sistema nervioso central al atravesar la barrera hematoencefálica. Finalmente, el fármaco se une al blanco molecular (una proteína, un receptor o un canal de iones) y ejerce la acción deseada.

Los parámetros farmacocinéticos se derivan de la medición de la concentración del fármaco en la sangre o plasma. Entre los más relevantes se encuentran:

- *Volumen de distribución (V_d)*. Es un concepto teórico que relaciona la dosis administrada con la concentración inicial real (C_0) presente en la circulación: $V_d = \text{Dosis} / C_0$.

La mayoría de los fármacos se unirán a varios tejidos y en particular a proteínas plasmáticas, tales como la albúmina. Ya que solo la fracción libre del fármaco se unirá al blanco molecular, se usa el concepto de volumen de distribución de la fracción libre: $V_{dl} = V_d / f_l$, donde f_l es la fracción de fármaco libre.

- *Aclaramiento plasmático (Cl)*. El aclaramiento o eliminación del fármaco del cuerpo tiene lugar principalmente por vía hepática (metabolismo o aclaramiento hepático, o excreción biliar) o renal (excreción renal).

Ploteando la concentración plasmática contra el tiempo, se obtiene el área bajo la curva (*ABC*) que relaciona la dosis, la biodisponibilidad y el aclaramiento: $ABC = F \times Dosis / Cl$.

- *Tiempo de vida media* ($t_{1/2}$). Es el tiempo requerido para reducir la concentración plasmática de un fármaco al 50%. El mismo es una función del aclaramiento y el volumen de distribución, y determina la frecuencia de administración de un fármaco: $t_{1/2} = 0.693 Vd / Cl$.

Los principales parámetros farmacocinéticos y su importancia para determinar los regímenes de dosis se muestran en la Figura 1.2 (79).

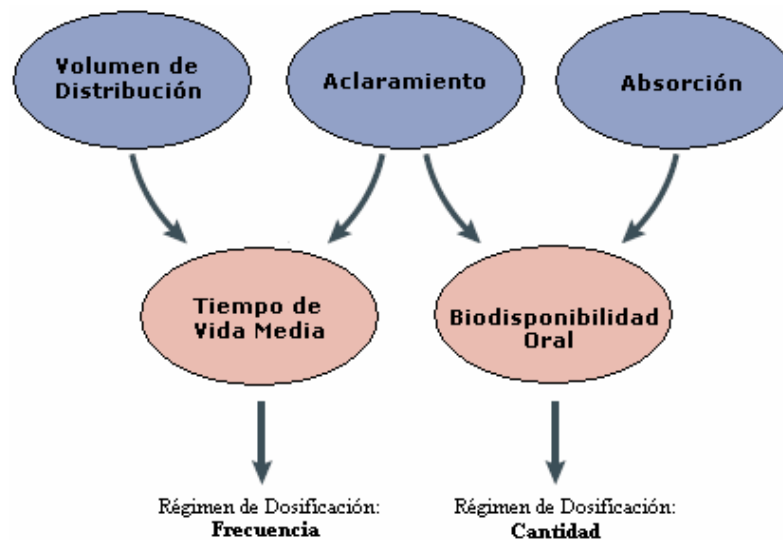


Figura 1.2 Principales parámetros farmacocinéticos y su importancia en la determinación los regímenes de dosificación. (Adaptado de: van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov.* 2003; 2:192-204).

Metabolismo. Eventualmente el cuerpo trata de eliminar los xenobióticos, incluidos los fármacos. Para muchos fármacos, esto requiere de procesos previos de

metabolismo o biotransformación, los que tienen lugar en parte en el intestino durante la absorción, pero principalmente en el hígado. La fracción de la dosis inicial que entra a la vena porta es la fracción absorbida, y la fracción que alcanza la circulación sanguínea luego del primer paso a través del hígado determina la biodisponibilidad del fármaco.

Tradicionalmente se ha hecho una distinción entre el metabolismo de fase I y de fase II, aunque esto no ocurra necesariamente de forma secuencial. En el metabolismo de fase I, la molécula es funcionalizada, por ejemplo, a través de procesos de oxidación, reducción, o hidrólisis. Las enzimas más importantes involucradas pertenecen a la superfamilia de citocromos P450. En el metabolismo de fase II, las moléculas de fármaco funcionalizadas son posteriormente transformadas mediante las conocidas reacciones de conjugación. Estas incluyen por ejemplo, reacciones de conjugación con el glutatión y con el ácido glucourónico. Se debe resaltar que el metabolismo en animales puede ser diferente del humano, y por tanto la predicción del metabolismo y la farmacocinética en humanos a partir de datos provenientes de animales puede no resultar sencilla.

1.1.2 Generalidades del proceso de descubrimiento de fármacos

El desarrollo de una terapia para una patología específica es un proceso usualmente articulado en tres pasos. El primer paso (identificación del blanco biológico o terapéutico) consiste en la identificación de una molécula biológica, mayormente proteínas, involucrada en algún mecanismo que conduce a un proceso patológico. El propósito del segundo paso es identificar una molécula con un perfil biológico interesante capaz de interferir con el blanco terapéutico antes

mencionado. Eventualmente, antes de que el candidato a fármaco entre al mercado, el tercer paso (validación clínica) debe probar su eficiencia y seguridad a través de una evaluación extensiva en animales y humanos (80-82).

Identificación del blanco biológico o terapéutico. El objetivo principal en la investigación terapéutica es interferir alguna vía o señal metabólica responsable de una enfermedad o proceso patológico. Las vías o señales metabólicas son cascadas de reacciones químicas intracelulares que llevan respectivamente a la formación de un producto metabólico que es usado por la célula, o a una alteración de la expresión de un gen debido a la activación de factores de transcripción. La tarea de la investigación terapéutica es encontrar una molécula de fármaco capaz de modificar esta vía mediante la alteración de una entidad clave, usualmente una proteína, involucrada en la cascada de reacciones correspondiente: el blanco terapéutico. La identificación del blanco involucra conocimientos tanto biológicos como químicos, con el objetivo de descubrir blancos potenciales y conocer en que medida este puede ser alterado por una molécula de fármaco, lo que se conoce en este ámbito como drugability (80).

Previo a la fase de descubrimiento de fármacos, el blanco terapéutico identificado debe ser validado con el objetivo de demostrar su papel determinante en la enfermedad. Esta validación usualmente involucra experimentos *in vitro* e *in vivo* (83, 84).

Descubrimiento de fármacos. En este segundo paso, el objetivo es encontrar una molécula pequeña, denominada ligando, capaz de unirse mediante fuerzas intermoleculares al blanco biológico y alterar su normal funcionamiento. Esta

interacción se dice que es directa cuando el fármaco se une al sitio activo del blanco y compite con su sustrato natural, o indirecta si el fármaco se une a un sitio secundario e induce cambios en la conformación química del blanco, modulando así su afinidad con el ligando natural (85).

Para cuantificar la actividad del ligando, correspondiente al grado de interacción con el blanco, se debe diseñar un procedimiento experimental denominado ensayo de la actividad biológica. La actividad de las moléculas candidato puede ser subsecuentemente ensayada con el objetivo de encontrar candidatos a fármaco, o compuestos líderes, capaces de interferir con el blanco a bajas concentraciones (80).

Teóricamente, el número de moléculas candidato con características de fármaco simples que sería posible sintetizar se encuentra en el orden de 10^{60} (8). La identificación de candidatos prometedores en esta vasta (casi infinita) cantidad de moléculas depende fuertemente de la pericia bioquímica y tradicionalmente se logra en un proceso iterativo, a menudo denominado como el ciclo de descubrimiento de fármacos que alterna entre los pasos de selección, síntesis y ensayo biológico de los candidatos, guiando este último al próximo paso de selección (9).

Durante los ensayos biológicos iniciales del ciclo de descubrimiento de fármacos son identificados los denominados hits. A esta fase de generación de hits le sigue la fase de generación de leads, donde los hits identificados son validados mediante ensayos confirmativos y refinados estructuralmente con el objetivo de incrementar su potencia con respecto al blanco. De lograrse una potencia suficiente, se pueden realizar ensayos biológicos adicionales para asegurar que el compuesto líder no

interacciona con proteínas homólogas del blanco, con el fin de limitar sus efectos secundarios (86).

Hasta este punto es posible identificar compuestos líderes con perfiles de unión al blanco adecuados. Sin embargo, el fármaco no solo debe interferir con el blanco terapéutico, sino que además debe poseer un perfil biológico favorable, específicamente una toxicidad baja de manera que no sea dañino para el organismo, y propiedades farmacocinéticas adecuadas.

La fase final del descubrimiento de fármacos es la fase de optimización del líder, donde se refina la estructura química del mismo de manera que cumpla con los criterios requeridos para convertirse en un fármaco. Este proceso de optimización es altamente iterativo y se considera la fase más crítica del proceso de descubrimiento de fármacos ya que es aquí donde ocurre la mayor cantidad de fallas. Una vez descubierto un compuesto líder con características de fármaco prometedoras, el paso final hacia la puesta en el mercado del fármaco es la fase de validación clínica (80).

Validación Clínica. Previo a la puesta en el mercado, el candidato a fármaco debe ser validado durante una fase de prueba extensiva, dirigida a demostrar su eficacia y seguridad para el organismo humano: la validación clínica. Esta fase comienza con la realización de pruebas preliminares de seguridad en animales, la etapa preclínica, y es subsecuentemente articulada en tres fases (1):

- Fase I (1 a 2 años): Inicialmente se llevan a cabo pruebas de seguridad con un número limitado (< 100) de personas sanas.

- Fase II (1 a 2 años): Seguidamente se llevan a cabo pruebas de seguridad y eficacia a una muestra mayor compuesta por cientos de personas que incluye grupos de sanos y enfermos.
- Fase III (2 a 3 años): Finalmente, el estudio se completa con la realización de pruebas de eficacia a gran escala, las que involucran una muestra mucho mayor de personas (miles) de diferentes áreas demográficas.

Eventualmente, una vez proveídos los resultados de este estudio clínico y concedida la aprobación gubernamental, la explotación comercial del fármaco puede comenzar. La aprobación gubernamental es concedida por ejemplo por la Administración de Alimentos y Medicamentos (Food and Drugs Administration, FDA) en los Estados Unidos de América, por la Agencia Europea para la Evaluación de Productos Médicos en Europa o para el caso particular de nuestro país por el Centro para el Control Estatal de la Calidad de los Medicamentos (CECMED).

1.1.3 Métodos computacionales

Debido a la necesidad de explotar las cantidades masivas de datos generados por las tecnologías de alto rendimiento, los métodos computacionales se han ido implementando de manera creciente en el proceso de descubrimiento de fármacos y de manera general en la química. Para unificar la combinación de los métodos computacionales y la química, F. K. Brown acuñó en 1998 el término *quimioinformática* definiéndola como “la combinación de aquellos recursos de información para transformar datos en información y la información en conocimiento con el propósito de tomar mejores y más rápidas decisiones en el área de la

identificación y optimización de compuestos líderes”(87). Esta definición general engloba múltiples aspectos, en particular, la representación, almacenamiento, recolección y análisis de la información química (88-90).

Diseño de bibliotecas. Tanto la academia como la industria farmacéutica no tardaron en percatarse que el número de compuestos que se podría generar por medio de la química combinatoria sería tan grande que la evaluación sistemática y aleatoria de bibliotecas combinatorias no era el camino óptimo para el descubrimiento de nuevos compuestos líderes. Esta observación motivó una racionalización del proceso de evaluación con el objetivo de limitar el número de ensayos HTS y así reducir el tiempo y el costo del paso de identificación del compuesto líder. Así, el centro de la investigación terapéutica moderna ha cambiado de la evaluación de grandes bibliotecas combinatorias hacia un proceso de diseño de bibliotecas dirigido a que las mismas sean más pequeñas pero que a la vez contengan más información (5, 9, 88).

Debido a que el objetivo de los ensayos de evaluación evoluciona durante el proceso de descubrimiento de fármacos, el diseño de las bibliotecas a ser evaluadas debe estar relacionado con los avances del proyecto. Con el objetivo de maximizar la probabilidad de identificar hits estructuralmente diferentes, los ensayos de evaluación iniciales deben involucrar bibliotecas diversas de manera que se cubra lo más ampliamente posible el espacio químico (3). Por el contrario, en el paso posterior “hits to lead”, deben ser diseñadas bibliotecas dirigidas constituidas por moléculas estructuralmente similares a los hits identificados (9). Bibliotecas diversas pueden definirse por ejemplo, mediante la selección de

elementos representativos de los distintos sectores de la partición del espacio químico o de grupos o clusters de moléculas, mientras que bibliotecas dirigidas pueden obtenerse directamente de los sectores o clusters a los que el hit pertenece (5, 88).

Tamizaje virtual o “Virtual Screening”. Tamizaje virtual o *in silico* es el término usado para denotar el análisis computacional de bases de datos de compuestos, dirigido a identificar candidatos que posean la actividad biológica deseada sobre un blanco terapéutico específico. Esto puede verse como una alternativa al desarrollo de ensayos experimentales con la principal ventaja de que pueden ser evaluadas cantidades arbitrarias de moléculas reales o virtuales. Con la identificación de compuestos potencialmente activos, el tamizaje virtual puede por tanto ayudar a la reducción del número de ensayos experimentales y motivar la obtención o síntesis de nuevas moléculas (5, 9). En la práctica, el tamizaje virtual requiere del conocimiento de la estructura del blanco terapéutico, usualmente obtenido por métodos cristalográficos, o bien de la actividad medida experimentalmente de un conjunto de compuestos.

Si la estructura del blanco es conocida, el enfoque más común para el tamizaje virtual son los estudios de acoplamiento o docking, los que consisten en la derivación de una puntuación o score de la actividad a partir del posicionamiento óptimo del ligando en el sitio activo del blanco (86, 91-100). Si se desconoce la estructura del blanco, los métodos de tamizaje virtual pueden derivarse de un pool de compuestos con actividad conocida obtenidos de ensayos experimentales previos. Estos métodos se conocen como enfoques del tamizaje virtual basados en

ligandos, en oposición al enfoque anterior basado en la estructura del blanco. Un método simple basado en ligandos consiste en el ordenamiento de las moléculas de una base de datos con respecto a su similitud con un conjunto de compuestos activos conocidos, y la selección de los candidatos mejor ordenados como los más prometedores (86). Alternativamente, el conjunto de compuestos activos puede usarse para derivar un *modelo farmacóforo* que puede usarse como un filtro para eliminar aquellos compuestos que no cumplan con las condiciones de actividad necesarias (86, 101). Un enfoque más general consiste en la construcción de un modelo que correlacione la estructura de las moléculas con sus respectivas actividades biológicas a partir de un pool de moléculas previamente evaluadas, integrando así información tanto de la actividad como de la inactividad de los compuestos. Este problema se conoce como la modelación de la relación estructura-actividad (REA) más comúnmente conocido por sus siglas en inglés QSAR, acrónimo de Quantitative Structure-Activity Relationships, e involucra métodos de los campos de la estadística y el aprendizaje de máquinas (machine learning).

Cuando se conoce la estructura del blanco, las aproximaciones basadas en la estructura, naturalmente constituyen la vía más racional para el tamizaje virtual (ver referencia (102) y referencias incluidas). Por su parte, las aproximaciones basadas en ligandos son de un interés considerable ya que los métodos de acoplamiento son muy costosos desde el punto de vista computacional, y difíciles de automatizar (2). En segundo lugar, la estructura del blanco es difícil, si no imposible, de obtener, siendo en este caso igualmente imposible de aplicar un estudio de acoplamiento.

Integración temprana del concepto “drug-likeness”. Las propiedades de una molécula que contribuyen a un perfil farmacológico, farmacocinético y toxicológico favorable se denomina en la investigación terapéutica moderna a través de un término originalmente definido en inglés: drug-likeness. En la actualidad es ampliamente aceptado que la principal causa de fallas en el proceso de descubrimiento de fármacos es precisamente un pobre perfil farmacocinético y toxicológico del candidato a fármaco (2, 3, 9, 103). Esta observación reveló que en la fase de optimización del líder, la capacidad de mejorar el perfil terapéutico del candidato seleccionado basándose solamente en su actividad farmacológica fue sobreestimada, lo que condujo a la integración del concepto drug-likeness en etapas anteriores a la optimización, durante la fase de identificación del líder (2, 9). Esta reconsideración del paradigma secuencial de las fases de identificación y optimización del líder marca un avance significativo hacia la racionalización del proceso de descubrimiento de fármacos, lo que representa una severa reducción del retraso de la incorporación a la clínica de nuevos agentes terapéuticos y como consecuencia también una reducción del costo global del proceso (1, 2).

Considerando el enfoque REA involucrado en el tamizaje virtual, los modelos estadísticos y basados en aprendizaje computarizado que correlacionan directamente la estructura de las moléculas con sus propiedades “drug-like” son también útiles en este contexto (78, 103). Estrictamente hablando, estos modelos no se relacionan con los análisis REA ya que el término actividad refiere a la noción de unión con el blanco, y la terminología adoptada usualmente es la de análisis de las relaciones estructura-propiedad (REP). No obstante, la integración del concepto

drug-likeness en el proceso de identificación del líder puede verse como la reformulación del problema del tamizaje virtual y los análisis REA en un marco multidimensional o multipropiedades más general (2, 3, 10), si se considera que los términos tamizaje virtual y REA a menudo ignoran este significado más amplio.

1.2 Temas Abiertos en Quimioinformática

1.2.1 Representación y codificación de la estructura molecular

Representación de la estructura molecular. Existen diferentes formas de representar la estructura molecular. La representación molecular más simple, en una dimensión (1D), se define a partir de la constitución atómica de la molécula. Por su parte, la representación molecular bidimensional (2D) se basa en un grafo denominado grafo molecular que representa los enlaces covalentes establecidos entre los átomos de la molécula. Finalmente, en el espacio tridimensional (3D), una molécula puede ser vista como un esqueleto, resultante de la introducción de coordenadas atómicas en el grafo molecular, o como varias superficies moleculares que representan la proyección de las propiedades fisicoquímicas de los átomos, tales como su potencial electrostático, hidrofobicidad, carga o volumen de Van der Waals, en una esfera que encierra a la molécula (90).

Codificación de la estructura molecular. Desde el punto de vista práctico, los métodos quimioinformáticos a menudo no son capaces de manejar directamente las representaciones moleculares, y se requiere encapsular la información estructural en un conjunto de descriptores capaces de codificar la información contenida en una representación molecular específica. Los *Descriptores o Índices Moleculares* (DMs o IMs) juegan un papel fundamental en la quimioinformática y en

particular para los estudios QSAR. En la actualidad un descriptor molecular se define como: *el resultado final de un procedimiento lógico-matemático el cual transforma la información química, codificada dentro de una representación simbólica de una molécula, en un número (o conjunto de estos) útil(es) o en el resultado de alguno de los experimentos ya estandarizados (104)*. De acuerdo a la naturaleza en su definición y a la complejidad de los rasgos moleculares estructurales que codifican, los DMs se clasifican de forma general según las dimensiones que abarcan en: DMs-0D (Descriptores Constitucionales), DMs-1D (Descriptores Unidimensionales), DMs-2D (Descriptores Bidimensionales o Invariantes de Grafos), DMs-3D (Descriptores Tridimensionales), y DMs-4D (Descriptores Tetradimensionales).

Los DMs-0D son descriptores que se obtienen directamente de la fórmula molecular y son independientes de cualquier conocimiento sobre la estructura molecular, i.e. número de átomos (A), el peso molecular (MW), conteo de átomos-tipo (N_x) o cualquier función de las propiedades atómicas. Los DMs-1D están basados en la representación unidimensional de la molécula (o representación que consiste en una lista de fragmentos estructurales de la molécula), aunque no requieren del conocimiento completo de la estructura molecular, i.e. descriptores de búsqueda y análisis subestructural, como los Descriptores de Conteo de Fragmentos. Los DMs-2D se basan en la representación bidimensional o *topológica* de la molécula, o sea, que consideran la conectividad de los átomos (vértices) en la molécula (pseudografo) en términos de la presencia y naturaleza de los enlaces químicos (aristas). Entre los DMs-2D más extendidos se encuentran los índices topológicos

(ITs), los cuales están relacionados con la noción de invarianza del grafo en la teoría de grafos. Ejemplos representativos de este tipo de descriptores incluyen los *Índices de Conectividad de Wiener y Randic* (90, 104). Los DMs-3D son derivados de la representación tridimensional de la molécula y se basan no solo en la naturaleza y conectividad de los átomos, sino también en la configuración espacial de la molécula. Una buena parte de estos descriptores requieren un paso preliminar de alineamiento molecular que consiste en colocar las moléculas en una orientación común en el espacio 3D por medio de operaciones de rotación y traslación (104). La calidad del alineamiento se cuantifica a través de una función de puntuación o scoring function, y se dice que las moléculas están alineadas cuando la puntuación es maximizada. Finalmente los DMs-4D son descriptores basados no solo en la configuración espacial de la molécula, sino también en los campos escalares de interacción que se originan como consecuencia de la distribución electrónica en dicha entidad química, i.e. Valores de la Energía de Interacción (104).

El número de descriptores moleculares propuestos en la literatura hasta el momento es realmente amplio. La presentación anterior esta lejos de ser exhaustiva, por lo que para una presentación detallada los lectores interesados pueden referirse a libros de texto que desarrollan este tema con profundidad (89, 90, 104). No obstante, la selección de descriptores “apropiados” para un problema específico continúa siendo una cuestión abierta.

La teoría de las cadenas de Markov como fuente de descriptores moleculares.

Cadenas de Markov es el nombre de una teoría o modelo matemático definido por Markov (31). Las cadenas de Markov constituyen uno de los modelos de la teoría

de probabilidades más usados en la investigación científica (105, 106), y en la actualidad son usadas como soporte de múltiples aplicaciones en áreas como la inteligencia artificial, la epidemiología, la medicina y la bioinformática (36, 107-120). Aunque las cadenas de Markov constituyen uno de los modelos de la teoría de probabilidades más usados, a continuación ofrecemos algunas notas básicas sobre estas que justifican la elección de las cadenas de Markov como fuente de descriptores moleculares.

- Las cadenas de Markov estudian fenómenos estocásticos, esto es: fenómenos que al ser estudiados mediante mediciones en el tiempo (t) el resultado obtenido no está determinado sino que puede obtenerse cada vez con cierta probabilidad uno de varios resultados posibles.
- En cadenas de Markov para referirse a un resultado determinado se habla de que el sistema ha ocupado un estado dado. En este trabajo hablaremos, además, de cómo cada estado está caracterizado por una magnitud fisicoquímica m_j , que depende del fenómeno en estudio.
- En la cadena de Markov se parte de una distribución inicial de probabilidades ${}^A p_0(j)$ que son las probabilidades absolutas iniciales ($t_0 = 0$) con que el sistema ocupa cada estado j . Dado un sistema con n estados, estas probabilidades pueden ordenarse en un vector de probabilidades iniciales ${}^0 \boldsymbol{\pi} = [{}^A p_0(1), {}^A p_0(2), {}^A p_0(3), \dots, {}^A p_0(n)]$.
- Las relaciones entre los estados del sistema expresadas por el grafo sobre el que se define la cadena de Markov pueden ser resumidas a través de la matriz

estocástica ${}^1\Pi$. Los elementos de esta matriz son las probabilidades ${}^1p_{ij}$ de que el sistema pase de un estado i en el tiempo $t_0 = 0$ a otro estado j en el tiempo $t_1 = 1$ o primer paso.

- Las cadenas de Markov cumplen con la condición de que las probabilidades de transición ${}^k p_{ij}$ tanto para el primer paso (${}^1 p_{ij}$) como para tiempos ulteriores $t_k = k > 1$ depende solamente del estado que el sistema ocupaba en el tiempo anterior $t_{k-1} = k-1$ pero no de tiempos anteriores (propiedad markoviana).
- Las cadenas de Markov cumplen con las ecuaciones de Chapman-Kolgomorov, por lo que las probabilidades absolutas ${}^A p_k(j)$, con que el sistema ocupa determinado estado j en el tiempo k , se determinan como los elementos de los vectores ${}^k \pi = {}^0 \pi \cdot {}^k \Pi = {}^0 \pi \cdot ({}^1 \Pi)^k$. Del mismo modo, las probabilidades de transición ${}^k p_{ij}$ pueden ser calculadas como elementos de las matrices ${}^1 \Pi = ({}^1 \Pi)^k$.
- Las probabilidades ${}^k p_{ii}$ para $i = j$ se denominan probabilidades de autoretorno donde el sistema regresa al estado inicial. Estas probabilidades se encuentran en la diagonal principal o traza (Tr) de las matrices estocásticas.
- Tanto las probabilidades absolutas iniciales ${}^A p_0(j)$ (elementos de ${}^0 \pi$), como las probabilidades ${}^1 p_{ij}$ (elementos de ${}^1 \Pi$) pueden ser determinadas a partir de las magnitudes fisicoquímicas m_j que caracterizan al estado.

De lo anterior se desprende que es posible derivar de la cadenas de Markov ciertos números que la caracterizan, ya que dependen de i) los estados presentes, ii) de su interconexión caracterizada por α_{ij} , y iii) de la tendencia del sistema a ocupar dichos estados caracterizados por su magnitud fisicoquímica m_j (71, 121). Por tanto, de ser

aplicada la cadena de Markov a un sistema molecular como los descritos anteriormente, dichos números podrán ser usados como descriptores moleculares para encontrar modelos tipo QSPR de una propiedad dada (20, 36-74). Entre los índices numéricos que se puede generar producto de la aplicación de las cadenas de Markov se encuentran:

Probabilidades absolutas en si (${}^A p_k$). Estas solo pueden ser usados como descriptores moleculares localmente definidos:

$${}^A p_k = {}^0 \boldsymbol{\pi} \cdot ({}^1 \boldsymbol{\Pi})^k = \sum_{S_X}^n {}^A p_k(j) \quad (1.1)$$

Momentos espectrales de la matriz estocástica (${}^{SR} p_k$), suma de las probabilidades de autoretorno:

$${}^{SR} p_k = \text{Tr}({}^k \boldsymbol{\Pi}) = \text{Tr}({}^1 \boldsymbol{\Pi})^k = \sum_j^k p_{jj} \quad (1.2)$$

Entropías de la distribución de probabilidades (Θ_k):

$$\Theta_k = -k_B \cdot \sum_{S_X}^n {}^A p_k(j) \cdot \log {}^A p_k(j) \quad (1.3)$$

Magnitudes medias ($[m]_k$) calculadas como sumas ponderadas de probabilidades absolutas:

$$[m]_k = \sum_j^A p_k(j) \cdot m_j \quad (1.4)$$

1.2.2 Determinación de la similitud molecular

La noción de similitud desempeña un papel importante en la quimioinformática debido a un paradigma central dentro de esta área conocido como el *principio de similitud*, el cual plantea que moléculas con estructuras similares tienden a exhibir propiedades biológicas similares (122). Este principio justifica los métodos que involucran algoritmos de agrupamiento, particionamiento y ordenamiento, y es por esto que el acceso a la similitud estructural ha venido despertando una atención considerable en la comunidad quimioinformática.

Una primera clase de algoritmos considera directamente las estructuras moleculares. La similitud entre pares de grafos moleculares (2D) puede ser definida a partir de la razón entre sus tamaños y el tamaño de su máximo sub-grafo común (MSC) (123). Este enfoque puede extenderse al caso 3D al considerar el mayor conjunto de átomos idénticos con distancias interatómicas coincidentes (124). También es posible obtener medidas de similitud entre estructuras 3D a través de sus puntuaciones de alineamiento (125-128). Alternativamente, se puede acceder a la similitud entre moléculas alineadas usando la metodología CoMFA, a partir de la comparación de los campos moleculares a través del índice de similitud de Carbó (129).

Sin embargo, debido a que tanto el alineamiento como la búsqueda del máximo sub-grafo común son procesos intensivos y costosos desde el punto de vista computacional, el enfoque principal se basa en definir directamente una medida de similitud a partir de la comparación de los descriptores moleculares.

Debido a la gran diversidad de descriptores moleculares, numerosos métodos han sido propuestos para cuantificar la similitud molecular, basados por ejemplo en índices de conectividad (130), pares o tripletes de átomos en las representaciones moleculares 2D y 3D (131), matrices de distancias interatómicas y mapeo atómico óptimo (124), o vectores de autocorrelación (132). El enfoque más extendido de cuantificación de la similitud molecular consiste en el conteo del número de características sub-estructurales compartidas. Varios *coeficientes de asociación* han sido definidos para este propósito. Los mismos se basan esencialmente en el producto entre los descriptores moleculares (131). Entre estos, el *coeficiente de Tanimoto*, definido para el par de descriptores moleculares (A , B) como

$$T_{AB} = \frac{A^T B}{A^T A + B^T B - A^T B} \quad (1.5)$$

ha emergido como una medida de similitud estándar (133).

Otra forma de cuantificar la similitud molecular desde el punto de vista estadístico es a través de *medidas de distancia*, particularmente útiles en el área del aprendizaje basado en casos (134-136) y el análisis de conglomerados (137, 138). Aunque existen otras opciones posibles, la *distancia euclidiana* es la más extendida. La distancia euclidiana E de un caso i con respecto a un caso j se define como:

$$E = \left[\sum (X_i - X_j)^2 \right]^{1/2} \quad (1.6)$$

donde X_i y X_j representan los respectivos valores del atributo X para el caso i y j , respectivamente.

Otras medidas de distancia pueden ser más apropiadas en circunstancias especiales. De cualquier modo, aun cuando se han realizado múltiples esfuerzos dirigidos a resolver el problema de la similitud molecular (139), acceder de forma eficiente a la similitud molecular continúa siendo un capítulo abierto de la quimioinformática.

1.2.3 Quimiometría y modelación de la relación estructura-actividad

Esta sección tratará lo referente a la construcción de modelos que correlacionen la actividad biológica de una molécula con su estructura a partir de un pool de compuestos de los cuales se conoce su actividad. El término actividad se emplea usualmente en la investigación terapéutica moderna en su sentido más amplio para referirse a cualquier propiedad biológica que determine el perfil terapéutico de un candidato a fármaco. Las décadas de investigación en el campo de la estadística y el aprendizaje de máquinas (Machine Learning) han provisto toda una plétora de métodos que han contribuido al desarrollo de lo que hoy se conoce como *quimiometría* (140-142).

Consideraciones Metodológicas. Los modelos pueden agruparse en dos categorías principales según la naturaleza de la propiedad a predecir. Los modelos empleados para predecir propiedades cuantitativas, como por ejemplo, el grado de unión a un blanco, se conocen como modelos de regresión. Por otra parte, los modelos de clasificación se encargan de predecir propiedades cualitativas (143).

Para construir el modelo, el pool de moléculas con actividad conocida es usualmente dividido en una serie de entrenamiento y una serie de prueba. La serie de entrenamiento se usa para el aprendizaje o entrenamiento del modelo. El

proceso de entrenamiento consiste en construir un modelo capaz de predecir correctamente la propiedad biológica de las moléculas incluidas en la serie de entrenamiento, sin llegar a “memorizarlas”, lo que se conoce como sobreajuste del modelo. Este fenómeno puede controlarse mediante el empleo de técnicas de validación cruzada que cuantifican la capacidad del modelo para predecir un subconjunto de la serie de entrenamiento que se dejó fuera durante la fase de entrenamiento. La serie de prueba es usada para evaluar la capacidad de generalización del modelo entrenado. En los métodos de regresión, la capacidad predictiva es cuantificada a través de la correlación entre los valores predichos y reales de la actividad. En los métodos de clasificación, un criterio estándar es la exactitud del clasificador, expresada como la fracción de compuestos clasificados correctamente (143), así como las razones de clasificación de verdaderos(falsos) positivos(negativos) (140, 143).

Puesto que usualmente requieren como entradas un conjunto limitado de variables no correlacionadas, la aplicación de modelos a la quimioinformática requiere condensar la información sobre las moléculas en un conjunto limitado de *rasgos*, lo cual dista de ser una tarea trivial debido a la gran cantidad de descriptores moleculares disponibles en la actualidad (89, 90, 104). Una manera bastante extendida de conducir este problema en la quimioinformática se basa en la aplicación de *análisis de componentes principales* (ACP) (144-146) o *mínimos cuadrados parciales* (MCP) (147-150), los cuales definen un conjunto limitado de variables independientes a partir de combinaciones lineales del pool inicial de

rasgos, de forma que sea posible conservar la mayor cantidad del contenido de información.

Alternativamente, los métodos de selección de rasgos pueden usarse para identificar un subconjunto de rasgos relevantes para la propiedad a predecir a partir de un pool inicial de rasgos. Entre los métodos de selección más usados en la modelación REA se encuentran la selección secuencial hacia delante (Sequential Feature Forward Selection), la eliminación secuencial hacia atrás (Sequential Feature Backward Elimination), el recorrido simulado (Simulated Annealing) y la selección basada en algoritmos genéticos, siendo esta última una de las más eficientes en el campo de modelación REA (29).

La elección de un modelo de entre la plétora de modelos existentes esta relacionada con el objetivo final del estudio, y mientras que modelos complejos pueden exhibir un gran potencial predictivo, estos a menudo va en detrimento de su interpretabilidad.

De forma resumida, los principios de la metodología REA pueden describirse mediante los siguientes pasos comunes (143): 1) Formulación del problema, en la cual se determina el objeto de análisis y el nivel de información requerido; 2) Parametrización cuantitativa de la estructura molecular de los compuestos; 3) Medición de la propiedad de interés; 4) Elección del tipo de modelo REA que se va a desarrollar; 5) Selección de los compuestos (diseño estadístico de la serie); 6) Análisis matemático de los datos y validación interna y externa de los modelos obtenidos; 7) Interpretación de los resultados y aplicación de los modelos desarrollados al diseño/descubrimiento de un nuevo compuesto líder, desarrollando

procedimientos de tamizaje virtual. Sin embargo, como ya se ha dicho, el desarrollo de cualquier estudio REA, es un ciclo iterativo.

Tendencias en la modelación REA. El primer modelo REA fue desarrollado en 1964 por Hansch y Fujita quienes aplicaron un análisis de regresión lineal múltiple (RLM) para correlacionar la actividad biológica de una molécula con un par de descriptores relacionados con su estructura electrónica y su hidrofobicidad (151, 152). Los modelos de regresión lineal múltiple aún son ampliamente usados en la modelación REA. Los factores y variables latentes generados por los métodos basados en ACP y MCP, respectivamente, son comúnmente usados como entradas en los denominados modelos de regresión de componentes principales (Principal Components Regression) o regresión de mínimos cuadrados parciales (Partial Least Squares Regression) (150). Este es, en particular, el caso de la metodología CoMFA (153). Más recientemente, los algoritmos evolutivos han sido introducidos para solucionar el problema de la selección de rasgos como una alternativa eficiente a los métodos tradicionales de selección (forward selection y backward elimination) (154). Enfoques lineales pueden aplicarse al problema de la clasificación a través del *análisis discriminante lineal* (ADL) (155).

Sin embargo, debido a que esta clase de modelos está limitada a establecer relaciones lineales, este enfoque puede ser demasiado restrictivo. Los estudios REA pueden beneficiarse de manera significativa del desarrollo de métodos no lineales de modelación, entre los que sobresalen las *redes neuronales artificiales*. Desde las primeras aplicaciones, las redes neuronales de propagación hacia atrás (back-propagation artificial neural networks) dieron cuenta de su superioridad en

cuanto a capacidad predictiva, y hoy en día constituyen una herramienta estándar de la modelación REA (156-161).

A pesar de su eficiencia predictiva, la mayor crítica a las redes neuronales artificiales es su falta de interpretabilidad. Una clase alternativa de modelos construye un clasificador expresado a través de una serie de reglas que relacionan la estructura molecular con la actividad biológica. Tales modelos pueden derivarse a través de algoritmos de generación de árboles de decisión (162, 163) o a partir de métodos desarrollados en el campo de la programación lógica inductiva (164, 165), aportando información valiosa sobre los rasgos estructurales responsables de la actividad de las moléculas.

Por otra parte, es conocida su tendencia al sobreajuste además de su difícil reproducibilidad debido a su inicialización aleatoria y posible convergencia hacia mínimos locales (166, 167). Estos aspectos teóricos son en alguna medida cubiertos por los algoritmos de maquinarias de vectores soportes (Support Vector Machines, SVM), conocidas en particular por evitar el problema de la convergencia hacia mínimos locales, prevenir el sobreajuste, y ofrecer un mejor control del error de generalización (168). En los últimos años los algoritmos SVM han demostrado ser una herramienta poderosa para el desarrollo de estudios REA, a menudo superando el desempeño de las redes neuronales artificiales en tareas de regresión y clasificación.

Métodos quimiométricos de modelación REA. La Quimiometría es una disciplina que recopila herramientas matemáticas y estadísticas para tratar con datos químicas complejas (169-175). La quimiometría es la herramienta más útil en los

estudios REA hoy en día interceptándose no solo con varios campos de la Matemática y la Estadística clásica sino también de la Inteligencia Artificial. Dentro de los métodos y técnicas generales que engloba se encuentran los siguientes:

Análisis Exploratorio de Datos. El Análisis Exploratorio de Datos es una colección de técnicas que examinan la estructura del conjunto de datos antes de calcular algún modelo estadístico (176). Su propósito es obtener información acerca de la distribución de los datos, de la presencia de outliers y conglomerados, y revelar relaciones y correlaciones *a priori* entre los objetos y/o variables. Además del uso clásico de estadísticas descriptivas y de gráficos, la exploración de datos se apoya hoy en técnicas de Descubrimiento de Conocimiento (Cubos OLAP (OnLine Analytical Processing)) y de reducción de datos, (clustering; y selección y transformación de rasgos como el ACP) (144, 177, 178).

Diseños Experimentales. Son procedimientos estadísticos para planificar un experimento, i.e. el coleccionar datos apropiados los cuales, luego de su análisis por métodos estadísticos, resultan en conclusiones válidas. El diseño incluye la selección de las unidades experimentales, la especificación de las condiciones experimentales, la especificación de los niveles de los factores involucrados y la combinación de tales factores, la selección de la respuesta a ser medida, y la elección del modelo estadístico para ajustar los datos (179, 180). Entre los diseños experimentales más conocidos se encuentran los Diseños Factoriales Completos, los Diseños Factoriales Fraccionados, los Diseños estratificados, los Diseños Compuestos, los Diseños Optimales y los Diseños de Mezclas (143).

Análisis de Regresión. Este término representa un conjunto de métodos estadísticos que usan ecuaciones matemáticas para modelar la relación entre una respuesta medida u observada y una o más variables predictivas. El objetivo de este análisis es doble: modelar y predecir. La relación (modelo de regresión) se describe en forma algebraica como $y = f(x) + e$ siendo f una función lineal en el caso más simple, de las componentes del vector x ; donde, x denota la(s) variable(s) predictiva(s), y la(s) variable(s) respuesta(s), $f(x)$ es la parte sistemática del modelo, y e es el error aleatorio o residual del modelo (181-184).

El análisis de regresión incluye no solo la estimación de los parámetros del modelo de regresión, sino también el cálculo de los estadísticos de la bondad de ajuste y la bondad de predicción, el diagnóstico de la predicción, el análisis de residuales, y el análisis de influencia (185).

Además de la bien conocida regresión lineal de Mínimos Cuadrados Ordinarios (MCO), también son importantes los modelos de regresión sesgada, regresión no-lineal y de regresión robusta. Dentro de los métodos sesgados, los métodos de regresión más populares son: Regresión de Componentes Principales, Regresión de Mínimos Cuadrados Parciales, Ridge Regression, Regresión Continua, Regresión por Pasos o Etapas.

Dentro de los métodos no lineales están además la regresión de mínimos cuadrados no lineal, i.e. Regresión Polinomial, el método de Mínimos Cuadrados Parciales No Lineal y las Expectativas Condicionales Alternantes. Sin embargo, algunas técnicas de redes neuronales artificiales también se han considerado

dentro de los métodos de regresión no-lineales tales como el Método de Propagación Hacia Atrás.

Métodos de Clasificación. Se utilizan para la asignación de objetos a una de varias clases basado en una regla de clasificación. Son métodos de aprendizaje supervisado, pues se “aprende” a partir de una serie de casos con variables predictivas y función “objetivo” o variable dependiente también conocida (esta serie de entrenamiento es el “maestro o supervisor”). Tal es el caso del método de regresión; solo que en los problemas de clasificación la función objetivo o variable dependiente es discreta, i.e. es una clase. El objetivo de tales técnicas es calcular una regla de clasificación (y, posiblemente, límites de clases, o probabilidades de pertenencia a una clase), basados en los objetos de la serie de entrenamiento y aplicar esta regla para asignar una de estas clases, a objetos de clases previamente desconocidas (186-188). Los métodos de clasificación son apropiados para modelar varias respuestas REA, como por ejemplo: compuestos activos/no-activos o compuestos de toxicidad baja/mediana/alta.

Los métodos de clasificación de origen estadístico más populares son: Análisis Discriminante Lineal; Análisis Discriminante Cuadrático; Análisis Discriminante Regularizado; Métodos de Árboles de Clasificación; Regression Decision Trees potentes árboles de decisión cuyas hojas finales no definen una clase para un segmento de la población, sino una ecuación de regresión, por ejemplo logística para clasificar a los elementos de ese segmento, los Clasificadores de Función Potencial; los Clasificadores Medios más Cercanos; los Clasificadores Medios más Cercanos Ponderados, esto últimos basados en el método de los K -ésimos Vecinos

más Cercanos y que basan su principio de funcionamiento en el aprendizaje basado en casos. Básicamente se almacenan los casos del entrenamiento y se usa una función de distancia para determinar el o los k miembros de conjunto de entrenamiento más cercanos a un caso de prueba desconocido. Una vez localizado el caso más cercano del conjunto de entrenamiento, se predice la clase del caso de prueba.

La regresión lineal múltiple también puede ser aplicada a problemas de clasificación en dominios con atributos numéricos al igualar a 1 los casos del entrenamiento pertenecientes a la clase y a 0 los que no pertenecen. El resultado es una expresión lineal para cada clase. Esta aproximación a menudo ofrece buenos resultados pero una de sus limitaciones principales es que la regresión por mínimos cuadrados asume que los errores no solo son estadísticamente independientes sino que además están normalmente distribuidos con la misma desviación estándar; una suposición que es palpablemente violada cuando el método es aplicado a problemas de clasificación dado que las observaciones solo pueden tomar valores iguales a 0 o 1. En condiciones teóricas desfavorables como la falta de normalidad u homogeneidad de varianzas, la regresión logística (RL) es un recurso fabuloso ya que en lugar de aproximar directamente los valores a 0 o 1 se construye un modelo lineal basado en una variable objetivo transformada. Para el caso de dos clases la regresión logística reemplaza la variable objetivo original ($\text{Pr}[1|a_1, a_2, \dots, a_k]$), la cual no puede ser aproximada de manera exacta por $(\log \text{Pr}[1|a_1, a_2, \dots, a_k] / 1 - \text{Pr}[1|a_1, a_2, \dots, a_k])$, generándose así valores que ya no están restringidos en el intervalo $[0,1]$ pero que están distribuidos a lo largo del intervalo $[-\infty, +\infty]$. Estas

variables transformadas son aproximadas usando una función lineal del mismo modo que en los modelos de regresión lineal múltiple (140). El modelo resultante es:

$$\text{Pr}[1|a_1, a_2, \dots, a_k] = 1/(1 + \exp(-w_0 - w_1 a_1 - \dots - w_k a_k))$$

Los pesos w_k son seleccionados de manera que ajusten apropiadamente los datos. En lugar de usar el cuadrado del error como medida de la bondad de ajuste, la RL emplea una transformación logarítmica de la máxima verosimilitud (log-likelihood) como criterio. Así, son seleccionados los pesos w_k que maximicen el criterio de log-likelihood (140).

Existen reportes de desempeños superiores incluso a una red neuronal artificial por medio de la regresión logística en problemas reales cuando la estructura de la red aconseja pocas neuronas y capas ocultas (189, 190).

Por otra parte, se pueden encontrar varios métodos de clasificación o de análisis supervisado dentro de la Inteligencia Artificial, en particular las redes neuronales artificiales. Las mismas representan un conjunto de métodos matemático-computacionales, modelos y algoritmos diseñados para imitar los métodos de adquisición del conocimiento y procesamiento de la información del cerebro humano. Las redes neuronales artificiales son especialmente apropiadas para tratar con relaciones y tendencias no-lineales y se han propuesto para enfrentar una gran variedad de problemas tales como: la exploración de datos, el reconocimiento de patrones, la modelación de respuestas continuas y categorizadas, problemas de respuestas múltiples, etc. (191-193).

Por otra parte, ante ciertos problemas no es posible lograr la respuesta óptima con un único clasificador y se plantea entonces la necesidad de “multiclasificadores” que combinen de alguna manera, las respuestas de varios clasificadores. Los multiclasificadores pueden ser enfocados desde varios puntos de vista, desde los más simples (ponderación estadística) hasta su combinación inteligente (194).

Análisis de Conglomerados (Cluster Analysis). Además de las técnicas de aprendizaje supervisado, se utilizan técnicas de aprendizaje no supervisado, en la cual se conforman “clases” a partir de variables predictivas conocidas sin que haya un conocimiento previo sobre la clase a que pertenece cada instancia, a veces, ni siquiera, sobre el número de clases a considerar. El Análisis de Conglomerados esta basado en la evaluación de la similitud/disimilitud de todos los pares de objetos del conjunto de datos. Esta información se reúne en la matriz de similitud o en la matriz de distancia de datos (137, 138). Los métodos de aglomerar más populares son los Métodos Aglomerantes Jerárquicos y los Métodos Divisivos Jerárquicos. Otros métodos populares son los Métodos no-Jerárquicos, tales como el Método de las *K*-Medias y el Método de Jarvis-Patrick. También la Inteligencia Artificial aporta métodos de análisis no supervisado análogos a las técnicas estadísticas de clustering pero de carácter esencialmente no jerárquico ni lineal, por ejemplo la bien conocida red de Cohonen y sus variantes (195).

Optimización. Las técnicas de optimización están orientados a encontrar el valor óptimo (mínimo o máximo) de una función numérica f , llamada función objetivo, con respecto a un conjunto de parámetros \mathbf{p} , $f(p_1, p_2, \dots, p_p)$.

Entre las técnicas de optimización más populares y de uso frecuente en nuestro campo, se encuentran: la Optimización Newton-Raphson, la Optimización del Ascenso más Abrupto, la Optimización del Descenso más Abrupto, la Optimización *Simplex*, la Optimización por Algoritmos Evolutivos, y el Recorrido Simulado (180, 196, 197). La Inteligencia Artificial aporta las llamadas técnicas de Aprendizaje Reforzado (198, 199).

Optimización Multiobjetivos. Los métodos de optimización multiobjetivos introducen una nueva filosofía para la optimización basada en la obtención de compromisos entre varios objetivos (11). Por medio de los métodos de optimización multiobjetivos es posible encontrar una solución óptima global mediante la optimización simultánea de múltiples propiedades interrelacionadas. El mayor beneficio de los métodos de optimización multiobjetivos es que el óptimo local correspondiente a un objetivo puede evitarse considerando simultáneamente todos los objetivos lo cual lleva a un proceso global más eficaz. Existen dos enfoques generales para la optimización multiobjetivos. Uno consiste en combinar las k funciones objetivo individuales en una única función compuesta. En este caso, la determinación de un único objetivo es posible mediante la suma ponderada de las funciones objetivo (SPFO) o a través del uso la teoría de utilidad o deseabilidad (200). Estos métodos ofrecen soluciones útiles, pero su principal desventaja reside en la selección apropiada de los pesos o funciones de deseabilidad (media geométrica) o utilidad (media aritmética) que caractericen adecuadamente las preferencias del investigador. El segundo enfoque consiste en determinar el conjunto total o un subconjunto de soluciones *pareto óptimas* (Métodos de optimización multiobjetivos

basados en pareto). El conjunto de soluciones pareto óptimas es un conjunto de soluciones que no son dominadas por ninguna otra solución. A medida que nos movemos de una solución pareto hacia otra, siempre existe cierto sacrificio en uno(o varios) objetivo(s) para lograr una cierta mejora en el(los) otro(s) objetivo(s). Un conjunto de soluciones pareto óptimas siempre es preferible a una única solución óptima ya que al considerar problemas reales la solución final del investigador siempre se basa en compromisos.

Un problema de optimización multiobjetivos puede formularse de la siguiente manera. Considere un problema de toma de decisiones donde el investigador desea optimizar k objetivos de tal modo que los objetivos no son conmensurables y el investigador no tiene una preferencia clara de un objetivo con respecto al resto. Sin pérdida de generalidad, todos los objetivos pueden considerarse como objetivos de minimización. Un objetivo de minimización puede transformarse en un objetivo de maximización simplemente multiplicando este por -1 .

Un problema de minimización multiobjetivos con k objetivos se define como sigue:

Dado un vector variable de decisión n -dimensional $\mathbf{x} = [x_1, \dots, x_n]$ en el espacio de soluciones \mathbf{X} , se pretende hallar un vector \mathbf{x}^* que minimice un conjunto dado de K funciones objetivo $z(\mathbf{x}^*) = \{z_1(\mathbf{x}^*), \dots, z_K(\mathbf{x}^*)\}$. El espacio de soluciones \mathbf{X} generalmente está limitado por una serie de restricciones, tales como $g_j(\mathbf{x}^*) = b_j$ para $j = 1, \dots, m$ y límites en las variables de decisión.

En muchos problemas de la vida real, los objetivos bajo consideración se encuentran en conflicto. Debido a esto, optimizar \mathbf{x} con respecto a un solo objetivo a menudo conlleva a resultados inaceptables con relación al resto de los objetivos.

Por tanto, es casi imposible obtener una solución multiobjetivos perfecta que optimice simultáneamente cada función objetivo. Una solución razonable a un problema multiobjetivos es investigar un conjunto de soluciones donde cada objetivo sea satisfecho a un nivel aceptable sin ser dominado por ninguna otra solución.

Si todas las funciones objetivo deben ser minimizadas, se dice que una solución factible \mathbf{x} domina a otra solución factible \mathbf{y} ($\mathbf{x} \succ \mathbf{y}$), si y solo si, $z_i(\mathbf{x}) \leq z_i(\mathbf{y})$ para $i = 1, \dots, K$ y $z_j(\mathbf{x}) < z_j(\mathbf{y})$ para al menos una función objetivo j . Se dice que una solución es *pareto óptima* si esta no es dominada por ninguna otra solución en el espacio de soluciones. Una solución *pareto óptima* no puede ser mejorada con respecto a cualquier objetivo sin afectar al menos uno de los objetivos restantes. El conjunto de todas las soluciones *no dominadas* factibles en \mathbf{X} se denomina como el *conjunto pareto óptimo*, y para un conjunto pareto óptimo dado, los valores de la función objetivo correspondiente en el espacio de los objetivos se denomina *frente pareto*. Para muchos problemas, el número de soluciones pareto óptimas es enorme (quizás infinitas).

El objetivo final de un algoritmo de optimización multiobjetivos es identificar las soluciones pertenecientes al conjunto pareto óptimo. Sin embargo, identificar la totalidad del conjunto pareto óptimo, para problemas de con gran cantidad de objetivos, es prácticamente imposible debido a su tamaño. Además, para muchos problemas, especialmente en problemas de optimización combinatoria, probar la optimalidad de la solución es computacionalmente inviable. Por tanto, un método práctico para la optimización multiobjetivos es investigar el conjunto de soluciones

que representan lo mejor posible al conjunto pareto óptimo (*el mejor conjunto pareto conocido*).

Varias aplicaciones de los métodos de optimización multiobjetivos han sido reportadas en la literatura. Las mismas van desde la minería de subestructuras hasta estudios de acoplamiento, incluyendo estudios QSAR y QSPR inverso (11, 201-207). Específicamente en el área de los estudios QSAR, las técnicas de optimización multiobjetivos han sido sólo recientemente aplicadas (75-77).

2. MATERIALES Y MÉTODOS

2 MATERIALES Y MÉTODOS

2.1 Estudios de Correlación Estructura-Propiedad Multisistemas Basados en el Uso de Índices Markovianos Sistema-Dependientes

2.1.1 Cálculo de los Índices Markovianos Sistema-Dependientes

El cálculo de todos los descriptores moleculares modificados basados en cadenas de Markov esta implementado en el programa MARCH-INSIDE (208). El mismo posee una interfaz gráfica que permite su uso de forma simple por usuarios con conocimientos computacionales básicos.

2.1.2 Conformación de la base de datos

2.1.2.1 Efectos adversos de fármacos sobre múltiples sistemas de órganos

La base de datos fue conformada por un conjunto de fármacos de uso frecuente en la practica clínica y reconocida capacidad para inducir diversos efectos adversos en diferentes sistemas de órganos en el hombre; siendo todos estos fármacos ampliamente probados en la practica clínica y sus efectos adversos reportados mediante estudios de farmacovigilancia. Para esto se empleó un reporte de efectos adversos de medicamentos recogido en INTERCOM 94' (209). La misma fue conformada por 301 fármacos de los que son reportados 39 efectos adversos que afectan 11 sistemas de órganos.

2.1.2.2 Coeficientes de partición de fármacos, compuestos químicos y contaminantes ambientales en múltiples sistemas de partición bifásicos

Inicialmente fueron colectados de la literatura (210-217) un total 413 compuestos que incluyen fármacos y compuestos químicos. Los compuestos se agruparon en

17 sistemas de partición bifásica (SPB) de relevancia biológica; medidos en diversas especies biológicas (ratones, ratas, conejos y humanos).

Debido a que los datos recolectados provienen de fuentes diversas y por lo tanto los parámetros han sido determinados por métodos diferentes y con diferentes niveles de error en las mediciones, en lugar de predecir directamente los coeficientes de partición de las moléculas sería más consistente en términos de exactitud clasificar las moléculas en dos grandes grupos denominados: *con tendencia a la partición (CTP)* y *sin tendencia a la partición (STP)* en los distintos sistemas bifásicos. Para simplificar el problema todos los coeficientes de partición fueron establecidos de manera que el proceso de partición ocurre de la fase más organizada a la menos organizada; esto es, del agua al aire, o del tejido al plasma (61). Cuando el valor experimental del coeficiente de partición es mayor que 1 ($\log P > 0$) la molécula es clasificada a priori como CTP, de otra forma se clasifica la molécula como STP.

Con el objetivo de adicionar sistemas de relevancia para la toxicología ambiental (efectos tóxicos de compuestos químicos sobre el hombre) la base de datos fue posteriormente ampliada a 1300 compuestos incluyendo ahora contaminantes ambientales (210-226). Los compuestos se agruparon en 38 sistemas de partición bifásicos de relevancia biológica o para la toxicología ambiental; medidos en diversas especies biológicas (peces, conejos, ratas y humanos).

En esta aplicación cuando el valor experimental del coeficiente de partición ($\log P$) del sistema bifásico es mayor que el punto de ruptura (PR) establecido para ese sistema la molécula es clasificada *a priori* como CTP, de otra forma se clasifica la

molécula como STP. El punto de ruptura de cada sistema se determina en este trabajo como $PR = (\log P_{\max} + \log P_{\min})/2$; donde $\log P_{\max}$ es el mayor valor del coeficiente de partición para un SPB específico y $\log P_{\min}$ es el menor valor.

2.1.3 Análisis estadístico

Haciendo uso de los *índices markovianos sistema-dependientes* es posible desarrollar una relación cuantitativa estructura-propiedad lineal simple con la siguiente fórmula general:

$$P_x = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_k \cdot X_k \quad (2.1.1)$$

Para la obtención de las funciones de clasificación se eligió el Análisis Discriminante Lineal implementado en el paquete estadístico STATISTICA 6.0 (227). Este modelo trata con la clasificación de un conjunto de diversos compuestos capaces de inducir diferentes efectos adversos en diferentes sistemas de órganos o sufrir un proceso de partición en diferentes SPB. La variable (P_x) indica la ocurrencia ($P_x = 1$) o ausencia ($P_x = -1$) del efecto adverso/proceso de partición en el sistema x de órganos/de partición bifásico en cuestión. Todas las variables a explorar fueron estandarizadas (143). El método Forward stepwise se fijó como estrategia de selección de variables (228).

La calidad de los modelos basados en Análisis Discriminante Lineal se determinó mediante la inspección de los porcentajes de buena clasificación del modelo, del estadístico U de Wilk como medida del poder discriminante del modelo, la razón F de Fisher y su nivel de significación estadística (p) como medidas de la significación estadística de la ecuación, y el cuadrado de la distancia de Mahalanobis (D^2) como

medida de la separación entre grupos. Se tuvo en cuenta además, la razón entre casos y variables incluidas en la ecuación así como el número de variables a ser explorado con el objetivo de evitar sobreajuste y/o correlaciones casuales (228, 229).

El poder predictivo de los modelos fue evaluado mediante una técnica de re-sustitución de casos en cuatro subconjuntos de predicción (4-fold Cross Validation) (143, 230, 231) donde se reportan los valores de exactitud para los subconjuntos de entrenamiento y predicción y para el conjunto total de compuestos en cada uno de los 4 experimentos así como los valores promediados de los subconjuntos de entrenamiento, predicción y para el conjunto total de compuestos relativos a los 4 experimentos.

Específicamente, en el caso de la derivación del modelo QSPR basado en los *energía libre de partición estocástica* para la modelación del perfil de distribución de los 1300 compuestos agrupados en 38 sistemas de partición bifásicos la selección de las variables mas relevantes fue conducida a través de varios métodos (forward stepwise, backward stepwise, forward entry, backward removal y best subset selection); con el objetivo de seleccionar el mejor modelo (228, 229). Se empleo el método de ortogonalización de Randic para eliminar el problema de la colinealidad (232-236).

En esta aplicación fue probada la capacidad de las redes neuronales artificiales para capturar relaciones no lineales (15, 16, 156, 237-240). Para esto se desarrollaron varios modelos de clasificación basados en redes neuronales artificiales usando las variables incluidas en el mejor modelo basado en Análisis

Discriminante Lineal encontrado. Los modelos de redes neuronales artificiales fueron desarrollados usando cinco arquitecturas de redes diferentes incluidas en el *Solucionador Inteligente de Problemas* implementado en el módulo de redes neuronales del paquete estadístico STATISTICA 6.0 (227).

2.1.4 Ensayos biológicos y animales de laboratorio

Determinación de la distribución del 2-bromo-5-[2-bromo-2-nitrovinil] furano en órganos de rata. Las muestras de sangre (100 μ L) fueron mezcladas con 1 ml de agua para provocar la hemólisis de las células rojas. Se procedió al conteo de las muestras de orina (1 ml) y plasma (100 μ L) añadiéndole 2 ml de cóctel de centelleo. Se adquirieron 32 ratas macho de la línea Sprague–Dawley pesando entre 280-300 g, las que se separaron en ocho grupos experimentales (4 ratas/tiempo de dosificación). Las ratas previamente pesadas recibieron una única dosis oral de [14 C]-2-bromo-5-[2-bromo-2-nitrovinil] furano en Miglyol[®] 810N (100 mg/Kg) mediante gavage oral. La dosis volumen fue de 5 ml/Kg a 0.5, 1, 2, 3, 6, 12, 24 y 48 horas después de la administración del fármaco. Las ratas fueron anestesiadas y la sangre fue colectada en tubos heparinizados. Las células rojas de la sangre y el plasma fueron separados por centrifugación. El nivel de radioactividad fue además determinado en la sangre total.

2.2 Empleo de Métodos de Optimización Multiobjetivos y Ordenamiento Basados en Funciones de Deseabilidad en Estudios Globales de Correlación Estructura-Actividad

Mejorar el perfil de una molécula para el proceso de descubrimiento de fármacos requiere de la optimización simultánea de numerosos objetivos que a menudo se

encuentran en conflicto. Esencialmente, el fármaco ideal debería poseer la mayor eficacia terapéutica y biodisponibilidad a la vez que la mínima toxicidad posible. Si desarrollar el fármaco ideal es sumamente difícil, encontrar el mejor compromiso posible entre objetivos que compiten como los antes mencionados es una opción más accesible y realista. De este modo, cuanto más deseable se torna el perfil de cada uno de estos objetivos más deseable será el candidato a fármaco (ver Figura 2.2.1).

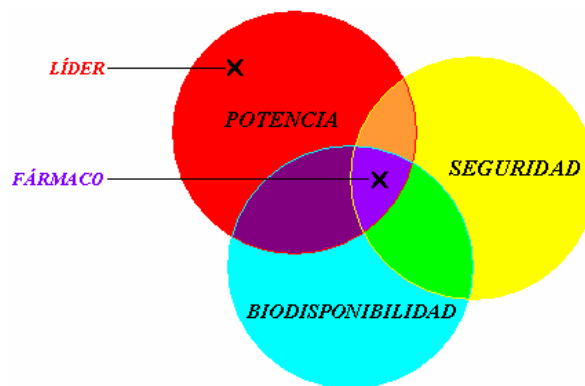


Figura 2.2.1. Representación gráfica del compromiso entre eficacia terapéutica (potencia), biodisponibilidad (propiedades ADME) y toxicidad (seguridad) requerido para alcanzar un fármaco exitoso.

Con este propósito, nos proponemos realizar una OMO basada en la estimación de la deseabilidad de varias respuestas interrelacionadas; lo que permitirá conducir estudios QSAR globales considerando simultáneamente el perfil farmacológico, farmacocinético y/o toxicológico de un conjunto de candidatos. La metodología propuesta (definida originalmente en inglés como **Multi-Objective OPTimization based on the Desirability Estimation of Several Interrelated REsponses [MOOP-DESIRE]**) encuentra la solución más deseable que simultáneamente optimiza un

problema multiobjetivos de química médica empleando las funciones de deseabilidad de Derringer (241), específicamente dirigida a conferir racionalidad al proceso de descubrimiento de fármacos.

2.2.1 Optimización multiobjetivos basada en la estimación de la deseabilidad de varias respuestas interrelacionadas

El proceso de optimización simultánea de múltiples propiedades de un candidato a fármaco es descrito a continuación:

1-Desarrollar un modelo de predicción (MP) para cada respuesta a optimizar (242).

Debido a la multiplicidad de factores involucrados en el perfil farmacéutico de una molécula no es irrazonable suponer que el mismo subconjunto de variables independientes no es capaz de explicar de manera óptima todas las respuestas para diferentes propiedades biológicas (especialmente propiedades competitivas como la potencia y la toxicidad). Para esta situación existe aun una forma de maximizar simultáneamente la deseabilidad de diferentes tipos de propiedades biológicas.

2-Calcular por separado los valores predichos de cada respuesta usando el MP respectivo y usar estos para ajustar un modelo que contenga todas las variables independientes usadas en la modelación de las respuestas originales.

Aquí las variables independientes usadas para calcular los valores predichos de las respuestas originales se corresponden con los usados para las respuestas originales. Las variables independientes no usadas en calcular los valores predichos de las respuestas originales tomaran valores igual a cero.

3-Especificar la función de deseabilidad para los valores predichos de cada respuesta.

Para cada respuesta Y_i , una función de deseabilidad $d_i(Y_i)$ asigna valores entre 0 y 1 para los posibles valores de Y_i . Esta respuesta transformada, llamada $d_i(Y_i)$, puede tomar varias formas diferentes. Independientemente de la forma, $d_i(Y_i) = 0$ representa un valor completamente indeseable de Y_i y $d_i(Y_i) = 1$ representa un valor de respuesta completamente deseable o ideal. Las deseabilidades individuales son entonces combinadas utilizando la media geométrica, generándose así la deseabilidad global D :

$$D = (d_1 \times d_2 \times \dots \times d_k)^{\frac{1}{k}} \quad (2.2.1)$$

donde k denota el número de respuestas.

Este valor D codifica de manera global la deseabilidad de la combinación de los niveles de respuesta. Claramente el rango de D estará también en el intervalo $[0, 1]$ y aumentará en la medida que el balance de las propiedades sea más favorable. Note que si cualquier respuesta Y_i es completamente indeseable [$d_i(Y_i) = 0$], entonces la deseabilidad global será cero. De esta forma, la máxima deseabilidad se corresponderá con los niveles de las variables independientes que simultáneamente maximicen D .

En dependencia de que una respuesta particular Y_i deba ser maximizada, minimizada, o le sea asignado un valor blanco o diana, podrán ser usadas diferentes funciones de deseabilidad $d_i(Y_i)$. En este trabajo son usadas las funciones de deseabilidad propuestas por Derringer y Suich (241).

L_i , U_i y T_i serán los valores mínimo, máximo, y diana, respectivamente, que son deseados para la respuesta Y_i , siendo $L_i \leq T_i \leq U_i$.

Si una respuesta es del tipo “un valor diana es el ideal”, entonces su función de deseabilidad individual será:

$$d_i(\hat{Y}_i) = \begin{cases} \left[\frac{\hat{Y}_i - L_i}{T_i - L_i} \right]^s & \text{si } L_i \leq \hat{Y}_i \leq T_i \\ \left[\frac{\hat{Y}_i - U_i}{T_i - U_i} \right]^t & \text{si } T_i < \hat{Y}_i \leq U_i \\ 0 & \text{si } \hat{Y}_i < L_i \text{ or } \hat{Y}_i > U_i \end{cases} \quad (2.2.2)$$

Si la respuesta debe ser maximizada, entonces su deseabilidad individual será definida como:

$$d_i(\hat{Y}_i) = \begin{cases} 0 & \text{si } \hat{Y}_i \leq L_i \\ \left[\frac{\hat{Y}_i - L_i}{T_i - L_i} \right]^s & \text{si } L_i < \hat{Y}_i < T_i \\ 1 & \text{si } \hat{Y}_i \geq T_i = U_i \end{cases} \quad (2.2.3)$$

T_i en este caso es interpretado como un valor suficientemente grande de la respuesta que puede coincidir con U_i .

Finalmente, si queremos minimizar una respuesta, entonces podríamos usar la siguiente función:

$$d_i(\hat{Y}_i) = \begin{cases} 1 & \text{si } \hat{Y}_i \leq T_i = L_i \\ \left[\frac{\hat{Y}_i - U_i}{T_i - U_i} \right]^s & \text{si } U_i < \hat{Y}_i < T_i \\ 0 & \text{si } \hat{Y}_i \geq U_i \end{cases} \quad (2.2.4)$$

T_i denota aquí un valor suficientemente pequeño de la respuesta que puede coincidir con L_i .

Aquí, los exponentes s y t determinan qué tan importante es alcanzar el valor diana T_i (241).

Finalmente, es posible predecir la deseabilidad global D de cada candidato a fármaco determinada por k respuestas Y , las que a su vez están determinadas por un conjunto específico de variables independientes X_n . Sin embargo, debido a que la función de deseabilidad de Derringer es construida usando las respuestas predichas (\hat{Y}_i) no es posible conocer cuan confiable es el valor predicho de deseabilidad global D de cada candidato.

Para superar esta limitación se propone un parámetro estadístico, el coeficiente de determinación de la deseabilidad global (R^2_D). Este es capaz de reflejar el efecto del conjunto de variables independientes (X_n) usado para ajustar las respectivas respuestas, en la reducción de la incertidumbre de la predicción de los valores D_i .

Si Y_i es estimada como una función continua de X_n ; $d_i(Y_i)$ es una función continua de Y_i (ver ecuaciones (2.2.2), (2.2.3) y (2.2.4)), y D es una función continua de $d_i(Y_i)$ (ver ecuación (2.2.1)); entonces D es también una función continua de X_n . Consecuentemente, R^2_D puede ser calculado de forma análoga al bien conocido coeficiente de determinación R^2 . Específicamente, R^2_D es calculado usando los

valores de deseabilidad global observados D_{Y_i} (calculados a partir de Y_i) y predichos $D_{\hat{Y}_i}$ (calculados a partir de \hat{Y}_i) en lugar de usar directamente los valores de respuesta observados (Y_i) y predichos (\hat{Y}_i).

$$R^2_D = 1 - \frac{SSE}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i})^2}{\sum (D_{Y_i} - \bar{D}_{Y_i})^2} \quad (2.2.5)$$

Donde D_{Y_i} y $D_{\hat{Y}_i}$, han sido previamente definidos. \bar{D}_{Y_i} es el valor medio de D para las respuestas Y_i de cada caso incluido en el conjunto de datos, $SSTO$ es la suma total de cuadrados, y SSE es la suma de cuadrados debido al error.

Así, R^2_D posee las mismas propiedades de R^2 . Estaría incluido en el rango $[0, 1]$ y mientras mayor sea R^2_D menor será la incertidumbre al predecir D usando un conjunto específico de variables independientes X_n (242).

Considerando que R^2_D mide la bondad de ajuste en lugar de la capacidad predictiva de un determinado modelo de predicción es aconsejable usar una medida análoga al coeficiente de determinación de la validación cruzada dejando un caso fuera (R^2_{LOO-CV} o Q^2) para determinar la confiabilidad del método en la predicción de D . Para esto, el coeficiente de determinación de la deseabilidad global debido a la validación cruzada (Q^2_D) puede ser definido en forma análoga a R^2_D :

$$Q^2_D = 1 - \frac{SSE_{LOO-CV}}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i(LOO-CV)})^2}{\sum (D_{Y_i} - \bar{D}_{Y_i})^2} \quad (2.2.6)$$

Donde SSE_{LOO-CV} y $D_{\hat{y}_i}(LOO-CV)$ son la suma de cuadrados debido al error en la validación cruzada y la deseabilidad global predicha mediante LOO-CV, respectivamente.

De este modo, podemos tener una medida de la confiabilidad de la optimización simultánea de las k respuestas sobre el dominio de variables independientes.

4-Optimizar simultáneamente las k respuestas (objetivos) maximizando la deseabilidad global D sobre el dominio de variables independientes.

La deseabilidad global D puede ser maximizada sobre el dominio de variables independientes X_n a través del método *simplex* de optimización de funciones (243-245). El resultado final se traduce en encontrar los niveles óptimos de las variables independientes que optimizan simultáneamente las k respuestas determinando la calidad final del producto. De este modo, el mejor compromiso posible entre las k respuestas es encontrado y consecuentemente la mayor deseabilidad global para el compuesto final es alcanzada.

2.2.2 Ordenamiento de soluciones basado en funciones de deseabilidad

El razonamiento basado en casos se basa fundamentalmente en la suposición de que casos con una descripción similar (determinada por determinados rasgos o características; en este trabajo los descriptores moleculares) debe tener una solución similar (el objetivo que se persigue en un estudio, en este trabajo las propiedades involucradas en el perfil terapéutico del candidato a fármaco) (246).

Este tipo de razonamiento posibilita proponer un método de ordenamiento basado en parámetros estimados a partir de la descripción de los casos. Específicamente

será posible ordenar candidatos provenientes de bibliotecas combinatorias de los cuales solo se conoce su estructura, con respecto a su similitud con el caso óptimo determinado mediante el proceso de optimización multiobjetivos descrito anteriormente.

Δ_i es el parámetro usado aquí para describir la similitud entre un caso i y el caso óptimo en función del conjunto de variables descriptivas empleadas en el proceso de optimización y se define como:

$$\Delta_i = \sum_{X=1}^m \delta_{i,X} \cdot w_X \quad (2.2.7)$$

donde $\delta_{i,X}$ es la distancia euclidiana del caso i con respecto al caso óptimo considerando el(los) parámetro(s) X y w_X representa el peso o influencia de la variable X sobre la deseabilidad global D del caso i .

Una vez determinado Δ_i , este parámetro es normalizado nuevamente mediante el empleo de las funciones de deseabilidad de Derringer (241) (específicamente a través de la ecuación (2.2.4)); de manera que los menores valores de Δ_i (que indican una mayor similitud con respecto al caso óptimo) tomen los valores mas cercanos a 1 y viceversa. Aquí L_i se corresponde con el mínimo valor de Δ_i (Δ_{iMIN}) y $U_i = \Delta_{iMAX}$. Esto se hace con el objetivo de llevar los valores de D_i y Δ_i a una misma escala de manera que la diferencia entre estos pueda ser miminizada para cada caso.

El conjunto óptimo de pesos w_X que minimizan la diferencia entre los valores de D_i y los valores normalizados de Δ_i se obtuvieron usando un algoritmo de optimización a

larga escala basado en un método de ajuste de curva no lineal (247, 248) implementado en la función "lsgcurvefit" del programa MATLAB, Versión 7.2 (249).

Una vez minimizada la diferencia entre D_i y los valores normalizados de Δ_i se obtiene una correspondencia entre la descripción (valores normalizados de Δ_i) y la solución de los casos (valores de D_i). Esto permite, bajo los principios del razonamiento basado en casos, ordenar nuevos candidatos de los cuales solo se conoce su estructura, en función de Δ_i . De este modo será posible identificar los candidatos mas prometedores los cuales serán ordenados en los primeros lugares (los de menores valores de Δ_i y por tanto los mas próximos al candidato óptimo) y descartar el resto de los candidatos que estén ordenados al final de la lista.

2.2.2.1 Validación del método de ordenamiento y estimación del índice de calidad del ordenamiento

Aun cuando el razonamiento basado en casos sugiera que el proceso de optimización (para encontrar los w_x óptimos) nos puede conducir a un ordenamiento adecuado, no es posible tener una idea de la calidad del ordenamiento logrado a través de dicho proceso. Debido a esto proponemos un nuevo método de validación del proceso de optimización a la vez que un criterio numérico de la calidad del ordenamiento obtenido. El método de validación del ordenamiento incluye los siguientes pasos:

- 1- Ordenar los casos de mayor a menor en función de D y asignarle un orden (1, 2, 3,, n) a cada caso (ordenamiento ideal o experimental).
- 2- Invertir el orden obtenido en el paso 1 (n ,, 3, 2, 1) (peor ordenamiento).

- 3- Ordenar los casos de menor a mayor en función de Δ_i y asignarle un orden (1, 2, 3,, n) a cada caso (ordenamiento real o predicho obtenido a través del método).
- 4- Normalizar mediante el empleo de las funciones de deseabilidad de Derringer (ecuación [2.2.4]) los valores asignados a cada caso en los pasos 1 al 3 donde $L_i = 1$ y $U_i =$ número de casos presentes en la serie (n). De este modo se obtienen los valores de ordenamiento normalizados para los ordenamientos ideal (${}^{IR}d_i$), peor (${}^{WR}d_i$) y real (${}^{RR}d_i$).
- 5- Usar los respectivos valores de ordenamiento normalizados para de determinar la diferencia del ordenamiento real con respecto al ordenamiento ideal (${}^{IR-RR}\delta_i$) y del peor ordenamiento con relación al ordenamiento ideal (${}^{IR-WR}\delta_i$):

$${}^{IR-RR}\delta_i = |{}^{IR}\delta - {}^{RR}\delta| \quad (2.2.8)$$

$${}^{IR-WR}\delta_i = |{}^{IR}\delta - {}^{WR}\delta| \quad (2.2.9)$$

- 6- Estimar la calidad del ordenamiento a través de la determinación del índice de calidad del ordenamiento (Ψ) el cual se define como el valor absoluto de la media de ${}^{IR-RR}\delta_i$ para el conjunto de datos a ordenar:

$$\Psi = \left| \frac{\sum_{i=1}^n {}^{IR-RR}\delta_i}{n} \right| \quad (2.2.10)$$

Este parámetro está incluido en el rango $[0, 0.5]$ donde para un ordenamiento ideal Ψ toma un valor igual a 0 y para el peor ordenamiento Ψ toma un valor aproximadamente igual a 0.5. Así, cuanto más próximo a 0 esté Ψ mayor será la calidad del ordenamiento.

Debido a que el valor de Ψ para el peor caso depende del tamaño del conjunto de datos este no es exactamente igual a 0.5. Considerando esto a Ψ se le aplica un factor de corrección ($F = 2/\Psi^{WR}$), donde Ψ^{WR} representa el índice de calidad del peor ordenamiento.

De esta forma se obtiene el índice de calidad del ordenamiento corregido (Ψ^*) el cual está incluido en el rango [0, 1]:

$$\Psi^* = \left| \frac{\sum_{i=1}^n IR-RR \delta_i}{n} \right| \cdot F = \left| \frac{\sum_{i=1}^n IR-RR \delta_i}{n} \right| \cdot \frac{2}{\Psi^{WR}} \quad (2.2.11)$$

Finalmente, es posible expresar Ψ^* como el porcentaje de calidad del ordenamiento ($R\% = (1 - \Psi^*) \times 100$).

2.2.3 Conformación de la base de datos

2.2.3.1 Perfil antiinflamatorio, analgésico y ulcerogénico de quinazolinonas

Los datos para la construcción de los distintos modelos de predicción fueron tomados de una biblioteca de quince quinazolinonas reportadas por Alagarsamy y colaboradores (250) según los cuales la actividad analgésica (A_n) (reportada como % A_n) de los quince compuestos fue medida experimentalmente usando el método del calentamiento de la cola en ratones albinos Wistar (251). La actividad antiinflamatoria (A_a) (reportada como % A_a) fue evaluada en ratas usando el ensayo de edema del talón inducido por carragenina (251). Igual que para A_n , los compuestos fueron administrados a una dosis de 20 mg·kg⁻¹ (la dosis máxima usada). El índice ulcerogénico (U) fue determinado por el método de Ganguly y

Bhatnagar (252). Las úlceras fueron inducidas en ratas usando el método descrito por Goyal y colaboradores (253). Los compuestos fueron administrados intraperitonealmente a una dosis de 20 mg·kg⁻¹.

2.2.3.2 Perfil antibacteriano y citotóxico de fluoroquinolonas

Los datos para la construcción de los distintos modelos de predicción fueron tomados de una biblioteca de 117 fluoroquinolonas reportada por Suto y colaboradores (254) según los cuales la citotoxicidad sobre células V79 de hámster chino fue medida como la concentración del compuesto que conduce a la supervivencia del 50% de las células comparado con células control no tratadas (IC₅₀) y expresada en µg/mL (254-261). La capacidad antibacteriana de los compuestos fue determinada mediante su evaluación sobre cinco microorganismos gram-negativos usando la técnica de microdilución estándar (262). Los datos usados representan la media geométrica de la mínima concentración inhibitoria (MIC) expresada en µg/mL para las bacterias gram-negativas *Enterobacter cloacae* MA 2646, *Escherichia coli* Vogel, *Klebsiella pneumonia* MGH-2, *Providencia rettgeri* M 1771 y *Pseudomonas aeruginosa* (254).

2.2.4 Especificaciones de las funciones de deseabilidad

La optimización de la deseabilidad global fue conducida mediante el método “Use general function optimization” implementado en el modulo de regresión general múltiple (RGM) del software STATISTICA 6.0 (227). El método “spline” (263, 264) fue usado para ajustar las funciones de deseabilidad. El nivel actual de cada variable independiente fue establecido como sus respectivos valores óptimos. Los

parámetros s y t fueron fijados a 1.00 bajo la hipótesis de que las funciones de deseabilidad aumentan linealmente hacia T_i para cada respuesta.

Optimización simultánea del perfil antiinflamatorio, analgésico y ulcerogénico de quinazolinonas. Se ajustó una función de deseabilidad para cada respuesta. Aquí A_n y A_a deben ser maximizadas. Para la estimación de los valores respectivos d_i el valor inferior L_i se fijó al 25 %, y el valor superior U_i que coincide con el valor diana T_i se fijó al 100 % para ambas respuestas. Contrariamente, U debe ser minimizado donde $L_i = T_i = 0$ y $U_i = 1.73$, coincidiendo este último con el índice ulcerogénico del ácido acetil salicílico (A.S.A).

Optimización simultánea del perfil antibacteriano y citotóxico de fluoroquinolonas. Se ajustó una función de deseabilidad para cada respuesta. La capacidad citotóxica debe ser minimizada. Esta propiedad se expresa a través de IC_{50} la cual debe maximizarse. Debido a que este valor fue transformado ($1/(1+IC_{50})$), esta propiedad transformada es en realidad minimizada (igualmente sucede con la capacidad antibacteriana). Para la estimación de los valores d_i correspondientes a los valores de la capacidad citotóxica el valor inferior L_i que coincide con el valor diana T_i se fijó en $1/(1+IC_{50}) = 0.002$ que se corresponde con un valor de $IC_{50} = 380 \mu\text{g/mL}$ exhibido por el compuesto menos citotóxico de la serie, y el valor superior U_i se fijó en $0.1 / 8 \mu\text{g/mL}$ exhibido por el compuesto más citotóxico de la serie. Contrariamente, la capacidad antibacteriana debe ser maximizada donde $L_i = (1/(1+MIC) = 0.038) = (MIC = 25 \mu\text{g/mL})$ correspondiéndose estos valores con el compuesto de menor actividad antibacteriana de la serie y $U_i =$

$T_i = (1/(1+MIC)) = 0.99 / MIC = 0.01 \mu\text{g/mL}$), coincidiendo este último con el compuesto mas activo de la serie.

2.2.5 Métodos computacionales y análisis estadístico

Las estructuras fueron inicialmente optimizadas método PM3 (265) implementado en el programa MOPAC 6.0 (266). Posteriormente se calcularon los 1664 DMs implementados en el software DRAGON 5.0 (267). DMs con valores constantes o cercanos a valores constantes fueron excluidos.

Para el caso del estudio de OMO de la biblioteca de quinazolinonas solo se usaron los 120 DMs de fragmentos centrados en átomos (ACF). Los descriptores moleculares de tipo ACF se seleccionaron debido a que su naturaleza simple ofrece una interpretación estructural clara. Se eliminaron además, descriptores con una correlación por pares mayor a 0.95; quedando constituido el conjunto de descriptores moleculares de tipo ACF a emplear en el proceso de selección de variables por solo 12 de estos. La selección de los descriptores moleculares relevantes se condujo por medio de una técnica basada en algoritmos genéticos (268-271) implementada en el programa BuildQSAR (272, 273) (ver detalles en la publicación original (274)).

La regresión lineal múltiple implementada en el módulo de regresión general múltiple del software STATISTICA 6.0 (227) se empleó como técnica de modelación. Para cada modelo de predicción, la bondad de ajuste se evaluó examinando el coeficiente de determinación (R^2), la desviación típica (s), el estadístico de Fisher (F), así como la proporción entre el número de casos y el número de parámetros ajustables en el modelo (ρ). La capacidad predictiva del

modelo se determinó mediante el coeficiente de determinación de la validación cruzada (Q^2_{LOO}) (275). El dominio de aplicación de los modelos de predicción se determinó mediante un ploteo de los valores de los residuales estandarizados contra los valores de leverage de los compuestos del entrenamiento (275, 276).

2.2.6 Análisis comparativo de distintos enfoques de tamizaje virtual

Se llevó a cabo un análisis comparativo de tres enfoques para el tamizaje virtual de una biblioteca de compuestos. Para esto se conduce un tamizaje virtual de la biblioteca de fluoroquinolonas descrita en el epígrafe 2.2.3.2 a través de tres enfoques distintos: secuencial, paralelo y multiobjetivos. La idea básica es seleccionar, a partir de una biblioteca de compuestos, un subconjunto de esta, tratando de maximizar el número de casos positivos en este subconjunto.

El tamizaje virtual secuencial se realiza a través del ordenamiento independiente de la biblioteca de compuestos de acuerdo a las dos propiedades en estudio relacionadas con el perfil terapéutico del candidato (eficacia antibacteriana (MIC) y perfil de seguridad (IC_{50})), y posterior filtrado de un porcentaje determinado (10%) de la biblioteca utilizando como criterio inicial de filtrado un umbral previamente especificado del valor de la actividad antibacteriana ($MIC = 0.115 \mu\text{g/mL} = d_{MIC} = 0.9 = 90\%$ de eficacia antibacteriana). Posteriormente se eliminan aquellos candidatos que no cumplen con el valor umbral, también previamente especificado, de un segundo criterio, que en este caso sería el perfil de seguridad del candidato ($IC_{50} = 82.5 \mu\text{g/mL} = d_{IC_{50}} = 0.9 = 90\%$ de seguridad).

El tamizaje virtual en paralelo se basa en el análisis simultáneo pero por separado de los dos criterios involucrados en el perfil terapéutico del candidato (IC_{50} y MIC).

En este caso se seleccionan aquellos candidatos que cumplen con el umbral pre-establecido para ambos criterios.

El tamizaje virtual multiobjetivos considera el perfil terapéutico global del candidato en lugar de tener en cuenta por separado cada propiedad relacionada con este. En este caso la deseabilidad global ($D_{MIC-IC50}$) del candidato es tomada como criterio del perfil terapéutico global del mismo. En nuestro caso la biblioteca es ordenada en función del criterio de similitud estructural con respecto al candidato óptimo previamente establecido. Igualmente se filtra el 10% de la biblioteca en busca de aquellos candidatos con un perfil terapéutico global igual o superior al 90% ($D_{MIC-IC50} = 0.9$).

Para evaluar el desempeño de cada uno de los enfoques, de manera que sea posible su comparación, se utilizan dos medidas de enriquecimiento conocidas: la razón de verdaderos positivos presentes en el porcentaje filtrado de la biblioteca ($VP_{10\%}$) y el factor de enriquecimiento para el porcentaje filtrado de la biblioteca ($FE_{10\%}$) (277). $VP_{10\%}$ nos da una medida de la probabilidad con que uno de n casos seleccionados es positivo. Por su parte $FE_{10\%}$ se define como:

$$FE_{10\%} = \frac{S_+ / S_{Total}}{N_+ / N_{Total}} \quad (2.2.12)$$

donde N_+ y S_+ representan el número total de casos positivos y el número de casos positivos presentes en el subconjunto filtrado, respectivamente. Por su parte N_{Total} y S_{Total} dan cuenta del número total de casos presentes en la biblioteca y en el subconjunto filtrado. Un esquema de tamizaje virtual que sea superior a una selección aleatoria de los candidatos debe devolver un valor superior a 1 (277).

3. RESULTADOS Y DISCUSIÓN

3 RESULTADOS Y DISCUSIÓN

Durante este capítulo se presentarán los principales resultados obtenidos en esta tesis. Así, es posible corroborar mediante su aplicación a problemas reales, la utilidad de los estudios de correlación estructura-actividad multiobjetivos en el descubrimiento y desarrollo racional de nuevos fármacos con perfiles farmacológico, farmacocinético y/o toxicológico simultáneamente mejorados.

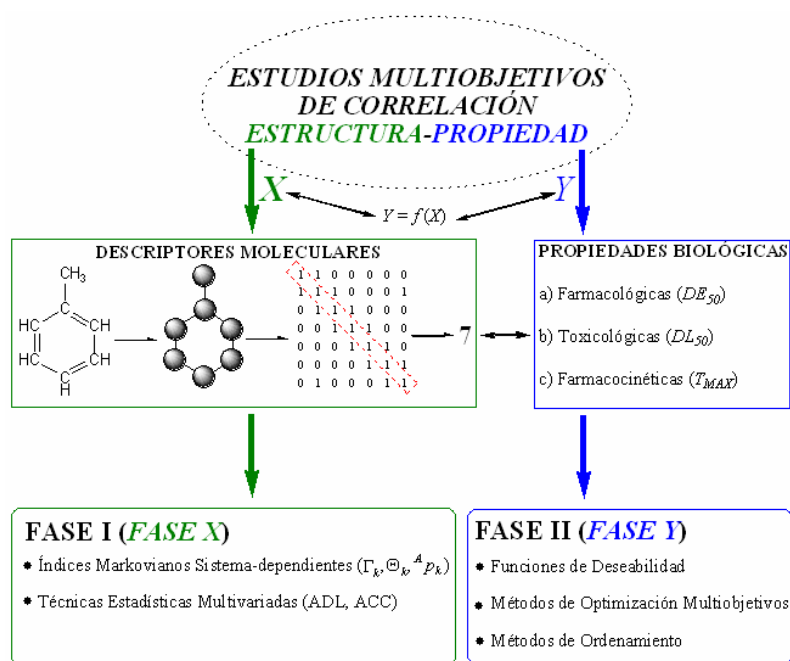


Figura 3.1 Representación esquemática de las 2 fases del proceso de desarrollo de los estudios de correlación estructura-propiedad multiobjetivos.

Debido a que la capacidad de estudiar múltiples propiedades de manera simultánea se logra mediante la modificación de los diferentes elementos involucrados en la relación estructura-propiedad, este trabajo se divide en dos etapas que se podrían llamar de modo informal Fases “X” y “Y”, en alusión al elemento que se transforma para lograr el objetivo final. La relación estructura-propiedad a que nos referimos

vista a través de un lente estadístico debe entenderse como la relación establecida entre la estructura molecular (descriptor molecular o variable independiente X) y la(s) propiedad(es) bajo estudio (variable dependiente Y). En una primera etapa se intenta conferir la propiedad “multiobjetivos” mediante la modificación de los descriptores moleculares (variable independiente X) de manera que estos sean capaces de codificar de manera diferente una misma molécula en dependencia del sistema que afecte. En una segunda etapa son las propiedades (variable dependiente Y) las que son transformadas de manera que sea posible su estudio de manera simultánea. La Figura 3.1 ilustra gráficamente esta división.

3.1 Estudios de Correlación Estructura-Propiedad Multisistemas Basados en el Uso de Índices Markovianos Sistema-Dependientes

3.1.1 Índices Markovianos Sistema-Dependientes

La teoría de las cadenas de Markov (31) se utilizará en nuestro trabajo como una aproximación sencilla a dos fenómenos moleculares:

- a) Interacción paso a paso de un fármaco con su receptor.
- b) Paso átomo por átomo de un fármaco desde una fase a otra (proceso de partición).

Particularmente en este trabajo nos proponemos conferir a los descriptores moleculares basados en cadenas de Markov previamente definidos (20, 36-74) la capacidad de codificar de manera diferente una misma molécula en dependencia del sistema con que interactúe. Esto se logra mediante el empleo de constantes

que aportan información sobre propiedades atómicas específicas en dependencia del sistema involucrado.

Así, para la modelación del proceso de interacción fármaco-receptor [caso a)] este problema es resuelto mediante la estimación de la constante Γ_j como la frecuencia de aparición del átomo j en moléculas de fármaco que inducen un efecto adverso determinado en un sistema de órganos específico.

Consideremos una situación hipotética en la cual una molécula de fármaco se encuentra libre en el espacio a un tiempo inicial arbitrario (t_0). Es entonces interesante desarrollar un modelo estocástico simple de la interacción paso-a-paso en el tiempo entre los átomos de una molécula de fármaco y un receptor molecular en la inducción de un efecto adverso.

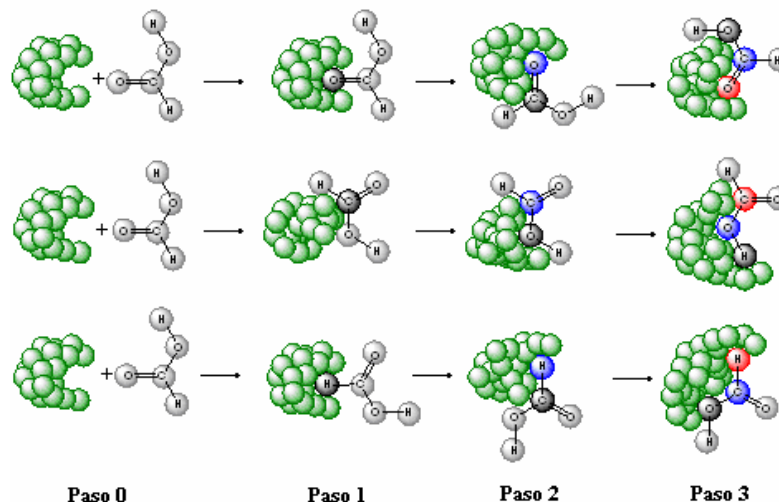


Figura 3.1.1 Modelo de interacción estocástica paso a paso fármaco-receptor.

La energía libre inicial de la interacción fármaco-receptor (0g_j) es una función de estados, por lo que un proceso reversible de interacción puede dividirse en varias interacciones elementales entre el átomo j y el receptor (278). Mas adelante en el

tiempo la interacción continúa y es necesario definir la energía libre de interacción entre el átomo j y el receptor dado que el átomo i ya ha interactuado a un tiempo previo t_k (${}^k g_{ij}$). En particular, inmediatamente después de la primera interacción ($t_0 = 0$) tiene lugar una interacción ${}^1 g_{ij}$ a tiempo $t_1 = 1$ y así sucesivamente a lo largo de periodos de tiempo discretos t_k ($k = 1, 2, 3, \dots, n$); ($k = 1$ en gris), ($k = 2$ en azul) y ($k = 3$ en rojo) a través del sistema de enlaces químicos. Existen varias formas alternativas en la que este proceso puede ocurrir. La Figura 3.1.1 ilustra esta idea.

La energía libre inicial del sistema ${}^0 g_j$ será considerado aquí como una función de la temperatura absoluta del sistema y la constante de interacción local en el equilibrio entre el átomo j y el receptor ${}^0 \Gamma_j$ (278). Adicionalmente, la energía ${}^1 g_{ij}$ puede ser definida por analogía en función de ${}^1 \Gamma_{ij}$:

$${}^0 g_j = R \cdot T \cdot \log {}^0 \Gamma_j \quad (3.1.1) \qquad {}^1 g_{ij} = R \cdot T \cdot \log {}^1 \Gamma_{ij} \quad (3.1.2)$$

$${}^0 \Gamma_j = \frac{n_j}{n} = e^{\frac{{}^0 g_j}{R \cdot T}} \quad (3.1.3) \qquad {}^1 \Gamma_{ij} = \alpha_{ij} \cdot \frac{n_j}{n} = e^{\frac{{}^1 g_{ij}}{R \cdot T}} \quad (3.1.4)$$

donde α_{ij} son los elementos de la matriz de adyacencia entre átomos, R es la constante de los gases y T es la temperatura absoluta del sistema.

La descripción del proceso de interacción a tiempos superiores $t_k > t_1$ permite el cálculo de la probabilidad de interacción fármaco-receptor en el tiempo hasta que el efecto bajo estudio es alcanzado. Este modelo trata con el cálculo de las probabilidades de interacción (${}^k p_{ij}$) de un átomo j cualquiera con el receptor molecular dado que otro átomo i ha interactuado antes; a través de periodos discretos de tiempo t_k .

El método ordena todos los valores de ${}^0\Gamma_j$ en un vector (Γ) y todas las constantes ${}^1\Gamma_{ij}$ como una matriz cuadrada de dimensión $n \times n$. Luego de la normalización de ambos podemos construir el vector de probabilidad absoluta inicial correspondiente ${}^0\pi$ y la matriz estocástica ${}^1\Pi$, la cual contiene los elementos ${}^A p_0(j)$ y ${}^1 p_{ij}$ respectivamente. Los elementos ${}^A p_0(j)$ de ${}^0\pi$ constituye la probabilidad absoluta con la cual el átomo j interactúa con el receptor a un tiempo inicial con respecto a cualquier átomo en la molécula:

$${}^A p_0(j) = \frac{{}^0\Gamma_j}{\sum_{a=1}^m {}^0\Gamma_a} = \frac{\frac{n_j}{n}}{\sum_{a=1}^m \frac{n_a}{n}} = \frac{\frac{1}{n} \cdot n_j}{\frac{1}{n} \cdot \sum_{a=1}^m n_a} = \frac{n_j}{\sum_{a=1}^m n_a} \quad (3.1.5)$$

donde, m representa todos los átomos en la molécula incluyendo al propio átomo j , n_a es la frecuencia de aparición de cualquier átomo incluyendo al átomo j con valor n_j . Por otra parte, la matriz ${}^1\Pi$ es denominada como la matriz estocástica de interacción fármaco-receptor en el primer paso. Los elementos (${}^1 p_{ij}$) de esta matriz son las probabilidades de interacción a tiempo $t_1 = 1$ del átomo j con el receptor dado que un átomo i ya se ha unido al receptor.

$${}^1 p_{ij} = \frac{{}^1\Gamma_{ij}}{\sum_{k=1}^{\delta+1} {}^1\Gamma_{ik}} = \frac{\alpha_{ij} \cdot \frac{n_j}{n}}{\sum_{k=1}^{\delta+1} \alpha_{ik} \cdot \frac{n_k}{n}} = \frac{\frac{1}{n} \cdot \alpha_{ij} \cdot n_j}{\frac{1}{n} \cdot \sum_{k=1}^{\delta+1} \alpha_{ik} \cdot n_k} = \frac{\alpha_{ij} \cdot n_j}{\sum_{k=1}^{\delta+1} \alpha_{ik} \cdot n_k} \quad (3.1.6)$$

donde δ representa la valencia del átomo j .

Mediante el empleo de estos elementos y las ecuaciones de Chapman-Kolmogorov es posible describir la futura evolución del sistema, mediante la determinación de

índices numéricos capaces de describir el proceso de interacción entre el átomo j y el receptor a instantes ulteriores de tiempo. Dada la capacidad de estos índices de codificar además de la estructura molecular, información relativa al efecto adverso inducido por una molécula en un sistema de órganos dado, consideramos apropiado denominar a esta familia de índices numéricos como *índices markovianos sistema-dependientes*.

Sumando las constantes de interacción para cada átomo podemos obtener la *Constante de Interacción Estocástica Promedio* (Γ_k) entre el fármaco y el receptor a instantes de tiempo específicos:

$$\Gamma_k = {}^0\boldsymbol{\pi} \cdot ({}^1\boldsymbol{\Pi})^k \cdot \boldsymbol{\Gamma} = \sum_{S_X}^n {}^A p_k(j) \cdot \Gamma_j = \sum_{S_X}^n {}^k \Gamma_j \quad (3.1.7)$$

Así mismo, se puede definir la *Entropía de Interacción Estocástica* Θ_k :

$$\Theta_k = {}^0\boldsymbol{\pi} \cdot ({}^1\boldsymbol{\Pi})^k \cdot \mathbf{Sh} = {}^k\boldsymbol{\pi} \cdot \mathbf{Sh} = -k_B \cdot \sum_{S_X}^n {}^A p_k(j) \cdot \log {}^A p_k(j) \quad (3.1.8)$$

así como la *Probabilidad Absoluta de Interacción* ${}^A p_k$ entre el fármaco y el receptor a instantes de tiempo específicos:

$${}^A p_k = {}^0\boldsymbol{\pi} \cdot ({}^1\boldsymbol{\Pi})^k = {}^k\boldsymbol{\pi} = \sum_{S_X}^n {}^A p_k(j) \quad (3.1.9)$$

donde ${}^k\boldsymbol{\pi}$ es un vector dependiente del tiempo cuyos componentes son las probabilidades absolutas ${}^A p_k(j)$ con las que un átomo interacciona con el receptor en instantes de tiempo ulteriores con respecto a cualquier átomo de la molécula. \mathbf{Sh} es la representación vectorial usada para describir las probabilidades ${}^A p_k(j)$ en términos de entropía y se expresa como su logaritmo: $\log {}^A p_k$. k_B es la constante de

Boltzman. S_X representa un grupo específico de átomos en la molécula. Cuando S_X contiene todos los átomos en la molécula el descriptor se convierte en un descriptor global. A su vez es posible calcular diferentes familias de descriptores moleculares mediante la selección de diferentes condiciones para S_X : Para este propósito los átomos son agrupados o clasificados en distintas clases s_x para describir aspectos locales de la estructura molecular: $s_1 = CSat$ = átomos de carbono saturados, $s_2 = CInst$ = átomos de carbono insaturados, $s_3 = Het$ = heteroátomos, $s_4 = HX$ = átomos de hidrógeno unidos a un heteroátomo.

De manera similar en el caso del proceso de partición [caso b)] es estimada la constante $\Gamma_{(j,sp,eb)}$ como el coeficiente de partición del átomo j en un sistema de partición sp dado y medido en una determinada especie biológica eb . $\Gamma_{(j,sp,eb)}$ es determinado como el cociente del número total de átomos j presentes en moléculas CTP en un sistema dado y medido en una especie específica $n_{CTP(j,sp,eb)}$ y este mismo valor para moléculas STP $n_{STP(j,sp,eb)}$.

$$\Gamma_{(j, sp, eb)} = \frac{n_{CTP(j, sp, eb)}}{n_{STP(j, sp, eb)}}$$

(3.1.10)

Este enfoque probabilístico introduce la idea de la descomposición paso a paso del proceso de partición de una molécula en un sistema bifásico como el paso átomo por átomo desde una fase a otra en intervalos discretos de tiempo t_k .

De forma similar a la anteriormente descrita es posible obtener la *Energía Libre de Partición Estocástica* ΔG_k de una molécula en un sistema bifásico (61):

$$\Delta G_k = -R \cdot T \cdot \log({}^k \Gamma) \quad (3.1.11)$$

donde R representa la constante universal de los gases y T la temperatura absoluta del sistema. Específicamente ${}^k \Gamma$ representa una constante cinética molecular promedio que caracteriza el proceso de partición a tiempo $t_k = k$. Cada ${}^0 \Gamma_{(j,sp,eb)}$ depende del sistema bifásico específico (sp) en que tiene lugar el proceso de partición y de la especie biológica en que fue medido (eb) a tiempo t_k :

$${}^k \Gamma = \sum_{j=1}^n A p_k(j, sp, eb) \cdot {}^0 \Gamma_{(j,sp,eb)} \quad (3.1.12)$$

$${}^0 \Gamma_{(j,sp,eb)} = \frac{n_{CTP(j,sp,eb)}}{n_{STP(j,sp,eb)}} = e^{\frac{g_{(j,sp,eb)}}{-R \cdot T}} \quad (3.1.13)$$

donde $n_{CTP(j,sp,eb)}/n_{STP(j,sp,eb)}$ es el número de átomos j presentes en moléculas con/sin tendencia a sufrir el proceso de partición en un sistema dado y medido en una especie biológica específica.

De manera similar es posible emplear los elementos de la diagonal principal o traza (Tr) de ${}^1 \mathbf{\Pi}$ para derivar otra familia de índices markovianos sistema-dependientes: los *Momentos Espectrales de Partición Estocástica* ${}^{SR} p_k$.

$${}^{SR} p_k = Tr({}^k \mathbf{\Pi}) = Tr({}^1 \mathbf{\Pi})^k = \sum_{i=j}^n {}^k p_{ij} \quad (3.1.14)$$

Como se podrá comprobar en los ejemplos desarrollados mas adelante, mediante el empleo de este tipo de descriptores será posible abordar de forma unificada múltiples aspectos de los perfiles toxicológicos y de biodistribución de fármacos y compuestos químicos orgánicos en general.

3.1.2 Predicción de múltiples efectos adversos de medicamentos

Artículo I. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. **Predicting Multiple Drugs Side Effects with a General Drug-Target Interaction Thermodynamic Markov Model.** *Bioorg. Med. Chem.* **2005**, *13*, 1119-1129.

Artículo II. Cruz-Monteagudo, M.; González-Díaz, H. **Unified Drug-Target Interaction Thermodynamic Markov Model using Stochastic Entropies to Predict Multiple Drugs Side Effects.** *Eur. J. Med. Chem.* **2005**, *40*, 1030-1041.

Artículo III. Cruz-Monteagudo, M.; González-Díaz, H.; Uriarte, E. **Simple Stochastic Fingerprints Towards Mathematical Modeling in Biology and Medicine. 2. Unify Markov Model for Drugs Side Effects.** *Bull. Math. Biol.* **2006**, *68*, 1527-1554.

En estos trabajos se desarrolla un método que permite el análisis simultáneo de varias propiedades para predecir la probabilidad de un fármaco de inducir un efecto adverso específico con relación a otro efecto adverso posible en un sistema de órganos dado. El empleo de los descriptores moleculares descritos en la sección anterior a través de una técnica de análisis multivariado como el análisis discriminante lineal permite predecir el perfil toxicológico de un conjunto de fármacos. Específicamente en el estudio se emplearon tres familias de descriptores moleculares los cuales fueron empleados en una base de datos de fármacos estructural y farmacológicamente diversos de los cuales se conoce *a priori* sus efectos adversos sobre varios sistemas de órganos.

Tabla 3.1.1 Desempeño de las tres familias de *índices markovianos sistema-dependientes* (Γ_k , Θ_k y ${}^A p_k$).

EFECTO ADVERSO	Γ_k			Θ_k			${}^A p_k$		
	Ent.	Val.	Estadísticos	Ent.	Val.	Estadísticos	Ent.	Val.	Estadísticos
Manifestaciones Respiratorias									
Infiltración pulmonar	92.3	98.1	N = 35	70.0	61.5	N = 35	84.6	72.7	N = 35
Broncoespasmo	100	100	$U = 0.114$	77.3	79.5	$U = 0.723$	72.7	78.4	$U = 0.705$
Total	97.1	97.5	$F = 255.7$ $p' = 2$	74.3	72.8	$F = 2.88$ $p' = 5$	77.1	75.7	$F = 3.13$ $p' = 5$
Manifestaciones Cardiovasculares									
Exacerbaciones anginosas	100	100		---	---		80.	77.5	
Arritmias	100	100		---	---		---	---	N = 37
Edema o retención líquida	100	98.9	N = 125	---	---	N = 27	---	---	$U = 0.407$
En insuficiencia cardíaca			$U = 0.003$			$U = 0.317$			$F = 4.39$
Hipertensión	100	100	$F = 158.6$	84.6	82.7	$F = 11.87$	76.9	76.9	$p' = 5$
Hipotensión	86.4	96.6	$p' = 4$	---	---	$p' = 5$	---	---	
Tromboembolismo	100	100		100	100		78.6	98.2	
Total	97.6	96.8		92.6	91.7		78.4	85.1	
Manifestaciones Hematológicas									
Agranulocitosis	85	98.8		75.0	76.3		---	---	
Anemia hemolítica	100	100		75.0	72.3		---	---	
Anemia hemolítica (en déficit de G6PD)	100	100	N = 103	---	---	N = 48	---	---	
Pancitopenia	90	100	$U = 0.012$	---	---	$U = 0.670$	---	---	---
Alteración de la agregación plaquetaria	100	100	$F = 90.5$	---	---	$F = 5.29$	---	---	
			$p' = 4$	---	---	$p' = 5$	---	---	
Total	95.1	99.1		75.0	74.0		---	---	
Manifestaciones Gastrointestinales									
Constipación o íleo	90.9	100		90.9	95.5		90.9	97.7	
Diarrea o colitis	100	100		---	---		---	---	
Daño hepatocelular difuso	95	98.8	N = 164	---	---	N = 42	---	---	N = 42
Sequedad de boca	86.7	100	$U = 0.002$	---	---	$U = 0.252$	---	---	$U = 0.259$
Nausea o vómito	100	100	$F = 190.9$	---	---	$F = 12.26$	---	---	$F = 11.88$
Pancreatitis	100	100	$p' = 4$	93.3	93.3	$p' = 4$	93.3	93.3	$p' = 4$
Úlcera péptica o hemorrágica	93.8	100		75.0	87.5		87.5	87.5	
Total	97	99.4		85.7	91.7		90.5	92.3	
Manifestaciones Dérmicas									
Acné	50	90		---	---		80.0	82.5	
Alopecia	100	100	N = 50	87.5	82.8	N = 28	75.0	89.1	N = 38
Eritema difuso	91.7	100	$U = 0.004$	---	---	$U = 0.484$	---	---	$U = 0.347$
			$F = 3920$	---	85.4	$F = 8.54$	---	---	$F = 4.32$
Fotodermatitis	100	100	$p' = 2$	91.7		$p' = 4$	91.7	85.4	$p' = 6$
Total	88	92.3		89.3	83.9		81.6	86.2	
Fenómenos Sistémicos									
Anafilaxia	100	100	N = 47	85.7	83.9	N = 26	85.7	82.1	N = 26
Lupus eritematoso	100	100	$U = .042$	91.7	87.5	$U = 0.437$	91.7	91.7	$U = 0.404$
Fiebre	100	100	$F = 506.1$	---	---	$F = 9.45$	---	---	$F = 7.76$
Total	100	100	$p' = 2$	88.5	85.6	$p' = 4$	88.5	86.5	$p' = 5$

p' : número de parámetros ajustables incluidos en cada función de clasificación.

Tabla 3.1.1 (continuación...)

EFECTO ADVERSO	Γ_k			Θ_k			$^A p_k$		
	Ent.	Val.	Estadísticos	Ent.	Val.	Estadísticos	Ent.	Val.	Estadísticos
Manifestaciones Metabólicas									
Galactorrea (amenorrea)	100	100		88.9	88.9		100	94.4	
Disminución de la libido e impotencia	100	100	N = 33 U = 0.12	---	---	N = 18 U = 0.529	---	---	N = 18 U = 0.309
Desordenes en tests de función tiroidea	88.9	94.4	F = 110 p' = 2	88.9	88.9	F = 6.68 p' = 3	100	97.2	F = 7.28 p' = 5
Total	97	95		88.9	88.9		100	95.8	
Hiperglicemia	100	100	N = 18	77.8	88.9	N = 18	---	---	
Hipopotasemia	100	100	U = 0.086	88.9	69.4	U = 0.586	---	---	---
Total	100	100	F = 170.8 p' = 2	83.3	79.2	F = 3.29 p' = 4	---	---	
Manifestaciones Neurológicas									
Convulsiones	100	100	N = 33	90.0	82.5	N = 33	90.0	87.5	N = 33
Efectos extrapiramidales	100	100	U = 0.026	70.0	71.1	U = 0.614	70.0	61.5	U = 0.632
Total	100	100	F = 1158.1 p' = 2	81.8	78.0	F = 3.40 p' = 6	81.8	77.3	F = 3.14 p' = 6
Manifestaciones Psiquiátricas									
Confusión o delirios	100	100	N = 55	---	---	N = 36	73.7	69.7	N = 41
Disfunciones del sueño	100	100	U = 0.04	85.7	85.7	U = 0.480	---	---	U = 0.660
Somnolencia	90.9	96.6	F = 66.5	90.9	89.8	F = 36.81	77.3	77.0	F = 9.79
Total	96.4	94.6	p' = 4	88.9	88.2	p' = 2	75.6	75.0	p' = 3
Manifestaciones Músculo-Esqueléticas									
Miopatía o mialgia	100	100	N = 23	---	---		---	---	
Osteoporosis	100	100	U = 0.027	---	---	---	---	---	---
Total	100	100	F = 361.4 p' = 3	---	---		---	---	
Manifestaciones Endocrinas									
Galactorrea (amenorrea)	100	100		88.9	88.9		100	94.4	
Disminución de la libido e impotencia	100	100	N = 33 U = 0.12	---	---	N = 18 U = 0.529	---	---	N = 18 U = 0.309
Desordenes en tests de función tiroidea	88.9	94.4	F = 110 p' = 2	88.9	88.9	F = 6.68 p' = 3	100	97.2	F = 7.28 p' = 5
Total	97	95		88.9	88.9		100	95.8	

Ent./Val.: serie de entrenamiento/validación; N: número de casos; U: Lambda de Wilk; F: razón de Fisher; p': número de parámetros ajustables incluidos en cada función de clasificación.

La capacidad de cada familia de descriptores moleculares previamente definidos (*Constante de Interacción Estocástica Promedio*, Γ_k ; *Entropía de Interacción Estocástica*, Θ_k y *Probabilidad Absoluta de Interacción*, $^A p_k$) para predecir el perfil toxicológico fue probada. Su desempeño se resume en la Tabla 3.1.1 en la que se muestran los parámetros estadísticos del análisis discriminante lineal junto con los porcentajes de buena clasificación exhibido por cada modelo en el entrenamiento y

en el proceso de validación. Todas las funciones de clasificación mostradas son estadísticamente significativas ($p < 0.05$).

De manera general, es fácil identificar a Γ_k como la familia de descriptores más eficiente en la descripción del problema. La misma fue capaz de explicar de manera satisfactoria los 39 efectos adversos agrupados en los 11 sistemas de órganos bajo estudio. Es posible notar que las funciones de clasificación emplean un número mínimo de variables (de 1 a 3 DMs); los porcentajes de compuestos bien clasificados para cada sistema de órganos en el entrenamiento y la validación (88% -100% y 92.3% - 100%, respectivamente); así como la separación entre los grupos de efectos adversos dentro de cada sistema de órganos ($U = 0.002 - 0.114$) fueron excelentes. Por su parte Θ_k fue capaz de explicar de manera satisfactoria 21 efectos adversos agrupados en 10 sistemas de órganos. Las funciones de clasificación obtenidas emplean de 1 a 5 descriptores moleculares). Se obtuvieron porcentajes de buena clasificación entre 74.3% y 92.6 % en el entrenamiento y entre 72.8% y 91.7 % para la validación. Así mismo, $^A p_k$ fue capaz de explicar de manera satisfactoria solo 19 efectos adversos agrupados en 8 sistemas de órganos. Las funciones de clasificación obtenidas emplean de 2 a 5 descriptores moleculares. Se obtuvieron porcentajes de buena clasificación entre 75.6% y 100 % en el entrenamiento y entre 75% y 95.8 % para la validación.

Finalmente, es posible afirmar que mediante la introducción de una familia de descriptores moleculares basados en cadenas de Markov capaces de codificar además de la estructura molecular del fármaco, la frecuencia de un efecto adverso determinado en un sistema de órganos específico, fue posible predecir el perfil

toxicológico (teniendo en cuenta múltiples efectos adversos asociados a varios sistemas de órganos) de un conjunto de fármacos.

Aún cuando el perfil toxicológico de un fármaco determina su perfil terapéutico, el perfil de biodistribución del mismo es igualmente determinante. Los siguientes trabajos tratan este aspecto mediante el uso de los índices markovianos sistema-dependientes.

3.1.3 Predicción del perfil de distribución de fármacos y compuestos químicos en sistemas de partición bifásicos de relevancia biológica

Artículo **IV**. González-Díaz, H.; Cabrera-Pérez, M. A.; Agüero-Chapín, G.; Cruz-Monteagudo, M.; Castañeda-Cancio, N.; del Rio M.A.; Uriarte, E. **Multi-Target QSPR Assemble of a Complex Network for the Distribution of Chemicals to Biphasic Systems and Biological Tissues**. *Chemomet. Intel. Lab. Systems*. **2008**, *94*, 160-165.

En este trabajo se pretende extender el uso de los momentos espectrales (${}^{SR}p_k$) de la matriz estocástica, incorporando ponderaciones atómicas relacionadas con el sistema de distribución específico y la especie biológica en que ocurre este proceso. Luego se muestra la posibilidad de analizar este conjunto de datos multisistemas a través de un único modelo QSPR. Además, se reportan nuevos resultados experimentales para el fármaco antibacteriano G-1 y se emplea el modelo desarrollado para el estudio de la distribución de este fármaco en diferentes tejidos de rata.

La función discriminante encontrada se muestra a continuación junto con información relativa al número de compuestos total, *CTP* y *STP* ($N(T)$, $N(CTP)$,

N(STP)), así como los porcentajes de compuestos clasificados correctamente tanto para el entrenamiento como para la validación cruzada. En la ecuación los términos *CInst*, *HX* y *CSat* indican que los momentos espectrales se calcularon sumando solamente los átomos de carbono insaturados, hidrógenos unidos a heteroátomos o carbonos saturados y no todos los átomos presentes en la molécula.

$$CTP = -7.21 + 0.84 \cdot^{SR} p_0(CInst) - 5.79 \cdot^{SR} p_1(HX) + 2.49 \cdot^{SR} p_5(CSat) \quad (3.1.15)$$

$N(T) = 413$	$\%(T) = 90.6$	$\%(Tcv) = 90.0$
$N(CTP) = 112$	$\%(CTP) = 95.5$	$\%(CTPcv) = 95.1$
$N(STP) = 301$	$\%(STP) = 88.7$	$\%(STPcv) = 88.0$

Tanto la significación estadística (nivel de significación $p < 0.01$ para una prueba de Fisher) como los valores de exactitud mostrados por el modelo son lo suficientemente apropiados como para aceptarlo como válido.

La predicción de compuestos CTP que posean el potencial de pasar desde el plasma hacia un tejido u órgano específico es un área específica pero muy interesante y útil de las ciencias farmacéuticas. Esto pudiera contribuir al desarrollo de lo que se conoce como vectorización de fármacos hacia un tejido específico (279). Los valores promediados de las probabilidades (expresados en %) con que el modelo clasifica a los compuestos CTP como tal están por encima del 85% en todos los casos. Este tipo de estudio pudiera ser denominado como vectorización *in silico*. Su mayor utilidad esta dada en la disminución de efectos adversos del fármaco al dirigir y concentrar su acción en el órgano diana (280).

El modelo también clasifica satisfactoriamente los compuestos STP tanto en el entrenamiento (88.7%) como en la validación cruzada (88.0%) El estudio del particionamiento de compuestos químicos orgánicos STP desde el aire hacia diversos órganos igualmente importante. La identificación de estos compuestos como probablemente seguros resulta de gran utilidad para la toxicología ambiental. Los valores promediados de las probabilidades con que el modelo clasifica a los compuestos como STP son notablemente mayores que el 95% para todos los casos.

Seguidamente, es ilustrado el uso del modelo obtenido para el cálculo de las probabilidades de distribución del fármaco antibacteriano G-1 (281) desde el plasma hacia diferentes tejidos de rata (piel, corazón y cerebro). La Tabla 3.1.2 resume los resultados obtenidos. Las clasificaciones del modelo coinciden con las clasificaciones experimentales deducidas de los coeficientes de partición determinados. Estos valores se determinaron tomando como referencia la concentración de [¹⁴C]-G1 en plasma luego de tres horas de la administración oral. Los resultados obtenidos apuntan hacia una pobre distribución del G-1 y presumiblemente a pocos efectos adversos en estos tejidos.

Tabla 3.1.2 Perfiles de distribución plasma-tejido experimental y teórico del G-1.

Sistema	Tejido	Valores Experimentales				Predicciones		
		C ^a	C _{error}	P ^b	logP	CE ^c	CT ^d	p% ^e
Plasma/Piel (Rata)	Piel	8.46	3.84	0.95	-0.02	STP	STP	45.5
Plasma/Corazón (Rata)	Corazón	8.39	3.94	0.94	-0.02	STP	STP	45.3
Plasma/Cerebro (Rata)	Cerebro	4.06	0.37	0.45	-0.34	STP	STP	32.4

^a Concentración plasmática (mg·mol⁻¹); ^b Coeficiente de partición; ^c Clasificación experimental; ^d Clasificación teórica; ^e Probabilidad de clasificación predicha (%).

Finalmente, basados en los resultados descritos se puede afirmar que es posible extender los momentos espectrales de la matriz estocástica de manera que estos codifiquen información relativa al ambiente molecular y emplearlos en la predicción del perfil de distribución de un compuesto en múltiples sistemas. A su vez, la coincidencia entre los estudios teóricos y experimentales del perfil de distribución del G-1 confirma la utilidad del modelo desarrollado.

3.1.4 Predicción del perfil de distribución de fármacos, compuestos químicos y contaminantes ambientales en sistemas de partición bifásicos de relevancia biológica y ambiental

Artículo V. Cruz-Monteagudo, M.; González-Díaz, H.; Agüero-Chapín, G.; Santana, L.; Borges, F.; Domínguez, E. R.; Podda, G.; Uriarte, E. **Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems.** *J. Comput. Chem.* **2007**, *28*, 1909-1923.

Este trabajo constituye una continuación del anterior y tiene como principal objetivo extender su aplicación a sistemas de interés para la toxicología ambiental y evaluar otras herramientas de modelación. Específicamente se desarrolla un modelo de clasificación unificado del proceso de partición basado en cadenas de Markov capaz de clasificar una base de datos estructuralmente diversa compuesta por 1300 fármacos y compuestos químicos de acuerdo a su comportamiento en 38 sistemas de partición bifásicos. El modelo de predicción de mejor desempeño se detalla a continuación:

$$CTP = 0.36 + 3.1\Delta G_5(T) - 1.01\Delta G_5(HX) + 1.54\Delta G_0(T) - 0.89\Delta G_1(CSat) \quad (3.1.16)$$

En la ecuación CTP es una variable dicotómica que toma el valor de 1 si el compuesto tiene tendencia a la partición; de otra forma toma el valor de -1. Los elementos $\Delta G_0(T)$ y $\Delta G_5(T)$ representan las energías libres de interacción estocástica inicial y a tiempo $t_k = 5$ de la molécula, respectivamente; y $\Delta G_1(\text{CSat})$ y $\Delta G_5(\text{HX})$ representan las energías libres de interacción estocástica a tiempo $t_k = 1$ y 5, de carbonos saturados e hidrógenos lábiles, respectivamente.

Para evaluar la capacidad predictiva del modelo se reporta además de los resultados alcanzados en los subconjuntos de entrenamiento y predicción, los resultados promediados de los cuatro subconjuntos de validación mediante una técnica de re-sustitución de casos. Los valores en todos los experimentos de validación permanecen estables alrededor de un valor promediado total de compuestos clasificados correctamente de 91.38%, 91.31% y 91.37% en los subconjuntos de entrenamiento, predicción y el conjunto total de compuestos, respectivamente (ver detalles en la Tabla A.1).

Tabla 3.1.3 Matriz de correlación entre variables relativa a la ecuación 3.1.16.

	$\Delta G_0(T)$	$\Delta G_5(T)$	$\Delta G_1(\text{CSat})$	$\Delta G_5(\text{HX})$
$\Delta G_0(T)$	1.000			
$\Delta G_5(T)$	0.798	1.000		
$\Delta G_1(\text{CSat})$	0.588	0.678	1.000	
$\Delta G_5(\text{HX})$	0.389	0.225	0.364	1.000

El cumplimiento de los supuestos paramétricos es un aspecto determinante en la aplicación de técnicas estadísticas multivariadas a estudios QSPR ya que la validez y significación estadística de un modelo está fuertemente condicionado esto (229, 282). A excepción de una ligera violación del supuesto de normalidad y un también

ligero grado de colinealidad entre las variables incluidas en el modelo, el modelo propuesto cumple con los principales supuestos paramétricos. Un análisis detallado de este aspecto puede encontrarse en la publicación original (283).

Teniendo en cuenta el ligero grado de colinealidad (284), las variables incluidas en el modelo son ortogonalizadas por el método de Randić (232-234).

$$CTP = 0.36 + 3.5^1 O_5(T) - 0.86^2 O_5(HX) + 1.55^3 O_0(T) - 0.89^4 O_1(CSat) \quad (3.1.17)$$

Aunque el modelo propuesto exhibe una adecuada capacidad predictiva, significación estadística y parsimonia; las pequeñas desviaciones de la normalidad de sus variables así como la colinealidad mostrada justifica la exploración de otras alternativas. Considerando la naturaleza no lineal y no paramétrica de las redes neuronales artificiales, se desarrollaron varios modelos con el objetivo de superar las limitaciones del modelo discriminante lineal y mejorar el desempeño del mismo.

Tabla 3.1.4 Resumen de los mejores modelos de redes neuronales artificiales basados en 5 arquitecturas de red diferentes.

Modelo	Arquitectura	Desemp. Ent.^a	Desemp. Sel.^b	Desemp. Pred.^c
<i>MLP3</i>	4:4-7-1:1	94.23	93.08	93.46
<i>MLP4</i>	4:4-10-7-1:1	93.46	92.69	93.08
<i>Entrenamiento Lineal de la Red</i>	4:1:1	91.28	92.69	89.23
<i>Red Neuronal Probabilística</i>	4:4-780-2-2:1	78.59	79.62	76.54
<i>RBF</i>	4:1:1	52.44	52.31	50.38
	Corr. (%)_{STP}^d	Incorr. (%)_{STP}^e	Corr. (%)_{CTP}^f	Incorr. (%)_{CTP}^g
<i>MLP3</i>	93.91	6.09	93.79	6.21
<i>MLP4</i>	93.26	6.74	93.21	6.79
<i>Entrenamiento Lineal de la Red</i>	91.12	8.88	91.19	8.82
<i>Red Neuronal Probabilística</i>	56.09	43.91	97.98	2.02
<i>RBF</i>	51.97	48.03	52.02	47.98

^{a/b/c} desempeño del modelo en el subconjunto de entrenamiento/selección/predicción; ^{d/e} Porcentaje de compuestos STP clasificados correctamente/incorrectamente; ^{f/g} Porcentaje de compuestos CTP clasificados correctamente/incorrectamente.

Como muestra la Tabla 3.1.4, el modelo basado en funciones de base radial (RBF) no muestra resultados prometedores. Sin embargo, el resto de las arquitecturas de redes exploradas muestran un desempeño satisfactorio.

Considerando la simpleza y la naturaleza lineal del modelo discriminante lineal (best subset selection) y comparando su desempeño con el resto de los modelos de clasificación obtenidos decidimos proponerlo como el modelo más confiable, predictivo y parsimonioso.

Como se mencionó anteriormente, el sistema de referencia ha sido definido de forma que el proceso de partición ocurre desde la fase más organizada hacia la menos organizada. En este trabajo se demuestra la estabilidad de los modelos propuestos ante la inversión de la dirección del proceso de partición. Para esto fue seleccionado uno de los cinco modelos basados en análisis discriminante lineal propuestos (modelo 2 de la Tabla A.2). Para esto se invirtió directamente en la ecuación el signo de las variables independientes (energías libres de partición ΔG_k) y consecuentemente la clasificación STP/CTP de las moléculas para un sistema de partición bifásico dado. Luego se recalcularon los coeficientes de cada variable independiente y el intercepto en la ecuación. Este proceso fue aplicado a cada uno de los 38 sistemas de partición bifásicos. Como se observa en la Tabla A.3, la variación de los coeficientes de la ecuación luego de la inversión de la dirección del proceso de partición en cada uno de los 38 sistemas de partición bifásicos es mínima. Los resultados de esta prueba sugieren que el modelo discriminante lineal es capaz de clasificar correctamente el 89.08% de los compuestos con un margen

de error de solo 0.17% independientemente de la dirección del proceso de partición usada para ajustar el modelo.

El análisis realizado en este estudio demuestra que el perfil de distribución de un amplio número de compuestos estructuralmente diversos (1300) en 38 sistemas de partición bifásicos de relevancia farmacéutica y/o ambiental, puede ser predicho a través de una ecuación de clasificación lineal simple y físicamente coherente.

3.2 Empleo de Métodos de Optimización Multiobjetivos y Ordenamiento Basados en Funciones de Deseabilidad en Estudios Globales de Correlación Estructura-Actividad

Como ha sido ilustrado en los trabajos anteriores, mediante el empleo de los índices markovianos sistema-dependientes es posible acceder a través de un enfoque unificado a múltiples aspectos del perfil toxicológico o de biodistribución de un candidato a fármaco. Aún cuando varios sistemas pueden ser analizados simultáneamente mediante este enfoque, su principal limitación es que no es posible analizar de forma integradora los posibles compromisos que son necesarios entre estos perfiles. Esta limitación es superada por medio del empleo de funciones de deseabilidad en experimentos de optimización multiobjetivos de varias propiedades involucradas en el perfil terapéutico de un fármaco.

3.2.1 Optimización Simultánea del perfil analgésico, antiinflamatorio y ulcerogénico de una familia de quinazolinonas

Artículo VI. Cruz-Monteaquedo, M.; Borges, F.; Cordeiro, M. N. **Desirability-Based Multi-Objective Optimization for Global QSAR Studies. Application to the**

Design of Novel NSAIDs with Improved Analgesic, Anti-inflammatory and Ulcerogenic Profiles. *J. Comput. Chem.* 2008, 29, 2445-2459.

Para probar la utilidad de la metodología de optimización multiobjetivos propuesta, esta fue aplicada a una biblioteca de quince quinazolinonas reportada por Alagarsamy y colaboradores (250) con el objetivo de optimizar simultáneamente sus propiedades analgésicas, antiinflamatorias y ulcerogénicas. Los detalles relativos a los compuestos usados en el entrenamiento pueden consultarse en la publicación original (274).

El primer paso para la optimización simultánea es desarrollar un modelo predictivo independiente para cada objetivo a optimizar. Consecuentemente, un modelo predictivo basado en regresión lineal múltiple fue desarrollado para cada propiedad. Como se puede apreciar en la Tabla 3.2.1, la calidad de los modelos desarrollados es palpable. La calidad general de los mismos es indicada por altos valores de F y pequeños valores de p , valores satisfactorios de ρ ($\rho = 5$), junto con valores de R^2 y $Adj.R^2$ (bondad del ajuste) que van desde 0.803 a 0.935 y 0.771 a 0.923, respectivamente; así como valores de Q^2_{LOO} (capacidad predictiva) entre 0.713 y 0.905.

Cada modelo de predicción desarrollado incluye tres parámetros ajustables. Por tanto, cualquier compuesto nuevo que sea predicho a través de uno de los tres modelos; con valores de influencia o leverage (h) mayor que el leverage crítico ($h^* = 0.6$) y/o valores de residual estandarizado mayor que dos esta fuera del dominio de aplicabilidad del modelo y consecuentemente tales predicciones deben considerarse con precaución. Otro factor que determina la fiabilidad de las

predicciones está relacionado con el cumplimiento de los supuestos paramétricos de la técnica de modelación usada (282). En esta caso no se detectó ninguna violación de tales supuestos (ver detalles en la publicación original (274)).

Tabla 3.2.1 Indicadores estadísticos de los modelos de predicción basados en descriptores moleculares de fragmentos centrados en átomos.

Modelo para la Actividad Analgésica (A_n)								
$A_n = 51.762(\pm 2.155) + 8.333(\pm 0.957) \cdot C - 001 - 6.929(\pm 1.534) \cdot C - 037$								
N	R	R^2	$R^2 \text{ Adj.}$	Q^2	$SPRESS$	ρ	F	P
15	0.967	0.935	0.923	0.905	3.143	5.000	85.15699	0.000000
Modelo para la Actividad Antiinflamatoria (A_a)								
$A_a = 36.708(\pm 1.789) + 5.527(\pm 1.232) \cdot C - 001 + 1.475(\pm 0.430) \cdot H - 046$								
N	R	R^2	$R^2 \text{ Adj.}$	Q^2	$SPRESS$	ρ	F	P
15	0.942	0.887	0.869	0.827	3.526	5.000	47.46719	0.000002
Modelo para el Índice Ulcerogénico (U)								
$U = 0.718(\pm 0.044) - 0.056(\pm 0.020) \cdot C - 001 + 0.137(\pm 0.032) \cdot C - 037$								
N	R	R^2	$R^2 \text{ Adj.}$	Q^2	$SPRESS$	ρ	F	P
15	0.896	0.803	0.771	0.713	0.065	5.000	24.56766	0.000057

Considerando los resultados anteriores la optimización simultánea de las propiedades analgésicas, antiinflamatorias y ulcerogénicas del conjunto de quinazolinonas puede ser llevado a cabo con un nivel adecuado de confiabilidad.

Ahora, el siguiente paso es usar los valores predichos de cada propiedad para ajustar un modelo conteniendo todas las variables independientes (C-001, C-037 y H-046) usadas en modelar las propiedades originales (A_n , A_a y U) que serán usados para especificar la función de deseabilidad respectiva de cada propiedad. Aquí, A_n y A_a son maximizadas y U minimizada con el objetivo de obtener un candidato que exhiba simultáneamente la mayor actividad analgésica y antiinflamatoria a la vez que el menor índice ulcerogénico posibles.

R^2_D está próximo a 1 lo que indica la calidad final del proceso de optimización simultánea. Por otra parte, un valor de 0.905 de Q^2_D indica una capacidad predictiva adecuada del método para predecir los valores de D .

Finalmente se procede a la optimización de la deseabilidad global. Como resultado del proceso de optimización es posible determinar los niveles de C-001, C-037 y H-046 que simultáneamente producen la combinación más deseable de las propiedades analgésica, antiinflamatoria y ulcerogénica.

La Figura 3.2.1 muestra la función de deseabilidad global basada en C-001, C-037 y H-046; así como las funciones de deseabilidad individual debido a la actividad analgésica, antiinflamatoria y el índice ulcerogénico determinadas por los respectivos pares de descriptores moleculares incluidos en cada modelo. La inspección visual de esta figura sugiere que es posible obtener un candidato con un 93.43% de actividad analgésica, un 82.04% de inhibición de la inflamación y un índice ulcerogénico igual a 0.44, lo que se traduce en un candidato con una deseabilidad global de 0.8; el cual pudiera obtenerse si exhibiese valores de C-001, C-037 y H-046 iguales a 5, 0 y 12, respectivamente. De cualquier modo, solo deben considerarse candidatos incluidos dentro del dominio de aplicabilidad de los modelos de predicción originales. Esto es debido al hecho que los compuestos del entrenamiento no muestran valores de C-001 superiores a 3 y por consiguiente candidatos nuevos con un valor de C-001 muy alejados de 3 podrían estar fuera del dominio de aplicabilidad.

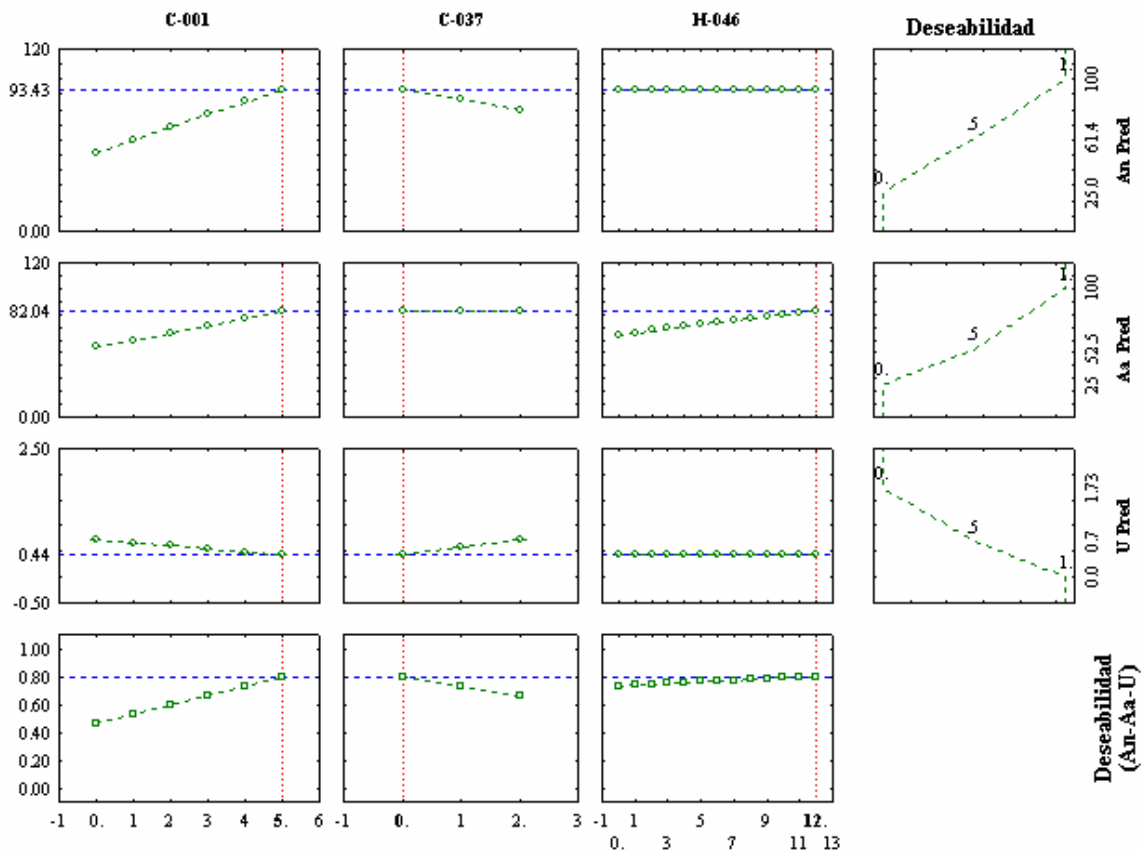


Figura 3.2.1 Función de discapacidad global a debido a la actividad analgésica, antiinflamatoria e índice ulcerogénico [$D(An-Aa-U)$] (última fila), y las respectivas funciones de discapacidad individual derivadas de los pares de variables predictivas incluidas en los tres modelos de regresión lineal múltiple (primeras tres filas).

De acuerdo con estos resultados, la variable más influyente sobre D es C-001. La interpretación estructural de este descriptor es muy simple: el número de grupos metilo en la molécula. Al mismo tiempo, la segunda variable más influyente, C-037, representa el número de heteroátomos unidos a un átomo de carbono Sp_2 que a su vez esta unido a un anillo aromático. Finalmente, la variable menos influyente es H-046, la cual representa el número de átomos del hidrógeno unidos a un carbono Sp_3 no unido a un heteroátomo unido a un átomo de carbono. Usando esta

información es fácil intuir las modificaciones necesarias para mejorar la deseabilidad global de los compuestos bajo estudio. Considerando la influencia positiva/negativa de C-001/C-037 sobre D es necesario incrementar el número de grupos metilo al tiempo que evitar la presencia de anillos aromáticos en la posición del C-2 del anillo de quinazolina. La forma más simple de aumentar el número de grupos metilo en esta posición es mediante sustituyentes alquílicos, por lo que se puede inferir un rol positivo del volumen de los sustituyentes en esta posición.

Una serie de nueve compuestos fue diseñada de manera tal que varios sustituyentes alquílicos son ligados a la posición C-2 del anillo de quinazolina, incrementando gradualmente el volumen del sustituyente. Las modificaciones hechas se muestran en la Tablas A.4 y A.5 junto con los valores predichos de An , Aa , U y D . Como se puede notar, algunos de los nuevos compuestos (**ASNEW4**, **ASNEW5**, **ASNEW6** y **ASNEW9**) están fuera del dominio de aplicabilidad de los modelos de predicción originales. Por esta razón se decide proponer el compuesto **ASNEW8** como el candidato más deseable y fiable de los diseñados en este estudio. El compuesto **ASNEW8** muestra un 93% de An , un 82% de Aa y un índice ulcerogénico de 0.44 lo que representa una excelente deseabilidad global predicha de 0.8.

Si los valores predichos de An , Aa y U de **ASNEW8** son comparados con el compuesto mas prometedor reportado por Alagarsamy y colaboradores (**AS3**) se puede observar una notable mejora. Específicamente, **ASNEW8** muestra valores de An y Aa 15% y 13% superior, respectivamente. Al mismo tiempo, **ASNEW8** mostró sólo el 78,6% de la capacidad ulcerogénica de **AS3**. Por otro lado, si

comparamos a **ASNEW8** con el diclofenaco es posible notar que **ASNEW8** supera dramáticamente el perfil analgésico, antiinflamatorio y ulcerogénico del diclofenaco (*An* y *Aa* 31% y 22% superior, respectivamente y casi una cuarta parte de la capacidad ulcerogénica del diclofenaco).

El principal resultado de la aplicación de estas transformaciones se puede resumir en una mejora significativa en el perfil ulcerogénico de los nuevos candidatos diseñados (aunque los perfiles analgésico y antiinflamatorio también mejoran simultáneamente). Este resultado sugiere un papel positivo de los sustituyentes alquílicos voluminosos en la posición C-2 del anillo de quinazolina sobre las propiedades ulcerogénicas de esta familia de compuestos. En cualquier caso, todos los resultados mostrados anteriormente son sólo teóricos y, lógicamente deben considerarse con cautela hasta una posterior corroboración experimental.

El uso de una base de datos pequeña y homogénea como la utilizada en este trabajo es más adecuado para fases posteriores del proceso de descubrimiento de fármacos una vez identificado el líder. En este punto, se pueden llevar a cabo modificaciones estructurales específicas sobre el líder siguiendo los resultados del proceso de optimización multiobjetivos. Es debido a esto que el uso de descriptores estructurales y/o fisicoquímicos claramente definidos e interpretables puede conducir a relaciones de estructura-deseabilidad que pueden ser utilizados en el diseño de nuevos candidatos con una mejoría simultánea de su perfil farmacológico, farmacocinético y/o toxicológico. Sin embargo, como se verá mas adelante, con la aplicación de la metodología MOOP-DESIRE también es posible manejar grandes y/o más diversos conjuntos de datos.

Finalmente, los resultados obtenidos en este estudio demuestran la utilidad de la metodología MOOP-DESIRE en el descubrimiento y desarrollo racional de nuevos fármacos con perfiles farmacológico y toxicológico simultáneamente mejorados.

3.2.2 Optimización simultánea del perfil antibacteriano y citotóxico de fluoroquinolonas

Artículo VII. Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N.; Fajin, J. L.; Morell, C.; Molina, R.; Cañizares-Carmenate, Y.; Rosa, E. **Desirability-Based Methods of Multi-Objective Optimization and Ranking for Global QSAR Studies. Filtering Safe and Potent Drug Candidates from Combinatorial Libraries.** *J. Comb. Chem.* **2008**, *10*, 897-913.

En este trabajo, se propone el empleo de los resultados del método de OMO propuesto anteriormente (274) como base para la aplicación de un nuevo método de ordenamiento también basado en las funciones de deseabilidad de Derringer (241). Para probar la utilidad de la metodología propuesta, la misma fue aplicada a una biblioteca de 95 fluoroquinolonas reportada por Suto y colaboradores (254) con el objetivo de optimizar simultáneamente sus propiedades antibacteriana frente a microorganismos gram-negativos (MIC) y citotóxica sobre células de mamíferos (IC₅₀) (261); ordenar los compuestos incluidos en esta biblioteca con respecto al candidato ideal determinado por medio del proceso de optimización multiobjetivos y filtrar los candidatos más prometedores de la misma.

Tabla 3.2.2 Coeficientes de la regresión y parámetros estadísticos de los modelos de regresión lineal múltiple obtenidos.

Modelo RLM de Actividad Antibacteriana ($MIC = 1/1+MIC$)									
$1/(1 + MIC) = 27.127(\pm 3.925) - 1.573(\pm 0.170) \cdot H4M - 13.504(\pm 1.969) \cdot BELp1$ $+ 0.071(\pm 0.012) \cdot RDF020e - 0.130(\pm 0.024) \cdot Mor05m - 0.006(\pm 0.001) \cdot G(F..F)$ $+ 5.670(\pm 1.097) \cdot HATS3m + 0.002(\pm 0.000) \cdot D / Dr06 - 0.234(\pm 0.064) \cdot Mor14v$									
<i>N</i>	<i>R</i>	<i>R</i> ²	<i>R</i> ² <i>Adj.</i>	<i>S</i>	<i>O</i> ²	<i>SPRESS</i>	<i>ρ</i>	<i>F</i>	<i>P</i>
95	0.883	0.779	0.753	0.096	0.725	0.107	8.636	29.601	0.0000
Modelo RLM de Actividad Citotóxica ($IC_{50} = 1/1+IC_{50}$)									
$1/(1 + IC_{50}) = -0.966(\pm 0.146) + 0.611(\pm 0.053) \cdot R5p - 0.135(\pm 0.012) \cdot GATS5p$ $- 0.147(\pm 0.018) \cdot H4m + 1.239(\pm 0.156) \cdot FDI + 0.002(\pm 0.000) \cdot G(F..F)$ $+ 0.114(\pm 0.019) \cdot Mor24v - 0.162(\pm 0.039) \cdot H6v + 0.183(\pm 0.045) \cdot MATS3e$									
<i>N</i>	<i>R</i>	<i>R</i> ²	<i>R</i> ² <i>Adj.</i>	<i>S</i>	<i>O</i> ²	<i>SPRESS</i>	<i>ρ</i>	<i>F</i>	<i>P</i>
95	0.867	0.750	0.721	0.014	0.686	0.016	8.636	25.313	0.0002

Se generó un modelo de predicción basado en regresión lineal múltiple para cada propiedad conteniendo un subconjunto de descriptores moleculares implementados en software DRAGON 5.1 previamente seleccionados por métodos de algoritmos genéticos. Como se puede apreciar en la Tabla 3.2.2, la calidad de los MP desarrollados es indicada por altos valores de *F* y pequeños valores de *p*, valores satisfactorios de *ρ* ($\rho > 5$), junto con valores de *R*² y *Adj.R*² que van desde 0.75 a 0.779 y 0.721 a 0.753, respectivamente; así como los valores *Q*²_{LOO} entre 0.686 y 0.725.

Se chequeó y demostró el cumplimiento de los principales supuestos paramétricos (282) relacionadas con los tres modelos usados para ajustar las funciones de deseabilidad (ver publicación original (285)).

Para especificar las respectivas funciones de deseabilidad la actividad antibacteriana (MIC) es maximizada mientras que la capacidad citotóxica (IC_{50}) es minimizada para obtener un candidato que exhiba simultáneamente la mayor actividad antibacteriana a la vez que la menor capacidad citotóxica posibles.

Un valor de R^2_D próximo a 0.7 indica una calidad adecuada del proceso de optimización simultánea. Por otra parte, un valor de 0.63 de Q^2_D indica una capacidad predictiva adecuada del método. Como resultado del proceso de optimización es posible determinar los niveles de los descriptores moleculares empleados que simultáneamente producen la combinación más deseable de las propiedades antibacteriana y citotóxica.

Tabla 3.2.3 Resultados del proceso de optimización simultanea multiobjetivos.

<i>Niveles Óptimos de las Variables</i>		
$JGI6 = 0.058539124$	$R4e+ = 0.215402953$	$RDF020e = 6.533512527$
$MATS3e = 0.097921819$	$R5p = 0.560622$	$RDF050e = 21.75996$
$GATS5p = 2.71639566$	$G(F..F) = -5.395274574$	$Mor05m = -6.618889553$
$FDI = 0.996478400$	$H4m = 0.836178947$	$Mor14v = -0.049636561$
$Mor24v = 0.095266$	$D/Dr06 = 202.3135$	$HATS3m = 0.049289$
$H6v = 0.266748712$	$BELp1 = 2.022804936$	$HATS3e = 0.242572857$

Los resultados del proceso de optimización multiobjetivos se muestran en la Tabla 3.2.3 donde se indican los valores de los descriptores moleculares necesarios para obtener un candidato de la familia de las fluoroquinolonas con el mejor compromiso posible entre las propiedades antibacterianas y citotóxicas ($D = 1.00$).

Como se describió anteriormente, es posible el empleo de la metodología MOOP-DESIRE para el ordenamiento y filtrado de bibliotecas de compuestos. Una vez

determinados los valores de los descriptores moleculares necesarios para la obtención del candidato ideal, estos son usados como patrón para ordenar la biblioteca de 95 fluoroquinolonas usadas en este trabajo. Debido a la naturaleza “outlier” de 10 de los 95 compuestos fue necesario reducir a 85 el conjunto de datos que será sometido a un segundo proceso de optimización donde los respectivos valores de Δ_i serán calculados y usados para realizar el ordenamiento de estos con respecto al candidato óptimo. Estos datos pueden consultarse en la Tabla A.6.

Basado en Δ_i es posible obtener un ordenamiento de la biblioteca de 85 fluoroquinolonas con un índice de calidad corregido Ψ^* de 0.313 lo que representa un porcentaje de calidad del ordenamiento ($R\%$) igual al 68.7%. El ordenamiento obtenido comparado con el ordenamiento ideal se muestra en la Figura 3.2.2.

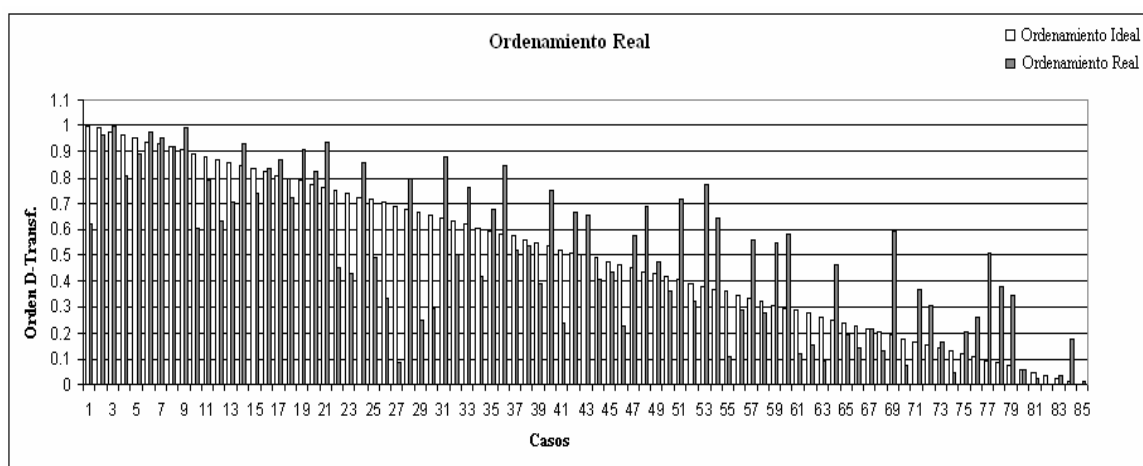


Figura 3.2.2 Ordenamiento de las 85 fluoroquinolonas en función de la su similitud con el candidato óptimo (Δ_i).

La calidad del ordenamiento obtenido ($R\% = 68.7$) es similar a los valores de predictividad obtenidos tanto para los modelos individuales como para el proceso global ($Q^2(\text{MIC}) = 0.693$, $Q^2(\text{IC}_{50}) = 0.686$, $Q^2_{D(\text{MIC-IC}_{50})} = 0.629$), lo que confirma que

la calidad tanto del proceso de optimización multiobjetivos como del ordenamiento depende de la calidad de los modelos de predicción iniciales.

Por otra parte, el objetivo de ordenar una biblioteca de compuestos en función de su similitud con un candidato óptimo es filtrar una parte de estos (los candidatos mas cercanos y por tanto mas semejantes al óptimo), los que serian propuestos para ser sintetizados y probar de manera experimental sus propiedades farmacológicas, toxicológicas y/o farmacocinéticas.

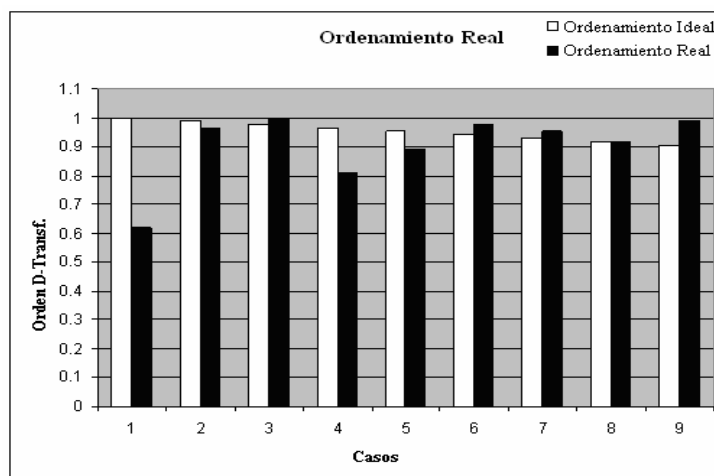


Figura 3.2.3 Ordenamiento obtenido para el 10% de la biblioteca de compuestos.

Consecuentemente, si tomáramos solamente el 10% (los mejores 9 compuestos) de nuestra base de datos para ser propuesta para su inclusión en el proceso de descubrimiento de fármacos se incrementaría la probabilidad de encontrar candidatos prometedores ya que esta fracción presenta un índice de calidad corregido Ψ^* de 0.173 ($R_{\%} = 82.74\%$). El ordenamiento obtenido solo para esta fracción de la biblioteca se muestra en la Figura 3.2.3.

El hecho de filtrar una fracción de los candidatos con la mejor combinación de propiedades farmacológicas, toxicológicas y farmacocinéticas le confiere al proceso de descubrimiento de fármacos un grado de racionalidad que no es posible alcanzar mediante el uso de los métodos QSAR tradicionales que optimizan cada propiedad de manera secuencial, lo que implica desconocer el perfil del resto de las propiedades involucradas en la calidad final de un fármaco o al menos dejar a la casualidad el hecho de encontrar un candidato con perfiles aceptables de estas propiedades de manera simultánea. Consecuentemente, un candidato potente identificado vía QSAR puede ser descartado posteriormente debido a un perfil toxicológico o farmacocinético negativo con el consecuente gasto inútil de recursos durante los procesos de síntesis y pruebas farmacológicas ya realizado (286). Igualmente, mediante el empleo de un jurado de modelos (modelos farmacológicos (QSAR), farmacocinéticos (QSPKR) y toxicológicos (QSTR)) la probabilidad de encontrar un candidato con perfiles aceptables de estas propiedades de manera simultánea es baja y de ocurrir esto, sería mas obra de la casualidad que fruto de una estrategia de diseño racional.

Finalmente, para probar lo anteriormente expuesto se lleva a cabo un experimento que a su vez sirve para probar la hipótesis fundamental planteada en este trabajo, esto es en esencia, la superioridad del enfoque multiobjetivos sobre el enfoque secuencial del descubrimiento de fármacos. Por medio del tamizaje secuencial ($VP_{10\%} = 11.112\%$ y $FE_{10\%} = 0.556$) solo uno de los nueve candidatos filtrados cumple con los requisitos de actividad y toxicidad requeridos para exhibir un perfil terapéutico aceptable. Algo similar sucede con el tamizaje paralelo ($VP_{10\%} =$

16.672% y $FE_{10\%} = 0.833$) donde solo tres de los 18 candidatos filtrados cumple con ambos requisitos. Por su parte, el tamizaje multiobjetivos ($VP_{10\%} = 77.778\%$ y $FE_{10\%} = 3.889$) muestra un desempeño superior al resto de los enfoques evaluados al mostrar 7 de los 9 compuestos filtrados un perfil terapéutico global adecuado. Como se puede notar en la Figura 3.2.4 y las Tablas A.7 a la A.9, el desempeño del tamizaje virtual multiobjetivos es superior al secuencial o al paralelo. El mismo supera significativamente a los dos restantes enfoques en los dos criterios de enriquecimiento empleados. Este resultado sumado al hecho de que a través de los enfoques secuencial y paralelo son descartados candidatos con perfiles terapéuticos globales aceptables (compuestos **028-24A**, **031-24E**, **032-24F**, **058-29E** y **065-30F**) refuerza la hipótesis de que un enfoque multiobjetivos del proceso de descubrimiento de fármacos puede conducir a un proceso global más racional y eficiente.

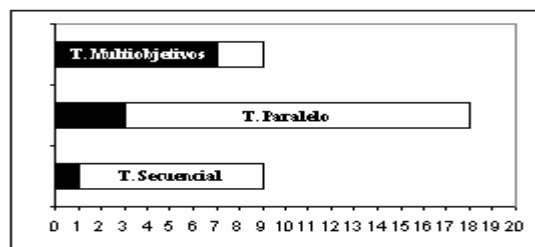


Figura 3.2.4 Grado de recuperación de casos positivos logrado por los diferentes métodos de tamizaje virtual. El número de casos verdaderos positivos se representan en negro mientras que los falsos positivos en blanco.

La metodología propuesta puede usarse como una estrategia racional para el filtrado de bibliotecas combinatorias; considerando únicamente aquellos candidatos incluidos dentro del dominio de aplicabilidad de los modelos en los que se basan

los procesos de optimización multiobjetivos y ordenamiento. En situaciones en la que el principal objetivo es el ordenamiento y filtrado, es aconsejable el uso de descriptores moleculares que conduzcan a relaciones estructura-deseabilidad altamente predictivas con el objetivo de asegurar la exactitud de las predicciones y consecuentemente, una estimación precisa de la deseabilidad global de la molécula. Este tipo de análisis es más apropiado para etapas iniciales del proceso de descubrimiento de fármacos, específicamente durante las fases de generación de hits e identificación del líder. Por el contrario, el uso de bases de datos reducidas y homogéneas es más adecuado en etapas posteriores del proceso (optimización del líder) una vez identificado el líder. En esta etapa se pueden llevar a cabo modificaciones estructurales específicas basadas en los resultados del proceso de optimización multiobjetivos. Para esto, el uso de descriptores moleculares físico-químicos o con una interpretación estructural clara pueden conducir a relaciones estructura-deseabilidad interpretables las que pueden usarse para el diseño de nuevos candidatos con un perfil terapéutico global mejorado (274). La Figura A.1 resume de forma esquemática el uso de la metodología MOOP-DESIRE como herramienta de apoyo en el proceso de descubrimiento y desarrollo racional de nuevos fármacos.

CONCLUSIONES

CONCLUSIONES

- ✓ Se introdujo una familia de descriptores moleculares basados en cadenas de Markov capaces de codificar de manera diferente una misma molécula en dependencia del sistema con que interactúe. El empleo de estos descriptores moleculares permitió la obtención de modelos de correlación estructura-actividad capaces de predecir el perfil de biodistribución o toxicológico de un fármaco en múltiples sistemas biológicos.
- ✓ La obtención de una metodología de optimización multiobjetivos basada en funciones de deseabilidad permitió determinar los niveles de las variables predictivas necesarios para obtener un candidato a fármaco con una combinación óptima de las propiedades involucradas en el perfil terapéutico global del mismo.
- ✓ La obtención de una metodología de ordenamiento basado en funciones de deseabilidad permitió ordenar candidatos a fármaco por su similitud respecto al candidato óptimo derivado del proceso de optimización multiobjetivos.

RECOMENDACIONES

RECOMENDACIONES

- ✓ El tratamiento simultáneo de objetivos en conflicto como la eficacia terapéutica y la toxicidad de un fármaco ha sido llevado a cabo con éxito. Sin embargo, ante la ausencia de datos, el desempeño de la metodología OMO propuesta no ha sido evaluado para problemas que además incluyan propiedades farmacocinéticas. Por esta razón, se recomienda evaluar el desempeño de esta metodología en bases de datos que incluyan entre sus objetivos propiedades farmacocinéticas.
- ✓ La metodología OMO propuesta solo emplea técnicas lineales de modelación (regresión lineal múltiple) y el método simplex de optimización. Consecuentemente, su aplicación esta limitada a problemas de naturaleza lineal. Considerando además de lo anterior, la plétora métodos de modelación y optimización existentes, se recomienda evaluar el desempeño de la metodología propuesta mediante el empleo de técnicas no lineales de modelación, así como de métodos evolutivos de optimización.
- ✓ El esquema de ordenamiento desarrollado emplea como criterio de similitud (distancia) la distancia euclidiana ya que la misma es aplicable a la mayora de los problemas. Sin embargo, otras medidas de distancia pueden ser mas apropiadas es circunstancias especiales. De aquí que se recomienda explorar el desempeño de otras medidas de similitud en el proceso de ordenamiento.

- ✓ Se recomienda la evaluación experimental de las propiedades analgésicas, antiinflamatorias y ulcerogénicas de los nuevas quinazolinonas diseñadas, de manera que sea posible corroborar las predicciones teóricas realizadas para estos candidatos.

REFERENCIAS BIBLIOGRÁFICAS

REFERENCIAS BIBLIOGRÁFICAS

1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ*2003 Mar;22(2):151-85.
2. Bleicher KH, Bohm HJ, Muller K, Alanine AI. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov*2003 May;2(5):369-78.
3. Xu J, Hagler A. Chemoinformatics and Drug Discovery. *Molecules*2002;7:566-700.
4. Armstrong JW. A review of high-throughput screening approaches for drug discovery. Application note. 1999.
5. Bajorath J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov*2002 Nov;1(11):882-94.
6. Lazo JS, Wipf P. Combinatorial chemistry and contemporary pharmacology. *J Pharmacol Exp Ther*2000 Jun;293(3):705-9.
7. Miertus S, Fassina G, Seneci PF. Concepts of Combinatorial Chemistry and Combinatorial Technologies. *Chem Listy*2000;94:1104-10.
8. O'Driscoll C. A Virtual Space Odyssey. *Nature Horizon: Charting Chemical Space*. 2004.
9. Manly CJ, Louise-May S, Hammer JD. The impact of informatics and computational chemistry on synthesis and screening. *Drug Discov Today*2001 Nov 1;6(21):1101-10.
10. Ekins S, Boulanger B, Swaan PW, Hupcey MA. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J Comput Aided Mol Des*2002;16(5-6):381-401.

11. Nicolaou AC, Brown N, Pattichis CS. Molecular optimization using computational multi-objective methods. *Curr Opin Drug Discov Devel*2007;10(3):316-24.
12. Brown N, Lewis RA. Exploiting QSAR methods in lead optimization. *Curr Opin Drug Discov Devel*2006;9(4):419-24.
13. Hansch C. On the structure of medicinal chemistry. *J Med Chem*1976 Jan;19(1):1-6.
14. Cruz-Monteagudo M, Borges F, Perez Gonzalez M, Cordeiro MN. Computational modeling tools for the design of potent antimalarial bisbenzamidines: overcoming the antimalarial potential of pentamidine. *Bioorg Med Chem*2007 Aug 1;15(15):5322-39.
15. Fernandez M, Caballero J. Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor. *J Mol Graph Model*2006 Dec;25(4):410-22.
16. Fernandez M, Caballero J, Tundidor-Camba A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. *Bioorg Med Chem*2006 Jun 15;14(12):4137-50.
17. Guo W, Hu X, Chu N, Yin C. Quantitative structure-activity relationship studies on HEPTs by supervised stochastic resonance. *Bioorg Med Chem Lett*2006 Mar 27.
18. Vasanthanathan P, Lakshmi M, Arockia Babu M, Gupta AK, Kaskhedikar SG. QSAR Study of 3-Phenyl-5-acyloxymethyl-2H,5H-furan-2-ones as Antifungal

- Agents: The Dominant Role of Electronic Parameter. *Chem Pharm Bull (Tokyo)*2006 Apr;54(4):583-7.
19. Saiz-Urra L, Gonzalez MP, Teijeira M. QSAR studies about cytotoxicity of benzophenazines with dual inhibition toward both topoisomerases I and II: 3D-MoRSE descriptors and statistical considerations about variable selection. *Bioorg Med Chem*2006 Nov 1;14(21):7347-58.
 20. Santana L, Uriarte E, Gonzalez-Diaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J Med Chem*2006 Feb 9;49(3):1149-56.
 21. Xie A, Sivaprakasam P, Doerksen RJ. 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2,5-diaminobenzophenone scaffold. *Bioorg Med Chem*2006 Nov 1;14(21):7311-23.
 22. Gonzalez MP, Teran C, Teijeira M, Helguera AM. QSAR Studies Using Radial Distribution Function for Predicting A(1) Adenosine Receptors Agonists. *Bull Math Biol*2006 Oct 24.
 23. Gonzalez MP, Teran C, Teijeira M, Helguera AM. Quantitative structure activity relationships as useful tools for the design of new adenosine receptor ligands. 1. Agonist. *Curr Med Chem*2006;13(19):2253-66.
 24. Morales AH, Perez MA, Combes RD, Gonzalez MP. Quantitative structure activity relationship for the computational prediction of nitrocompounds carcinogenicity. *Toxicology*2006 Mar 1;220(1):51-62.

25. Marrero-Ponce Y, Torrens F, Alvarado YJ, Rotondo R. Bond-based global and local (bond, group and bond-type) quadratic indices and their applications to computer-aided molecular design. 1. QSPR studies of diverse sets of organic chemicals. *J Comput Aided Mol Des* 2006 Oct;20(10-11):685-701.
26. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, Rotondo R. Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification. *J Mol Graph Model* 2006 Sep 26.
27. Casanola-Martin GM, Marrero-Ponce Y, Khan MT, Ather A, Sultan S, Torrens F, et al. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: Evaluation of different classification model combinations using bond-based linear indices. *Bioorg Med Chem* 2006 Nov 2.
28. Montero-Torres A, Garcia-Sanchez RN, Marrero-Ponce Y, Machado-Tugores Y, Nogal-Ruiz JJ, Martinez-Fernandez AR, et al. Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. *Eur J Med Chem* 2006 Apr;41(4):483-93.
29. Dudek AZ, Arodz T, Gálvez J. Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Comb Chem High Throughput Screen* 2006;9:1-16.

30. Boik JC, Newman RA. Structure-activity models of oral clearance, cytotoxicity, and LD50: a screen for promising anticancer compounds. *BMC Pharmacol*2008;8:12.
31. Markov AA. Extension de la loi de grands nombres aux événements dependants les uns des autres. *Bull Soc Phys Math Kasan*1906;15:135-56.
32. Marrero-Ponce Y, Torrens F. Bond, Bond-Type, and Total Linear Indices of the non-Stochastic and Stochastic Edge Adjacency Matrix. 1. Theory and QSPR Studies. *Proceedings of the 9th International Electronic Conference on Synthetic Organic Chemistry*. 2005. URL: <http://www.usc.es/congresos/ecsoc/9/CC/q015/q015.pdf>
33. Diudea M. Cluj matrix invariants. *J Chem Inf Comput Sci*1997;37:300-5.
34. Guha R. On the interpretation and interpretability of quantitative structure-activity relationship models. *J Comput Aided Mol Des*2008 Dec;22(12):857-71.
35. Randic M, Balaban AT, Basak SC. On structural interpretation of several distance related topological indices. *J Chem Inf Comput Sci*2001 May-Jun;41(3):593-601.
36. Cruz-Monteagudo M, Gonzalez-Diaz H, Borges F, Dominguez ER, Cordeiro MN. 3D-MEDNEs: An Alternative "in Silico" Technique for Chemical Research in Toxicology. 2. Quantitative Proteome-Toxicity Relationships (QPTR) based on Mass Spectrum Spiral Entropy. *Chem Res Toxicol*2008 Mar;21(3):619-32.

37. Gonzalez-Diaz H, Gonzalez-Diaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics*2008 Feb;8(4):750-78.
38. Gonzalez-Diaz H, Prado-Prado FJ. Unified QSAR and network-based computational chemistry approach to antimicrobials, Part 1: Multispecies activity models for antifungals. *J Comput Chem*2008;29(4):656-67.
39. Prado-Prado FJ, Gonzalez-Diaz H, Santana L, Uriarte E. Unified QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorg Med Chem*2007 Jan 15;15(2):897-902.
40. Gonzalez-Diaz H, Olazabal E, Santana L, Uriarte E, Gonzalez-Diaz Y, Castanedo N. QSAR study of anticoccidial activity for diverse chemical compounds: Prediction and experimental assay of trans-2-(2-nitrovinyl)furan. *Bioorg Med Chem*2007 Jan 15;15(2):962-8.
41. González-Díaz H, Pérez-Bello A, Cruz-Monteagudo M, González-Díaz Y, Santana L, Uriarte E. Chemometrics for QSAR with low sequence homology: Mycobacterial promoter sequences recognition with 2D-RNA entropies. *Chemomet Intel Lab Syst. [Research]*. 2007 15 January 2007;85(1):20-6.
42. Gonzalez-Diaz H, Saiz-Urra L, Molina R, Santana L, Uriarte E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J Proteome Res*2007 Feb;6(2):904-8.
43. Gonzalez-Diaz H, Saiz-Urra L, Molina R, Gonzalez-Diaz Y, Sanchez-Gonzalez A. Computational chemistry approach to protein kinase recognition

- using 3D stochastic van der Waals spectral moments. *J Comput Chem* 2007 Apr 30;28(6):1042-8.
44. Gonzalez-Diaz H, Perez-Castillo Y, Podda G, Uriarte E. Computational chemistry comparison of stable/nonstable protein mutants classification models based on 3D and topological indices. *J Comput Chem* 2007 Sep;28(12):1990-5.
45. Gonzalez-Diaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: predicting antifungal activity against different species. *Bioorg Med Chem* 2006 Sep 1;14(17):5973-80.
46. Cruz-Monteagudo M, Gonzalez-Diaz H, Borges F, Gonzalez-Diaz Y. Simple stochastic fingerprints towards mathematical modeling in biology and medicine. 3. Ocular irritability classification model. *Bull Math Biol* 2006 Oct;68(7):1555-72.
47. Agüero-Chapin G, Gonzalez-Diaz H, Molina R, Varona-Santos J, Uriarte E, Gonzalez-Diaz Y. Novel 2D maps and coupling numbers for protein sequences. The first QSAR study of polygalacturonases; isolation and prediction of a novel sequence from *Psidium guajava* L. *FEBS Lett* 2006 Feb 6;580(3):723-30.
48. Gonzalez-Diaz H, Sanchez-Gonzalez A, Gonzalez-Diaz Y. 3D-QSAR study for DNA cleavage proteins with a potential anti-tumor ATCUN-like motif. *J Inorg Biochem* 2006 Jul;100(7):1290-7.

49. Gonzalez-Diaz H, Perez-Bello A, Uriarte E, Gonzalez-Diaz Y. QSAR study for mycobacterial promoters with low sequence homology. *Bioorg Med Chem Lett* 2006 Feb;16(3):547-53.
50. Gonzalez-Diaz H, Vina D, Santana L, de Clercq E, Uriarte E. Stochastic entropy QSAR for the in silico discovery of anticancer compounds: prediction, synthesis, and in vitro assay of new purine carbanucleosides. *Bioorg Med Chem* 2006 Feb 15;14(4):1095-107.
51. Gonzalez-Diaz H, Uriarte E. Proteins QSAR with Markov average electrostatic potentials. *Bioorg Med Chem Lett* 2005 Nov 15;15(22):5088-94.
52. Gonzalez-Diaz H, Agüero-Chapin G, Varona-Santos J, Molina R, de la Riva G, Uriarte E. 2D RNA-QSAR: assigning ACC oxidase family membership with stochastic molecular descriptors; isolation and prediction of a sequence from *Psidium guajava* L. *Bioorg Med Chem Lett* 2005 Jun 2;15(11):2932-7.
53. Saiz-Urra L, Gonzalez-Diaz H, Uriarte E. Proteins Markovian 3D-QSAR with spherically-truncated average electrostatic potentials. *Bioorg Med Chem* 2005 Jun 1;13(11):3641-7.
54. Gonzalez-Diaz H, Cruz-Monteagudo M, Vina D, Santana L, Uriarte E, De Clercq E. QSAR for anti-RNA-virus activity, synthesis, and assay of anti-RSV carbonucleosides given a unified representation of spectral moments, quadratic, and topologic indices. *Bioorg Med Chem Lett* 2005 Mar 15;15(6):1651-7.
55. Gonzalez-Diaz H, Torres-Gomez LA, Guevara Y, Almeida MS, Molina R, Castanedo N, et al. Markovian chemicals "in silico" design (MARCH-INSIDE),

- a promising approach for computer-aided molecular design III: 2.5D indices for the discovery of antibacterials. *J Mol Model (Online)*2005 Mar;11(2):116-23.
56. Gonzalez-Diaz H, Tenorio E, Castanedo N, Santana L, Uriarte E. 3D QSAR Markov model for drug-induced eosinophilia--theoretical prediction and preliminary experimental assay of the antimicrobial drug G1. *Bioorg Med Chem*2005 Mar 1;13(5):1523-30.
57. Gia O, Marciani Magno S, Gonzalez-Diaz H, Quezada E, Santana L, Uriarte E, et al. Design, synthesis and photobiological properties of 3,4-cyclopentenepsoralens. *Bioorg Med Chem*2005 Feb 1;13(3):809-17.
58. Gonzalez-Diaz H, Uriarte E. Biopolymer stochastic moments. I. Modeling human rhinovirus cellular recognition with protein surface electrostatic moments. *Biopolymers*2005 Apr 5;77(5):296-303.
59. Gonzalez-Diaz H, Uriarte E, Ramos de Armas R. Predicting stability of Arc repressor mutants with protein stochastic moments. *Bioorg Med Chem*2005 Jan 17;13(2):323-31.
60. Gonzalez-Diaz H, Molina R, Uriarte E. Recognition of stable protein mutants with 3D stochastic average electrostatic potentials. *FEBS Lett*2005 Aug 15;579(20):4297-301.
61. Gonzalez-Diaz H, Agüero G, Cabrera MA, Molina R, Santana L, Uriarte E, et al. Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological

- species: distribution of the antimicrobial G1 on rat tissues. *Bioorg Med Chem Lett*2005 Feb 1;15(3):551-7.
62. González-Díaz H, Pérez-Bello A, Uriarte E. Stochastic molecular descriptors for polymers. 3. Markov electrostatic moments as polymer 2D-folding descriptors: RNA-QSAR for mycobacterial promoters. *Polymer*2005;46:6461–73.
63. Gonzalez-Diaz H, Molina R, Uriarte E. Markov entropy backbone electrostatic descriptors for predicting proteins biological activity. *Bioorg Med Chem Lett*2004 Sep 20;14(18):4691-5.
64. Gonzalez-Diaz H, Bastida I, Castanedo N, Nasco O, Olazabal E, Morales A, et al. Simple stochastic fingerprints towards mathematical modelling in biology and medicine. 1. The treatment of coccidiosis. *Bull Math Biol*2004 Sep;66(5):1285-311.
65. Gonzalez MP, Dias LC, Morales AH, Rodriguez YM, de Oliveira LG, Gomez LT, et al. TOPS-MODE based QSARs derived from heterogeneous series of compounds. Applications to the design of new anti-inflammatory compounds. *Bioorg Med Chem*2004 Aug 15;12(16):4467-75.
66. Ramos de Armas R, Gonzalez-Diaz H, Molina R, Perez Gonzalez M, Uriarte E. Stochastic-based descriptors studying peptides biological properties: modeling the bitter tasting threshold of dipeptides. *Bioorg Med Chem*2004 Sep 15;12(18):4815-22.
67. Ramos de Armas R, Gonzalez-Diaz H, Molina R, Uriarte E. Markovian Backbone Negentropies: Molecular descriptors for protein research. I.

- Predicting protein stability in Arc repressor mutants. *Proteins*2004 Sep 1;56(4):715-23.
68. Gonzalez-Diaz H, de Armas RR, Molina R. Vibrational Markovian modelling of footprints after the interaction of antibiotics with the packaging region of HIV type 1. *Bull Math Biol*2003 Nov;65(6):991-1002.
69. Gonzalez-Diaz H, de Armas RR, Molina R. Markovian negentropies in bioinformatics. 1. A picture of footprints after the interaction of the HIV-1 Psi-RNA packaging region with drugs. *Bioinformatics*2003 Nov 1;19(16):2079-87.
70. Gonzalez-Diaz H, Marrero Y, Hernandez I, Bastida I, Tenorio E, Nasco O, et al. 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis. *Chem Res Toxicol*2003 Oct;16(10):1318-27.
71. Gonzalez-Diaz H, Gia O, Uriarte E, Hernandez I, Ramos R, Chaviano M, et al. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: discovery of anticancer compounds. *J Mol Model (Online)*2003 Dec;9(6):395-407.
72. Gonzalez-Diaz H, Sanchez IH, Uriarte E, Santana L. Symmetry considerations in Markovian chemicals 'in silico' design (MARCH-INSIDE) I: central chirality codification, classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Comput Biol Chem*2003 Jul;27(3):217-27.
73. Gonzalez-Diaz H, Olazabal E, Castanedo N, Sanchez IH, Morales A, Serrano HS, et al. Markovian chemicals "in silico" design (MARCH-INSIDE), a

- promising approach for computer aided molecular design II: experimental and theoretical assessment of a novel method for virtual screening of fasciolicides. *J Mol Model (Online)*2002 Aug;8(8):237-45.
74. Gonzalez-Diaz H, Ramos de Armas R, Uriarte E. In silico markovian bioinformatics for predicting ¹ha-nmr chemical shifts in mouse epidermis growth factor (megf). *Online Journal of Bioinformatics*2002;1(1):83-95.
75. Nicolotti O, Gillet VJ, Fleming PJ, Green DV. Multiobjective optimization in quantitative structure-activity relationships: deriving accurate and interpretable QSARs. *J Med Chem*2002 Nov 7;45(23):5069-80.
76. Stockfisch TP. Partially unified multiple property recursive partitioning (PUMP-RP): A new method for predicting and understanding drug selectivity. *J Chem Inf Comput Sci*2003;43(5):1608-13.
77. Rao SN, Stockfisch TP. Partially unified multiple property recursive partitioning (PUMP-RP) analyses of cyclooxygenase (COX) inhibitors. *J Chem Inf Comput Sci*2003;43(5):1614-22.
78. Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. In silico prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15. *Eur J Pharm Sci*2002 Dec;17(4-5):183-93.
79. Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J Med Chem*2002;45:2867-76.

80. Drews J. Drug discovery: a historical perspective. *Science* 2000 Mar 17;287(5460):1960-4.
81. Kubinyi H. Strategies and recent technologies in drug discovery. *Pharmazie* 1995 Oct;50(10):647-62.
82. Warne P, Page C. Is there a best strategy for drug discovery? *Drug News Perspect* 2003 Apr;16(3):177-82.
83. Chanda SK, Caldwell JS. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov Today* 2003 Feb 15;8(4):168-74.
84. Debouck C, Goodfellow PN. DNA microarrays in drug discovery and development. *Nat Genet* 1999 Jan;21(1 Suppl):48-50.
85. Ren J, Stammers DK. HIV reverse transcriptase structures: designing new inhibitors and understanding mechanisms of drug resistance. *Trends Pharmacol Sci* 2005 Jan;26(1):4-7.
86. Jorgensen WL. The many roles of computation in drug discovery. *Science* 2004;303(5665):1813-8.
87. Brown FK. Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery 1998.
88. Hann M, Green R. Chemoinformatics--a new name for an old problem? *Curr Opin Chem Biol* 1999 Aug;3(4):379-83.
89. Leach AR, Gillet VJ. An introduction to chemoinformatics: Kluwer Academic Publishers; 2003.
90. Gasteiger J, Engel T. Chemoinformatics : a Textbook. Gasteiger J, Engel T, editors: Wiley-VCH; 2003.

91. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004 Nov;3(11):935-49.
92. Apostolakis J, Caflisch A. Computational ligand design. *Comb Chem High Throughput Screen* 1999 Apr;2(2):91-104.
93. Graves AP, Brenk R, Shoichet BK. Decoys for docking. *J Med Chem* 2005 Jun 2;48(11):3714-28.
94. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. *J Med Chem* 2005 Feb 24;48(4):962-76.
95. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997 Apr 4;267(3):727-48.
96. Moitessier N, Henry C, Maignet B, Chapleur Y. Combining pharmacophore search, automated docking, and molecular dynamics simulations as a novel strategy for flexible docking. Proof of concept: docking of arginine-glycine-aspartic acid-like compounds into the alphavbeta3 binding site. *J Med Chem* 2004 Aug 12;47(17):4178-87.
97. Murcia M, Ortiz AR. Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. *J Med Chem* 2004 Feb 12;47(4):805-20.
98. Glick M, Grant GH, Richards WG. Docking of flexible molecules using multiscale ligand representations. *J Med Chem* 2002 Oct 10;45(21):4639-46.

99. Miteva MA, Lee WH, Montes MO, Villoutreix BO. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J Med Chem* 2005 Sep 22;48(19):6012-22.
100. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004 Mar 25;47(7):1739-49.
101. Finn P, Muggleton S, Page D, Srinivasan A. Pharmacophore discovery using the inductive logic programming language Progol. *Machine Learning* 1998;30:241-70.
102. Klebe G. Recent developments in structure-based drug design. *J Mol Med* 2000;78(5):269-81.
103. Butina D, Segall MD, Frankcombe K. Predicting ADME properties in silico: methods and models. *Drug Discov Today* 2002;7(11 Suppl):S83-S8.
104. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. 1st ed. Mannhold R, Kubinyi H, Timmerman H, editors. D-69469 Weinheim, Federal Republic of Germany: WILEY-VCH Verlag GmbH; 2000.
105. Bharucha-Reid AT. *Elements of theory of markov process on the application*. McGraw-Hill Series in Probability and Statistics. New York: McGraw-Hill; 1960. p. 167-434.
106. Freund JA, Poschel T. *Stochastic processes in physics, chemistry, and biology*. Lect Notes Phys. Berlin, Germany: Springer-Verlag; 2000.

107. Papasaikas PK, Bagos PG, Litou ZI, Hamodrakas SJ. A novel method for GPCR recognition and family classification from sequence alone using signatures derived from profile hidden Markov models. *SAR QSAR Environ Res*2003 Oct-Dec;14(5-6):413-20.
108. Dwyer MG, Bergsland N, Saluste E, Sharma J, Jaisani Z, Durfee J, et al. Application of hidden Markov random field approach for quantification of perfusion/diffusion mismatch in acute ischemic stroke. *Neurol Res*2008 Oct;30(8):827-34.
109. Cao Y, Kuske R, Li H. Direct observation of markovian behavior of the mechanical unfolding of individual proteins. *Biophys J*2008 Jul;95(2):782-8.
110. Titman AC, Sharples LD. A general goodness-of-fit test for Markov and hidden Markov models. *Stat Med*2008 May 30;27(12):2177-95.
111. Waagepetersen R, Ibanez-Escriche N, Sorensen D. A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. *Genet Sel Evol*2008 Mar-Apr;40(2):161-76.
112. Berezhkovskii AM, Weiss GH. Propagators and related descriptors for non-Markovian asymmetric random walks with and without boundaries. *J Chem Phys*2008 Jan 28;128(4):044914.
113. Rodrigue N, Philippe H, Lartillot N. Uniformization for sampling realizations of Markov processes: applications to Bayesian implementations of codon substitution models. *Bioinformatics*2008 Jan 1;24(1):56-62.
114. Lladser ME, Betterton MD, Knight R. Multiple pattern matching: a Markov chain approach. *J Math Biol*2008 Jan;56(1-2):51-92.

115. Vergne N. Drifting Markov models with polynomial drift and applications to DNA sequences. *Stat Appl Genet Mol Biol*2008;7:Article6.
116. Allefeld C, Bialonski S. Detecting synchronization clusters in multivariate time series via coarse-graining of Markov chains. *Phys Rev E Stat Nonlin Soft Matter Phys*2007 Dec;76(6 Pt 2):066207.
117. Winters-Hilt S, Baribault C. A novel, fast, HMM-with-Duration implementation - for application with a new, pattern recognition informed, nanopore detector. *BMC Bioinformatics*2007;8 Suppl 7:S19.
118. Truong K, Ikura M. Identification and characterization of subfamily-specific signatures in a large protein superfamily by a hidden Markov model approach. *BMC Bioinformatics*2002;3:1.
119. Markovic S, Markovic Z, McCrindle RI. Spectral moments of phenylenes. *J Chem Inf Comput Sci*2001 Jan-Feb;41(1):112-9.
120. Maruyama T. A Markov process of gene frequency change in a geographically structured population. *Genetics*1974;76(2):367-77.
121. González-Díaz H. Nuevos descriptores moleculares basados en cadenas de markov: Estudios de relación estructura-propiedad y bioinformática [PhD Thesis]. Santiago de Compostela: Universidad de Santiago de Compostela; 2005.
122. Johnson MA, Maggiora GM, editors. *Concepts and Applications of Molecular Similarity*: John Wiley & Sons; 1990.
123. Bunke H, Shearer K. A graph distance metric based on the maximal common subgraph. *Pattern Recog Lett*1998;19(3):255-9.

124. Pepperrell CA, Willett P. Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances. *J Comput Aided Mol Des*1991 Oct;5(5):455-74.
125. Thimm M, Goede A, Hougardy S, Preissner R. Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J Chem Inf Comput Sci*2004 Sep-Oct;44(5):1816-22.
126. Bultinck P, Kuppens T, Girones X, Carbo-Dorca R. Quantum similarity superposition algorithm (QSSA): a consistent scheme for molecular alignment and molecular similarity based on quantum chemistry. *J Chem Inf Comput Sci*2003 Jul-Aug;43(4):1143-50.
127. Arakawa M, Hasegawa K, Funatsu K. Application of the novel molecular alignment method using the Hopfield Neural Network to 3D-QSAR. *J Chem Inf Comput Sci*2003 Sep-Oct;43(5):1396-402.
128. Arakawa M, Hasegawa K, Funatsu K. Novel alignment method of small molecules using the Hopfield Neural Network. *J Chem Inf Comput Sci*2003 Sep-Oct;43(5):1390-5.
129. Carbó R, Leyda L, Arnau M. How Similar is a Molecule to Another - an Electron-Density Measure of Similarity Between 2 Molecular Structures. *Int J Qantum Chem*1980;17:1185-9.
130. Basak SC, Magnuson VR, Niemi GJ, Regal RR. Determining structural similarity of chemicals using graph-theoretic indices. *Discrete Applied Mathematics*1988;19(1):17-44.

131. Willett P. Chemical Similarity Searching. *J Chem Inf Comput Sci*1998;38:983-96.
132. Bauknecht H, Zell A, Bayer H, Levi P, Wagener M, Sadowski J, et al. Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. *J Chem Inf Comput Sci*1996;36(6):1205-13.
133. Salim N, Holliday J, Willett P. Combination of fingerprint-based similarity coefficients using data fusion. *J Chem Inf Comput Sci*2003 Mar-Apr;43(2):435-42.
134. Martin B. Instance-based learning: Nearest neighbour with generalisation. New Zealand: University of Waikato; 1995.
135. Cleary JG, Trigg LE, editors. K*: An instance-based learner using an entropic distance measure. Proceedings of the Twelfth International Conference on Machine Learning; 1995; Tahoe City, CA. San Francisco: Morgan Kaufmann.
136. Aha DW. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*1992;36(2):267-87.
137. Massart DL, Kaufman DL. The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis. New York (NY): Wiley; 1983.
138. Willett P. Similarity and Clustering in Chemical Information Systems. Letchworth (UK): Research Studies Press; 1987.
139. Horvath D, Jeandenans C. Neighborhood behavior of in silico structural spaces with respect to in vitro activity spaces-a novel understanding of the

- molecular similarity principle in the context of multiple receptor binding profiles. *J Chem Inf Comput Sci*2003 Mar-Apr;43(2):680-90.
140. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. Gray J, editor. San Francisco: Morgan Kaufmann; 2005.
141. Duda RO, Hart PE, Stork DG. *Pattern Classification*. Second ed: Wiley-Interscience; 2001.
142. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*: Springer; 2001.
143. Van Waterbeemd H. *Chemometric methods in molecular design*. Manhnhold R, Krosgaard-Larsen P, Timmerman H, Van Waterbeemd H, editors. New York: Wiley-VCH; 1995.
144. Jackson JE. *A User's Guide to Principal Components*. New York (NY): Wiley; 1991.
145. Natarajan R, Nirdosh I, Basak SC, Mills DR. QSAR modeling of flotation collectors using principal components extracted from topological indices. *J Chem Inf Comput Sci*2002 Nov-Dec;42(6):1425-30.
146. Hemmateenejad B, Akhond M, Miri R, Shamsipur M. Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines (nifedipine analogous). *J Chem Inf Comput Sci*2003 Jul-Aug;43(4):1328-34.
147. Moro S, Bacilieri M, Cacciari B, Spalluto G. Autocorrelation of molecular electrostatic potential surface properties combined with partial least squares

- analysis as new strategy for the prediction of the activity of human A(3) adenosine receptor antagonists. *J Med Chem* 2005 Sep 8;48(18):5698-704.
148. Eriksson L, Damborsky J, Earll M, Johansson E, Trygg J, Wold S. Three-block bi-focal PLS (3BIF-PLS) and its application in QSAR. *SAR QSAR Environ Res* 2004 Oct-Dec;15(5-6):481-99.
149. Tantishaiyakul V. Prediction of aqueous solubility of organic salts of diclofenac using PLS and molecular modeling. *Int J Pharm* 2004 May 4;275(1-2):133-9.
150. Saxena AK, Prathipati P. Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR QSAR Environ Res* 2003 Oct-Dec;14(5-6):433-45.
151. Hansch C, T. F. ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* 1964;86:1616-26.
152. Hansch C, Muir RM, T. F, Maloney PP, Geiger F, Streich M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammet constants and partition coefficients. *J Am Chem Soc* 1963;85:2817-24.
153. Kubinyi H. Comparative Molecular Field Analysis. In: Gasteiger J, editor. *Handbook of Chemoinformatics From Data to Knowledge*. Weinheim: Wiley-VCH,; 2003. p. 1555-74.
154. Rogers D, Hopfinger AJ. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J Chem Inf Comput Sci* 1994;34(4):854-62.

155. Martin YC, Holland JB, Jarboe CH, Plotnikoff N. Discriminant analysis of the relationship between physical properties and the inhibition of monoamine oxidase by aminotetralins and aminoindans. *J Med Chem* 1974 Apr;17(4):409-13.
156. Zupan J, Gasteiger J. *Neural networks in chemistry and drug design*. 2nd ed. Weinheim: Wiley-VCH; 1999.
157. Devillers J. *Neural Networks in QSAR and Drug Design*. London: Academic Press; 1996.
158. Schneider G, Wrede P. Artificial neural networks for computer-based molecular design. *Prog Biophys Mol Biol* 1998;70(3):175-222.
159. Aoyama T, Suzuki Y, Ichikawa H. Neural networks applied to quantitative structure-activity relationship analysis. *J Med Chem* 1990 Sep;33(9):2583-90.
160. Andrea TA, Kalayeh H. Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J Med Chem* 1991 Sep;34(9):2824-36.
161. Egolf LM, Jurs PC. Prediction of Boiling Points of Organic Heterocyclic Compounds Using Regression and Neural Networks Techniques. *J Chem Inf Comput Sci* 1993;33:616-35.
162. A-Razzak M, Glen RC. Applications of rule-induction in the derivation of quantitative structure-activity relationships. *J Comput Aided Mol Des* 1992;6(4):349-83.

163. Rusinko A, 3rd, Farnen MW, Lambert CG, Brown PL, Young SS. Analysis of a large structure/biological activity data set using recursive partitioning. *J Chem Inf Comput Sci*1999 Nov-Dec;39(6):1017-26.
164. King RD, Muggleton S, Lewis RA, Sternberg MJ. Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. *Proc Natl Acad Sci U S A*1992;89(23):11322-6.
165. King RD, Muggleton SH, Srinivasan A, Sternberg MJ. Structure-activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. *Proc Natl Acad Sci U S A*1996;93(1):438-42.
166. Manallack DT, Livingstone DJ. Neural networks in drug discovery: have they lived up to their promise? *European Journal of Medicinal Chemistry*1999;34(3):195-208.
167. Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput Chem*2001 Dec;26(1):5-14.
168. Vapnik VN. *Statistical Learning Theory*. New-York: Wiley; 1998.
169. Brereton RG. *Chemometrics*. Horwood E, editor. Chichester (UK): Ellis Horwood; 1990.
170. Devillers J, Karcher W. *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers J, Karcher W, editors. Dordrecht (The

- Netherlands): Kluwer Academic Publishers for the European Communities; 1991.
171. Frank IE, Todechini R. The Data Analysis Handbook. Amsterdam (The Netherland): Elsevier; 1994.
172. van de Waterbeemd H, Constantino G, Clementi S, Cruciani G, Valigi R. Experimental Design in Synthesis Planning and Structure-Property Correlations. Disjoint Principal Properties of Organic Substituents. In: van de Waterbeemd H, editor. Chemometric Methods in Molecular Design. New York (NY): VCH Publishers; 1995. p. 103-12.
173. Massart DL, Vandeginste BGM, Buydens L, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics. Part A. Amsterdam (The Netherland): Elsevier; 1997.
174. Massart DL, Vandeginste BGM, Buydens L, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics. Part B
Amsterdam (The Netherland): Elsevier; 1998.
175. Legendre P, Legendre L. Numerical Ecology. Amsterdam (The Netherland): Elsevier; 1998.
176. Krzanowski WJ. Principles of Multivariate Analysis. New York (NY): Oxford Univ. Press; 1988.
177. Jolliffe IT. Principal Component Analysis. New Yor (NY): Springer-Verlag; 1986.
178. Basilevsky A. Statistical Factor Analysis and Related Methods. New York (NY): Wiley; 1994.

179. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters*. New York (NY): Wiley; 1978.
180. Carlson R. *Design and Optimization in Organic Synthesis*. Amsterdam (The Netherlands): Elsevier; 1992.
181. Draper N, Smith H. *Applied Regression Analysis*. New York (NY): Wiley; 1998.
182. Frank IE, Friedman JH. A Statistical View of Some Chemometrics Regression Tools. *Technometrics*1993;35:109-35.
183. Ryan TP. *Modern Regression Methods*. New York (NY): Wiley; 1997.
184. Wold S. PLS for Multivariate Linear Modelling. In: van der Waterbeemd H, editor. *Chemometric Methods in Molecular Design*. Weinheim, Germany: VCH Publishers 1994. p. 195-218.
185. Atkinson AC. *Plots, Transformations, and Regression*. Oxford (UK): Clarendon Press; 1985.
186. Hand DJ. *Discrimination and Classification*. Chichester (UK): Wiley; 1981.
187. Frank IE, Friedman JH. Classifications: Oldtimers and Newcomers *J Chemom*1989;3:463-75.
188. Hand DJ. *Construction and Assessments of Classification Rules*. Chichester (UK): Wiley; 1997.
189. Kawakami S, Numao N, Okubo Y, Koga F, Yamamoto S, Saito K, et al. Development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. *Eur Urol*2008;54(3):601-11.

190. Kattan MW. Editorial comment on: development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. *Eur Urol*2008;54(3):611.
191. Klopman G. Artificial Intelligence Approach to Structure-Activities Studies. Computer Automated Structure Evaluation of Biological Activity of Organic Molecules. *J Am Chem Soc*1984;106:7315-21.
192. Zupan J, Gasteiger J. *Neural Networks for Chemistry and Drug Design*. Weinheim (Germany): Wiley-VCH Publishers; 1999.
193. Anzali S, Barnickel G, Krug M, Sadowski J, Teckentrup A, Wagener M. The Use of Self-Organizing Neural Networks in Drug Design. In: Kubinyi H, Folkers G, Martin YC, editors. *3D QSAR in Drug Design*. Dordrecht (The Netherlands): Kluwer/ESCOM; 1998. p. 273-99.
194. Rodríguez-Abed A. Nuevo sistema multclasificador jerárquico. Posibilidades de aplicación. Santa Clara, Cuba: Universidad Central "Martha Abreu" de Las Villas; 2007.
195. Ultsch A, editor. *Emergence in Self-Organizing Feature Maps*. Workshop on Self-Organizing Maps (WSOM '07) 2007 Bielefeld, Germany: Springer-Verlag.
196. Holland J. *Adaptation in Artificial and Neural Systems*. Ann Arbor (MI): University of Michigan Press; 1975.

197. Papadopoulos MC, Dean PM. Molecular Structure Matching by Simulated Annealing. IV. Classification of Atom Correspondences in Sets of Dissimilar Molecules *J Comput Aided Mol Des*1991;5:119-33.
198. Verbeeck K. Coordinated Exploration in Multi-agent Reinforcement Learning. Brussel, Belgium: Vrije Universiteit Brussel; 2004.
199. Grau R, Chavez MC, Sanchez R, Morgado E, Casas G, Bonet I, editors. Boolean Algebraic Structures of the Genetic code. Possibilities of Applications. KDECB 2006, LNBI 4366; 2007; Heidelberg: Springer-Verlag Berlin.
200. Konak A, Coit DW, Smith AE. Multi-objective optimization using genetic algorithms: A tutorial. *Reliab Eng Syst Safety*2006;91:992-1007.
201. Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J Comput Aided Mol Des*1995;9(6):532-49.
202. Handschuh S, Wagener M, Gasteier J. Superposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. *J Chem Inf Comput Sci*1998;38(2):220-32.
203. Shepphird JK, Clarck RD. A marriage made in torsional space: Using GALAHAD models to drive pharmacophore multiple searches. *J Comput Aided Mol Des*2006;20(12):735-49.
204. Janson S, Merkle D. A new multi-objective particle swarm optimization algorithm using clustering applied to automated docking. In: Blesa MJ, Blum C, Roli A, Sampels M, editors. *Hybrid Metaheuristics Second International*

- Workshop, HM 2005; August 29-30; Barcelona, Spain: Springer-Verlag; 2005. p. 128-41.
205. Zoete V, Grosdidier A, Michielin O, editors. EADock: A new approach to the docking of small molecules to protein active sites. High Performance Computing for the Life Sciences Symposium; 2005; Lausanne, Switzerland.
206. Brown N, Mckay B, Gasteiger J. A novel workflow for the inverse QSPR problem using multiobjective optimization. *J Comput Aided Mol Des*2006;20(5):333-41.
207. Lameijer EW, Kok JN, Back T, Ijerman AP. The molecule avoluator. An interactive evolutionary algorithm for the design of drug-like molecules. *J Chem Inf Model*2006;46(2):545-52.
208. Hernandez I, Gonzales-Diaz H. MARCH-INSIDE version 1.0 (Markovian Chemicals In Silico Design). 1.0 ed: Chemicals Bio-actives Center, Central University of Las Villas, Cuba.; 2002. p. MARCH-INSIDE version 1.0 (Markovian Chemicals In Silico Design).
209. Garcia AG, Horga de la Parte JF. Índice de especialidades farmacéuticas. Prescripción racional de fármacos. INTERCON. Madrid: Editores Médicos S.A. EDIMSA.; 1994.
210. Poulin P, Krishnan K. A biologically-based algorithm for predicting human tissue: blood partition coefficients of organic chemicals. *Hum Exp Toxicol*1995 Mar;14(3):273-80.
211. Abraham MH, Kamlet MJ, Taft RW, Doherty RM, Weathersby PK. Solubility properties in polymers and biological media. 2. The correlation and prediction

- of the solubilities of nonelectrolytes in biological tissues and fluids. *J Med Chem*1985 Jul;28(7):865-70.
212. Kamlet MJ, Doherty RM, Abboud JL, Abraham MH, Taft RW. Linear solvation energy relationships: 36. Molecular properties governing solubilities of organic nonelectrolytes in water. *J Pharm Sci*1986 Apr;75(4):338-49.
213. Paterson S, Mackay D. Correlation of tissue, blood, and air partition coefficients of volatile organic chemicals. *Br J Ind Med*1989 May;46(5):321-8.
214. Abraham MH, Weathersby PK. Hydrogen bonding. 30. Solubility of gases and vapors in biological liquids and tissues. *J Pharm Sci*1994 Oct;83(10):1450-6.
215. Katritzky AR, Kuanar M, Fara DC, Karelson M, Acree WE, Jr. QSPR treatment of rat blood:air, saline:air and olive oil:air partition coefficients using theoretical molecular descriptors. *Bioorg Med Chem*2004 Sep 1;12(17):4735-48.
216. Moss GP, Cronin MT. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. *Int J Pharm*2002 May 15;238(1-2):105-9.
217. Degim T, Pugh JW, Hadgraft J. Skin permeability data: anomalous results. *Int J Pharm*1998;170:129-33.
218. Gramatica P, Corradi M, Consonni V. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere*2000;41 763-77.

219. Escuder-Gilabert L, Martín-Biosca Y, Sagrado S, Villanueva-Camañas RM, Medina-Hernández MJ. Biopartitioning micellar chromatography to predict ecotoxicity. *Anal Chim Acta*2001;448:173-85.
220. Katritzky AR, Dobchev DA, Hur E, Fara DC, Karelson M. QSAR treatment of drugs transfer into human breast milk. *Bioorg Med Chem*2005 Mar 1;13(5):1623-32.
221. Burg P, Fydrych P, Abraham MH, Matt M, Gruber R. The characterization of an active carbon in terms of selectivity towards volatile organic compounds using an LSER approach. *Fuel*2000;79:1041-5.
222. Roth CM, Goss K, Schwarzenbach RP. Sorption of diverse organic vapors to snow. *Environ Sci Technol*2004;38:4079-84.
223. Martín-Biosca Y, Molero-Monfort M, Sagrado S, Villanueva-Camañas RM, Medina-Hernández MJ. Rapid in vitro test to predict ocular tissue permeability based on biopartitioning micellar chromatography. *Eur J Pharm Sci*2003;20 209-16.
224. Abraham MH, Acree Jr WE. Characterisation of the water-isopropyl myristate system. *Int J Pharm* 2005;294:121-8.
225. Duchowicz PR, Fernández FM, Castro EA. Alternative statistical and theoretical analysis of fluorophilicity. *J Fluor Chem*2004;125:43-8.
226. Patel H, Schultz TW, Cronin MTD. Physico-chemical interpretation and prediction of the dimyristoyl phosphatidyl choline-water partition coefficient. *J Mol Struct (THEOCHEM)*2002;593:9-18.
227. STATISTICA for Windows release 6.0. Statsoft Inc. 2001.

228. Van Waterbeemd H. Discriminant Analysis for Activity Prediction. In: Van Waterbeemd H, editor. Chemometric methods in molecular design. New York: Wiley-VCH; 1995. p. 265-82.
229. Bisquerra Alzina R. Introducción conceptual al análisis multivariante: Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD. Barcelona: PPU; 1989.
230. Cruz-Monteagudo M, Gonzalez-Diaz H. Unified drug-target interaction thermodynamic Markov model using stochastic entropies to predict multiple drugs side effects. *Eur J Med Chem* 2005 Oct;40(10):1030-41.
231. Gonzalez-Diaz H, Cruz-Monteagudo M, Molina R, Tenorio E, Uriarte E. Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model. *Bioorg Med Chem* 2005 Feb 15;13(4):1119-29.
232. Randic M. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. *J Chem Inf Comput Sci* 1991;31(2):311-20.
233. Randic M. Orthogonal Molecular Descriptors. *New J Chem* 1991;15:517-25.
234. Randic M. Correlation of enthalphy of octanes with orthogonal connectivity indices. *J Mol Struct (THEOCHEM)* 1991;233:45-9.
235. Lucic B, Nikolic S, Trinajstic N, Juric D. The Structure-Property Models Can Be Improved Using the Orthogonalized Descriptors. *J Chem Inf Comput Sci* 1995;35(3):532-8.
236. Estrada E, Perdomo I, Torres-Lavandeira J. Combination of 2D-, 3D-Connectivity and Quantum Chemical Descriptors in QSPR. Complexation of

- a- and b-Cyclodextrin with Benzene Derivatives. *J Chem Inf Comput Sci*2001;41(6):1561-8.
237. Havel J, Madden E, Haddad PR. Prediction of retention times for anions in ion chromatography using artificial neural networks. *Chromatographia*1999;49:481-8.
238. Fernandez M, Caballero J, Helguera AM, Castro EA, Gonzalez MP. Quantitative structure-activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds. *Bioorg Med Chem*2005 May 2;13(9):3269-77.
239. Caballero J, Fernandez M. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. *J Mol Model (Online)*2005 Oct 21:1-14.
240. Fernandez M, Tundidor-Camba A, Caballero J. Modeling of cyclin-dependent kinase inhibition by 1H-pyrazolo[3,4-d]pyrimidine derivatives using artificial neural network ensembles. *J Chem Inf Model*2005 Nov-Dec;45(6):1884-95.
241. Derringer G, Suich R. Simultaneous Optimization of Several Response Variables. *J Quality Technol*1980;12(4):214-9.
242. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. Fifth ed. New York: McGraw Hill; 2005.
243. Nelder JA, Mead R. A Simplex method for function minimization. *Computer Journal*1965;7:308-13.

244. Fletcher R, Reeves CM. Function minimization by conjugate gradients. *Computer Journal*1964;7:149-54.
245. Hooke R, Jeeves TA. Direct search solution of numerical and statistical problems. *J Assoc Comp Machin*1961;8:212-29.
246. Watson I, Marir F. Case-based reasoning: a review. *The knowledge engineering review*1994;9(4):Cambridge, UK: Cambridge University Press.
247. Coleman TF, Li Y. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J Optim*1996;6:418-45.
248. Coleman TF, Li Y. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. *Math Program*1994;67(2):189-224.
249. MATLAB. 7.2 ed: The MathWorks, Inc.; 2006.
250. Alagarsamy V, Dhanabal K, Parthiban P, Anjana G, Deepa G, Murugesan B, et al. Synthesis and pharmacological investigation of novel 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-ones as analgesic and anti-inflammatory agents. *J Pharm Pharmacol*2007;59:669–77.
251. Arulmozhi DK, Veeranjanyulu A, Bodhankar SL, Arora SK. Investigations into the antinociceptive activity of *Sapindus trifoliatus* in various pain models. *J Pharm Pharmacol*2004;56:655–61.
252. Ganguly AK, Bhatnagar OP. Effect of bilateral adrenalectomy on production of restraint ulcers in the stomach of albino rats. *Can J Physiol Pharmacol* 1973;51:748–50.

253. Goyal RK, Chakrabarti A, Sanyal AK. The effect of biological variables on the anti-ulcerogenic effect of vegetable plantain banana. *Planta Med*1985;29:85–8.
254. Suto MJ, Domagala JM, Roland GE, Mailloux GB, Cohen MA. Fluoroquinolones: Relationships between Structural Variations, Mammalian Cell Cytotoxicity, and Antimicrobial Activity. *J Med Chem*1992;35:4745-50.
255. Holden HH, Barret JF, Huntington CM, Muehlbauer PA, Wahrenburg MG. Genetic Profile of a Nalidixic Acid Analog: A Model for the Mechanism of Sister Chromatid Exchange. *Environ Mol Mutagen*1989;13:238-25.
256. Hoshino K, Sato K, Akahane K, Yoshida A, Hayakawa I, Sato M, et al. Significance of the Methyl Group on the Oxazine Ring of Ofloxacin Derivatives in the Inhibition of Bacterial and Mammalian Type II Topoisomerases. *Antimicrob Agents Chemother*1991;35:309-12.
257. Moreau NJ, Robaux N, Baron L. Inhibitory Effects of Quinolones on Pro- and Eukaryotic DNA Topoisomerases I and II. *Antimicrob Agents Chemother*1990;34:1955-60.
258. Kohlbrenner WE, Wideburg N, Weigl D, Saldivar A, Chu DTW. Induction of Calf Thymus Topoisomerase II-Mediated DNA Breakage by the Antibacterial Isothiazoloquinolones A-65281 and A-66282. *Antimicrob Agents Chemother*1992;36:81-6.
259. Fernandes PB. Mode of Action, In Vitro and In Vivo Activities of the Fluoroquinolones. *J Clin Pharmacol*1988;28:156-68.

260. Domagala JM, Hanna LD, Heifetz CM, Hutt MP, Mich TF, Sanchez JP, et al. New Structure-Activity Relationships of the Quinolone Antibacterials Using the Target Enzyme. The Development and Application of a DNA Gyrase Assay. *J Med Chem*1986;29:894-404 and references therein.
261. Gracheck SJ, Mychalonka M, Gambino L, Cohen M, Roland GE, Ciaravino V, et al., editors. Correlations of Quinolone Inhibition Endpoints vs Bacterial DNA Gyrase, Topoisomerase I and II, Cell Clonogenic Survival and In Vitro Micronuclei Induction. 31st Interscience Conference on Antimicrobial Agents and Chemotherapy; 1991; Chicago, Illinois.
262. Cohen MA, Griffen TJ, Bien PA, Heifetz CL, Domagala JM. In Vitro Activity of C1-934, a Quinolone Carboxylic Acid Active Against Gram-positive and - Negative Bacteria. *Antimicrob Agents Chemother*1985;28:766-72.
263. De Boor C. A practical guide to splines. New York: Springer-Verlag; 1978.
264. Gerald CF, Wheatley PO. Applied numerical analysis. 4th ed. Reading, MA: Addison Wesley; 1989.
265. CambridgeSoft. ChemDraw Ultra. 9.0 ed2004.
266. Frank J. MOPAC. 6.0 ed: Seiler Research Laboratory, US Air Force Academy, Colorado Springs, CO.; 1993.
267. Todeschini R, Consonni V, Pavan M. DRAGON Software. 2.1 ed. Milano: Milano Chemometrics; 2002.
268. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J Chemom*1992;6:267-81.

269. Yasri A, Hartsough D. Toward an optimal procedure for variable selection and QSAR model building. *J Chem Inf Comput Sci*2001 Sep-Oct;41(5):1218-27.
270. Hou TJ, Wang JM, Liao N, Xu XJ. Applications of genetic algorithms on the structure-activity relationship analysis of some cinnamamides. *J Chem Inf Comput Sci*1999 Sep-Oct;39(5):775-81.
271. Hasegawa K, Kimura T, Funatsu K. GA strategy for variable selection in QSAR studies: application of GA-based region selection to a 3D-QSAR study of acetylcholinesterase inhibitors. *J Chem Inf Comput Sci*1999 Jan-Feb;39(1):112-20.
272. Barbosa de Oliveira D, Gaudio AC. BuildQSAR. Vitória ES, Brasil: Physics Department-CCE, University of Espírito Santo; 2000.
273. Barbosa de Oliveira D, Gaudio AC. BuildQSAR: A new computer program for QSAR analysis. *Quant Struct-Act Relat*2000;19:599-601.
274. Cruz-Monteagudo M, Borges F, Cordeiro MN. Desirability-based multiobjective optimization for global QSAR studies: Application to the design of novel NSAIDs with improved analgesic, antiinflammatory, and ulcerogenic profiles. *J Comput Chem*2008 May 1;29(14):2445-59.
275. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect*2003 Aug;111(10):1361-75.

276. Atkinson AC. Plots, Transformations and Regression: Oxford:Clarendon Press; 1985.
277. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection--what can we learn from earlier mistakes? *J Comput Aided Mol Des*2008 Mar-Apr;22(3-4):213-28.
278. Villa A, Zangi R, Pieffet G, Mark AE. Sampling and convergence in free energy calculations of protein-ligand interactions: the binding of triphenoxypyridine derivatives to factor Xa and trypsin. *J Comput Aided Mol Des*2003 Oct;17(10):673-86.
279. Joly V, Yeni P. Vectorization of drugs. A winsome concept with delicate practical applications. *Rev Prat*1993;43:2100-3.
280. Temsamani J, Bonnafous C, Rousselle C, Fraisse Y, Clair P, Granier LA, et al. Improved brain uptake and pharmacological activity profile of morphine-6-glucuronide using a peptide vector-mediated strategy. *J Pharmacol Exp Ther*2005;313:712-9.
281. Blondeau JM, Castañedo N, Gonzalez O, Medina R, Silveira E. Antimicrob Agents Chemother1999;11:1663.
282. Stewart J, Gill L. Econometrics. 2nd edition ed. Allan P, editor. London: Prentice Hall; 1998.
283. Cruz-Monteagudo M, Gonzalez-Diaz H, Agüero-Chapin G, Santana L, Borges F, Dominguez ER, et al. Computational chemistry development of a unified free energy Markov model for the distribution of 1300 chemicals to 38

- different environmental or biological systems. *J Comput Chem*2007 Aug;28(11):1909-23.
284. Fernandez FM, Duchowicz PR, Castro EA. About orthogonal descriptors in QSPR/QSAR theories. *MATCH, Comm Math Comp Chem*2004;51:39-57.
285. Cruz-Monteagudo M, Borges F, Cordeiro MNDS, Cagide Fajin JL, Morell C, Molina Ruiz R, et al. Desirability-Based Methods of Multiobjective Optimization and Ranking for Global QSAR Studies. Filtering Safe and Potent Drug Candidates from Combinatorial Libraries. *J Comb Chem*2008;10(6):897-913.
286. Drews J. Innovation deficit revisited: Reflections on the productivity of pharmaceutical R&D. *Drug Discov Today*1998;3:491-4.

ANEXOS

PRODUCCIÓN CIENTÍFICA DEL AUTOR SOBRE EL TEMA DE LA TESIS

Listado de Publicaciones Originales

Los principales resultados de esta tesis están resumidos en las siguientes publicaciones las cuales serán referidas por sus números romanos:

- I. Predicting Multiple Drugs Side Effects with a General Drug-Target Interaction Thermodynamic Markov Model.**
Humberto González-Díaz, Maykel Cruz-Monteagudo, Reinaldo Molina, Esvieta Tenorio and Eugenio Uriarte.
Bioorganic & Medicinal Chemistry. **2005**, *13*, 1119-1129.
 - II. Unified Drug-Target Interaction Thermodynamic Markov Model using Stochastic Entropies to Predict Multiple Drugs Side Effects.**
Maykel Cruz-Monteagudo, Humberto González-Díaz.
European Journal of Medicinal Chemistry. **2005**, *40*, 1030-1041.
 - III. Simple Stochastic Fingerprints towards Mathematical Modeling in Biology and Medicine. 2. Unify Markov model for drugs side effects.**
Maykel Cruz-Monteagudo, Humberto González-Díaz and Eugenio Uriarte.
Bulletin of Mathematical Biology. **2006**, *68*, 1527-1554.
 - IV. Multi-Target QSPR Assemble of a Complex Network for the Distribution of Chemicals to Biphasic Systems and Biological Tissues.**
Humberto González-Díaz, Miguel A. Cabrera-Pérez, Guillermin Agüero-Chapín, Maykel Cruz-Monteagudo, Nilo Castañeda-Cancio, Miguel A. del Río, Eugenio Uriarte.
Chemometrics and Intelligent Laboratory Systems. **2008**, *94*, 160-165.
 - V. Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems.**
Maykel Cruz-Monteagudo, Humberto González-Díaz, Guillermin Agüero-Chapín, Lourdes Santana, Fernanda Borges, Elena Rosa Domínguez, Gianni Podda, Eugenio Uriarte.
Journal of Computational Chemistry. **2007**, *28*, 1909-1923.
 - VI. Desirability-Based Multi-Objective Optimization for Global QSAR Studies. Application to the Design of Novel NSAIDs with Improved Analgesic, Anti-inflammatory and Ulcerogenic Profiles.**
Maykel Cruz-Monteagudo, Fernanda Borges, M. Natália D.S. Cordeiro.
Journal of Computational Chemistry. **2008**, *29*, 2445-2459.
 - VII. Desirability-Based Methods of Multi-Objective Optimization and Ranking for Global QSAR Studies. Filtering Safe and Potent Drug Candidates from Combinatorial Libraries.**
Maykel Cruz-Monteagudo, Fernanda Borges, M. Natália D.S. Cordeiro, J. Luis Cagide Fajin, Carlos Morell, Reinaldo Molina Ruiz, Yudith Cañizares-Carmenate, Elena Rosa Domínguez.
Journal of Combinatorial Chemistry, **2008**, *10*, 897-913.
-

Listado de Otras Publicaciones Relacionadas con la Tesis o con el Autor

- VIII. Markovian Chemicals “*in silico*” Design (MARCH-INSIDE), a Promising Approach for Computer-Aided Molecular Design I: Discovery of Anticancer Compounds.**
Humberto González-Díaz, Ornella Gia, Eugenio Uriarte, Ivan Hernández, Ronal Ramos, Mayrelis Chaviano, Santiago Seijo, Juan A. Castillo, Lázaro Morales, Lourdes Santana, Delali Akpaloo, Enrique Molina, Maikel Cruz, Luis A. Torres, Miguel A. Cabrera.
Journal of Molecular Modelling, **2003**, 9, 395-407.
- IX. QSAR for Anti-RNA-Virus Activity, Synthesis, and Assay of Anti-RSV Carbonucleosides given an Unify Representation of Spectral Moments, Quadratic, and Topologic Indices.**
Humberto González-Díaz, Maykel Cruz-Monteagudo, Dolores Viña, Lourdes Santana, Eugenio Uriarte and Erik De Clercq.
Bioorganic & Medicinal Chemistry Letters. **2005**, 15, 1651-1657.
- X. Simple Stochastic Fingerprints towards Mathematical Modeling in Biology and Medicine. 3. Ocular Irritability Classification Model.**
Maykel Cruz-Monteagudo, Humberto González-Díaz, Fernanda Borges and Yenny González-Díaz.
Bulletin of Mathematical Biology. **2006**, 68, 1555-1572.
- XI. Chemometrics for QSAR with Low Sequence Homology: Mycobacterial Promoter Sequences Recognition with 2D-RNA Entropies.**
Humberto González-Díaz, Alcides Pérez-Bello, Maykel Cruz-Monteagudo, Yenny González-Díaz, Lourdes Santana, Eugenio Uriarte.
Chemometrics and Intelligent Laboratory Systems. **2007**, 85, 20-26.
- XII. Computational Modeling Tools for the Design of Potent Antimalarial Bisbenzamidines. Overcoming the Antimalarial Potential of Pentamidine.**
Maykel Cruz-Monteagudo, Fernanda Borges, Maykel Pérez González, Natalia Cordeiro.
Bioorganic & Medicinal Chemistry. **2007**, 15, 5322-5339.
- XIII. Computational Chemistry Approach for the Early Detection of Drug-Induced Idiosyncratic Liver Toxicity.**
Maykel Cruz-Monteagudo, M. Natalia D. S. Cordeiro, Fernanda Borges.
Journal of Computational Chemistry, **2008**, 29, 533-549.
- XIV. 3D-MEDNEs: An Alternative “*in Silico*” Technique for Chemical Research in Toxicology. 2. Quantitative Proteome-Toxicity Relationships (QPTR) based on Mass Spectrum Spiral Entropy.**
Maykel Cruz-Monteagudo, Humberto González-Díaz, Fernanda Borges, Elena Rosa Dominguez, M. Natália D.S. Cordeiro.
Chemical Research in Toxicology, **2008**, 21, 619-632.
- XV. Quantitative Proteome-Property Relationships (QPPRs). 1. Finding Biomarkers of Organic Drugs with Mean Markov Connectivity Indices of Spiral Networks of Blood Mass Spectra.**
Maykel Cruz-Monteagudo, Cristian Robert Munteanu, Fernanda Borges, M. Natália D.S. Cordeiro, Eugenio Uriarte, Humberto González-Díaz.
Bioorganic & Medicinal Chemistry. **2008**, 16, 9684-9693.
- XVI. Stochastic Molecular Descriptors for Polymers. 4. Study of Complex Mixtures with Topological Indices of Mass Spectra Spiral and Star Networks: The Blood Proteome Case.**
Maykel Cruz-Monteagudo, Cristian Robert Munteanu, Fernanda Borges, M. Natália D.S. Cordeiro, Eugenio Uriarte, Kou Chen Chou, Humberto González-Díaz.
Polymer. **2008**, 49, 5575-5587.
-

Participación en Eventos Científicos y Premios Obtenidos

1. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 1er Simposium Internacional de Ciencias Biológicas Charles T. Ransdem in memoriam. Universidad de Oriente, Santiago de Cuba, Cuba. 10/04.
 2. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 3er Encuentro de las Provincias Centrales de la Sociedad Cubana de Matemática y Computación COMPUMAT 2004. UCLV, Santa Clara, Cuba. 10/04.
 3. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 5To Taller sobre Ecología y Medio Ambiente PROECO 2005. IBP, UCLV, Cuba. 03/05.
 4. Cruz-Monteagudo, M.; González-Díaz, H.; Molina, R.; Tenorio, E.; Uriarte, E. Predicting Multiple Drugs Side Effects with a General Drug-Target Interaction Thermodynamic Markov Model. Experimental Corroboration of Some Theoretic Predictions for G-1 (2-Bromo-5-[2-Bromo-2-Nitrovinyl] Furan). XI Convención Internacional de la informática y las comunicaciones INFORMÁTICA 2005. Palacio de las convenciones, La Habana, Cuba. 05/05.
 5. Cruz-Monteagudo, M.; González-Díaz, H.; Molina, R.; Tenorio, E.; Uriarte, E. Predicción de Múltiples Reacciones Adversas de Medicamentos a través de un Modelo Markoviano Termodinámico de Interacción Fármaco-Receptor. Una Nueva Herramienta a Tener en Cuenta en el Proceso de Farmacovigilancia. Taller Internacional de Servicios Farmacéuticos Clínicos SERVIFARMA 2005. Centro de Convenciones "Bolívar", Santa Clara, Cuba. 05/05.
 6. González-Díaz, H. Cruz-Monteagudo, M.; Viña, D.; Santana, L.; Uriarte, E.; De Clercq, E. QSAR for Anti-RNA-Virus Activity, Synthesis, and Assay of Anti-RSV Carbonucleosides given an Unify Representation of Spectral Moments, Quadratic, and Topologic Indices. The Ninth International Electronic Conference on Synthetic Organic Chemistry, ECSOC-9. Universidad de Santiago de Compostela, Santiago de Compostela, España. 11/05.
 7. Ramos, R.; González-Díaz, H.; Molina, R.; Uriarte, E.; Cruz-Monteagudo, M. Markovian Backbone Negentropies: Molecular Descriptors for Protein Research. Predicting Protein Stability in ARC Represor Mutants. 6th National Congress of the Cuban Society of Pharmacology and Therapeutics. 3rd Workshop of Pharmacoepidemiology and 1st Cuban Congress of Clinic Trials. ISCM, Santiago de Cuba, Cuba. 11/05.
 8. González-Díaz, H.; Ramos, R.; Molina, R.; Cruz-Monteagudo, M. Markovian Negentropies in Bioinformatics. A Picture of Footprints after the Interaction of the HIV-1 Ψ -RNA Packaging Region with Drugs. 18 Conferencia de Química. Santiago de Cuba, Cuba. 12/05.
 9. Cruz-Monteagudo, M.; González-Díaz, H. Simple Stochastic Fingerprints towards Mathematical Modeling in Biology and Medicine. Unify Drug-Target Interaction Thermodynamic Markov model for Drugs Side Effects. Simposio Cuba-UK en Química y Ciencias de la Vida. La Habana, Cuba. 01/06.
 10. Cruz-Monteagudo, M.; González-Díaz, H. Predicting Multiple Drugs Side Effects through a Unified Drug-Target Interaction Thermodynamic Markov Model. First International Workshop on Bioinformatics Cuba-Flanders '2006. UCLV, Santa Clara, Cuba. 02/06.
 11. Cruz-Monteagudo, M.; Borges, F.; Molina Ruiz, R. *In silico* prediction of the idiosyncratic hepatotoxic potential of drugs. 4º Encontro da Rede de Química e Tecnologia (REQUIMTE). Fátima, Portugal. 03/06.
 12. Cruz-Monteagudo, M.; Borges, F.; Molina Ruiz, R. *In silico* prediction of the idiosyncratic hepatotoxic potential of drugs. Chemistry in the 21st Century. University of Lisbon, Lisbon, Portugal. 10/06.
 13. Cruz-Monteagudo, M.; González-Díaz, H.; Borges, F. Modelo QSAR Alternativo para Predecir el Potencial Efecto Irritante Ocular de Sustancias Químicas. XXVII Congreso Latinoamericano de Química. La Habana, Cuba. 10/06.
 14. Cruz-Monteagudo, M.; Borges, F.; D. S. Cordeiro, M. N. Addressing idiosyncratic hepatotoxicity in early drug discovery. Annual Meeting of the Portuguese Society of Pharmacology, Faculty of Medicine, University of Porto, Porto, Portugal. 12/06.
 15. Cruz-Monteagudo, M.; D. S. Cordeiro, M. N.; Borges, F. Addressing idiosyncratic hepatotoxicity in early drug discovery. III Simposio Internacional de Química. UCLV, Santa Clara, Cuba. 06/07.
 16. Cruz-Monteagudo, M.; González-Díaz, H.; Agüero-Chapín, G.; Santana, L.; Borges, F.; Rosa
-

-
- Domínguez, E.; Podda, G.; Uriarte, E. Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems. XXXIII Congreso de Químicos Teóricos de Expresión Latina. UH, La Habana, Cuba. 09/07.
17. González-Díaz, H.; Cruz-Monteagudo, M.; Borges, F.; Uriarte, E. Predicting Proteome-Early Drug Induced Cardiac Toxicity Relationships (Pro-EDICToRs) with Node Overlapping Parameters (NOPs) of a new class of Blood Mass-Spectra graphs. The 11th International Electronic Conference on Synthetic Organic Chemistry, ECSOC-11. Universidad de Santiago de Compostela, Santiago de Compostela, España. 11/07.
 18. Cruz-Monteagudo, M.; D. S. Cordeiro, M. N.; Borges, F. Addressing idiosyncratic hepatotoxicity in early drug discovery. IV Congreso Internacional de la Sociedad Cubana de Toxicología. La Habana, Cuba. 12/07.
 19. Cruz-Monteagudo, M.; González-Díaz, H.; Borges, F.; Rosa Domínguez, E.; D. S. Cordeiro, M. N. Predictive Approaches and Drug-Induced Toxicity: From Mass Spectrum Blood Proteome to Drug-Induced Cardiac Toxicity. International Courses on Toxicology 2008. "Metabolic Toxicology: From Pathway to Organism". Center for Neurosciences and Cell Biology, University of Coimbra, Portugal. 04/08.
 20. Cruz-Monteagudo, M.; González-Díaz, H.; Borges, F.; Rosa Domínguez, E.; D. S. Cordeiro, M. N. Graph theory in Toxicoproteomics: Quantitative Proteome-Toxicity Relationships (QPTR) based on Mass Spectrum Spiral Entropies. MATH/CHEM/COMP 2008 CONFERENCE. International Academy of Mathematical Chemistry. Verbania-Intra, Italy. 06/08.
 21. Cruz-Monteagudo, M.; Borges, F.; D. S. Cordeiro, M. N. MOOP-DESIRE-Based Simultaneous Optimization of the Analgesic, Anti-Inflammatory and Ulcerogenic Profiles of 3-(3-Methylphenyl)-2-Substituted Amino-3H-Quinazolin-4-ones. The 12th International Electronic Conference on Synthetic Organic Chemistry, ECSOC-12. Universidad de Santiago de Compostela, Santiago de Compostela, España. 11/08.
 22. Cruz-Monteagudo, M.; Cagide Fajin, J. L.; Molina Ruiz, R.; D. S. Cordeiro, M. N.; Borges, F. Filtering Safe and Potent Drug Candidates from Combinatorial Libraries throughout Desirability-Based Methods of Multi-Objective Optimization and Ranking. 1^o Encontro Nacional de Química Terapêutica. Universidade do Porto, Porto, Portugal. 11/08.
 23. 3er Encuentro de las Provincias Centrales de la Sociedad Cubana de Matemática y Computación COMPUMAT 2004. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 10/04. (*Mención*).
 24. Encuentro Científico Juvenil de la Universidad Central "Marta Abreu" de Las Villas, JOVEN CIENCIA 2004. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 12/04. (*Premio*).
 25. 3er Encuentro Científico Juvenil del Instituto Superior de Ciencias Médicas, CENTROMED III. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 01/05. (*Premio*).
 26. Premio Nacional ACC (como colaborador): "De la estructura molecular a la actividad biológica: Aplicaciones y potencialidades de novedosos métodos de diseño computacional en estudios QSAR, QSPR, QSPkR, QSTR, proteómica y bioinformática". 02/05.
 27. Premio Provincial del Concurso Científico Técnico Juvenil Provincial de las Brigadas Técnicas Juveniles 200. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer. 03/05.
 28. 5^{to} Taller sobre Ecología y Medio Ambiente PROECO 2005. González-Díaz, H.; Cruz-Monteagudo, M.; Castillo-Garit, J. A. Diseño Markoviano *in silico* de Compuestos Anticancer.. 03/05. (*Premio*).
 29. Premio Nacional XII Exposición "Forjadores del Futuro". Cruz-Monteagudo, M. Predicción de Múltiples Reacciones Adversas de Medicamentos a través de un Modelo Markoviano Termodinámico de Interacción Fármaco-Receptor. 03/07.
 30. *Bioorganic & Medicinal Chemistry Most Cited Paper 2005-2008 Award*: "Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model". H. González-Díaz, M. Cruz-Monteagudo, R. Molina, E. Tenorio, E. Uriarte. *Bioorganic & Medicinal Chemistry*, Volume 13, Issue 4 (2005), Pages 1119-1129.
-

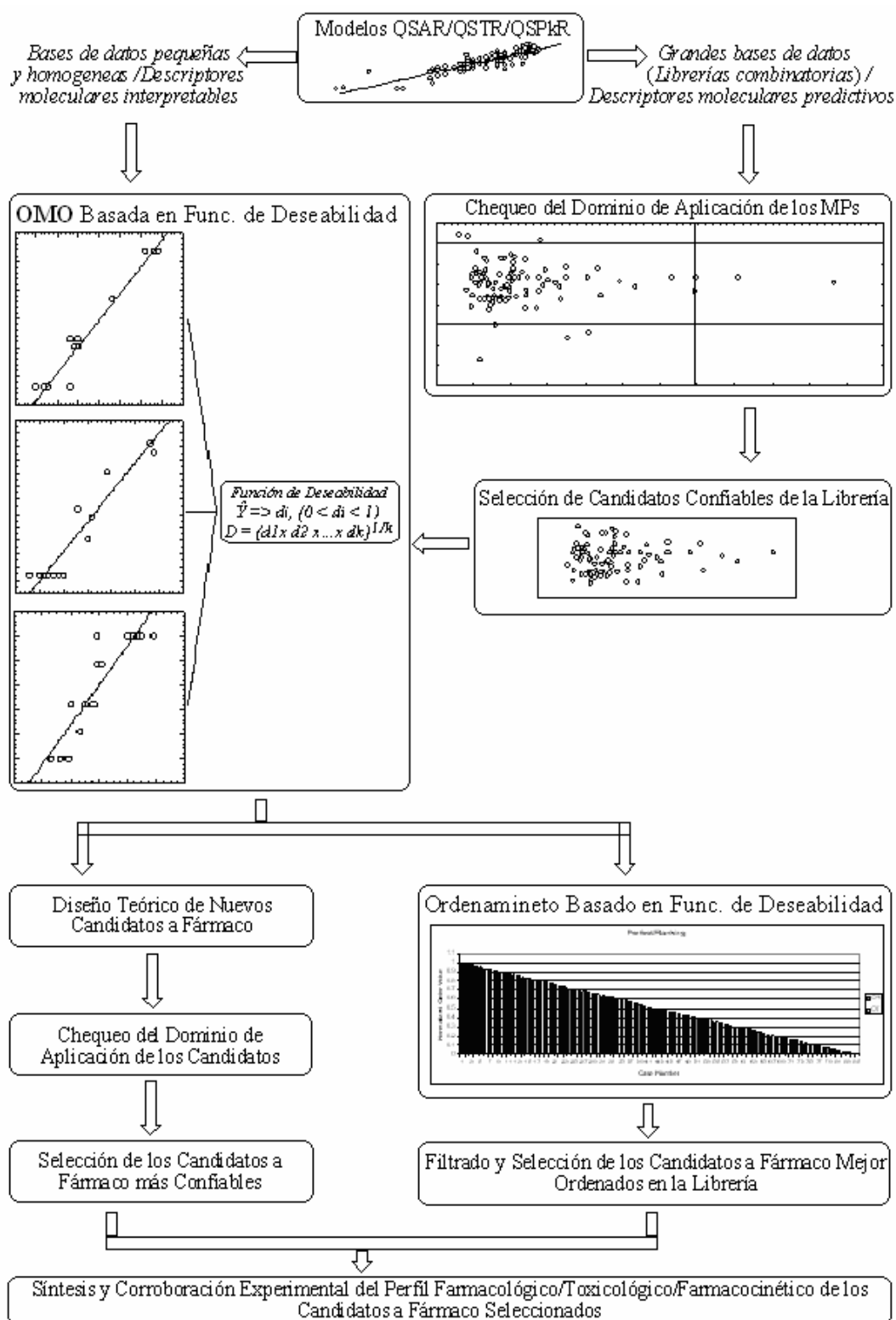


Figura A.1 Descubrimiento y desarrollo racional de nuevos fármacos basado en la metodología MOOP-DESIRE.

Tabla A.1 Resultados de la validación del modelo ADL - best subset selection (ecuación 3.1.2) mediante una técnica de re-sustitución de casos en cuatro subconjuntos de predicción.

<i>Conjunto Total</i>			<i>Subconjunto de Entrenamiento</i>			<i>Subconjunto de Predicción</i>		
Experimento 1								
<i>(1300, 692, 608)^a</i>			<i>(975, 522, 453)^a</i>			<i>(325, 170, 155)^a</i>		
<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>
91.33	90.79	91.08	91.95	91.61	91.79	89.41	88.39	88.92
Experimento 2								
<i>(1300, 692, 608)^a</i>			<i>(975, 518, 457)^a</i>			<i>(325, 174, 151)^a</i>		
<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>
91.33	91.12	91.23	90.93	91.25	91.08	92.53	90.73	91.69
Experimento 3								
<i>(1300, 692, 608)^a</i>			<i>(975, 517, 458)^a</i>			<i>(325, 175, 150)^a</i>		
<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>
91.18	91.78	91.46	91.88	91.27	91.59	89.14	93.33	91.08
Experimento 4								
<i>(1300, 692, 608)^a</i>			<i>(975, 519, 456)^a</i>			<i>(325, 173, 152)^a</i>		
<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>
90.90	92.60	91.69	89.98	92.32	91.08	93.64	93.42	93.54
Resultados Promediados de los 4 Experimentos de Validación								
<i>Conjunto Total</i>			<i>Subconjunto de Entrenamiento</i>			<i>Subconjunto de Predicción</i>		
<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>	<i>%_{CTP}</i>	<i>%_{STP}</i>	<i>%_{Total}</i>
91.18	91.57	91.37	91.18	91.61	91.38	91.18	91.47	91.31
Parámetros Estadísticos								
<i>N</i>	<i>D²</i>	<i>F</i>	<i>p</i>	<i>U</i>	<i>R_{CAN}</i>	<i>Autovalor</i>		
1300	5.88	355.32	0.0000	0.4100	0.7709	1.4652		

^a (Número total de compuestos, número de compuestos CTP, número de compuestos STP)

Tabla A.2 Resumen de los mejores modelos ADL obtenidos por 5 métodos diferentes de selección de variables.

Variables								
Modelo	B_0	$\Delta G_5(T)$	$\Delta G_1(CSat)$	$\Delta G_5(HX)$	$\Delta G_0(T)$	$\Delta G_0(HX)$	$\Delta G_0(Hal)$	$\Delta G_0(CSat)$
1*	0.36	3.10	-0.89	-1.01	1.54	---	---	---
2	0.35	3.80	-1.56	-0.94	---	---	---	1.41
3	0.35	3.80	-1.56	-0.94	---	---	---	1.41
4	0.38	3.16	-1.04	-1.47	1.22	0.94	0.40	---
5	0.38	3.16	-1.04	-1.47	1.22	0.94	0.40	---

Parámetros Estadísticos									
Modelo	N	D^2	F	p	U	R_{CAN}	<i>Autovalor</i>	$\%T_{Ent.}$	$\%T_{Pred.}$
1*	1 300	5.88	355.32	0.00	0.41	0.7709	1.4652	91.79	88.92
2	1 300	5.84	352.78	0.00	0.41	0.7698	1.4548	89.85	86.77
3	1 300	5.84	352.78	0.00	0.41	0.7698	1.4548	89.85	86.77
4	1 300	6.42	258.09	0.00	0.38	0.7844	1.5997	91.79	89.85
5	1 300	6.42	258.09	0.00	0.38	0.7844	1.5997	91.79	89.85

1: Modelo ADL - Best Subset Selection; 2: Modelo ADL - Forward Stepwise Selection; 3: Modelo ADL - Forward Entry Selection; 4: Modelo ADL - Backward Stepwise Selection; 5: Modelo ADL - Backward Removal Selection; $\%T_{Ent.}$: Porcentaje total de compuestos clasificados correctamente en el subconjunto de entrenamiento; $\%T_{Pred.}$: Porcentaje total de compuestos clasificados correctamente en el subconjunto de predicción; * Mejor modelo ADL de acuerdo con el principio de parsimonia.

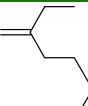
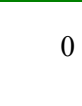
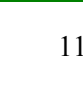
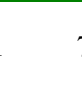
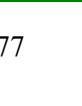

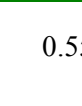
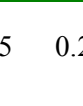
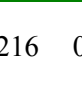
Tabla A.3 Detalles del experimento de inversión del sistema.

Modelo ADL-Forward Stepwise Selection							
CTP = 0.35 + 3.8 ΔG₅(T) + 1.41 ΔG₀(CSat) - 1.56 ΔG₁(CSat) - 0.94 ΔG₅(HX)							
N = 1300 %T = 89.08							
SPB INVERTIDO	b₀	ΔG₅(T)	ΔG₀(CSat)	ΔG₁(CSat)	ΔG₅(HX)	%T	Res.
Aire/Sangre ³	0.33	3.80	1.41	-1.56	-0.94	89.08	0.00
Plasma/Cerebro ⁴	0.34	3.80	1.41	-1.56	-0.95	89.08	0.00
Plasma/Corazón ⁵	0.34	3.81	1.40	-1.56	-0.95	89.08	0.00
Aire/Hígado ⁴	0.33	3.80	1.41	-1.56	-0.94	89.08	0.00
Aire/Músculo ³	0.32	3.80	1.40	-1.56	-0.94	89.08	0.00
Plasma/Músculo ⁵	0.35	3.80	1.40	-1.55	-0.94	89.08	0.00
Aire/Solución salina ¹	0.40	3.82	1.44	-1.60	-0.94	89.08	0.00
Plasma/Piel ⁵	0.36	3.80	1.41	-1.55	-0.95	89.08	0.00
Plasma/Piel ⁴	0.32	3.80	1.41	-1.55	-0.94	89.08	0.00
Epitelio+Estroma/Humor vítreo ⁵	0.34	3.79	1.41	-1.55	-0.94	89.08	0.00
Plasma/Cerebro ⁵	0.35	3.81	1.40	-1.56	-0.95	89.15	-0.08
Plasma/Leche materna ³	0.29	3.83	1.41	-1.55	-0.92	89.15	-0.08
Plasma/Corazón ⁴	0.38	3.81	1.40	-1.55	-0.95	89.15	-0.08
Aire/Hígado ³	0.33	3.79	1.41	-1.55	-0.94	89.15	-0.08
Aire/Músculo ⁴	0.33	3.80	1.40	-1.56	-0.94	89.00	0.08
Plasma/Músculo ⁴	0.39	3.81	1.40	-1.56	-0.96	89.00	0.08
Aire/Piel ³	0.50	3.82	1.42	-1.57	-0.98	89.15	-0.08
Plasma/Baso ⁴	0.34	3.80	1.41	-1.55	-0.94	89.00	0.08
Estroma/Humor vítreo ⁵	0.35	3.80	1.41	-1.55	-0.95	89.00	0.08
Dimiristoil fosfatidil colina/Agua ¹	0.36	3.82	1.40	-1.58	-0.94	89.15	-0.08
Aire/Carbón activado ¹	0.31	3.78	1.42	-1.55	-0.94	89.23	-0.15
Aire/Tejido adiposo ³	0.35	3.79	1.42	-1.56	-0.94	89.23	-0.15
Plasma/Hueso ⁵	0.34	3.79	1.40	-1.55	-0.94	89.23	-0.15
Agua/Pez ²	0.31	3.75	1.40	-1.53	-0.97	89.23	-0.15
Plasma/Pulmón ⁵	0.36	3.81	1.40	-1.55	-0.95	89.23	-0.15
Aire/Suelo ¹	-0.19	3.75	1.38	-1.53	-0.92	89.23	-0.15
Plasma/Intestinos ⁴	0.33	3.79	1.41	-1.56	-0.94	88.92	0.15
Plasma/Pulmón ⁴	0.32	3.80	1.41	-1.55	-0.94	88.92	0.15
Isopropil miristato/Agua ¹	0.77	4.00	1.40	-1.62	-0.97	88.92	0.15
Retención BMC en 0.04M Brij35 ¹	0.26	3.78	1.40	-1.54	-0.94	89.31	-0.23
Aire/Superficie de la nieve ¹	0.27	3.76	1.39	-1.54	-0.92	89.31	-0.23
Toda la córnea/Humor vítreo ⁵	0.29	3.78	1.40	-1.54	-0.94	89.31	-0.23
Aire/Tejido adiposo ⁴	0.31	3.77	1.41	-1.55	-0.94	89.38	-0.31
CH ₃ C ₆ H ₅ /CF ₃ C ₆ F ₁₁ ¹	0.34	3.80	1.36	-1.51	-0.94	89.38	-0.31
Aire/Agua ¹	0.27	3.77	1.40	-1.55	-0.93	89.38	-0.31
Aire/Sangre ⁴	0.27	3.73	1.45	-1.55	-0.93	89.46	-0.38
Agua/Aceite vegetal ¹	0.25	3.78	1.39	-1.53	-0.93	89.46	-0.38
Aire/Aceite de oliva ¹	0.52	3.87	1.63	-1.63	-0.97	88.62	0.46
<i>Media</i>	<i>0.34</i>	<i>3.80</i>	<i>1.41</i>	<i>-1.56</i>	<i>-0.94</i>	<i>89.14</i>	<i>---</i>
<i>Desviación Estándar</i>	<i>0.12</i>	<i>0.04</i>	<i>0.04</i>	<i>0.02</i>	<i>0.01</i>	<i>0.17</i>	<i>---</i>

Retención BMC: Retención en Cromatografía de Bio-partición Micelar, un sistema cromatográfico constituido por una fase móvil micelar de polioxetileno (23) lauril éter, Brij35 y una fase estacionaria invertida C₁₈. Brij35: Polidocanol, una mezcla de éteres monolaurílicos de polioxietileno glicoles usada como aditivo farmacéutico (surfactante). %T: Porcentaje total de

compuestos clasificados correctamente. ¹ Abiótico; ² Medido en Peces; ³ Medido en Humanos; ⁴ Medido en Ratas; ⁵ Medido en Conejos

Tabla A.4 Estructura, descriptores moleculares ACF calculados (C-001, C-037 y H-046), valores predichos de la actividad analgésica (An), antiinflamatoria (Aa) y ulcerogénica (U) y valores de leverage (h) de los nuevos compuestos diseñados.

Compuesto	R	C-001	C-037	H-046	An_{pred} (%)	Aa_{pred} (%)	U_{pred} (%)	$h(An)$	$h(Aa)$	$h(U)$
ASNEW1		3	0	11	77	70	0.55	0.216	0.361	0.216
ASNEW2		3	0	13	77	72	0.55	0.216	0.496	0.216
ASNEW3		4	0	12	85	77	0.49	0.403	0.453	0.403
ASNEW4*		5	0	15	93	86	0.44	0.573	0.614	0.573
ASNEW5*		6	0	18	102	96	0.38	0.695	0.724	0.695
ASNEW6*		7	0	21	110	106	0.33	0.777	0.796	0.777
ASNEW7		4	0	9	85	72	0.49	0.403	0.401	0.403
ASNEW8		5	0	12	93	82	0.44	0.573	0.562	0.573
ASNEW9*		5	0	15	93	86	0.44	0.573	0.614	0.573

* Compuestos fuera del dominio de aplicabilidad de los MPs; valores de $h > h^*$ se resaltan en negrita.

Tabla A.5 Valores predichos de deseabilidad debido a la actividad analgésica [$d(An)$], antiinflamatoria [$d(Aa)$], índice ulcerogénico [$d(U)$], y deseabilidad global [$D(An-Aa-U)$] de los nuevos compuestos diseñados.

Compuesto	$d(An)_{pred}$	$d(Aa)_{pred}$	$d(U)_{pred}$	$D(An-Aa-U)_{pred}$
ASNEW1	0.69	0.60	0.73	0.67
ASNEW2	0.69	0.63	0.73	0.68
ASNEW3	0.80	0.69	0.65	0.71
ASNEW4*	0.91	0.81	0.75	0.82
ASNEW5*	1.00	0.95	0.78	0.90
ASNEW6*	1.00	1.00	0.81	0.93
ASNEW7	0.80	0.63	0.72	0.71
ASNEW8	0.91	0.76	0.75	0.80
ASNEW9*	0.91	0.81	0.75	0.82

* Compuestos fuera del dominio de aplicabilidad de los MPs.

Tabla A.6 Valores de Δ_i de los 85 compuestos usados en el ordenamiento.

Compuesto	Δ_i	$^D\Delta_i$	Pred. $D_{(MIC-IC50)}$	Compuesto	Δ_i	$^D\Delta_i$	Pred. $D_{(MIC-IC50)}$
004-4-Ciprofloxacin	0.305	0.993	0.956	064-30E	1.221	0.766	0.793
006-6-Tosufloxacin	0.330	0.987	0.968	065-30F	0.718	0.891	0.885
010-10	2.764	0.382	0.452	066-31A	0.359	0.980	0.882
014-15	0.801	0.870	0.751	067-31B	1.241	0.761	0.717
015-16	0.927	0.839	0.788	068-31C	0.871	0.853	0.733
016-17	1.416	0.717	0.776	070-31E	0.947	0.834	0.769
018-19	0.463	0.954	0.943	071-31F	0.765	0.879	0.780
019-20	0.510	0.943	0.959	073-32B	1.130	0.788	0.796
020-21	1.274	0.753	0.793	074-32C	1.123	0.790	0.709
021-22	0.919	0.841	0.901	075-32D	0.970	0.828	0.826
022-23A	0.528	0.938	0.806	077-32F	0.708	0.893	0.848
023-23B	1.132	0.788	0.777	078-33B	1.205	0.770	0.820
024-23C	0.411	0.967	0.904	079-34B	2.903	0.348	0.699
025-23D	1.040	0.811	0.761	080-35B	0.988	0.824	0.894
027-23F	0.680	0.900	0.856	081-36B	1.729	0.640	0.715
028-24A	0.730	0.888	0.930	082-37B	1.703	0.646	0.695
029-24C	0.576	0.926	0.879	083-38A	1.046	0.809	0.857
030-24D	0.829	0.863	0.882	084-38B	1.589	0.674	0.803
031-24E	1.060	0.806	0.823	085-39A	2.044	0.561	0.596
032-24F	0.701	0.895	0.896	086-39B	4.303	0.000	0.358
033-25A	1.004	0.820	0.790	088-41A	1.117	0.792	0.763
034-25B	1.713	0.644	0.508	090-42A	1.214	0.768	0.729
037-25E	1.425	0.715	0.699	092-48	0.745	0.884	0.770
038-25F	0.859	0.856	0.713	093-49	0.486	0.949	0.920
040-26D	1.658	0.657	0.737	094-50	1.120	0.791	0.771
041-26E	1.904	0.596	0.756	095-51	0.672	0.902	0.929
042-26F	0.631	0.912	0.811	096-52	1.279	0.751	0.664
043-27A	1.723	0.641	0.707	098-54	0.444	0.959	0.957
044-27B	2.595	0.424	0.000	100-56	0.746	0.884	0.895
046-27D	1.405	0.720	0.647	102-58	1.183	0.775	0.738
047-27E	1.572	0.679	0.667	103-59	0.656	0.906	0.838
048-27F	1.359	0.731	0.685	104-60	0.680	0.900	0.890
049-28A	1.912	0.594	0.753	105-61	0.825	0.864	0.641
052-28D	1.509	0.694	0.707	106-62	2.219	0.518	0.446
054-28F	1.784	0.626	0.789	107-63	1.159	0.781	0.840
055-29B	1.132	0.788	0.835	110-70	1.630	0.664	0.637
056-29C	1.012	0.818	0.791	111-71	1.050	0.808	0.712
057-29D	1.061	0.806	0.822	112-72	1.142	0.785	0.753
058-29E	0.279	1.000	0.811	113-73	1.205	0.770	0.655
059-29F	0.711	0.893	0.905	114-74	1.631	0.664	0.754
061-30B	1.191	0.773	0.872	115-75	1.495	0.698	0.675
062-30C	1.278	0.752	0.800	118-78	0.739	0.886	0.775
063-30D	0.945	0.834	0.860				

Tabla A.7 Lista ordenada de los candidatos, obtenida a través del tamizaje secuencial del 10% de la biblioteca de 85 fluoroquinolonas.

TAMIZAJE VIRTUAL SECUENCIAL								
<i>(Perfil Antibacteriano (MIC) y Perfil de Seguridad (IC₅₀) ≥ 90%)</i>								
ID Candidato	1/1+MIC Predicho	d(MIC) Predicho	Clase MIC	1/1+IC ₅₀ Predicho	d(IC ₅₀) Predicho	Clase IC ₅₀	D _{MIC-IC50}	Clase Control (D _{MIC-IC50})
034-25B	1.016	1.000	+	0.075	0.258	-	0.404	-
055-29B	1.009	1.000	+	0.032	0.698	-	0.872	-
056-29C	1.016	1.000	+	0.039	0.626	-	0.862	-
028-24A	0.973	0.982	+	0.014	0.882	-	0.902	+
107-63	0.973	0.982	+	0.030	0.718	-	0.879	-
103-59	0.960	0.968	+	0.029	0.725	-	0.891	-
073-32B	0.951	0.959	+	0.035	0.66	-	0.860	-
019-20	0.947	0.955	+	0.006	0.962	+	0.938	+
032-24F	0.944	0.951	+	0.017	0.844	-	0.948	+
<i>Medidas de Enriquecimiento</i>		VP _{10%} = 11.112%				FE _{10%} = 0.556		

Tabla A.8 Lista ordenada de los candidatos, obtenida a través del tamizaje en paralelo del 10% de la biblioteca de 85 fluoroquinolonas.

TAMIZAJE VIRTUAL EN PARALELO								
(Perfil Antibacteriano (MIC) y Perfil de Seguridad (IC₅₀) ≥ 90%)								
Según MIC								
ID Candidato	1/1+MIC Predicho	d(MIC) Predicho	Clase MIC	1/1+IC ₅₀ Predicho	d(IC ₅₀) Predicho	Clase IC ₅₀	D _{MIC-IC50}	Clase Control (D _{MIC-IC50})
034-25B	1.016	1.000	+	0.075	0.258	-	0.404	-
055-29B	1.009	1.000	+	0.032	0.698	-	0.872	-
056-29C	1.016	1.000	+	0.039	0.626	-	0.862	-
028-24A	0.973	0.982	+	0.014	0.882	-	0.902	+
107-63	0.973	0.982	+	0.030	0.718	-	0.879	-
103-59	0.960	0.968	+	0.029	0.725	-	0.891	-
073-32B	0.951	0.959	+	0.035	0.66	-	0.860	-
019-20	0.947	0.955	+	0.006	0.962	+	0.938	+
032-24F	0.944	0.951	+	0.017	0.844	-	0.948	+
Según IC₅₀								
065-30F	0.784	0.783	-	-0.019	1.000	+	0.917	+
014-15	0.576	0.565	-	-0.011	1.000	+	0.791	-
004-4	0.908	0.914	+	-0.010	1.000	+	0.954	+
Ciprofloxacin								
064-30E	0.637	0.629	-	-0.007	1.000	+	0.714	-
006-6	0.931	0.938	+	-0.006	1.000	+	0.932	+
Tosufloxacin								
058-29E	0.664	0.658	-	-0.002	1.000	+	0.914	+
080-35B	0.799	0.799	-	-0.002	1.000	+	0.853	-
031-24E	0.683	0.677	-	0.001	1.000	+	0.908	+
093-49	0.844	0.847	-	0.001	1.000	+	0.797	-
<i>Medidas de Enriquecimiento</i>		VP _{10%} = 16.67%				FE _{10%} = 0.833		

Tabla A.9 Lista ordenada de los candidatos, obtenida a través del tamizaje multiobjetivos del = 10% de la biblioteca de 85 fluoroquinolonas.

TAMIZAJE VIRTUAL MULTI OBJETIVOS (<i>Perfil Terapéutico</i> ($D_{MIC-IC50} \geq 90\%$))				
ID Candidato	$D_{MIC-IC50}$ Predicho	Δ_i	$D_{MIC-IC50}$	Clase Control ($D_{MIC-IC50}$)
058-29E	0.811	0.279	0.914	+
004-4 Ciprofloxacin	0.956	0.305	0.954	+
006-6 Tosufloxacin	0.968	0.330	0.932	+
066-31A	0.882	0.359	0.880	-
024-23C	0.904	0.411	0.934	+
098-54	0.957	0.444	0.924	+
018-19	0.943	0.463	0.941	+
093-49	0.920	0.486	0.797	-
019-20	0.959	0.510	0.938	+
<i>Medidas de Enriquecimiento</i>	$VP_{10\%} = 77.778\%$		$FE_{10\%} = 3.889$	