

UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS

“50 Aniversario del Triunfo de la Revolución”



Aplicación de la Metodología MOOP-DESIRE a la Identificación de Inhibidores No Nucleósidos de la Transcriptasa Inversa del VIH Tipo 1 con perfiles de Eficacia y Seguridad Simultáneamente Optimizados

Tesis para optar por el Título de Licenciado en Ciencias Farmacéuticas



HAI PHAM THE

Curso 2008 - 2009

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



**“Aplicación de la Metodología MOOP-DESIRE a la
Identificación de Inhibidores No Nucleósidos de la
Transcriptasa Inversa del VIH Tipo 1 con perfiles
de Eficacia y Seguridad Simultáneamente
Optimizados”**

Tesis para optar por el Título de Licenciado en Ciencias Farmacéuticas

Autor: Hai Pham The

Tutor: Lic. Maykel Cruz Monteagudo

Santa Clara, Cuba

2009



Dedico este trabajo:

*A mi familia
y a todos aquellos que siempre están a mi lado*

Mis sinceros agradecimientos a los que me han brindado su estímulo y apoyo incondicional:

A mi tutor Maykel Cruz Monteagudo por todas sus atenciones, apoyo durante la realización de esta tesis, por el tiempo dedicado así como la formación científica que he adquirido gracias a su dedicación y seriedad en el trabajo.

A mi familia, muy especialmente a mi mamá, mi hermana y mi cuñado por su gran apoyo siempre.

A Huong por confiar en mí y estar siempre a mi lado.

A los profesores de facultad y departamento por enseñarme en todos estos años.

A mis compañeros tanto vietnamitas como cubanos por los buenos y malos momentos.

A todos, muchas gracias.

SÍNTESIS

El número de personas que viven con el virus de inmunodeficiencia humana (VIH) o con SIDA en el mundo alcanzó los 33.2 millones a finales del año 2007 y se estima que será responsable de la muerte de más de la mitad de los adultos en países subdesarrollados del continente africano, asiático y Latinoamérica. En la actualidad no existe un tratamiento eficaz de la enfermedad, y la rápida aparición de cepas de virus resistentes a los fármacos actuales, agravan la situación. En este sentido, los inhibidores no nucleosídicos de la transcriptasa inversa (INNTIs), han jugado un importante rol debido a su potencia antiviral, alta especificidad y baja toxicidad comparada con sus análogos nucleosídicos. Aun así, los INNTIs no están exentos de efectos tóxicos, siendo el principal de estos su nivel de toxicidad sobre células sanguíneas. De aquí la necesidad de encontrar un agente antiretroviral efectivo con un perfil de seguridad adecuado. 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 de gran utilidad para el tamizaje virtual de moléculas empleado en el desarrollo racional de fármacos. A pesar de la disponibilidad de numerosos objetivos de optimización, las técnicas OMO han sido sólo recientemente aplicadas al área de los estudios QSAR. Lo anterior apunta hacia la necesidad de proponer una estrategia de tamizaje virtual basada en estudios QSAR globales que consideren simultáneamente los perfiles farmacológicos y toxicológicos de candidatos tipo INNTI, para la identificación y selección racional de candidatos con perfiles de eficacia y seguridad simultáneamente optimizados. Esto es implementado mediante el empleo de una metodología de OMO y ordenamiento basado en las funciones de deseabilidad de Derringer denominada *MOOP-DESIRE*, acrónimo de *MultiObjective Optimization based on the Desirability Estimation of Several Interrelated REsponses*. La metodología propuesta fue aplicada a una biblioteca de 109 compuestos reportados en literatura con el objetivo de optimizar simultáneamente su actividad inhibitoria sobre la TI del VIH-1 y la toxicidad sobre células sanguíneas MT4; ordenar los compuestos incluidos en esta librería con respecto al candidato ideal identificado durante el proceso de OMO y filtrar los candidatos más prometedores de la misma. Los valores obtenidos de las medidas de la calidad estadística de la función de deseabilidad global ($R^2_D/Q^2_D \sim 0.69/0.62$) indican que la optimización simultánea de los perfiles farmacológicos y toxicológicos de candidatos del conjunto de 109 INNTIs puede ser llevado a cabo con un nivel adecuado de confiabilidad. Basados en el porcentaje de calidad ($R\%$) del esquema de ordenamiento obtenido, se infiere que su aplicación a nuevos análogos tipo INNTI ordenará correctamente el 76,16% de estos. Todos estos resultados sugieren que la aplicación de la metodología *MOOP-DESIRE*, puede derivar en una herramienta útil para la identificación racional de inhibidores no nucleosidos de la TI del VIH-1 con perfiles de eficacia y seguridad simultáneamente optimizados.

TABLA DE CONTENIDOS

SINTESIS

TABLA DE CONTENIDOS

GLOSARIO

INTRODUCCIÓN	1
1. REVISIÓN BIBLIOGRÁFICA.....	5
1.1 VIH/SIDA. GENERALIDADES	5
1.1.1 Elementos clínicos, epidemiológicos y transmisión del VIH/SIDA.....	5
1.1.2 Estructura y morfología del VIH-1	8
1.1.3 Organización del genoma Viral	9
1.1.4 Ciclo de replicación	10
1.1.5 Capacidad mutagénica y resistencia del VIH	10
1.2 TERAPIA ANTI-VIH	12
1.2.1 La Transcriptasa inversa (TI) del VIH, blanco biológico principal de los fármacos antiretrovirales...	13
1.2.2 Fármacos Inhibidores No Nucleosídicos de la Transcriptasa Inversa (INNTI).....	15
1.3 TENDENCIAS ACTUALES.....	18
1.4 QUIMIOINFORMÁTICA, QUIMIOMETRÍA Y DESARROLLO RACIONAL DE FÁRMACOS.....	20
1.4.1 Generalidades del proceso de descubrimiento de fármacos	20
1.4.2 Tecnologías de alto rendimiento (<i>High-throughput technologies</i>).....	21
1.4.3 Métodos computacionales	22
2. MATERIALES Y MÉTODOS	36
2.1 EMPLEO DE MÉTODOS DE OPTIMIZACIÓN MULTI OBJETIVOS (OMO) Y ORDENAMIENTO BASADOS EN FUNCIONES DE DESEABILIDAD EN ESTUDIOS GLOBALES DE CORRELACIÓN ESTRUCTURA-PROPIEDAD	36
2.1.1 Optimización multiobjetivos basada en la estimación de la deseabilidad de varias respuestas interrelacionadas	37
2.1.2 Ordenamiento de soluciones basado en funciones de deseabilidad.....	41
2.1.3 Validación del método de ordenamiento y estimación del índice de calidad del ordenamiento	43
2.2 CONFORMACIÓN DE LA BASE DE DATOS	45
2.3 ESPECIFICACIONES DE LAS FUNCIONES DE DESEABILIDAD	47
3. RESULTADOS Y DISCUSIÓN	51
3.1 OPTIMIZACIÓN SIMULTÁNEA DE LA ACITIVIDAD INHIBITORIA SOBRE LA TI DEL VIH-1 Y LA TOXICIDAD SOBRE CÉLULAS SANGUÍNEAS MT-4 DE COMPUESTOS TIPO INNTI	51
CONCLUSIONES.....	69
RECOMENDACIONES	70
REFERENCIAS BIBLIOGRÁFICAS.....	72

GLOSARIO

2',3'-ddN	2',3'-Didesoxinucleósidos
ADG	Análisis Discriminante General
ADL	Análisis Discriminante Lineal
ADME	Absorción, Distribución, Metabolismo y Excreción
ADN	Ácido desoxirribonucleico
AG	Algoritmo Genético
ARN	Ácido ribonucleico
ART	Antiretroviral
AZT	3'-Azido-3'-desoxitimidina
CC ₅₀	Concentración citotóxica media
CECMED	Centro para el Control Estatal de la Calidad de los Medicamentos
CENPALAB	Centro Nacional de Animales de Laboratorio
CoMFA	<i>Comparative Molecular Field Analysis</i>
CR	<i>Continuum Regresión</i>
ddI	2',3'-Didesoxiinosina
ddI	2',3'-Didesoxiinosina
DM	Descriptor Molecular
EC ₅₀	Concentración efectiva media
F	Fisher's statistics
FDA	<i>The US-American Food and Drug Administration</i>
HAART	Highly active antiretroviral therapy
HTS	<i>High-Throughput Screening</i>
IM	Índice Molecular
IT	Índice Topológico
LOO-CV	Procedimiento de validación cruzada dejando un caso fuera, <i>leave-one-out cross validation</i>
MCO	Mínimos Cuadrados Ordinarios
MCP	Mínimos Cuadrados Parciales
MIC	Mínima Concentración Inhibitoria
MLG	Modelos Lineales Generales
MOOP-DESIRE	Multi-Objective OPTimization based on the Desirability Estimation of Several Interrelated Responses
MP	Modelo de predicción
MRG	Modelos de Regresión Generales
MSC	Máximo sub-grafo común
N	Número de compuestos
NEM	Nueva Entidad Molecular
INTI	Inhibidores nucleosídicos de la transcriptasa inversa
INNTI	Inhibidores no nucleosídicos de la transcriptasa inversa
OMO	Optimización multiobjetivos
OMS	Organización Mundial de la Salud
ONUSIDA	Programa Conjunto de las Naciones Unidas sobre VIH/SIDA
PFC	<i>Potencial Function Clasifiers</i>

GLOSARIO

<i>QDA</i>	<i>Quadratic Discriminant Analysis</i>
<i>QSAR</i>	<i>Quantitati Structure-Activity Relatioships</i>
<i>QSPR</i>	<i>Quantitative Structure-Property Relationships</i>
R^2	Determination coefficient or corelation coefficient
R^2 adj.	Adjusted determination coefficient
<i>RDA</i>	<i>Regularized Discriminant Analysis</i>
REA	Relación Estructura-Actividad
REP	Relación Estructura-Propiedad
RLM	Regresión Lineal Múltiple
TI	Enzima viral Transcriptasa Inversa
S	Desviación estándar
SIDA	Síndrome de InmunoDeficiencia Adquirida
VIH	Virus de inmunodeficiencia humana
VIH-1	Virus de inmunodeficiencia humana tipo 1
VIS	Virus de Inmunodeficiencia en Simios
<i>WNMC</i>	<i>Weighted Nearest Mean Classifiers</i>

INTRODUCCIÓN

INTRODUCCIÓN

“The beginning is the most important part of the work.”

Platon

En mayo de 1981, los primeros casos de pacientes homosexuales afectados por enfermedades que hasta entonces habían sido raras, como la neumonía por *Pneumocystis* (PCP) y el sarcoma de Kaposi, fueron publicados (1-3). En esa etapa Robert Gallo en Estados Unidos y Luc Montagné en Francia, aportaron casi simultáneamente la evidencia de que un virus era la causa de la nueva enfermedad (4, 5). Pronto quedó claro que esta enfermedad afectaba también a otros grupos de la población, mayormente consumidores de drogas inyectables. Transcurrieron casi dos años hasta que en 1983 fue aislado el virus de la inmunodeficiencia humana (VIH) tipo 1 (VIH-1) (6-8), y el tipo 2 (VIH-2) en 1986. Estos dos virus se han identificado durante casi 20 años como la causa primaria del Síndrome de Inmunodeficiencia Adquirida (SIDA).

Hoy en día el SIDA se ha convertido en una epidemia mucho más grave y extensa de lo que predecían las estadísticas hace casi tres décadas. De acuerdo con el informe titulado “La epidemia del SIDA: situación en diciembre de 2007”, dado a conocer por el Programa Conjunto de las Naciones Unidas sobre VIH/SIDA (ONUSIDA) y la Organización Mundial de la Salud (OMS) (9, 10), se estima que el número de personas que viven con el virus de inmunodeficiencia humana (VIH) o con SIDA en el mundo alcanzó los 33.2 millones a finales del año 2007. Desgraciadamente, el SIDA será responsable de la muerte de más de la mitad de los adultos en algunos países subdesarrollados del continente africano, asiático y Latinoamérica.

Ciertamente, la epidemia ya no está fuera de control en los países desarrollados, gracias a las nuevas combinaciones de medicamentos contra el VIH continúan reduciendo de manera significativa las defunciones a causa del SIDA. Sin embargo, en la actualidad no existe cura para el SIDA, y lo que es más grave, la terapia del SIDA se enfrenta al problema de la rápida aparición de cepas de virus resistentes en tratamientos tanto en monoterapia como en terapias combinadas. Sin embargo, no todo es negativo, el desarrollo constante de la terapia antirretroviral nos indica que la prevención y un tratamiento efectivo de la infección por VIH serían factibles y abriría nuevas páginas en la historia de la medicina.

El aislamiento del VIH como agente etiológico del SIDA (6, 11), y su identificación (1984) como miembro de la familia de los retrovirus, ha estimulado enormemente la búsqueda de agentes capaces de bloquear la replicación del virus y frenar así la enfermedad. Los retrovirus son así denominados porque en su ciclo replicativo, la información fluye del ARN al ADN, en una dirección inversa a la habitual. En este proceso está implicada una enzima viral denominada Transcriptasa Inversa (TI). Debido a su asociación única a retrovirus, la TI es considerada uno de los objetivos más atractivos en el diseño de compuestos selectivos frente a retrovirus y más concretamente frente al VIH (12, 13). Al estar presente la TI únicamente en las células infectadas por el virus, la inhibición de dicha enzima supone un punto de ataque selectivo sobre la replicación del VIH (14).

Recientemente, la estructura del complejo de un inhibidor con en el sitio activo de la TI ha sido determinado, lo que unido a la creciente sensibilidad social ocasionada por la fatal epidemia de SIDA da un ímpetu adicional para el diseño racional y el screening virtual con el objetivo de encontrar en el menor tiempo y

costo posibles, inhibidores que tengan perspectivas de llegar a fármacos y ser utilizados en la terapia antiretroviral.

En particular, los inhibidores no nucleosídicos de la transcriptasa inversa (INNTIs) desempeñan un papel esencial en la represión de la replicación del VIH-1. A diferencia de los inhibidores nucleosídicos de la transcriptasa inversa (INTIs), los INNTIs generalmente muestran una baja toxicidad, ya que no se unen a las polimerasas celulares, sino que interactúan con un sitio alostérico específico adyacente al sitio de la polimerasa del VIH-1 dando lugar a un mecanismo no-competitivo. Aun así, los INNTIs no están exentos de efectos tóxicos, siendo el principal de estos su nivel de citotoxicidad (principalmente sobre células sanguíneas tipo MT4 y CEM) (15, 16). En este sentido, en los dos últimos decenios se han realizado esfuerzos para desarrollar fármacos efectivos y seguros para el tratamiento de la infección por VIH. Actualmente se han identificado y aprobado para su uso clínico más de 30 clases estructuralmente diversas de INNTIs.

Como cualquier medicamento, desarrollar un fármaco antiretroviral exitoso es el resultado del descubrimiento del mejor compromiso entre numerosos objetivos que muy a menudo compiten entre sí. El fracaso de un candidato a fármaco con una potencia adecuada durante el proceso de desarrollo es usualmente debido a múltiples factores involucrados; principalmente a una pobre biodisponibilidad, y/o toxicidad (17). En una forma muy 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 y desarrollo de fármacos.

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 (18-21). Esto 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 (18, 21).

Todo lo anterior ha llevado tanto a la academia como a la industria farmacéutica a una reconsideración del paradigma secuencial de las fases de identificación y optimización del líder a favor de un enfoque multiobjetivos del proceso de descubrimiento y desarrollo de nuevos fármacos. Este cambio de paradigma marca un avance significativo hacia la racionalización del proceso (18, 22). 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 (23).

A su vez, 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 y desarrollo de nuevos fármacos (24), redireccionando nuestro pensamiento a una reestructuración de la química médica (25). Sin embargo, a

pesar de los avances logrados en el campo de la quimioinformática, la naturaleza multiobjetiva de los medicamentos todavía es ignorada por estos métodos al optimizar secuencialmente cada propiedad involucrada en el perfil terapéutico global de un candidato (26-41).

A pesar de la disponibilidad de numerosos objetivos de optimización, las técnicas OMO han sido sólo recientemente aplicadas al área de los estudios QSAR y han estado dirigidas a la optimización del modelo. Realmente, muy poca información existe acerca de la aplicación de métodos OMO a estudios QSAR (42-44). Hasta hoy es raro el reporte de estudios QSAR aplicados a 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 fármacos tipo INNTIs empleados en la actualidad en la terapia del VIH-1, aunque eficaces, no están exentos de efectos tóxicos, siendo el principal de estos su nivel de toxicidad sobre células sanguíneas; lo que sugiere que la optimización del perfil terapéutico del candidato anti-VIH-1, basándose solamente en su actividad farmacológica se ha sobreestimado.

Como vía para solucionar el problema científico se formula la siguiente hipótesis:

La aplicación de una estrategia de tamizaje virtual basada en estudios QSAR globales que consideren simultáneamente el perfil farmacológico y toxicológico de una familia de candidatos, puede derivar en una herramienta útil para la identificación racional de inhibidores no nucleósidos de la TI del VIH-1 con perfiles de eficacia y seguridad simultáneamente optimizados.

Para demostrar la anterior hipótesis y dar respuesta al problema científico planteado, se propone el siguiente objetivo general:

Objetivo General

Proponer una estrategia de tamizaje virtual basada en estudios QSAR globales que consideren simultáneamente la actividad inhibitoria sobre la TI del VIH-1 y la toxicidad sobre células sanguíneas MT4 de una familia estructuralmente diversa de análogos no nucleosídicos, para la identificación y selección racional de candidatos con perfiles de eficacia y seguridad simultáneamente optimizados.

Objetivos específicos

- Conformar una base de datos consistente, estructuralmente diversa y representativa de inhibidores no nucleosídicos de la TI del VIH-1 que contenga información relativa a su actividad inhibitoria sobre la TI del VIH-1 y su toxicidad sobre células sanguíneas MT4.
- Determinar los niveles de las variables predictivas necesarios para obtener un análogo no nucleosídico con una combinación óptima entre su actividad inhibitoria sobre la TI del VIH-1 y su toxicidad sobre células sanguíneas MT4, mediante el empleo de un método de OMO basado en funciones de deseabilidad de Derringer.

- Ordenar los compuestos inhibidores no nucleosídicos de la TI del VIH-1 incluidos en la serie de entrenamiento con respecto a su similitud estructural con el candidato óptimo identificado durante el proceso de OMO mediante el empleo de un método de ordenamiento basado en funciones de deseabilidad.
- Utilizar el criterio estructural de ordenamiento obtenido, como esquema de ordenamiento y filtrado de nuevos candidatos tipo INNTI con perfiles de eficacia y seguridad simultáneamente optimizados.

1 REVISIÓN BIBLIOGRÁFICA

1. REVISIÓN BIBLIOGRÁFICA

1.1 VIH/SIDA. GENERALIDADES

“Since the measuring device has been constructed by the observer, we have to remember that what we observe is not nature in itself, but nature exposed to our method of questioning.”

Werner Karl Heisenberg

1.1.1 Elementos clínicos, epidemiológicos y transmisión del VIH/SIDA

Desde la descripción de los primeros casos en 1981, el SIDA se ha convertido en la primera pandemia del siglo XXI. A pesar de los avances terapéuticos, que han modificado el espectro de la enfermedad en los países más desarrollados hasta convertirla en una "infección crónica manejable", la epidemia sigue avanzando a nivel mundial (45).

Tres décadas han pasado, más de 60 millones de personas se han infectado por el VIH. Actualmente es la primera causa de muerte en África subsahariana y la cuarta en el ámbito mundial. En muchas partes del mundo en desarrollo, la mayoría de las nuevas infecciones ocurren en adultos jóvenes, con un aumento creciente de la prevalencia en mujeres jóvenes. Aproximadamente una tercera parte de las personas que actualmente viven con el VIH tienen entre 15 y 24 años de edad, y la mayoría de ellas desconocen que son portadoras de la infección (46).

Nuevas estimaciones arrojan cambios considerables en el número estimado de las personas viviendo con VIH en todo el mundo. La cifra estimada es de 33.2 millones, una reducción de 16% comparado con lo publicado en 2006 <http://www.unaids.org/>.

Tabla 1.1: La epidemia del SIDA*

	Adultos y niños vivían con VIH	Prevalencia del VIH en adultos (15-49) [%]	Nueva infecciones por día**	Defunciones diarias por causa del AIDS**
África Subsahariana	22.5 millón	5.0	4,700	4,400
Asia de Sur y Sureste	4.0 millón	0.3	900	740
Europa Oriental y Asia Central	1.6 millón	0.9	410	210
América Latina	1.6 millón	0.5	270	160
Asia Oriental	800,000	0.1	250	90
África del Norte y Oriente Medio	380,000	0.3	100	70
América del Norte	1.3 millón	0.6	130	60
Caribe	230,000	1.0	50	30
Europa Occidental y Central	760,000	0.3	80	30
Australia, Nueva Zelanda y Región Pacífico	75,000	0.4	40	3
Total	33.2 millón	0.8	6900	5800

*Adaptado de WHO: Epidemia de SIDA Actualizado en diciembre de 2007, <http://hiv.net/link.php?id=227>

** En adultos y niños

La transmisión del VIH puede ocurrir a través del contacto con fluidos corporales infectados. Se asume que las principales vías de transmisión del virus son los fluidos celulares como la sangre, el semen y las secreciones vaginales; transmisión de madre a hijo (transmisión vertical) y mediante la lactancia materna siendo más efectivos estos que fluidos deficientes de células como saliva, orina y lágrimas. La transmisión puede ocurrir a través de membranas mucosas o la piel dañada durante el contacto sexual (heterosexual y homosexual), pero esta también puede ocurrir vía intravenosa con agujas o productos sanguíneos infectados.

La manera en que este virus causa la enfermedad es a través de infección y eventual destrucción de los linfocitos TCD4+. Estos tienen una función central en el desarrollo de las respuestas inmunitarias contra agentes infecciosos; pero no sólo se involucran en el reconocimiento de los agentes infecciosos, sino que también estimulan y coordinan la actividad de todas las células implicadas en la respuesta inmunológica. De alguna manera, este compartimiento celular se ha comparado con el director de una sinfónica, ya que el resto de los elementos de la inmunidad –que serían los músicos– obedecen en armonía las señales de las células TCD4+. Al destruirse éstas, el sistema inmune se vuelve anárquico y, aunque puede demostrarse la integridad de muchas de sus funciones efectoras –como la producción de anticuerpos–, la respuesta ante otros organismos patógenos simplemente no es eficiente. Así, en pacientes en quienes ya se han destruido una gran cantidad de células TCD4+, concurren infecciones por gérmenes oportunistas.

El virus VIH tiene un gran número de trucos que lo ayudan a evadir la defensa del organismo, que incluye una rápida y alta capacidad de mutación. Esto implica que una vez que el VIH está dentro, el sistema inmune nunca logra deshacerse totalmente de él.

La infección por VIH no presenta una sintomatología evidente. Una persona infectada por VIH puede sentirse perfectamente durante muchos años sin saber que está infectada. El problema más grave, desde el punto de vista epidemiológico y social, es que por cada individuo con la forma cadena de infección, existen de tres a cuatro personas, cuando menos, que están infectadas por el virus, que no tienen manifestaciones clínicas, que muchas de las veces ignoran estar infectadas, pero que sí son capaces de transmitir la infección a sujetos sanos. Las personas asintomáticas infectadas por el VIH son las mayores responsables de la diseminación de la epidemia, máxime cuando son ignorantes de su situación, de cuáles son los factores de riesgo y los mecanismos de transmisión de la enfermedad, dado que no adoptan cambios en su conducta sexual, ni en su actitud hacia la donación de sangre.

En esta etapa de la enfermedad, la infección sólo es detectable mediante exámenes de laboratorio que, en su mayoría, buscan anticuerpos contra el VIH en la sangre, cuya presencia indica la exposición a este virus. Las pruebas para investigar estos anticuerpos son de dos tipos y deben hacerse en forma secuencial. En primera instancia, en individuos con factores de riesgo, está indicada la realización de pruebas de escrutinio o tamizaje, la más acreditada es la prueba conocida con el acrónimo de ELISA, que corresponde a las siglas en inglés del método que se usa para su investigación (Enzyme Linked Immuno Sorbent Assay), pero que no es sinónimo – como popularmente se emplea – de un examen para detectar el SIDA. Las pruebas de escrutinio se emplean para descartar la infección, por lo que un resultado negativo permite afirmar, con una certeza mayor al 98%, que el

sujeto no está infectado por el VIH. El resultado positivo, sin embargo, lejos de considerarse como diagnóstico de infección, indica la necesidad de ratificar el resultado y, de ser repetidamente positivo, se procede a realizar una prueba confirmatoria. El resultado positivo de la prueba confirmatoria sí establece el diagnóstico de infección por el VIH. En ocasiones excepcionales, o circunstancias especiales, pueden requerirse otras pruebas – como la investigación de antígenos o genes del VIH– para establecer el diagnóstico de infección de manera definitiva (46).

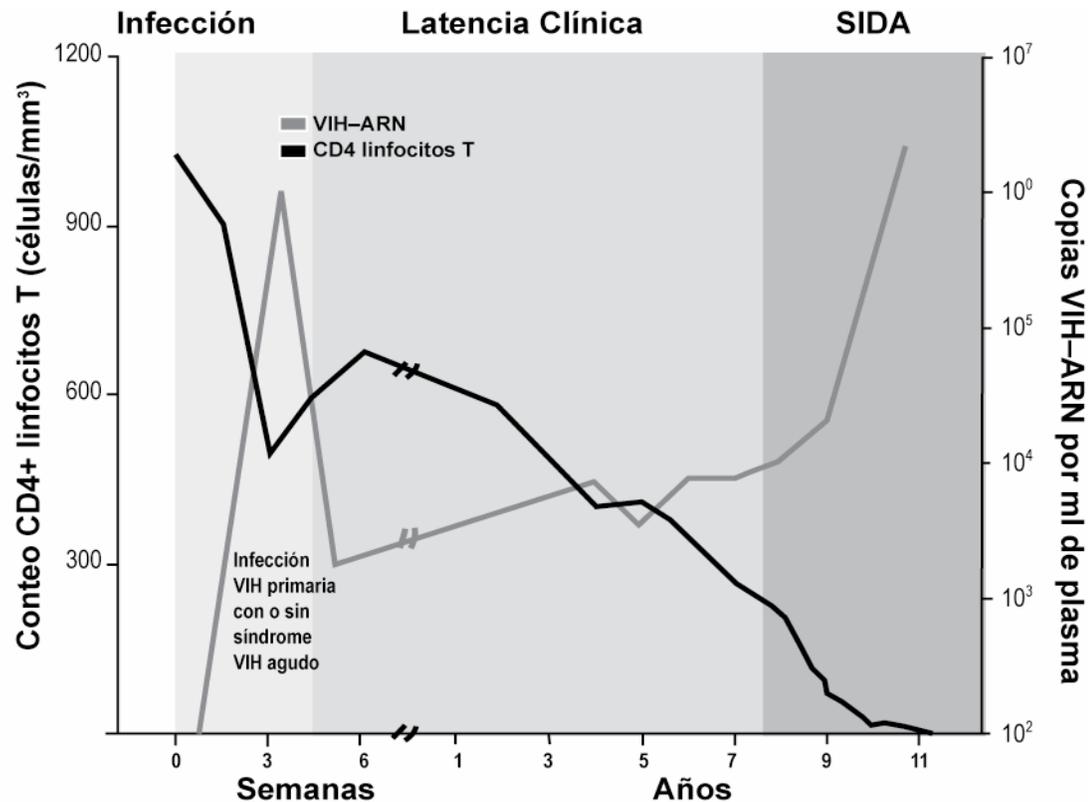


Figura 1.1. Historia Natural de la Infección por VIH

La gravedad de la enfermedad es determinada por la cantidad de virus en el cuerpo (carga viral) y el grado de supresión inmune (recuentos de CD4 descendentes). A mayor carga viral, más rápida es la supresión inmune. Hay dos formas en que el virus del SIDA ocasiona la muerte: indirecta (permitiendo las infecciones, el cáncer...) y directamente. La primera comprende infecciones oportunistas relacionadas con el VIH. Los gérmenes oportunistas se conocen a todos aquellos microorganismos de baja o nula virulencia que en las personas sanas causan, si acaso, infecciones banales; sin embargo, en estos pacientes, o en los afectados por otras inmunodeficiencias, producen formas graves de enfermedad.

Las infecciones que caracterizan al paciente con SIDA son la candidiasis en las vías respiratorias y el esófago, el *Herpes zoster* recurrente, la listeriosis, nocardiosis y criptococosis extrapulmonar, la coccidioidomicosis diseminada, la criptosporidiosis intestinal que dura más de un mes, la infección por el virus

citomegálico fuera del bazo, hígado y ganglios linfáticos, formas diseminadas de *Herpes simple*, histoplasmosis diseminada, infecciones por *Mycobacterium avium* y *Mycobacterium kansasii*, la neumonía por *Pneumocystis carinii*, y la toxoplasmosis cerebral. Aun cuando no es una infección oportunista, la tuberculosis se considera como una infección asociada al SIDA pues, antes del inicio de la pandemia, prácticamente ya se había erradicado en el ámbito mundial (46).

En un reciente artículo de la revista New York presuntuosamente titulado "La Última Palabra en lo que Concierna a Evitar el SIDA", el autor declara abiertamente: "La gente no muere del SIDA, muere de otras enfermedades contra las cuales no pueden luchar". Sin embargo, los investigadores saben que el virus del SIDA puede matar directamente, al afectar seriamente al cerebro, sin la presencia de ninguna infección oportunista y sin destruir el sistema inmunológico. El virus del SIDA por si sólo destruye las células del cerebro y causa la enfermedad progresiva de éste (Encefalopatía por el virus del SIDA). Los reportes médicos indican que la enfermedad cerebral producida por el SIDA es uno de los aspectos más importantes de la infección de esta enfermedad (46, 47).

Debido a que actualmente el VIH-1 es la principal causa de SIDA a nivel mundial, nuestra discusión se centra principalmente en la infección por VIH-1.

1.1.2 Estructura y morfología del VIH-1

El VIH-1 es un *retrovirus* que pertenece a la familia de los lentivirus (46, 48). En general, estos virus tienen una serie de características específicas que son determinantes en la compleja patogenia de la infección por el VIH:

- Gran diversidad genética (virus ARN) y genoma muy complejo (*lentivirus*)
- En su ciclo vital hay 2 fases: virión infectante (ARN) y provirus (ADN). Esta fase intermedia de integración en el genoma huésped le permite prolongados periodos asintomáticos (latencia), a pesar de una viremia persistente.
- Se replica mediante un mecanismo inverso al habitual en los virus ARN.
- Sus células huésped son los linfocitos CD4+, macrófagos, células nerviosas de la microglía y células dendríticas residentes en mucosas (células de Langerhans).

Las partículas virales del VIH-1 tienen un diámetro de aproximadamente 145 nm y están rodeadas por una membrana de lipoproteína que contiene varias partes: (i) una interna que contiene dos moléculas de ARN, la nucleoproteína y algunas enzimas virales, la transcriptasa inversa, la proteasa y la integrasa (TI, PR, IN); (ii) una intermedia, la cápside, constituida por un núcleo icosaédrico y (iii) una más externa que forma la envoltura.

Cada partícula viral contiene 72 complejos de glicoproteína que están integrados en esta membrana lipídica y cada uno está compuesto de trímeros de la glucoproteína externa gp120 y de la proteína transmembranal gp41, imprescindibles para la entrada del virus en la célula (49). La unión entre gp120 y gp41 no es fuerte, por tanto la gp120 se puede liberar espontáneamente y se puede detectar en el suero (50) así como también en el tejido linfático de los pacientes infectados por VIH (51).

La proteína de matriz p17 está anclada en el interior de la membrana lipoprotéica viral. El antígeno p24 de la cápside contiene dos copias del ARN del VIH-1. Este es parte de un complejo de proteína-ácido nucleico que está compuesto por la nucleoproteína p7 y la transcriptasa inversa p66. La partícula viral contiene todo el equipo enzimático necesario para la replicación: la TI p66, la IN p32 y la PR p11 (52).

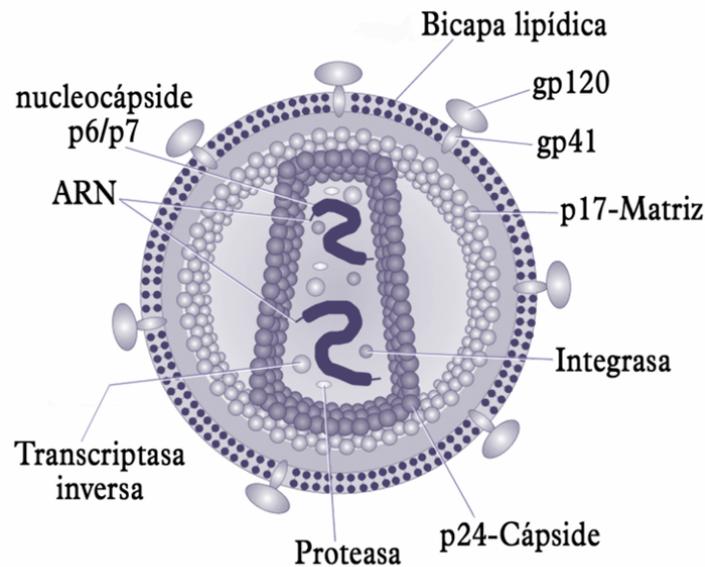


Figura 1.2. Estructura del VIH-1

1.1.3 Organización del genoma Viral

La mayoría de los retrovirus competentes para la replicación dependen de tres genes: *gag*, *pol* y *env*: **gag** significa “antígeno de grupo”, **pol** representa la “polimerasa” y **env** se refiere a “envoltura” (Figura 1.3.). El esquema estructural “clásico” del genoma retroviral es: 5’LTR-*gag-pol-env*-LTR 3’. Las regiones LTR (“*long terminal repeat*”) son los dos extremos del genoma viral, que están conectados con el ADN de la célula huésped después de la integración y no codifican para ninguna proteína viral. Los genes *gag* y *env* codifican para la nucleocápside y para las glucoproteínas de la membrana viral; el gene *pol* codifica para la transcriptasa inversa y para otras enzimas. Además, el VIH-1 contiene seis genes (*vif*, *vpu*, *vpr*, *tat*, *rev* y *nef*) en su RNA de 9kB, que contribuyen a su complejidad genética (53).

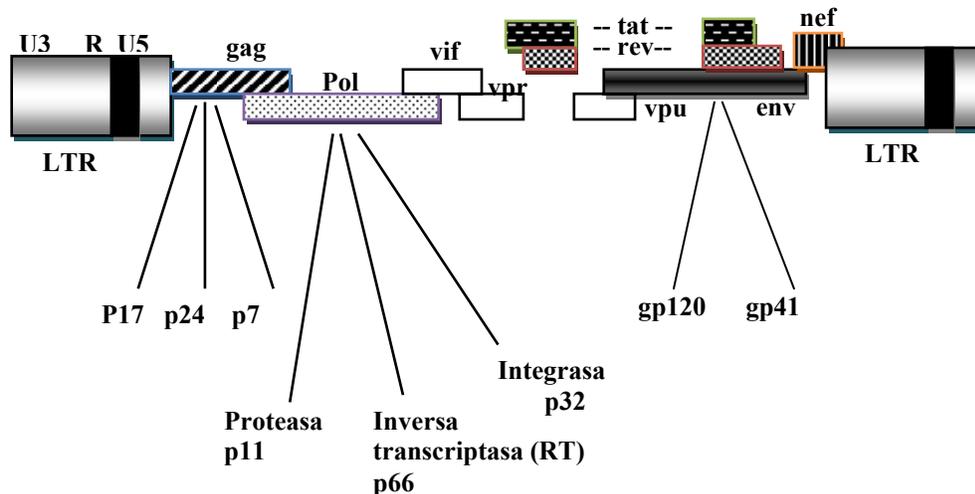


Figura 1.3. Organización del genoma del VIH-1

1.1.4 Ciclo de replicación

- i) *Absorción, fusión e internalización del virión en el citoplasma celular*: Mediante la interacción de las proteínas de la envoltura (gp120 y gp41) con el receptor celular CD4 y el correceptor de la célula T del VIH-1.
- ii) *Transcripción reversa del genoma ARN en ADN* por la enzima transcriptasa inversa
- iii) *Translocación al núcleo celular e integración del genoma viral en la célula* por la enzima integrasa.
- iv) *Expresión temprana de genes reguladores (Tat, Vpu, Nef)* durante la transcripción de genes celulares.
- v) *Expresión tardía de genes estructurales* en forma de precursores poliproteicos (*gag*, *gag-pol* y *env*), las glicoproteínas de superficie (SU) y transmembranales (TM).
- vi) *Salida del virión y maduración proteica completa* de *gag* y *gag-pol* para dar lugar a las diferentes proteínas estructurales, proceso llevado a cabo por la proteasa dando lugar a viriones con completa capacidad infectiva (54).

1.1.5 Capacidad mutagénica y resistencia del VIH

La mutación genética es una característica inherente a todos los virus ARN y el caso del VIH-1 no es una excepción. El desarrollo rápido de las variantes resistentes se debe al elevado recambio del VIH - cada día se producen aproximadamente 10 millones de partículas virales nuevas (55). También se debe a la tasa de error excepcionalmente alta de la transcriptasa inversa del VIH. Esto causa una tasa de mutación elevada y una producción constante de cepas virales nuevas, aún en ausencia de tratamiento. En presencia de medicamentos antirretrovirales las cepas resistentes son seleccionadas como especies dominantes (56). El desarrollo de cepas virales resistentes es una de las principales razones del fracaso de la terapia antirretroviral.

La relevancia clínica de las mutaciones primarias de resistencia al inicio del tratamiento es evidente. La prevalencia de mutaciones en pacientes sin tratamiento previo varía entre las diferentes regiones demográficas. Las mutaciones de resistencia transmitidas pueden limitar las opciones de tratamiento en el futuro. A principios

del 2005, un paciente en New York sorprendió a todo el mundo por haberse infectado con un tipo de virus que contiene 7 mutaciones relevantes para INTIs, 2 mutaciones para INNTIs y 12 mutaciones para fármacos inhibidores de la enzima Proteasa (IP). Después de 4 a 20 meses (no se sabe exactamente el tiempo de infección), el número de linfocitos CD4 decreció a 80 célula/ μ L. La capacidad de replicación del virus resistente fue comparable con el virus tipo silvestre. Solamente 2 fármacos tipo ART, el T-20 (Enfuvirtide) y el efavirenz continuaron siendo activos. Aunque la transmisión de virus multi-resistentes y la progresión clínica tan rápida son eventos raros, este caso reportado es una muestra de las consecuencias posibles de la resistencia primaria a múltiples fármacos (57).

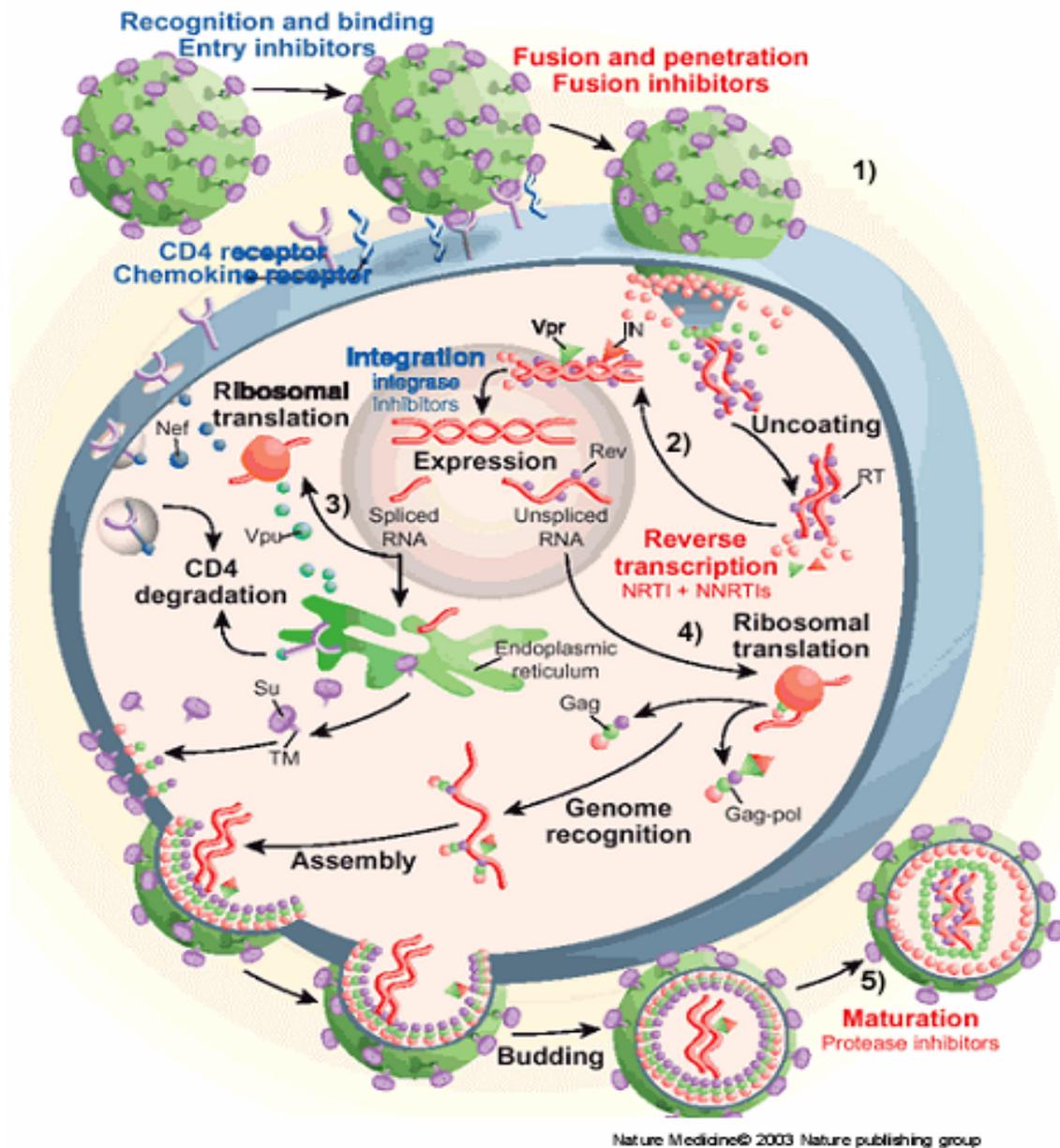


Figura 1.4. Ciclo replicativo del VIH-1.

En rojo figuran los inhibidores de replicación viral. En azul están indicados los inhibidores del ciclo de replicación en fases de desarrollo clínico o preclínico.

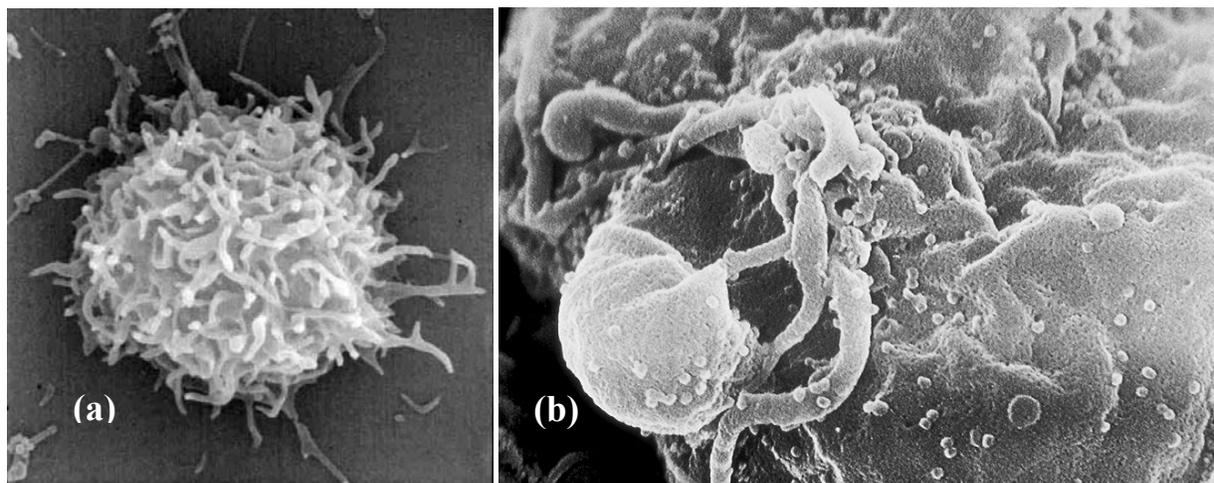


Figura 1.5. Micrografía electrónica de una célula T no infectada (a) e infectada crónicamente (b) por el VIH-1 (IIB).

1.2 TERAPIA ANTI-VIH

Los medicamentos contra el VIH (conocidos también como medicamentos antirretrovirales) se usan para controlar la reproducción del virus y para detener la progresión de la enfermedad relacionada con la infección por el VIH. La terapia antirretroviral altamente activa (TARAA), es la recomendada para la infección por el VIH. La TARAA combina tres o más medicamentos contra el VIH en un régimen diario. Por supuesto los medicamentos contra el VIH no curan la infección del mismo y las personas que toman estos medicamentos pueden todavía transmitir el VIH a otras personas. Además, están surgiendo nuevos problemas relacionados con la toxicidad a corto y a largo plazo de los tratamientos con medicamentos y la frecuencia de mutaciones de resistencia tanto en los virus circulantes como en los transmitidos. El curso de esta pandemia también dependerá de la disponibilidad de una vacuna efectiva de profilaxis en el futuro próximo.

Los fármacos antirretrovirales empleados en la terapia anti-VIH pueden dividirse en cuatro familias: (i) los Inhibidores de la entrada (ii) los Inhibidores de la Transcriptasa Inversa con dos subfamilias; los análogos nucleosídicos (INTIs) y los no nucleosídicos (INNTIs); (iii) Inhibidores de la integrasa (INs) y los (iv) los Inhibidores de la Proteasa (IPs), incorporados recientemente a las terapias antirretrovirales. La parte del ciclo viral sobre la que actúa cada uno de ellos aparece representada en la Figura 1.6. Simultáneamente, las mutaciones de resistencia también pueden clasificarse atendiendo al tipo de fármaco antirretroviral sobre el que se ha desarrollado la resistencia.

Los "Medicamentos para el tratamiento de infección por el VIH" (2008) aprobados por la FDA, están expuestos en la página web <http://www.fda.gov/oashi/aids/virals.html>.

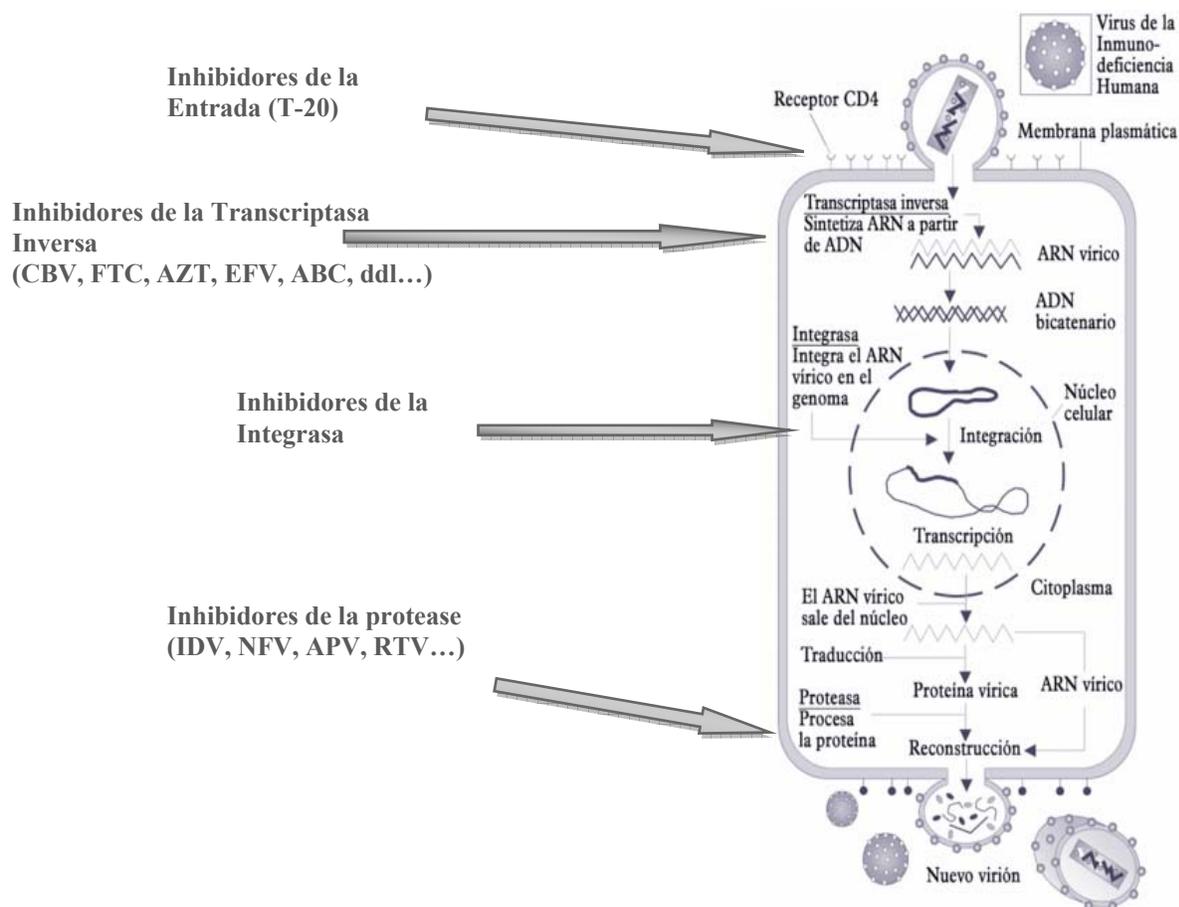


Figura 1.6. Principales blancos de los fármacos anti-VIH

1.2.1 La Transcriptasa inversa (TI) del VIH, blanco biológico principal de los fármacos antiretrovirales

Después de la entrada del virión VIH-1 en la célula, las partículas del núcleo viral son liberadas y el ARN viral es convertido en una doble cadena lineal de ADN por la transcriptasa inversa viral. Una vez que este ADN fue duplicado y circularizado se integra en los cromosomas de la célula huésped. Este proceso representa un estadio crítico en el ciclo vital del virus. El complejo ADN-núcleo viral (ó complejo de pre-integración) es muy grande para pasar de modo simple por los poros de la membrana nuclear en el proceso de **importación al núcleo** celular. Para el VIH-1, la importación al núcleo celular es un proceso activo, que depende de la presencia de **señales de localización nuclear** dentro de algunas proteínas del virión. Esto da como resultado que los VIH-1 pueden infectar las células que no se dividen como los macrófagos.

La transcriptasa inversa (**una enzima multifuncional**) es una ADN polimerasa, dependiente del ARN, que sintetiza ADN a partir de una plantilla de ARN. También posee actividad hidrolasa, lo que le permite degradar el ARN de los híbridos ARN. La estructura cristalina de la TI ha sido estudiada y se han logrado producir varios inhibidores específicos de ésta enzima. Una de las propiedades más importantes de la TI es que

no tiene actividad de corrección de lectura. Por ende, no puede corregir los errores que hace durante la síntesis de ADN, y por tanto la replicación del VIH-1 tiene una *frecuencia alta de mutaciones*.

Dicha enzima es un heterodímero formado por dos subunidades, una subunidad de 66 KDa y una subunidad de 51 KDa (denominadas p66 y p51, respectivamente). En la subunidad p66 residen las actividades polimerasa y RNasa H mientras que la subunidad p51 actúa como soporte estructural de la subunidad p66, estabilizando la conformación activa de dicha subunidad facilitando así la unión del ADN a la misma.

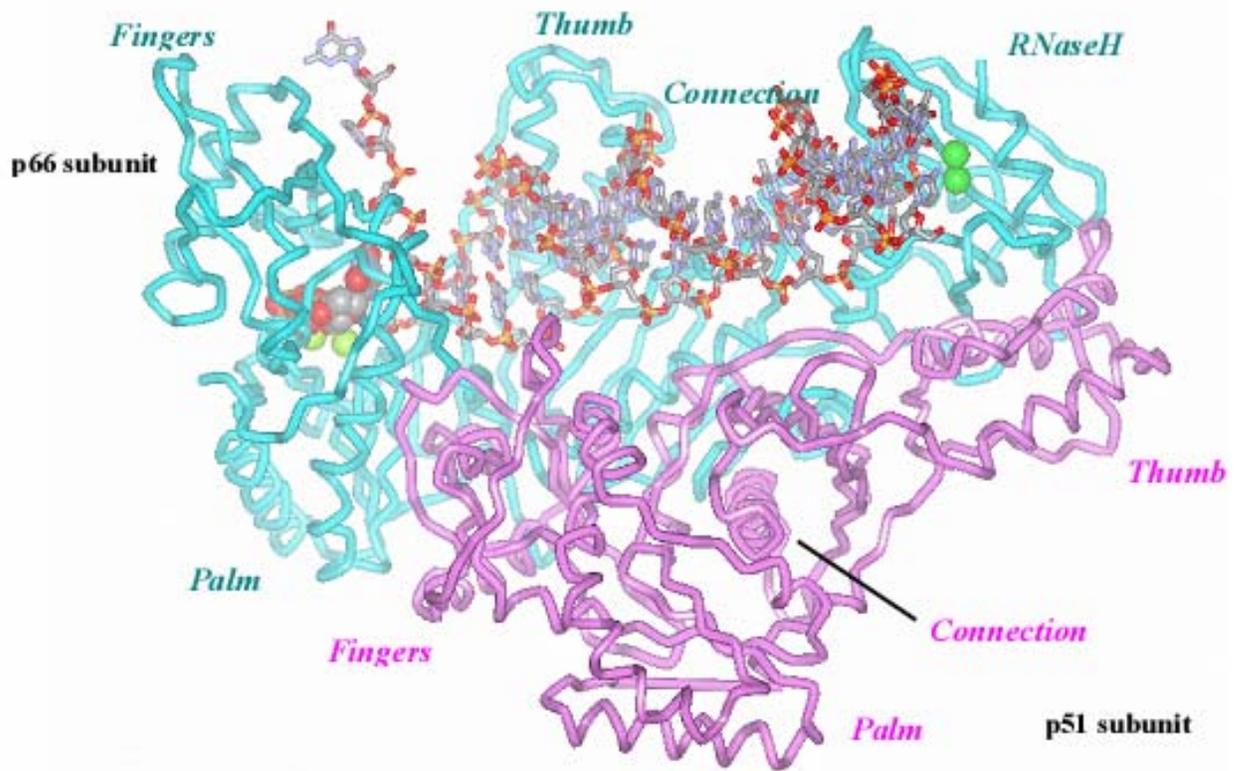


Figura 1.7. Estructura cristalina de la TI del VIH-1

El dominio polimerasa de la subunidad p66 puede ser comparado anatómicamente con una mano derecha, y está formado a su vez por cuatro subdominios, denominados dedos, palma, pulgar y conexión (ver Figura 1.7.). El centro activo, (sitio de unión del sustrato) de esta subunidad, localizado en el dominio de la palma, está formado por una tríada catalítica de residuos de ácido aspártico (Asp 110, Asp 185, Asp 186). Se ha propuesto que al menos dos iones Mg^{2+} asisten al proceso de polimerización, coordinándose con los carboxilos de los restos de Asp del centro activo (58, 59) (ver Figura 1.8.).

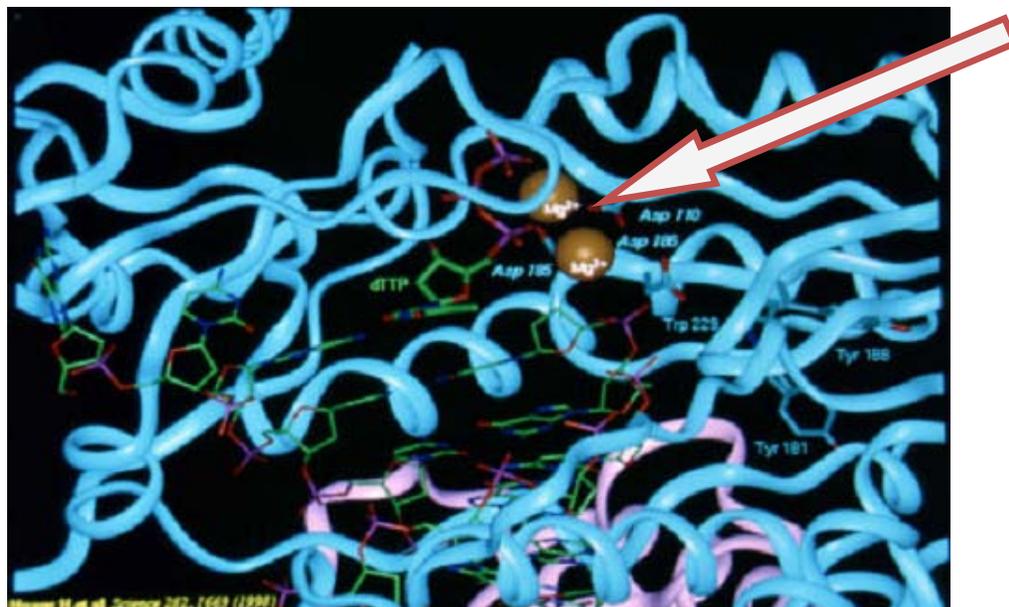


Figura 1.8. Centro catalítico de la TI del VIH-1

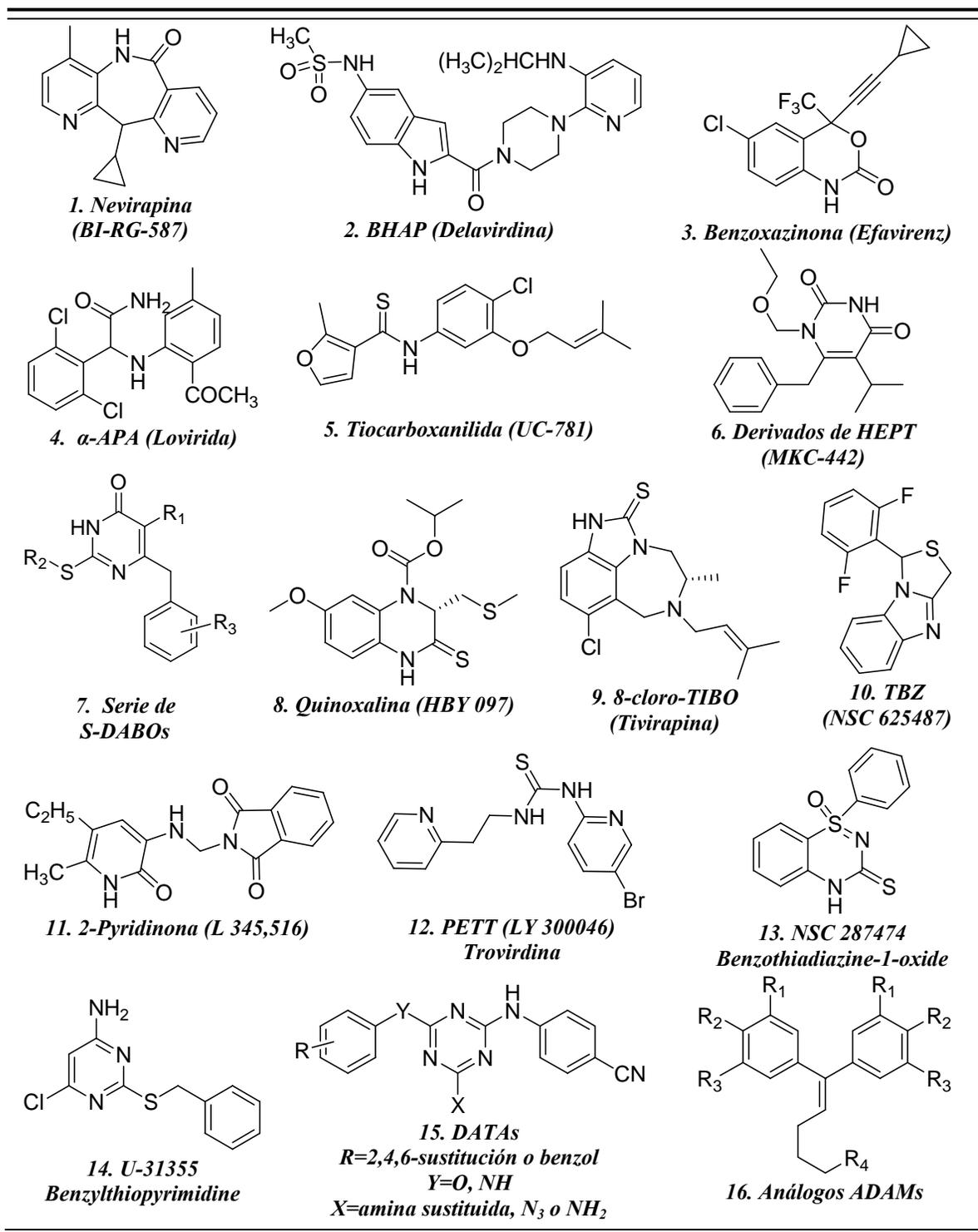
1.2.2 Fármacos Inhibidores No Nucleosídicos de la Transcriptasa Inversa (INNTI)

La enzima TI de VIH-1 es un blanco atractivo para desarrollo de nuevos fármacos en el proceso de la lucha contra el SIDA. En la investigación, los Inhibidores específicos no nucleosídicos de la transcriptasa inversa (INNTI) han obtenido un importante y definitivo lugar debido a su propia potencia antiviral, alta especificidad y baja toxicidad.

Desde 1990 y como resultado de amplios programas de screening, se han descrito más de 30 familias de compuestos que son inhibidores potentes y altamente específicos de la TI del VIH-1 (60-64). Dichos compuestos, a diferencia de los 2',3'-Didesoxinucleósidos (2',3'-ddN)- inhibidores de la transcriptasa inversa análogos de nucleósidos (INTI), no afectan a la TI del VIH-2 o a otras ADN polimerasas. Por otro lado, no inhiben a la TI con un sustrato alternativo, sino que se unen directamente y de manera no competitiva con la enzima. El complejo que resulta bloquea el sitio de unión, activado por catalizador, de la TI, el cual entonces se puede unir con menos nucleósidos y la polimerización se hace significativamente más lenta. En contraste con los INTIs, los INNTIs no necesitan de activación dentro de las células.

Algunos de estos compuestos tales como la Nevirapina (BI-RG-587, 1), el BHAP (Delavirdina, 2) y la Benzoxazinona (Efavirenz, 3) están aprobados para el tratamiento clínico del SIDA (65) en terapias combinadas con ddN y/o inhibidores de proteasa (60, 62, 65). Otros se encuentran en fases clínicas avanzadas como el a-APA (Lovirida, 4), la Tiocarboxanilida (UC-781, 5), el derivado de la familia HEPT (MKC-442, 6), derivados de serie S-DABOs, la Quinoxalina (HBY 097, 8) y el 8-Cloro-TIBO (Tivirapina, 9) (66-68) (ver Tabla 1.2.).

Tabla 1.2. Líderes tipo INNTI del VIH-1 aprobados para el tratamiento del SIDA o en fases clínicas avanzadas



En los últimos años se han descrito las estructuras cristalográficas tanto de la apoenzima como de la enzima formando complejos con distintos inhibidores no-nucleosídicos. Estas estructuras han revelado como se forma el sitio de unión para los INNTIs y las sutiles pero importantes diferencias en su geometría “butterfly-like”

(69) dependiendo de la naturaleza del inhibidor determinada por los estudios de acoplamiento y dinámica molecular entre la TI y sus ligandos.

El sitio de unión de los INNTIs es una cavidad hidrófoba, flexible y próxima (aunque distinta) al centro catalítico de la enzima (aproximadamente a 10-15 Å del sitio activo de la polimerasa). Este bolsillo no existe como tal cuando el inhibidor no está interaccionando con la enzima. La flexibilidad y tamaño de esta cavidad permite que compuestos estructuralmente tan diferentes ejerzan un mismo tipo de inhibición, una inhibición de tipo alostérico.

Estudios tanto estructurales (70) como cinéticos (59) revelan que la unión de los INNTIs a la TI hacen que la cadena molde/cebador (*template/primer*) se recolocque en la proteína dando lugar a un distanciamiento de aproximadamente 2 Å entre el canal de unión y el sitio activo. Esto provoca una marcada disminución de la actividad enzimática. Por tanto, los inhibidores específicos inhiben la TI de VIH-1 debido a una distorsión de la tríada catalítica de aspartato (sitio activo) en el dominio polimerasa.

Los INNTIs se caracterizan por su baja toxicidad, su alta hidrofobicidad y su elevada potencia antiviral, lo que les convierte en atractivos agentes anti-VIH. Sin embargo, al igual que los INTIs inducen el rápido desarrollo de cepas de virus resistentes. El riesgo de resistencia cruzada es muy alto y ésta puede desarrollarse muy rápidamente; Una mutación puntual en la posición 103 (K103N) del sitio de unión hidrofóbico es suficiente para eliminar a la clase completa de fármacos (71).

Tabla 1.3. Mutaciones más comunes en aislados clínicos, implicados en resistencia del VIH-1 a INNTIs.

Mutaciones de la TI que confieren resistencia a los NNIs	Comentarios
K103	Mutación normalmente seleccionada en terapias con efavirenz. Ocasionalmente seleccionada en terapias con nevirapina. Confiere una elevada resistencia a todos INNTIs
V108I Y181C Y188C	Mutaciones seleccionadas en presencia de nevirapina. Confiere un alto nivel de resistencia a nevirapina pero bajo a efavirenz. Y188L, a diferencia de Y188C, se observa principalmente en terapias con efavirenz.
L100I V106A G190A, G191S	Mutaciones que se acumulan durante terapia prolongada infectiva a la mayor parte de los INNTIs.

En particular, como un tema que está en proceso de desarrollo, los INNTIs conocidos y/o nuevos encontrados se están ensayando tanto "in vitro" como "in vivo". Se conocen muchos fármacos que tienen la capacidad de inhibir al VIH, aunque pocos son capaces de demostrar una eficacia clínica que haga aconsejable su empleo en los pacientes con VIH/SIDA debido principalmente al nivel de toxicidad mostrado sobre células

sanguíneas, mayormente del tipo MT4 y CEM (15, 16). Entre ellos los que se tiene mayor experiencia clínica se encuentran los análogos de HEPT, DATAs, N-acylthiocarbamate (ATC) y los S-DABOs. Ninguno de los análogos antes citados son capaces de inhibir o inactivar por completo la expresión de VIH “in vitro”, como lo pone de manifiesto los ensayos realizados sobre células MT4, línea de células linfoides T transformadas con HTLV-1 (Human T-Lymphotropic Virus-1), en las que se ha evaluado (i) el efecto citopatogénico en dichas células cuando se sometían a una sobre-exposición con VIH tratado con una composición antiviral de las proporcionadas por esta invención y (ii) la expresión de VIH medida en los sobrenadantes de los cultivos celulares como expresión de actividad transcriptasa reversa y como detección y medida del antígeno P25. El antígeno P25 se libera cuando se produce la lisis del VIH, por lo que puede detectarse mediante una reacción de complejamiento con un anticuerpo anti-P25 y a partir de ahí evaluar la cantidad de virus (15, 16, 72).

Aunque se han ido aportando soluciones y se prevén nuevas alternativas, sigue existiendo la necesidad de encontrar un agente antiretroviral que sea efectivo y carezca de los efectos tóxicos indeseables previamente citados.

1.3 TENDENCIAS ACTUALES

Comparado con la rápida aparición (semanas o meses) de cepas virales resistentes a los diferentes fármacos empleados en monoterapias, el virus se puede suprimir durante largos períodos de tiempo (meses o incluso años) cuando se utilizan terapias combinadas (politerapias) en lo que hoy se conoce como terapia antiretroviral de alta eficacia (TARAA) (73, 74).

Tabla 1.4. Fármacos (no combinados) INNTIs aprobados por la Administración de Alimentos y Medicamentos EE.UU (FDA)

Nombre genérico	Marca y otros nombres	Fabricante	Fecha de aprobación por el FDA
Delavirdina	Rescriptor [®] ; DLV	Pfizer	Abril 4, 1997
Efavirenz	Sustiva [®] , EFV	Bristol-Myers Squibb	Sept. 17, 1998
Entravirina	Intelence [®] , Celsentri [®] , TMC125, ETR	Biotec	Enero 18, 2008
Nevirapina	Viramune [®] , NVP	Boehringer Ingelheim	Junio 21, 1996

La reducción en la carga viral, conseguida en pacientes tratados en fases tempranas de la enfermedad con politerapia, hace que la enfermedad no progrese, ya que implica un menor número de linfocitos T infectados y permite, por tanto, una recuperación del sistema inmune(75, 76). Por tanto, hoy en día es evidente que, al igual que en la quimioterapia del cáncer y de algunas Enfermedades bacterianas, la estrategia a seguir en la lucha contra el SIDA es la terapia combinada (77, 78). Una combinación racional de fármacos puede dar lugar, por un lado a una menor toxicidad, ya que se emplean concentraciones inferiores de los fármacos que las empleadas

individualmente, por otro a la prevención o retraso en la aparición de resistencias y por último puede aumentar la eficacia de los agentes terapéuticos debido a efectos sinérgicos o al menos aditivos (79).

Sin embargo, el empleo de terapias combinadas no está exento de problemas. El primero de ellos es que a pesar de que en muchos casos la carga viral es indetectable, el virus “se esconde” en ciertos reservorios del organismo (linfocitos CD4⁺, macrófagos, células dendríticas, vasos linfáticos, etc.), donde está a salvo de la acción de los fármacos antiretrovirales, manteniendo su capacidad de replicarse y por tanto de mutar y desarrollar resistencias frente a dichos fármacos.

Además, no siempre la combinación de fármacos da lugar a beneficios terapéuticos. Existen numerosos ejemplos que muestran el efecto antagonista, el aumento de toxicidad, la interferencia metabólica o incluso una aceleración en el desarrollo de resistencias, cuando se combinan ciertos fármacos (79). Por tanto, la elección y combinación de fármacos debe hacerse de un modo racional.

Otro de los problemas de la terapia combinada surge a la hora de decidir cuál es el cóctel de fármacos más adecuado. A medida que aumenta el número de compuestos aprobados para el tratamiento del SIDA, aumentan las posibilidades de combinación y, por tanto se hace más difícil elegir la mejor combinación.

Los estudios de resistencia cruzada, utilizando el arsenal de compuestos anti-VIH/SIDA hoy disponibles, y la caracterización de cepas mutantes, aportan una información muy valiosa para el diseño de estas terapias combinadas.

La elección de la combinación de fármacos anti-VIH se realiza, de un modo racional, siguiendo distintas aproximaciones (77, 78, 80-82). (i) La combinación de compuestos que se unen a distintos sitios de la misma proteína viral, por ejemplo, dentro de los inhibidores de la TI, se combinan análogos de 2',3'-ddN con INNTIs. (ii) La combinación de compuestos que actúan sobre enzimas implicadas a diferentes niveles del ciclo replicativo del virus, como por ejemplo, la combinación de INNTIs e IPs. (iii) Se han diseñado terapias combinadas que incluyen compuestos que actúan sobre un mismo punto de la enzima pero que presentan un perfil de sensibilidad/resistencias complementario.

El éxito o el fracaso de la terapia anti-VIH/SIDA está determinado por la interrelación de tres factores, como son el virus, el tratamiento y el enfermo. Esto hace que sea inmensamente complicado predecir el resultado de un tratamiento y hacer extensible a otros pacientes aquel tratamiento que tuvo éxito con un paciente determinado.

Los avances de la comunidad científica en el conocimiento, prevención y lucha contra el SIDA han sido espectaculares, sin embargo, quedan todavía muchas metas por alcanzar. La aparición de una tercera cepa de VIH (83), no detectable por los métodos de detección empleados en la actualidad, ha complicado el panorama y ha puesto de manifiesto la importancia de seguir investigando en este campo. Asimismo, se ha demostrado la transmisión de cepas de virus resistentes (84) a los fármacos que se emplean en clínica. Esto indica que el “arsenal” actual del que disponemos no es suficiente para controlar la replicación viral y luchar así contra la enfermedad, por tanto, es necesario el diseño de fármacos más potentes y de menor toxicidad en nuestra lucha contra el SIDA.

1.4 QUIMIOINFORMÁTICA, QUIMIOMETRÍA Y DESARROLLO RACIONAL DE FÁRMACOS

1.4.1 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, en este caso es la TI, y su mecanismo que conduce a un proceso patológico.
- segundo paso El propósito es identificar una molécula (usualmente moléculas pequeñas) 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.
- Tercer paso (validación clínica) debe probar su eficiencia y seguridad a través de una evaluación extensiva en animales y humanos(85-87).

Durante los ensayos biológicos iniciales del ciclo de descubrimiento de fármacos son identificados los denominados *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. 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(88).

Vale la pena señalar que la investigación terapéutica es un proceso en extremo complejo y costoso en términos de tiempo y dinero. Un estudio reciente que incluyó una muestra de 68 fármacos aprobados en la década de los 90's en los Estados Unidos de América, mostró que el tiempo y el costo promedio para el desarrollo de un fármaco, desde la identificación del blanco hasta la obtención de la aprobación por la FDA, estaba alrededor de los 15 años y US\$ 900 millones(22). En las pasadas dos décadas 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.

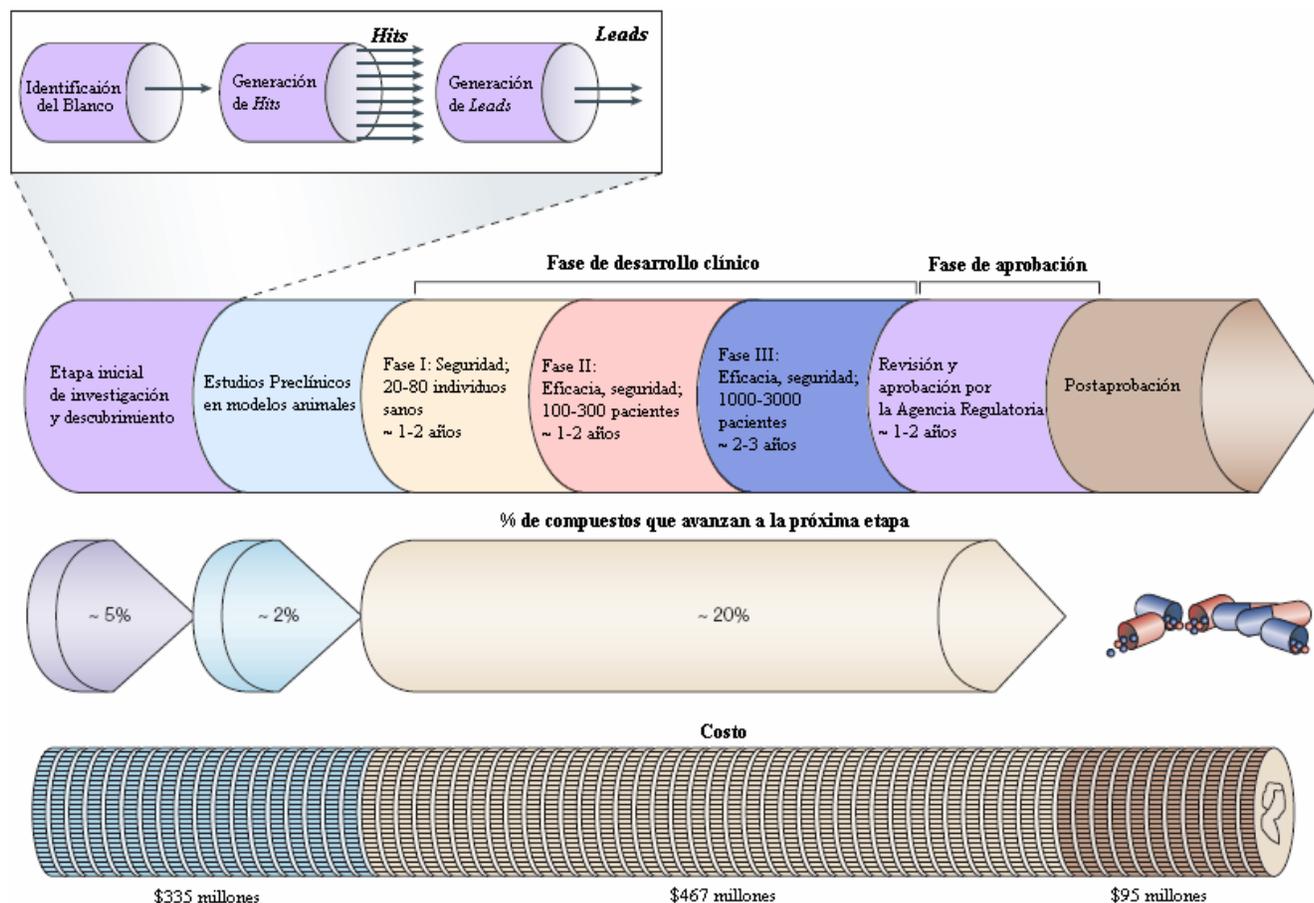


Figura 1.9. Representación esquemática del proceso de descubrimiento y desarrollo de fármacos.

1.4.2 Tecnologías de alto rendimiento (*High-throughput technologies*)

A partir de esta sección los elementos desarrollados se centran en el problema del descubrimiento de fármacos como tal, esto es, la identificación de candidatos clínicamente promisorios para un blanco terapéutico validado. Hasta los años 80, el paso de generación de *hits* constituía el principal cuello de botella del proceso de descubrimiento de fármacos debido al costo (en ambos, tiempo y dinero) de la síntesis y evaluación de nuevas moléculas(18, 19). Los avances en las áreas de miniaturización y robotización en los años 80 condujeron a la aparición de dispositivos de alta capacidad de ensayo o *high-throughput screening (HTS)*, permitiendo la evaluación masiva en paralelo de cientos o miles de moléculas. Hoy en día, los dispositivos *ultra-HTS* permiten a las grandes compañías farmacéuticas el procesamiento de hasta 100 000 moléculas por día(89, 90).

En respuesta a estas nuevas tecnologías, la investigación química desarrolló al mismo tiempo métodos para sintetizar moléculas en paralelo, permitiendo así la generación de grandes librerías de moléculas. En

particular, la química combinatoria hace posible la generación de un gran número de moléculas a partir de la combinación de bloques estructurales elementales a un costo relativamente bajo(91, 92).

En la práctica sin embargo, pronto se evidenció que este método a no gran escala era la respuesta al problema del descubrimiento de fármacos. 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 cuello de botella del descubrimiento de fármacos no era la generación de *hits*, sino los pasos de identificación y optimización del compuesto líder (18, 85). Esta observación condujo progresivamente a una reconsideración y racionalización progresiva del proceso de descubrimiento de fármacos, en la cual los métodos computacionales han venido a desempeñar un papel determinante(21).

1.4.3 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 mezcla de los métodos computacionales y la química, F. K. Brown acuñó en 1998 el término *quimioinformática* definiéndola como “la mezcla 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”(93).

El campo de aplicación de los métodos quimioinformáticos en la investigación terapéutica es muy amplio. En esta sección, se introducen los tres principales pasos en el camino a transitar hacia la racionalización del proceso de descubrimiento de fármacos. Para una revisión más extensa de los métodos quimioinformáticos el lector puede consultar las referencias(94, 95).

Diseño de librerías. 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 librerías 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 librerías combinatorias hacia un proceso de diseño de librerías dirigido a que las mismas sean más pequeñas pero que a la vez contengan más información(21, 90, 96).

Debido a que el objetivo de los ensayos de evaluación evoluciona durante el proceso de descubrimiento de fármacos, el diseño de las librerías 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 librerías diversas de manera que se cubra lo más ampliamente posible el espacio

químico(19). Por el contrario, en el paso posterior “*hits to lead*”, deben ser diseñadas librerías dirigidas constituidas por moléculas estructuralmente similares a los *hits* identificados(21).

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 aun sin sintetizar e incluso moléculas 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(21, 90). 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(88, 97). Los algoritmos de *docking* consisten en dos componentes centrales: un algoritmo de búsqueda que se encarga del posicionamiento del ligando, y una función de evaluación que cuantifica la fuerza de la unión resultante en términos de energía a través de la evaluación de las fuerzas intermoleculares. El algoritmo de búsqueda necesita optimizar la configuración tridimensional del ligando, su conformación, con respecto a las estructuras espaciales y electrostáticas del sitio activo. El enfoque del tamizaje virtual basado en *docking* consiste en acoplar todas las moléculas de una base de datos y seleccionar aquellas con más alta puntuación como las moléculas activas más prometedoras. Debido a la presencia de enlaces rotacionales, las moléculas usualmente poseen un gran número de conformaciones por lo que el acoplamiento puede ser un proceso realmente complejo en la práctica(98-100).

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(88). Este método es esencialmente similar a los algoritmos de agrupamiento y particionamiento empleados para la construcción de librerías dirigidas, que fueron descritos en la sección anterior. Sin embargo, este enfoque es más general ya que el simple criterio de proximidad en el espacio químico puede reemplazarse por esquemas de ordenamiento más evolucionados que consideren simultáneamente todo el conjunto de compuestos activos(101). Alternativamente, el conjunto de compuestos activos puede usarse para derivar un *modelo farmacóforo* que representa un arreglo tridimensional del fragmento molecular responsable de la actividad (o al menos se cree que lo es). Este modelo farmacóforo puede entonces usarse como un filtro para eliminar del conjunto de compuestos aquellos que no cumplan con las condiciones de actividad necesarias(88, 102).

Debido al hecho de concentrarse solamente en los compuestos activos, las anteriores aproximaciones pueden ser criticadas por el hecho de no aprovechar al máximo los resultados de los ensayos experimentales(21). 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, la cual es ignorada por las aproximaciones anteriores. 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 las maquinarias de aprendizaje (*machine learning*). Debido a que este tipo de estudio constituye el pilar fundamental de los resultados presentados en este trabajo, el mismo se abordará con más profundidad más adelante.

Finalmente, es posible notar una fuerte sinergia entre el tamizaje virtual y los ensayos experimentales: los ensayos experimentales suministran datos válidos para construir o refinar los modelos provenientes del tamizaje virtual, los que a su vez optimizan los ensayos experimentales. La combinación del tamizaje virtual y los ensayos de alta capacidad en una forma iterativa también conduce a un proceso de evaluación secuencial(21, 90, 103). Mientras esta sinergia es particularmente obvia en las aproximaciones basadas en ligandos, este esquema secuencial puede también ser útil para refinar los modelos de acoplamiento, donde se sabe que la optimización de la función de evaluación es una tarea difícil en la práctica(97).

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(18-21). 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(18, 21). 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(18, 22).

Un ejemplo decisivo de la integración temprana del concepto *drug-likeness* lo constituye el trabajo de Lipinski y colaboradores quienes propusieron un conjunto simple de reglas, conocida como “reglas del cinco” definida originalmente en inglés como “*rules of five*”. Estas reglas relacionan la propensión de una molécula a ser absorbida por el organismo con propiedades moleculares generales tales como el peso molecular y el número de donadores y aceptores de enlaces de hidrógeno (104). Estas reglas se usan comúnmente como un filtro para remover compuestos durante el proceso de descubrimiento de fármacos. Generalmente es poco probable que una

molécula sea absorbida si viola al menos dos de estas reglas. Filtros similares han sido propuestos para considerar otras propiedades involucradas en el concepto *drug-likeness*, en particular en el contexto de la toxicidad, donde moléculas potencialmente tóxicas pueden ser identificadas mediante la detección de toxicóforos, esto es, grupos reactivos responsables de la toxicidad de la molécula(21, 90).

Considerando el enfoque REA involucrado en el tamizaje virtual, los modelos estadísticos y basados en maquinarias de aprendizaje que correlacionan directamente la estructura de las moléculas con sus propiedades “*drug-like*” son también útiles en este contexto(20, 105). 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 multi-propiedades más general(17-19), si se considera que los términos tamizaje virtual y REA a menudo ignoran este significado más amplio.

Finalmente se debe resaltar que estos métodos son altamente complementarios dentro del proceso de identificación de candidatos líderes con un perfil terapéutico global (farmacológico, farmacocinético y toxicológico) promisorio. La Figura 1.10 ofrece una visión esquemática de la posible integración de estos métodos en el proceso de descubrimiento de fármacos en los casos en que la estructura del blanco terapéutico es desconocida.

Las moléculas con pobres propiedades “*drug-like*” son removidas inicialmente del pool de compuestos disponibles mediante el uso de modelos de relación cuantitativa estructura-propiedad (*Quantitative Structure-Property Relationships, QSPR*) y/o filtros para propiedades “*drug-like*”. Un pool de tamizaje primario es subsecuentemente definido a través de la elección de un subconjunto diverso de candidatos. El resultado del tamizaje facilita la construcción de modelos QSAR, los cuales en conjunción con modelos QSPR, pueden ser usados en un tamizaje secuencial para el descubrimiento de compuestos líderes con un perfil farmacéutico global promisorio.

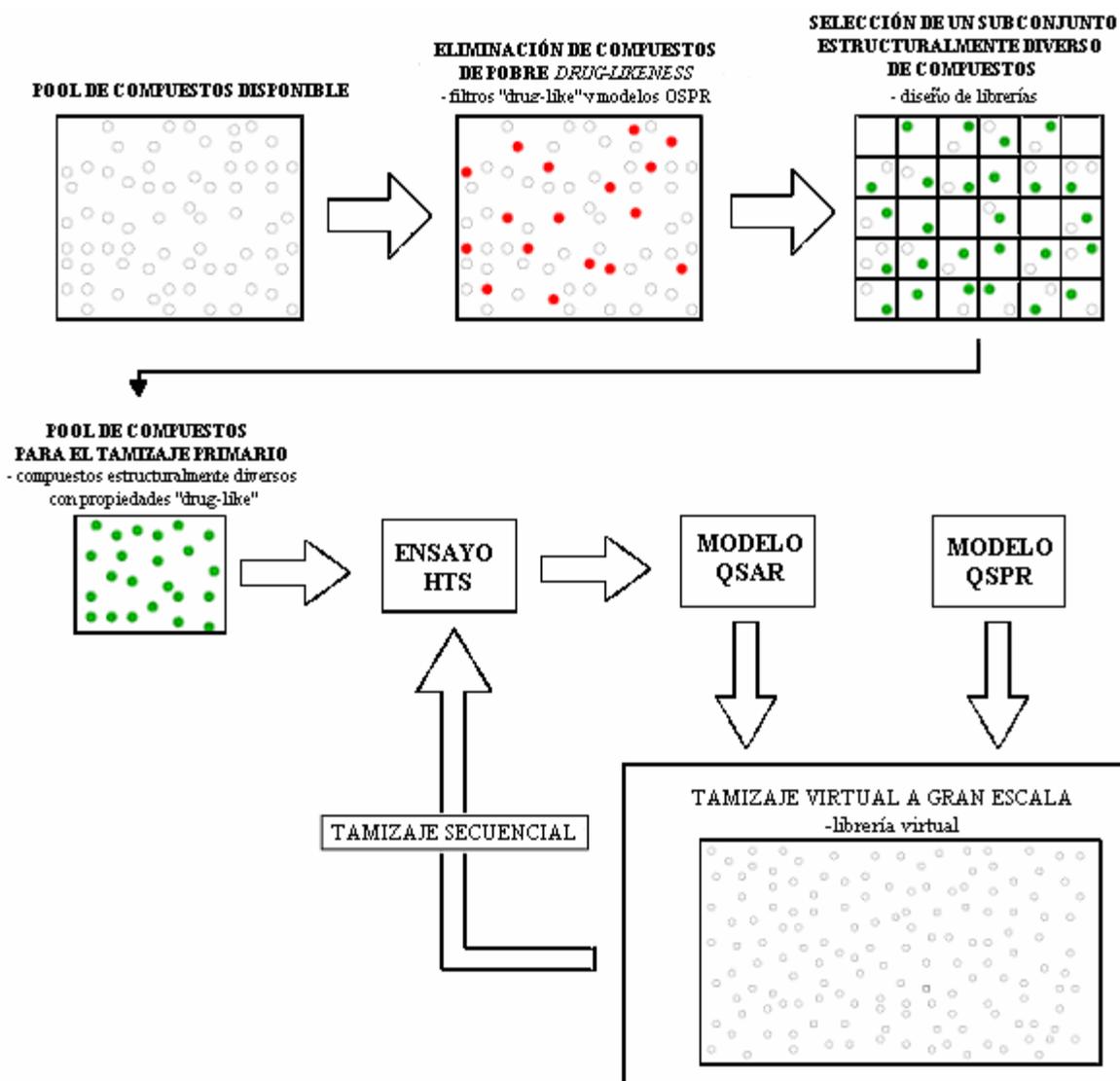


Figura 1.10. Posible integración de métodos computacionales en el proceso de descubrimiento de fármacos.

1.4.4 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.
- 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(95). Estas superficies son en particular de gran interés en estudios de acoplamiento (*docking*) para cuantificar la complementariedad entre un ligando y un blanco terapéutico. Sin embargo, las moléculas no son elementos inertes, y debido a la presencia de enlaces rotacionales, estas pueden llegar a tomar varias configuraciones espaciales de baja energía conocidas como conformaciones. Por tanto, si se considera la estructura tridimensional de una molécula, surge entonces el problema del análisis conformacional, lo que en la práctica conduce a representar las moléculas como un juego de conformaciones, en un enfoque denominado *multi-confórmeros* o bien como una estructura flexible en cuyo caso se introducen grados de libertad estructural para cada enlace rotacional en la molécula.

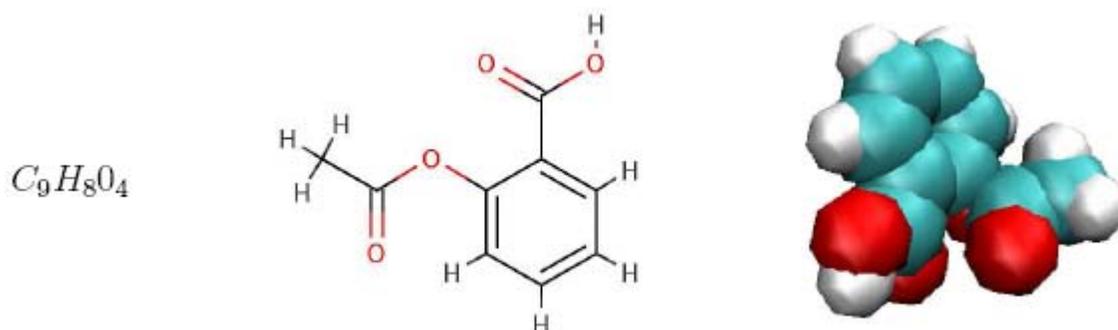


Figura 1.11. Representaciones 1D, 2D y 3D de la molécula de aspirina.

Codificación de la estructura molecular. La Figura 1.11 muestra diferentes representaciones de la molécula de aspirina. Desde el punto de vista práctico, los métodos quimioinformáticos a menudo no son capaces de manejar directamente estas representaciones, 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*(106). El término *útil* tiene aquí un significado doble: significa que estos números pueden brindar una visión más amplia en la interpretación de las propiedades moleculares y/o son capaces de tomar parte en un modelo para la predicción de propiedades moleculares interesantes. En este sentido, si la interpretación del descriptor fuera débil, provisional, o carente completamente de significado, aún el mismo pudiera estar fuertemente correlacionado con alguna propiedad molecular, lo cual sería útil en la generación de modelos con poderes altamente predictivos. Por el contrario, los descriptores con un poder predictivo pobre se pudieran retener de manera útil en los modelos cuando están bien fundados teóricamente y son, por ende, interpretables debido a su capacidad de codificar la información química estructural (106). 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)
- 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.

Estos descriptores (0D y 1D) ofrecen poca información sobre la estructura molecular y se usan esencialmente para derivar filtros tales como las reglas del cinco de Lipinski, o en combinación con otros descriptores.

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*, definidos respectivamente a partir de la longitud del camino más corto entre pares de átomos, y sus grados en el grafo molecular(95, 106). En un enfoque relacionado, los vectores de autocorrelación topológica miden la autocorrelación de las propiedades fisicoquímicas, tales como la carga o la polaridad, de pares de átomos separados a una distancia topológica dada, la cual se expresa como la longitud del camino más corto que conecta a los átomos en el grafo molecular(107).

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(106). 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. Las funciones de puntuación típicas consideran el número de átomos idénticos superpuestos en la representación del esqueleto molecular (108), o el solapamiento de las nubes de electrones que rodean la molécula (109). Para poder realizar análisis conformacionales, el alineamiento puede ser flexible en cuyo caso se introducen grados de libertad adicionales para manejar los enlaces rotacionales, o rígido y basado en el alineamiento óptimo de pares de multi-

conformeros. El alineamiento molecular puede resultar un proceso complejo, por lo que el lector interesado en los detalles de las técnicas existentes puede consultar a Lemmen y Lengauer (110). Una vez que las moléculas están alineadas, los DMS-3D pueden definirse, por ejemplo, mediante el muestreo de la superficie molecular de acuerdo a los rayos que emanan del centro de masa de las moléculas alineadas (111, 112), o en el caso del Análisis Comparativo de Campos Moleculares (*Comparative Molecular Field Analysis, CoMFA*) midiendo la interacción entre las moléculas y un átomo problema (por ejemplo, un átomo cargado o lipófilo) en cada punto de una caja discreta que contiene las moléculas (113).

Un enfoque opuesto consiste en extraer los descriptores independientemente de la orientación molecular. A excepción de los descriptores globales de forma como el volumen de Van der Waals de la molécula o el área de superficie molecular, la mayoría de los descriptores moleculares independientes del alineamiento se basan en distancias interatómicas. Por ejemplo, un estudio desarrollado por Pepperrell y Willett propone la caracterización de la estructura molecular a través de su matriz de distancias interatómicas(114). Si bien estos autores proponen varios métodos para comparar tales matrices, este enfoque no es conveniente del todo ya que no conduce a una representación fija de las moléculas. Las representaciones vectoriales estándar pueden derivarse considerando pares de átomos de la molécula. Los vectores de autocorrelación topológica por ejemplo, pueden ser extendidos a vectores de autocorrelación tridimensional mediante el cálculo de la autocorrelación de propiedades atómicas a partir de pares de átomos incluidos dentro de un rango de distancia Euclídeana específico, en lugar de una distancia topológica dada en el grafo molecular(115). Otras representaciones se basan en el conteo del número de veces que se encuentran pares de átomos de un tipo específico dentro de rangos de distancias predefinidos en la estructura tridimensional de la molécula (116-118). La consideración de tripletes o conjuntos más grandes de átomos conduce a la noción de *farmacóforo*. Un farmacóforo se define usualmente como un arreglo tridimensional de átomos o grupo de átomos responsable de la actividad biológica de una molécula de fármaco (119, 120). Los átomos que poseen propiedades particulares (por ejemplo, cargas positivas o negativas, alta hidrofobicidad, capacidad para donar o aceptar hidrógenos, anillos aromáticos) pueden constituir características farmacofóricas típicas de interés(121). En este contexto, las huellas farmacofóricas o *pharmacophore fingerprints* han sido propuestas como la contraparte tridimensional de las huellas de fragmentos moleculares o *molecular fragment fingerprints*. Las huellas farmacofóricas representan una molécula a través de una pequeña secuencia que codifica su contenido farmacofórico, usualmente definido como una lista exhaustiva de los tripletes o características farmacofóricas encontradas dentro de un conjunto predefinido de rangos de distancia en su estructura tridimensional(122-125), aunque también existen extensiones a farmacóforos basados en cuartetos(126). Por tanto, las huellas farmacofóricas codifican posibles farmacóforos de las moléculas, y debido a que el número de farmacóforos potenciales puede ser muy grande, estas son usualmente fragmentadas(124) o comprimidas(127, 128).

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 (106).

1.5 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. Este modelo puede ser subsecuentemente usado en aplicaciones de tamizaje virtual para predecir la actividad de compuestos que ni siquiera tienen existencia física. 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 global de un candidato a fármaco, tales como sus propiedades farmacológicas, toxicológicas o farmacocinéticas. Las décadas de investigación en el campo de la estadística y las maquinarias de aprendizaje (*Machine Learning*) han proveído toda una plétora de métodos útiles en el establecimiento de las relaciones entre la actividad de una molécula y su estructura, y en general en la investigación química, lo que ha contribuido al desarrollo de lo que hoy se conoce como *quimiometría*. Particularmente, en esta sección se pretende ofrecer una consideración metodológica e histórica general sobre la aplicación de estos métodos a la quimioinformática, así como un resumen del basamento teórico que soportan los métodos quimiométricos empleados en la investigación terapéutica moderna, lo que dista de ser una descripción detallada de estos métodos. Por esta razón se invita al lector interesado en una introducción más profunda a esta temática a consultar los libros de texto clásicos(129-131).

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. En los estudios REA, la mayoría de las propiedades consideradas son en esencia cuantitativas, pero el problema de la predicción es a menudo enfocado como un problema de clasificación binaria mediante la introducción de un límite sobre el cual las moléculas se consideran globalmente activas e inactivas bajo este. El término clasificación se usa comúnmente para referirse a la clasificación binaria(132).

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, que se corresponde con su habilidad para hacer predicciones correctas de un conjunto de nuevas moléculas que no fueron empleadas en la fase de entrenamiento. Existen diferentes criterios para este tipo de evaluación. 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(132). Sin embargo, si una de las dos clases esta sobre-representada en la serie de entrenamiento, y/o los costos de clasificaciones erróneas son diferentes, es más seguro considerar las razones de clasificación de verdaderos (falsos) positivos (negativos). La razón de verdaderos positivos(negativos) ofrece información referente a la fracción de compuestos positivos(negativos) que son clasificados correctamente mientras que la razón de falsos positivos(negativos) representa la fracción de compuestos positivos(negativos) que son clasificados incorrectamente (129, 132). En las aplicaciones del tamizaje virtual donde normalmente no es deseable clasificar incorrectamente un compuesto potencialmente activo, se favorecen los modelos con razones pequeñas de falsos negativos, incluso a expensas de un incremento de la razón de falsos positivos.

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(94, 95, 106). Una manera bastante extendida de conducir este problema en la quimioinformática se basa en la aplicación de *análisis de componentes principales* (ACP)(133-135) o *mínimos cuadrados parciales* (MCP)(136-139), 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. La diferencia entre estos dos métodos proviene del hecho de que el ACP se basa únicamente en los valores de los rasgos, mientras que el método MCP considera además su correlación con la propiedad a predecir.

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, específicamente para reducir el número de rasgos a seleccionar, se encuentran los basados en la teoría de la información a partir del criterio de información mutua o ganancia de información (140) y en criterios estadísticos como la razón de Fisher o el grado de correlación entre pares de rasgos. Para la selección de rasgos en sí, los métodos más extendidos incluyen 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 (41).

La elección de un modelo de entre la plétora de modelos existentes está 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 (132): 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 (efectos biológicos u otros); 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 (141, 142). Los modelos RLM aun 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*) (139). Este es, en particular, el caso de la metodología CoMFA, donde la regresión basada en MCP es el modelo estándar para la predicción de la actividad biológica de una molécula a partir de una vasta cantidad de puntos en un enrejado que determinan los campos moleculares (113). Más recientemente, los algoritmos genéticos han sido introducidos para solucionar el problema de la selección de rasgos como una alternativa eficiente a los métodos tradicionales de selección hacia delante y eliminación hacia atrás (*forward selection* y *backward elimination*, respectivamente) (143). Enfoques lineales pueden aplicarse al problema de la clasificación a través del *análisis discriminante lineal* (ADL) (144).

Sin embargo, debido a que esta clase de modelos está limitada a establecer relaciones lineales, este enfoque puede ser demasiado restrictivo para predecir eficientemente el espectro de propiedades biológicas. Mientras que el desempeño de los modelos puede mejorarse a través de la aplicación de transformaciones no lineales de las variables de entrada (145), los estudios REA pueden beneficiarse de manera significativa del desarrollo de métodos no lineales de modelación, entre los que sobresalen las *redes neuronales artificiales* (RNAs). 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, en comparación con las técnicas estándar de RLM (146-148). Estudios posteriores han confirmado la capacidad de las RNAs para predecir propiedades biológicas, y hoy en día constituyen una herramienta estándar de la modelación REA (149-151).

Métodos quimiométricos de modelación REA. La Quimiometría es una disciplina que recopila herramientas matemáticas y estadísticas para tratar con datos químicos complejas (152-158). En esta disciplina, las estrategias de caracterización principales son: los enfoques multivariados al problema en cuestión, la búsqueda de información relevante, la validación de los modelos para derivar objetos matemáticos con altos poderes predictivos, la comparación de los resultados obtenidos por diferentes métodos, y, la definición y uso de criterios capaces de medir la calidad de la información extraída y los modelos obtenidos.

La quimiometría es la herramienta más útil en los estudios REA, áreas donde esta disciplina brinda una batería bastante completa de métodos disímiles y, por tanto, una base firme para el análisis de los datos y la

modelación en general. No obstante, un aspecto relevante de la filosofía de la Quimiometría es la atención que presta al poder predictivo de los modelos (el cual se estima con técnicas de validación), a la complejidad del modelo, y, a la continua búsqueda de parámetros adecuados para evaluar la calidad de los mismos, tales como los parámetros de clasificación y parámetros de regresión.

La Quimiometría hoy en día se intercepta no solo con varios campos de la Matemática y la Estadística clásica sino también de la Inteligencia Artificial. Dentro de los métodos y técnicas de interés se describen por consiguientes:

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 predictoras. El objetivo de este análisis es doble: modelar y predecir. La relación se describe en forma algebraica como:

$y = f(x) + e$ con f una función en el caso más simple lineal, de las componentes del vector x ,

En el caso lineal, el problema puede ser expresado en forma matricial $\mathbf{Y} = \mathbf{B} \cdot \mathbf{X} + \mathbf{E}$

Donde, x denota la(s) variable(s) predictor(a)s, y la(s) variable(s) respuesta(s), $f(x)$ es la parte sistemática del modelo, y e es el error aleatorio, también conocido como error o residual del modelo. En la forma generalizada contigua: \mathbf{Y} , es el vector de las respuestas para cada uno de los datos de aprendizaje; \mathbf{B} es el vector constante de coeficientes de regresión a ser estimados; \mathbf{X} se denomina usualmente la matriz de datos, i.e. sus columnas son las variables independientes usadas en el modelo de regresión y sus filas representan los datos de aprendizaje; \mathbf{E} , es el vector de los errores aleatorios intrínsecos al modelo correspondiente. La ecuación matemática usada para describir la relación entre la(s) variable(s) respuesta(s) y la variable(s) predictor(a)s se denomina modelo de regresión (159-162).

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 (163).

Toma de Decisiones Multi-Criterio. El enfoque de la Toma de Decisiones Multi-Criterio se basa en métodos orientados a categorizar los objetos estudiados (eventos, moléculas, casos, etc.) sobre la base de criterios múltiples (164, 165). En particular, las Funciones de Conveniencia y las Funciones de Utilidad son funciones de decisión multicriterios capaces de asignar un puntaje a cada objeto de acuerdo a los intereses del usuario.

Las funciones de decisión multicriterios se pueden usar para representar la información química mediante la “armonización” de la información estructural, el conocimiento experimental y otras características específicas del problema estudiado tales como parámetros de salud y medioambientales.

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} ,

$f(p_1, p_2, \dots, p_p)$. Si los valores que los parámetros pueden tomar están restringidos, el procedimiento se llama optimización restringida u optimización con restricciones.

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 (*Steepest Ascent Optimization*), la Optimización del Descenso más Abrupto (*Steepest Descent Optimization*), la Optimización *Simplex*, la Optimización por Algoritmos Genéticos, el Recorrido Simulado (*Simulated Annealing*). (165-167). Las técnicas de reducción y selección de variables también se consideran en cierto sentido, técnicas de optimización, porque buscan la reducción de datos maximizando, por ejemplo, la “varianza” que los datos reducidos son capaces de explicar.

La Inteligencia Artificial aporta las llamadas técnicas de Aprendizaje Reforzado (*Reforceiment Learning*). Ellas combinan en cierto sentido, técnicas de aprendizaje supervisado y no supervisado con técnicas de optimización. Por ejemplo: los sistemas multiagentes (*Multiagent Systems*) son capaces de aprender por la acción coordinada de múltiples “agentes”, sin un supervisor o maestro general, pero hay un “ambiente” o “entorno” con los cuales intercambian información (y que de hecho se convierte en un ambiente supervisor) con vistas a optimizar una cierta función de “ganancia de información”. En muchos problemas complejos, por ejemplo, de la bioinformática, donde ya se refirió la necesidad de multi-clasificadores, el telescopio de búsqueda de herramientas apunta hacia los sistemas multiagentes (168, 169).

Optimización Multiobjetivos. Los métodos de optimización multiobjetivos (OMO) introducen una nueva filosofía para la optimización basada en la obtención de compromisos entre varios objetivos (23). Por medio de los métodos OMO es posible encontrar una solución óptima global mediante la optimización simultánea de múltiples propiedades interrelacionadas. El mayor beneficio de los métodos OMO 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. Actualmente, la mayoría de los métodos OMO desarrollados para su aplicación en el campo del descubrimiento y desarrollo de fármacos están basados principalmente en dos métodos: La suma ponderada de funciones objetivos (SPFO) (170) y métodos basados en pareto (23). Varias aplicaciones de los métodos OMO han sido reportadas en la literatura. Las mismas van desde la minería de subestructuras hasta estudios de acoplamiento, incluyendo estudios QSAR y QSPR inverso (23).

En el área de la minería de subestructuras, las aplicaciones de los métodos OMO se han concentrado en el alineamiento molecular y la identificación de farmacóforos. Ejemplos de métodos que atacan el problema de la minería de subestructuras desde una perspectiva multiobjetivos incluyen el programa de similitud molecular basado en algoritmos genéticos (un método basado en SPFO) (171) y algunos métodos basados en pareto tales como el algoritmo genético para alineamiento molecular múltiple (probablemente la primera aplicación en la quimioinformática de un enfoque basado en pareto) (172), y el algoritmo genético con asignación lineal para el alineamiento hiper molecular de bases de datos (173).

Con relación a los estudios de acoplamiento, varios grupos de investigación son particularmente activos en el uso de métodos OMO basados en pareto. Por ejemplo, Janson y colaboradores han introducido una aplicación

a un estudio de optimización del acoplamiento denominado *ClustMPSO*, basado en un algoritmo de optimización inspirado en enjambre de partículas, el cual minimiza simultáneamente la energía intermolecular entre la proteína y el ligando y la energía intramolecular del ligando (174). También Zoete y colaboradores han usado un algoritmo evolutivo multiobjetivos en su programa computacional de estudios de acoplamiento *EADock* (175).

Recientemente, los métodos OMO han sido aplicados a la optimización de nuevas entidades químicas a través del diseño molecular *de novo* y estudios QSPR inverso. En esta área existen aplicaciones importantes como el método *CoG* introducido por Brown y colaboradores para resolver el problema QSPR inverso (176) así como el “*Molecule Evaluator*” propuesto por Lameijer y colaboradores, donde el usuario asume el papel de la función de ajuste mediante la selección de moléculas candidatas para las próximas evoluciones luego de cada iteración (177).

Finalmente, las técnicas OMO han sido sólo recientemente aplicadas al área de los estudios QSAR. En este sentido, Nicolotti (42) empleó una variante de un algoritmo evolutivo llamado programación genética multiobjetiva que usa un ordenamiento basado en un Pareto para optimizar modelos QSAR. Un número de objetivos en conflicto incluyendo la exactitud del modelo, el número de parámetros, la complejidad interna e interpretabilidad de los descriptores usados en el modelo, fueron considerados. Por otra parte, Stockfisch (43) propuso una técnica multiobjetiva no evolutiva llamada partición recursiva de múltiples propiedades parcialmente unificadas (*PUMP-RP*) para la generación de modelos QSAR. Este método no basado en Pareto se usó exitosamente en la construcción de modelos para analizar relaciones de selectividad entre inhibidores de ciclooxigenasa (COX) 1 y 2 (44).

2 MATERIALES Y MÉTODOS

2. MATERIALES Y MÉTODOS

"Everything you can imagine is real."
Pablo Picasso

2.1 EMPLEO DE MÉTODOS DE OPTIMIZACIÓN MULTI OBJETIVOS (OMO) Y ORDENAMIENTO BASADOS EN FUNCIONES DE DESEABILIDAD EN ESTUDIOS GLOBALES DE CORRELACIÓN ESTRUCTURA-PROPIEDAD

Mejorar el perfil de una molécula para el proceso de descubrimiento y desarrollo de fármacos requiere de la optimización simultánea de numerosos objetivos que a menudo se encuentran en conflicto. En una forma muy simplificada, el fármaco ideal debería poseer la mayor eficacia terapéutica y biodisponibilidad a la vez que la mínima toxicidad posible. Debido a la relación competitiva entre las propiedades antes mencionadas, alcanzar tal fármaco es casi una quimera y de ser posible, una tarea sumamente difícil, costosa y consumidora de tiempo. Sin embargo, si desarrollar el fármaco ideal es sumamente difícil, encontrar el mejor compromiso posible entre objetivos que compiten como la eficacia terapéutica, la biodisponibilidad y la toxicidad 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 2.1.).

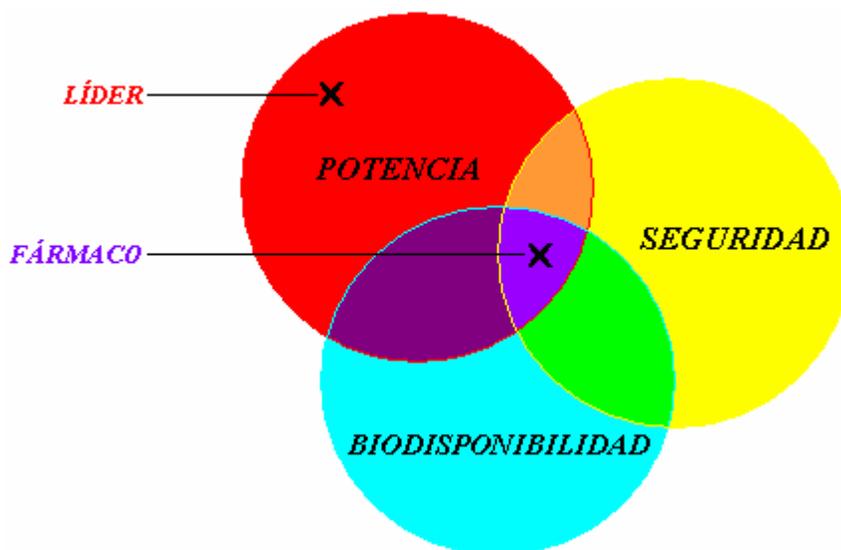


Figura 2.1. Representación gráfica del compromiso entre eficacia terapéutica (potencia), biodisponibilidad (propiedades ADME) y toxicidad (seguridad) requerida para alcanzar un fármaco exitoso.

Con este propósito, nos proponemos realizar una optimización multiobjetivos basada en la estimación de la deseabilidad de varias respuestas interrelacionadas; lo que permitirá conducir estudios QSPR globales considerando simultáneamente el perfil farmacológico, farmacocinético y 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 (178), específicamente dirigida a conferir racionalidad al proceso de desarrollo de fármacos.

2.1.1 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. Con objetivos de generalización, en lo adelante los términos de “variables respuesta” y “variables independientes” en el contexto de descubrimiento y desarrollo de fármacos vía métodos QSPR deben ser entendidos como cualquier propiedad farmacológica, farmacocinética o toxicológica a ser optimizada y cualquier serie de descriptores moleculares utilizados para modelar cada propiedad, respectivamente.

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

Cada variable respuesta (Y_i) se relaciona con n variables independientes (X_n) por una relación funcional desconocida la cual se aproxima a una función polinómica (usualmente lineal) con el objetivo de encontrar la respuesta estimada o predicha (\hat{Y}_i). Cada respuesta \hat{Y}_i es estimada por técnicas de regresión basadas en mínimos cuadrados (179).

En algunas situaciones donde están involucradas varias respuestas, el MP para alguna de las respuestas puede compartir las mismas variables independientes con el MP para otra respuesta, pero con coeficientes diferentes. En esta situación atípica, encontrar el mejor compromiso entre las respuestas a considerar se torna más simple. Sin embargo, debido a la multiplicidad de factores involucrados en el perfil farmacéutico de una molécula no es irrazonable suponer que el mismo subconjunto de factores (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 produzcan la máxima deseabilidad, dado los MPs originales usados para predecir cada respuesta original.

En dependencia de que una respuesta particular Y_i deba ser maximizada, minimizada, o asignado un valor blanco o diana (un valor único que se supone sea el ideal), 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 (178).

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:

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

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

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

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:

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

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 . Para $s = t = 1$, la función de deseabilidad aumenta de forma lineal hacia el valor T_i . Valores grandes de s y t serían seleccionados si fuera muy deseable que el valor \hat{Y}_i estuviera muy próximo a T_i o aumentara rápidamente por encima de L_i . Por otro lado, valores pequeños de s y t serían especificados si casi cualquier valor de \hat{Y}_i por encima L_i , y debajo de U_i fueran aceptables o si el tener valores de \hat{Y}_i considerablemente por encima de L_i no fuera de una importancia crítica (178).

De este modo, es posible finalmente 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 estimadas o predichas (\hat{Y}_i) no existe ninguna manera de conocer cuán 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 la variable respuesta Y_i es estimada como una función continua de las variables independientes X_n ; las deseabilidades individuales $d_i(Y_i)$ son una función continua de las variables respuesta Y_i (ver ecuaciones (2.2.2), (2.2.3) y (2.2.4)), y la deseabilidad global 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) y 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.

De manera similar a R^2 , el coeficiente de determinación de la *deseabilidad* global ajustado ($Adj.R^2_D$) puede ser calculado si consideramos los grados de libertad totales ($n-1$) y los grados de libertad debido al error ($n-2$), siendo n el número de casos usados para ajustar los MPs.

$$Adj.R^2_D = 1 - \frac{SSE}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i})^2}{\frac{n-2}{\sum (D_{Y_i} - \bar{D}_{Y_i})^2} (n-1)} \quad (2.2.6)$$

Así, ambos R^2_D y $Adj.R^2_D$ poseen las mismas propiedades de R^2 y $Adj.R^2$. Ambos estarían incluidos en el rango $[0, 1]$ y mientras mayor sea $R^2_D/Adj.R^2_D$ menor será la incertidumbre al predecir D usando un conjunto específico de variables independientes X_n (179).

Considerando que R^2_D y $Adj.R^2_D$ miden la bondad de ajuste en lugar de la capacidad predictiva de un determinado MP 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 :

$$Q^2_D = 1 - \frac{SSE_{LOO-CV}}{SSTO} = 1 - \frac{\sum (D_{Y_i} - D_{\hat{Y}_i}(LOO-CV))^2}{\sum (D_{Y_i} - \bar{D}_{Y_i})^2} \quad (2.2.7)$$

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.

Como se demostró arriba, la función de *deseabilidad* condensa un problema de optimización multivariado en uno univariado. Consecuentemente, la *deseabilidad* global D puede ser maximizada sobre el dominio de variables independientes X_n . Para esto, puede ser usada la opción *Response/Desirability Profiler* de cualquiera de los módulos de Modelos Lineales Generales (MLG), Modelos de Regresión Generales (MRG) o de Análisis Discriminante General (ADG) implementados en el software STATISTICA 7.0 (180). La *deseabilidad* global D es optimizada mediante el método *Use general function optimization* (el método *simplex* de optimización de funciones (181-183)) o el método *Optimum desirability at exact grid points* para realizar búsquedas exhaustivas de la *deseabilidad* óptima. La opción *Use general function optimization* es usualmente más rápida, pero el método predefinido es *Optimum desirability at exact grid points*, excepto cuando el número de valores predichos que deben ser calculados para realizar la búsqueda exhaustiva excede los 200 000, en cuyo caso el método *Use general function optimization* se convierte en el método predeterminado.

Un beneficio adicional del método es la capacidad para representar gráficamente D como una función de una o más variables independientes. Esto permite al usuario encontrar una tendencia en la relación entre las respuestas y las variables independientes considerando la forma de la función de deseabilidad relacionada a cada variable independiente lo que permite establecer un rango óptimo para cada variable independiente sobre los valores óptimos determinados mediante el proceso de optimización.

El resultado final se traduce en encontrar los niveles óptimos (o un rango óptimo) de las variables independientes que optimizan simultáneamente las k respuestas determinando la calidad final del producto (en nuestro caso un medicamento). 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 (o sea, el candidato a fármaco más deseable).

Teniendo en cuenta que D es maximizada directamente sobre el dominio de variables independientes y al mismo tiempo los valores predichos de D dependen de la serie inicial de MPs es necesario considerar el dominio de aplicación de cada MP (específicamente los respectivos valores de *leverage* y residuales estandarizados) para determinar el nivel óptimo de cada variable independiente así como para la selección de las soluciones óptimas (moléculas o candidatos a fármaco).

2.1.2 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). Por tanto, teóricamente es posible determinar la solución de un caso basándose solamente en su descripción (184).

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 librerías combinatorias de los cuales solo se conoce su estructura, con respecto a su similitud con el caso óptimo determinado mediante el proceso de optimización simultánea de múltiples objetivos 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:

$$\Delta_i = \sum_{X=1}^m \delta_{i,X} \cdot w_X \quad (2.2.8)$$

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 .

La distancia euclidiana E de un caso i con respecto a un caso j cuando se considera más de un rasgo o característica se define como:

$$E = \left[\sum (X_i - X_j)^2 \right]^{1/2} \quad (2.2.9)$$

En nuestro caso decidimos determinar la similitud entre un caso i y el caso óptimo considerando cada variable X por separado en lugar de determinarla en función del total de variables que describen el caso. Esto permitirá una mayor libertad en la aplicación de un segundo proceso de optimización empleado para la determinación del conjunto óptimo de pesos w_X asociados a las respectivas variables X . Por otra parte, de los resultados de este segundo proceso de optimización se podrá inferir la influencia relativa de cada variable X sobre la deseabilidad global D para un caso i .

En un caso como el nuestro donde solo se considera un rasgo o variable, la distancia euclidiana entre dos casos coincide con el valor absoluto de la diferencia entre estos dos casos. Consecuentemente, $\delta_{i,j}$ se define como:

$$\delta_{i,X} = |X_i - X_{OPT}| \quad (2.2.10)$$

donde X_i y X_{OPT} son los respectivos valores del parámetro X para el caso i y el caso óptimo.

Una vez determinado Δ_i , este parámetro es normalizado nuevamente mediante el empleo de las funciones de deseabilidad de Derringer(178). Específicamente, los respectivos valores de Δ_i son minimizados 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_{MIN}) y $U_i = \Delta_{MAX}$. 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 minimizada para cada caso.

El conjunto optimo de pesos w_X que minimizan la diferencia entre los valores de D_i y los valores normalizados de Δ_i son determinados mediante un proceso de ajuste de datos no lineal basado en mínimos cuadrados. Específicamente, los pesos se obtuvieron usando un algoritmo de optimización a larga escala basado en un método de ajuste de curva no lineal(185, 186) implementado en la función "*lsqcurvefit*" del programa MATLAB, Versión 7.2(187). Este proceso se corrió sobre Windows a un costo computacional muy bajo (~ 1 seg.).

Una vez minimizada la diferencia entre D_i y los valores normalizados de Δ_i se obtiene una correspondencia entre la descripción (expresada a través de los valores normalizados de Δ_i , determinados estos por la estructura molecular la que a su vez es representada por el conjunto de descriptores moleculares empleados) y la solución de los casos (determinada por los respectivos valores de D_i los que representan la combinación de las k propiedades involucradas en la calidad final del candidato a fármaco). 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 y dado que existe una correspondencia entre Δ_i y D_i será posible por tanto conseguir un ordenamiento en función de D_i . De este modo será posible identificar los candidatos más prometedores los cuales serán ordenados en los primeros lugares (los de menores valores de Δ_i y por tanto los más próximos al candidato optimo determinado por el proceso de OMO) y descartar el resto de los candidatos que estén ordenados al final de la lista.

2.1.3 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 en este trabajo un nuevo método de validación del proceso de optimización a la vez que un criterio numérico de la calidad del ordenamiento obtenido. Para esto es necesario utilizar el conjunto de datos usados en el entrenamiento ya que se conocen tanto la descripción (valores de los descriptores moleculares empleados) como la solución (valores de Y_i , $d(Y_i)$, y D) de cada caso.

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. El orden de los casos se corresponde aquí con el ordenamiento ideal o experimental de los mismos.
- 2- Invertir el orden obtenido en el paso 1 ($n, \dots \dots, 3, 2, 1$). El orden de los casos se corresponde aquí con el peor ordenamiento posible.
- 3- Ordenar los casos de menor a mayor en función de Δ_i y asignarle un orden (1, 2, 3,, n) a cada caso. El orden de los casos se corresponde aquí con el ordenamiento real o predicho obtenido a través del método.
- 4- Normalizar mediante el empleo de las funciones de deseabilidad de Derringer (específicamente empleando la 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 determinar la diferencia del ordenamiento real con respecto al ordenamiento ideal ($^{IR-RR}\delta_i$):

$$^{IR-RR}\delta_i = \left| ^{IR}\delta - ^{RR}\delta \right| \quad (2.2.11)$$

y del peor ordenamiento con relación al ordenamiento ideal ($^{IR-WR}\delta_i$):

$$^{IR-WR}\delta_i = \left| ^{IR}\delta - ^{WR}\delta \right| \quad (2.2.12)$$

La diferencia ideal sería igual a cero para todos los casos y se correspondería con un ordenamiento ideal. La Figura 2 muestra la forma que presentaría el peor ordenamiento y un ordenamiento ideal.

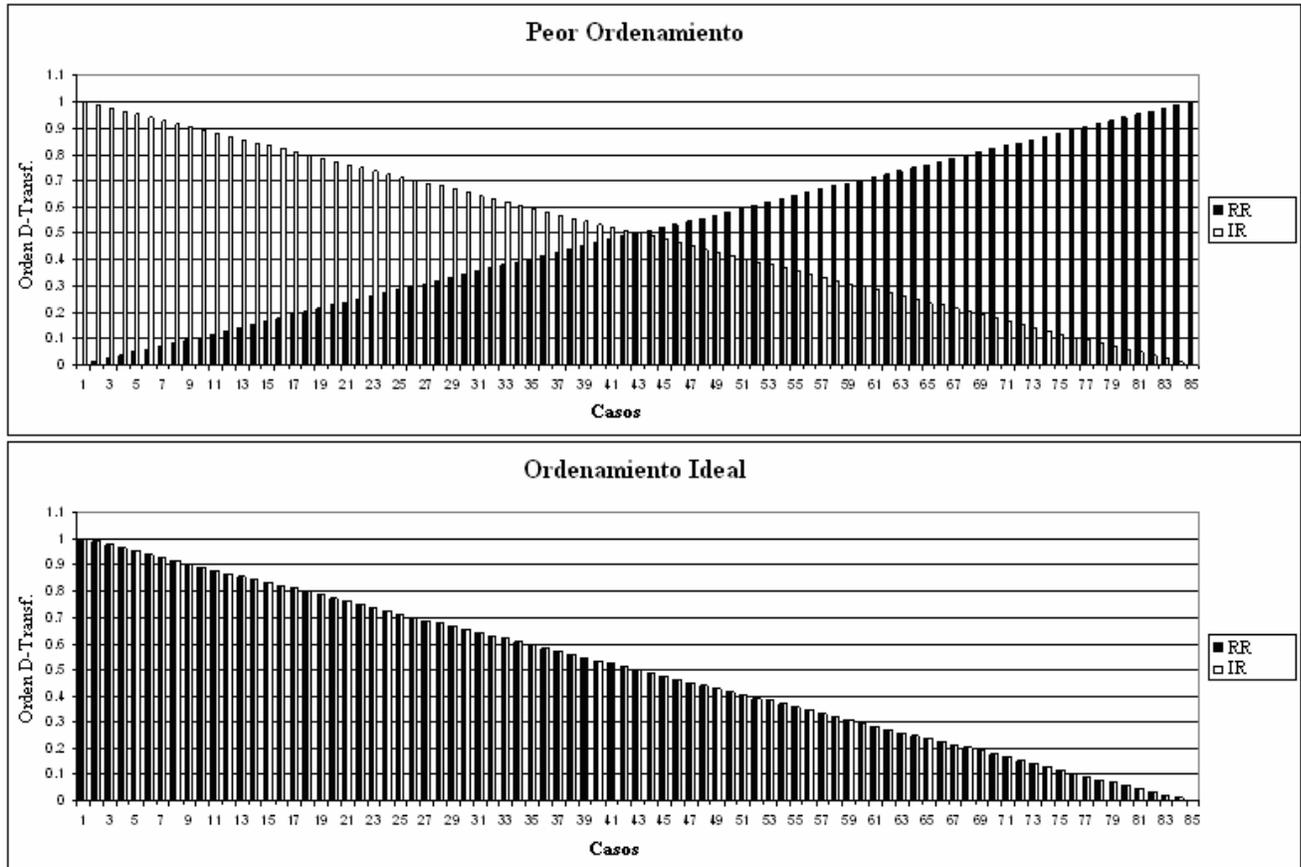


Figura. 2.2. Representación gráfica del peor ordenamiento (arriba) y el ordenamiento ideal (debajo).

- 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:

$$\Psi = \left| \frac{\sum_{i=1}^n {}^{IR-RR}\delta_i}{n} \right| \quad (2.2.13)$$

Este parámetro se mueve en un 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. De este modo, cuanto más cercano a 0 sea el valor de Ψ de un determinado ordenamiento mayor será la calidad del mismo mientras que valores cercanos a 0.5 indican una baja calidad del ordenamiento. Debido a que el valor de Ψ para el peor caso depende del tamaño del conjunto de datos a ordenar este no es exactamente igual a 0.5. Por otra parte un rango $[0, 1]$ al contrario de $[0, 0.5]$ sería un indicador más claro de la calidad del ordenamiento. Considerando estas dos cuestiones a Ψ se le aplica un factor de corrección (F):

$$F = \frac{2}{\Psi^{WR}} \quad (2.2.14)$$

donde Ψ^{WR} representa el índice de calidad del peor ordenamiento.

F es empleado aquí con el objetivo de obtener un indicador Ψ más representativo de la calidad de un ordenamiento a la vez que se logra que el mismo se mueva en un rango exacto $[0,1]$ donde el valor de Ψ para el peor ordenamiento sea de este modo exactamente igual a 1. De esta forma se obtiene el índice de calidad del ordenamiento corregido (Ψ^*):

$$\Psi^* = \left| \frac{\sum_{i=1}^n IR-RR \delta_i}{n} \right| \cdot F = \left| \frac{\sum_{i=1}^n IR-RR \delta_i}{n} \right| \cdot \frac{2}{\Psi^{WR}} \quad (2.2.15)$$

Finalmente, es posible expresar Ψ^* como el porcentaje de calidad del ordenamiento:

$$R_{\%} = (1 - \Psi^*) \cdot 100 \quad (2.2.16)$$

2.2 CONFORMACIÓN DE LA BASE DE DATOS

Para el desarrollo de los modelos de predicción (MPs) relativos la actividad inhibitoria sobre la TI del VIH-1 y tóxica sobre células sanguíneas MT4; el proceso de OMO de ambas propiedades y la posterior construcción del esquema de ordenamiento se conformó una librería de compuest por 109 compuestos tipo INNTI reportados en la literatura (72, 188-190).

Para asegurar una adecuada diversidad estructural se conformó una base de datos compuesta por las 4 principales familias de compuestos inhibidores de la TI del VIH-1. Las estructuras de estos compuestos son mostradas en las Tablas 2.1 - 2.4 (72, 188-190). La identificación de cada uno de los 109 compuestos utilizados en este trabajo con sus respectivas referencias, sus estructuras, así como sus respectivos valores de EC_{50} y CC_{50} puede consultarse en la sección de anexos.

Tabla 2.1. Estructuras diversas de 4 familias de los compuestos utilizados

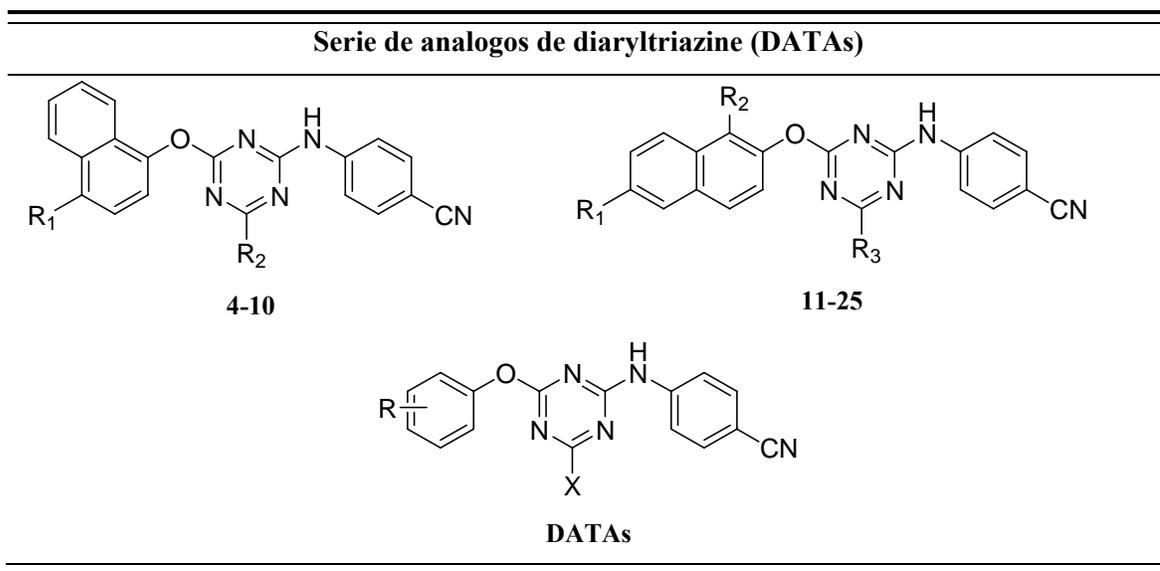
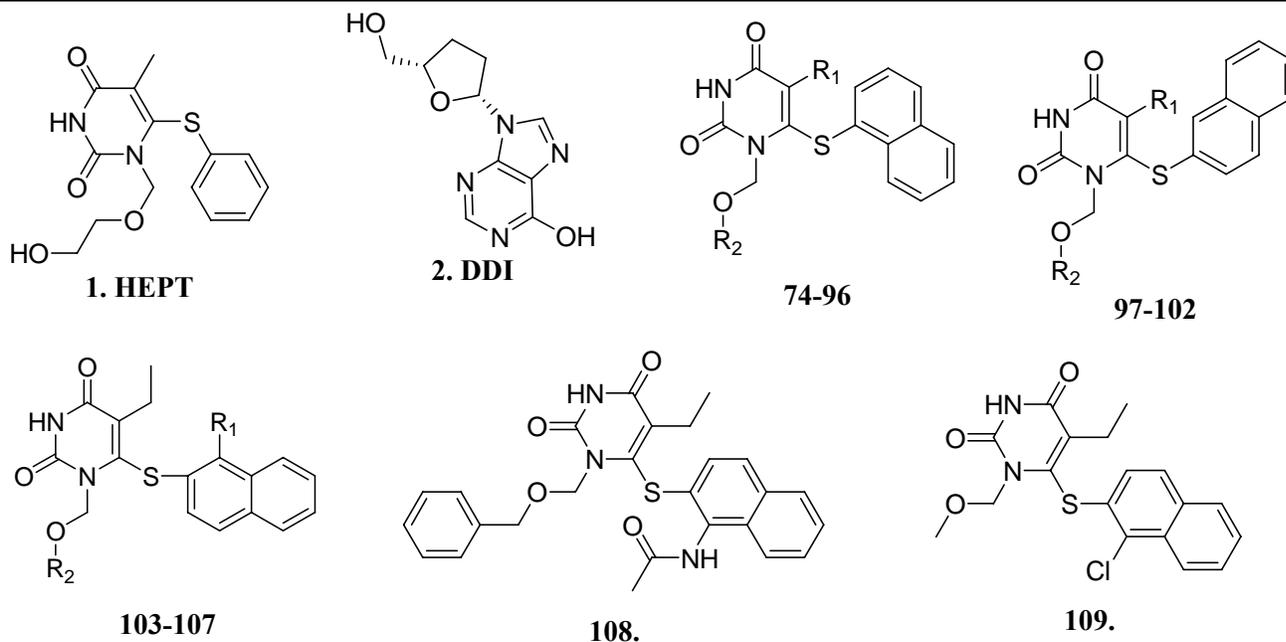
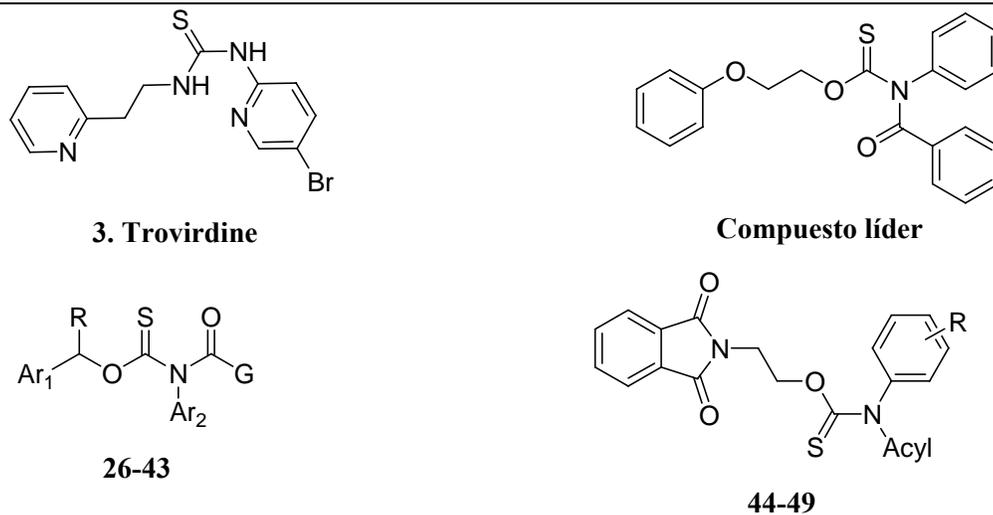


Tabla 2.1 (Continuación...)

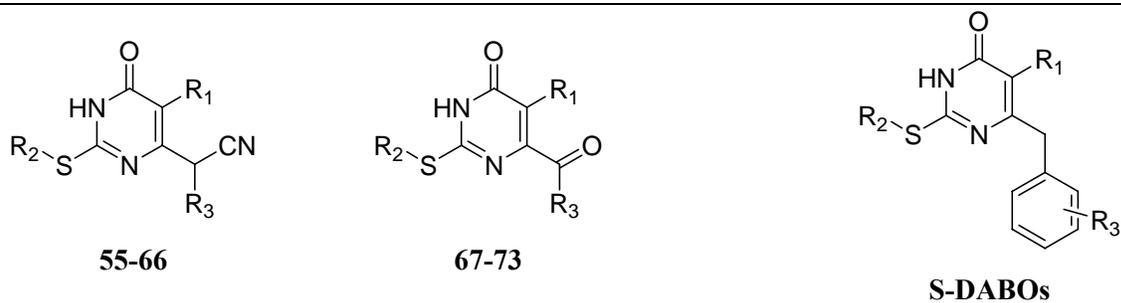
Serie de los derivados de 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT)



Serie de O-sust-N-sust-N-acylthiocarbamates (ATCs)



Serie de análogos de S-DABO (S-DABOs) 2-alkoxy-3,4-dihydro-6-benzyl-4(3H)-pyrimidin-4-ones (DABOs)



La toxicidad sobre células sanguíneas MT-4 es reportada como la concentración del compuesto que conduce a la supervivencia del 50% de las células infectadas en los diferentes medios cultivos con respecto con las células de control no tratadas (CC_{50}) y expresada en $\mu\text{g/mL}$ (191-198). La actividad inhibitoria sobre la TI del VIH-1 se reporta como la concentración efectiva (EC_{50}) la cual se define como la concentración de compuestos necesarios para inhibir la formación de protoplasma inducida por el virus hasta un 50% del número total de células. Tanto la EC_{50} como de CC_{50} de todos los compuestos fue medida sobre la cepa silvestre del VIH-1 tipo IIIB y sobre la línea de células sanguíneas MT-4 utilizando el método MTT (199, 200).

Como consecuencia del reporte inexacto de los valores de EC_{50} y/o CC_{50} (valores reportados como $<$, $>$ o \leq) cuyo uso reduce significativamente la calidad del ajuste del modelo de regresión lineal múltiple (RLM), 53 de los 162 compuestos reportados no fueron usados para el desarrollo de los respectivos modelos de relación cuantitativa estructura-actividad (QSAR). Por otra parte, los valores de EC_{50} y CC_{50} de los 109 compuestos usados para ajustar los modelos de RLM fueron transformados ($EC_{50} \rightarrow 1/(1+ EC_{50})$ y $CC_{50} \rightarrow \sqrt[3]{CC_{50} + 1/2}$) con el objetivo de obtener un mejor ajuste.

2.3 ESPECIFICACIONES DE LAS FUNCIONES DE DESEABILIDAD

La optimización de la deseabilidad global fue conducida mediante el metodo *Optimum desirability at exact grid points* para el conjunto de datos relativo a la familia de quinazolinonas; mientras que para la familia de fluoroquinolonas se usó el método *Use general function optimization*. Ambos métodos están implementados en el modulo MRG del software STATISTICA 7.0 (180). El método *spline* (201, 202) 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.

Se ajustaron dos funciones de deseabilidad, uno para cada respuesta. Específicamente, la toxicidad sobre células sanguíneas MT4 debe ser minimizada. En este trabajo esta propiedad se expresa a través de CC_{50} que por la información que aporta debe ser maximizada de manera que el compuesto con mayor valor de CC_{50} (transformada como $\sqrt[3]{CC_{50} + 1/2}$) es el más deseable ($d_i = 1$). Para la estimación de los valores d_i correspondientes a los valores de toxicidad sobre células sanguíneas MT4 el valor superior U_i que coincide con el valor diana T_i se fijó en $\sqrt[3]{CC_{50} + 1/2} = 23,011$ que se corresponde con un valor de $CC_{50} = 529 \mu\text{M}$ exhibido por el compuesto menos tóxico de la serie, y el valor inferior L_i se fijó en $\sqrt[3]{CC_{50} + 1/2} = 2,796 / CC_{50} = 7,32 \mu\text{M}$ exhibido por el compuesto más tóxico de la serie. Contrariamente, la actividad inhibitoria sobre la TI del VIH-1 (expresado como EC_{50} y transformada esta en $1/(1+EC_{50})$) debe ser maximizada. Aquí $U_i = T_i = 1/(1+EC_{50}) = 0.998 / EC_{50} = 0.002 \mu\text{gM}$ correspondiéndose estos valores con el compuesto más activo de la serie y $L_i = 1/(1+EC_{50}) = 0.017 / EC_{50} = 58 \mu\text{M}$, coincidiendo este último con el compuesto de menor actividad inhibitoria de la serie.

2.4 MÉTODOS COMPUTACIONALES Y ANÁLISIS ESTADÍSTICO

Las estructuras moleculares de todos los compuestos fueron representadas en ChemDraw Ultra 9.0(203). Las estructuras fueron inicialmente optimizadas por un método semi-empírico de orbitales moleculares implementado en el programa Hyperchem 8.05. El método Hamiltoniano PM3(203) fue usado para obtener estructuras optimizadas con geometrías correspondientes a su mínimo energético.

Posteriormente, los 1664 descriptores moleculares implementados en el software DRAGÓN 7.0 (204) fueron calculados. Para reducir información redundante que podría conducir a correlaciones casuales, descriptores con valores constantes o cercanos a valores constantes fueron excluidos del posterior proceso de selección de variables. La clase de descriptores moleculares (0D, 1D, 2D y 3D), las familias de descriptores incluidas dentro de cada clase y el número de descriptores de cada familia que fueron calculados a través del software DRAGÓN 7.0 son mostrados en la Tabla 2.5.

Tabla 2.2 Descriptores moleculares implementados en el software DRAGÓN 5.0.

FAMILIA	NÚMERO
DESCRIPTORES 0D	
Descriptores Constitucionales	48
DESCRIPTORES 1D	
Descriptores de Carga	14
Propiedades Moleculares	29
Conteo de Grupos Funcionales	154
Fragmentos Centrados en Átomos	120
DESCRIPTORES 2D	
Descriptores Topológicos	119
Conteo de Pasos y Caminos	47
Índices de Conectividad	33
Índices de Información	47
Autocorrelaciones 2D	96
Índices de Adyacencia entre Aristas	107
Autovalores de Burden	64
Índices Topológicos de Carga	21
Índices basados en Autovalores	44
DESCRIPTORES 3D	
Perfiles Moleculares de Randic	41
Descriptores Geométricos	74
Descriptores RDF	150
Descriptores 3D-MoRSE	160
Descriptores WHIM	99
Descriptores GETAWAY	197

Debido a que en los estudios QSPR la selección de descriptores apropiados es una tarea complicada, ya que no hay una regla absoluta que gobierne esta selección, fue aplicada una técnica de optimización - Algoritmo

Genético (AG)- para la selección de variables (205-208) usando el paquete informático BuildQSAR (209, 210). El mismo parte de ecuaciones polinómicas lineales con un máximo de 10 variables a incluir para cada propiedad dentro, de una población de 3000 posibles ecuaciones con una probabilidad de mutación del 30% y donde el criterio de optimización usado fue la maximización del coeficiente de correlación (R). En la Tabla 2.6, se muestran los descriptores moleculares seleccionados y usados finalmente para la optimización simultánea de la actividad inhibitoria sobre la TI del VIH-1 y toxicidad sobre células sanguíneas MT-4 de los 109 compuestos tipo INNTI incluidos en el conjunto de entrenamiento.

Tabla 2.3. Descriptores moleculares seleccionados para ser empleados en el proceso de optimización simultánea.

SÍMBOLO	DEFINICIÓN	FAMILIA	CLASE	PROPIEDAD
MATS8e	Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2D	EC ₅₀
GATS3m	Geary autocorrelation - lag 3 / weighted by atomic masses	2D autocorrelations	2D	EC ₅₀
piID	conventional bond-order ID number	Walk and path counts	2D	EC ₅₀
EEig09x	Eigenvalue 09 from edge adj. matrix weighted by edge degrees	Edge adjacency indices	2D	EC ₅₀
EEig10x	Eigenvalue 10 from edge adj. matrix weighted by edge degrees	Edge adjacency indices	2D	EC ₅₀
BEHe2	highest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2D	EC ₅₀
JGI4	mean topological charge index of order 4	Topological charge indices	2D	EC ₅₀
RDF130v	Radial Distribution Function - 13.0 / weighted by atomic van der Waals volumes	RDF Descriptors	3D	EC ₅₀
HTu	H total index / unweighted	GETAWAY Descriptors	3D	EC ₅₀
R7e	R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY Descriptors	3D	EC ₅₀
BEHm8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic atomic masses	Burden eigenvalues	2D	CC ₅₀
BELe4	lowest eigenvalue n. 4 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2D	CC ₅₀
RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF Descriptors	3D	CC ₅₀
RDF135m	Radial Distribution Function - 13.5 / weighted by atomic masses	RDF Descriptors	3D	CC ₅₀
RDF150m	Radial Distribution Function - 15.0 / weighted by atomic masses	RDF Descriptors	3D	CC ₅₀
G1e	1st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM Descriptors	3D	CC ₅₀
Ks	K global shape index / weighted by atomic electrotopological states	WHIM Descriptors	3D	CC ₅₀
H7v	H autocorrelation of lag 7 / weighted by atomic van der Waals volumes	GETAWAY Descriptors	3D	CC ₅₀
R8v+	R maximal autocorrelation of lag 8 / weighted by atomic van der Waals volumes	GETAWAY Descriptors	3D	CC ₅₀
R3p+	R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY Descriptors	3D	CC ₅₀

En lo que se refiere a la técnica de modelación, se optó por un enfoque basado en técnicas de RLM; en este caso, los coeficientes de regresión y los parámetros estadísticos fueron obtenidos mediante el módulo de Hai Pham The- Aplicación de la Metodología MOOP-DESIRE ...

regresión general múltiple (RGM) implementado en el software STATISTICA 7.0 (180). Para cada MP, la bondad de ajuste se evaluó examinando el coeficiente de determinación (R^2), el coeficiente de determinación ajustado ($\text{Adj.}R^2$), 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, conocido como el estadístico ρ . La robustez y la capacidad predictiva del modelo fueron determinadas considerando el coeficiente de determinación de la validación cruzada dejando un caso fuera de cada MP (Q^2_{LOO}) (211). El procedimiento de la validación cruzada dejando un caso fuera consiste en quitar un caso de la serie de entrenamiento y reconstruir el modelo sólo con la serie de entrenamiento remanente y luego predecir el caso removido. Todos los casos de la serie de entrenamiento fueron predichos y de este modo determinado el Q^2_{LOO} .

El dominio de aplicación de los MPs se determinó mediante un ploteo de los valores de los residuales estandarizados contra los valores de *leverage* de los compuestos del entrenamiento (211, 212). El *leverage* (h) de un compuesto mide la influencia sobre el modelo del espacio de variables originales incluidas en el MP y se define como:

$$h_i = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \quad (i = 1, \dots, n) \quad (2.2.17)$$

donde \mathbf{x}_i es el vector de descriptores del compuesto considerado y \mathbf{X} es la matriz de los valores de descriptores del modelo derivada de la serie de entrenamiento. El *leverage* crítico h^* se define a continuación:

$$h^* = 3 \times p' / n \quad (2.2.18)$$

Aquí n es el número de casos usados en el entrenamiento y p' es el número de parámetros ajustables incluidos en el modelo.

3 RESULTADOS Y DISCUSIÓN

3. RESULTADOS Y DISCUSIÓN

“Following the light of the sun, we left the old world.”
Christopher Columbus

3.1. OPTIMIZACIÓN SIMULTÁNEA DE LA ACITIVIDAD INHIBITORIA SOBRE LA TI DEL VIH-1 Y LA TOXICIDAD SOBRE CÉLULAS SANGUÍNEAS MT-4 DE COMPUESTOS TIPO INNTI

En este trabajo se propone el empleo de los resultados del método de OMO propuesto anteriormente (213) como base para la aplicación de un nuevo método de ordenamiento también basado en las funciones de deseabilidad de Derringer (178). Para probar la utilidad de la metodología propuesta, la misma fue aplicada a una biblioteca de 109 compuestos reportados en literatura (72, 188-190) con el objetivo de optimizar simultáneamente sus propiedades antiretroviral frente a VIH-1 (EC_{50}) y tóxicas sobre células MT-4 (CC_{50}); ordenar los compuestos incluidos en esta librería con respecto al candidato ideal determinado por medio del proceso de optimización simultánea y filtrar los candidatos más prometedores de la misma.

Se generó un MP basado en RLM para cada propiedad conteniendo un subconjunto de descriptores moleculares implementados en software DRAGON 7.0 previamente seleccionados por métodos de algoritmos genéticos.

Tabla 3.1 Coeficientes de la regresión y parámetros estadísticos de los modelos RLM obtenidos.

Modelo RLM de Actividad Anti-VIH-1 ($EC_{50} = 1/1+EC_{50}$)									
$1/1 + EC_{50} = -13.6517(\pm 3.307011) + 0.0097(\pm 0.000942) \cdot pilD + 1.0075(\pm 0.167701) \cdot MATS8e$ $+ 0.7729(\pm 0.165765) \cdot GATS3m + 0.4509(\pm 0.154286) \cdot EEig09x - 0.9091(\pm 0.152694) \cdot EEig10x$ $+ 3.9872(\pm 0.88408) \cdot BEHe2 - 42.1335(\pm 4.842091) \cdot JGI4 - 0.1333(\pm 0.021748) \cdot RDF130v$ $- 0.0666(\pm 0.00748) \cdot HTu + 1.3208(\pm 0.151555) \cdot R7e$									
<i>N</i>	<i>R</i>	<i>R</i> ²	<i>R</i> ² Adj.	<i>S</i>	<i>Q</i> ²	<i>SPRESS</i>	ρ	<i>F</i>	<i>p</i>
109	0.890	0.793	0.772	0.166	0.745	0.184	9.909	37.516	0.0000
Modelo RLM de Toxicidad sobre Células Sanguíneas MT4 ($CC_{50} = (CC_{50}+0.5)^{0.5}$)									
$\sqrt{0.5 + CC_{50}} = 75.931(\pm 9.29826) - 15.783(\pm 2.41401) \cdot BEHm8 - 12.364(\pm 3.64943) \cdot BELe4$ $+ 0.797(\pm 0.12065) \cdot RDF030m + 1.404(\pm 0.27743) \cdot RDF135m - 4.123(\pm 1.22895) \cdot RDF150m$ $- 95.007(\pm 30.53947) \cdot Gle + 14.356(\pm 1.97627) \cdot Ks + 18.96(\pm 5.96789) \cdot H7v$ $- 716.154(\pm 80.40825) \cdot R8v^+ + 149.816(\pm 34.72703) \cdot R3p^+$									
<i>N</i>	<i>R</i>	<i>R</i> ²	<i>R</i> ² Adj.	<i>S</i>	<i>Q</i> ²	<i>SPRESS</i>	ρ	<i>F</i>	<i>p</i>
109	0.839	0.705	0.674	2.196	0.645	4.197	9.909	23.372	0.0000

Como se puede apreciar en la Tabla 3.1, 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

0.705 a 0.793 y 0.674 a 0.772, respectivamente; así como los valores Q^2_{LOO} entre 0.645 y 0.745. Por otra parte todas las variables incluidas en los modelos tienen una influencia significativa sobre las respectivas actividades, como se muestra en la Figura 3.2.

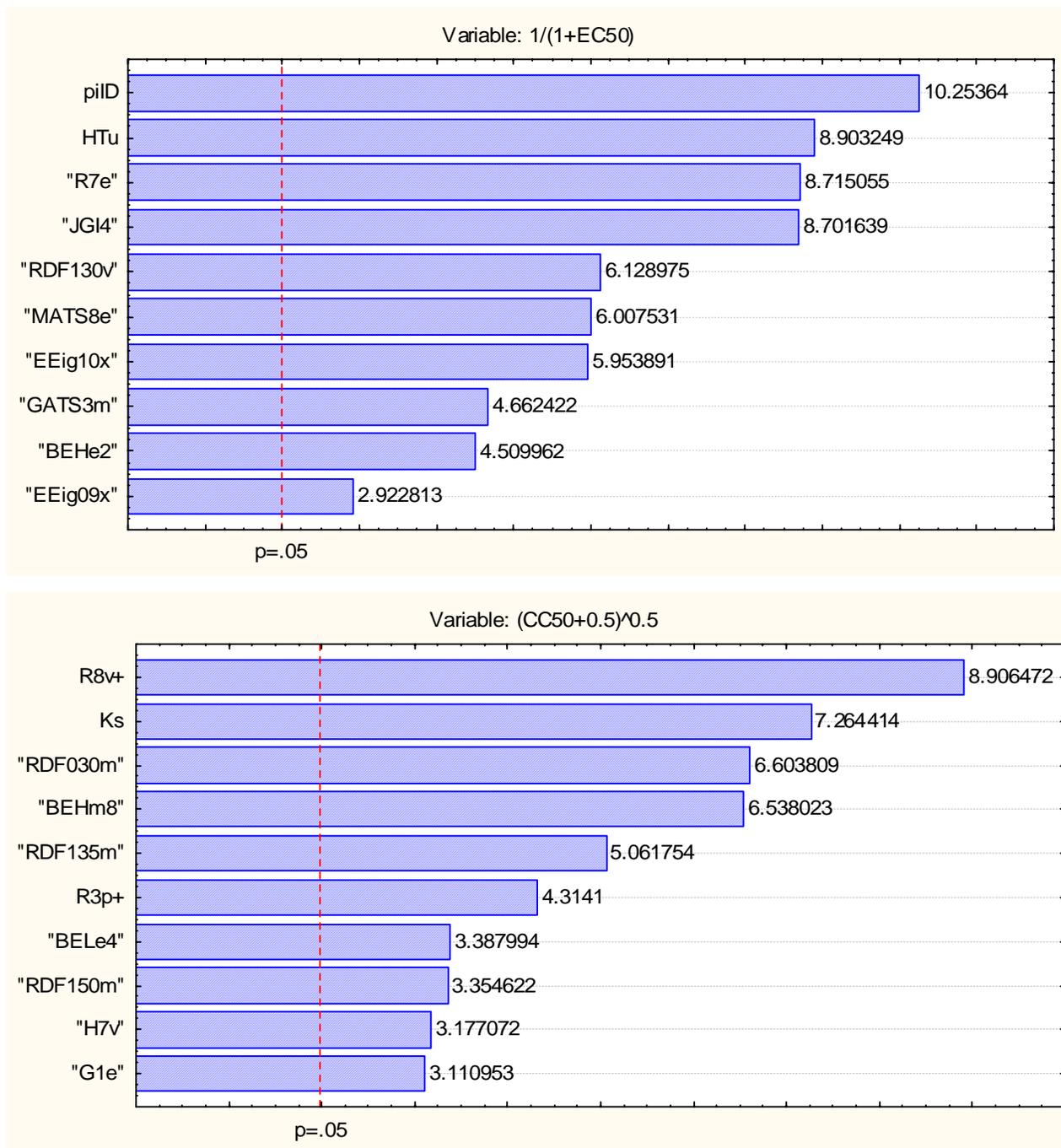


Figura 3.1. Tabla pareto de los valores t (test de Student de significación estadística) relativos a las variables incluidas en los modelos RLM de las propiedades antiVIH-1 (arriba) y citotóxicas (debajo) de la serie de los INNTIs usados en el estudio.

Cada MP desarrollado incluye once parámetros ajustables (diez descriptores moleculares y el intercepto) para ajustar los 109 compuestos empleados en el entrenamiento. Por tanto, cualquier predicción basada en

cualquiera de los dos MPs desarrollados de nuevos compuestos con valores de $h > h^* = 0.303$ está fuera del dominio de aplicabilidad del modelo y consecuentemente tales predicciones deben considerarse con precaución. Los dominios de aplicación de los MPs para la actividad antiVIH-1 (EC_{50}) y la toxicidad sobre células MT-4 (CC_{50}), respectivamente se muestran en la Figura 3.2.

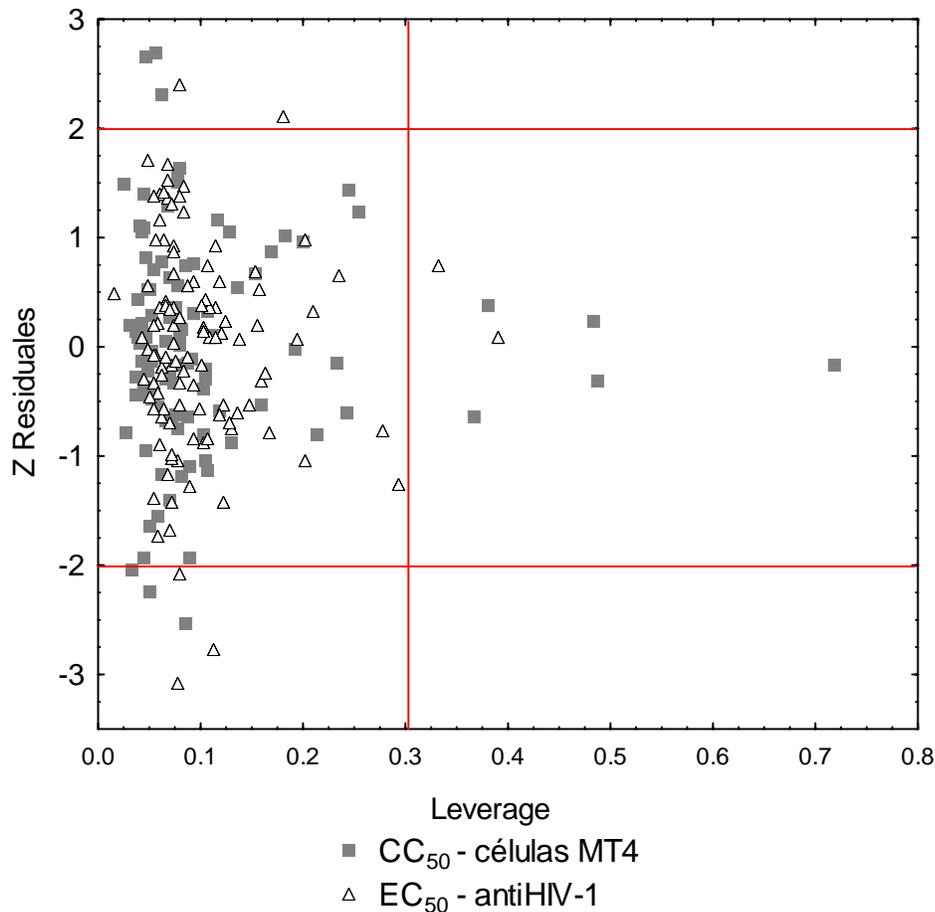
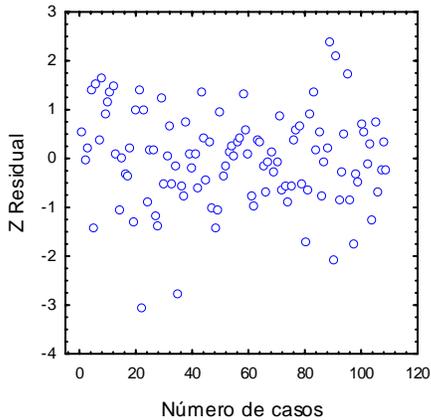
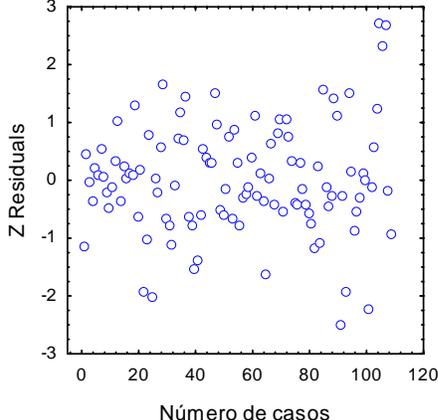
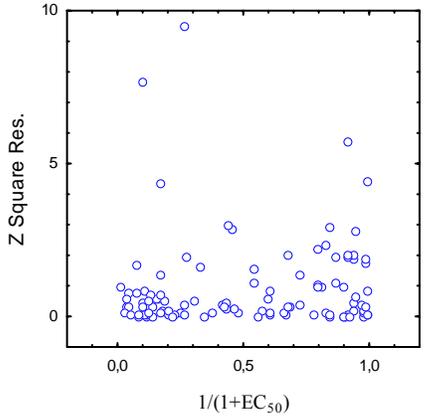
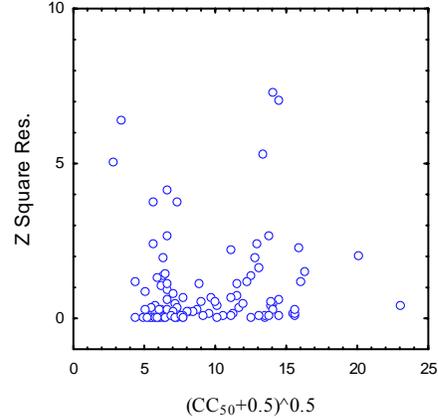
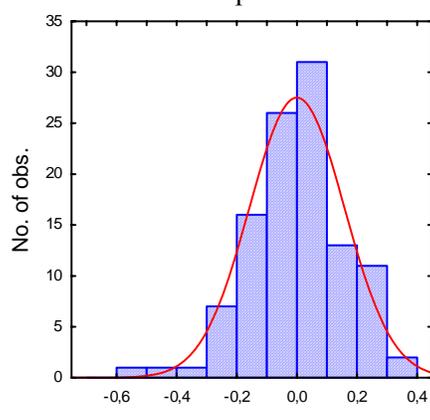
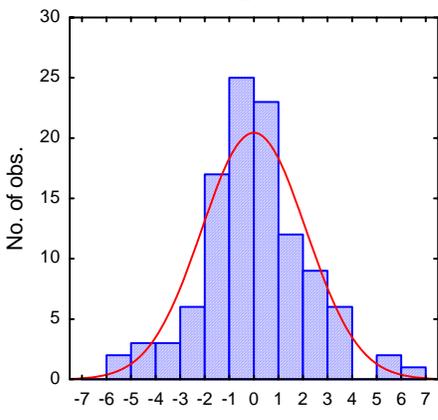


Figura 3.2. Dominios de aplicación de los respectivos modelos RLM.

Otro factor que determina la fiabilidad de las predicciones está relacionado con la técnica de modelación usada, específicamente, el método de estimación de parámetros usado (mínimos cuadrados ordinarios). Como en cualquier método paramétrico, es necesario hacer varias suposiciones para la estimación de los parámetros, los denominados supuestos paramétricos. Violaciones severas de uno o varios de estos supuestos podrían comprometer la fiabilidad de las predicciones resultantes de tales técnicas (214).

Primeramente se comprobó la hipótesis de linealidad mediante la inspección de los residuales estandarizados de los casos incluidos en la serie de entrenamiento. El hecho de que ningún ploteo de residuales (ver Tabla 3.2, fila 1) muestre un patrón específico de distribución refuerza la hipótesis formulada a priori de que ninguno de los dos modelos exhiben una dependencia no lineal entre las variables independientes y la variable dependiente (214).

Tabla 3.2. Chequeo de los principales supuestos paramétricos relacionados con los dos MPs basados en RLM usados para ajustar las funciones de deseabilidad.

Supuestos Paramétricos de la RLM		
	Modelo RLM (EC_{50})	Modelo RLM (CC_{50})
Relación Lineal entre la Variable Dependiente (Y) y las Variables Independientes (X)		
Homogeneidad de Varianzas		
	K-S d = 0.04052, p > 0.20 Lilliefors p > 0.20	K-S d = 0.06401, p > 0.20 Lilliefors p > .20
Residuales Normalmente Distribuidos		
	X <= Category Boundary	X <= Category Boundary
Media Res. = 0	0.000000	0.000000
Nº Colinealidad	Mayor $R_{X_iX_j} = 0.94$	Mayor $R_{X_iX_j} = 0.73$

La hipótesis de homoscedasticidad (homogeneidad de varianzas) puede ser confirmada mediante el ploteo de los residuales estandarizados contra las respectivas propiedades o variables dependientes (214) (ver Tabla 3.2.3, fila 2). Otra vez ningún ploteo muestra patrones de distribución específico que indique presencia de heteroscedasticidad. Este tipo de ploteo provee información sobre la autocorrelación de los residuales lo cual no se evidencia tampoco en estos gráficos.

La hipótesis de normalidad de los residuales puede ser fácilmente confirmada mediante el empleo del test estadísticos de normalidad de Kolmogorov-Smirnov (ver Tabla 3.2, fila 3). El valor de $p > 0.05$ mostrado sugiere que no existen razones significativas para rechazar la hipótesis de que los residuales están normalmente distribuidos. Además, debido a que el término relativo al error (representado por los residuales) no se incluye en el modelo RLM, este debe ser cero lo que ocurre en realidad como puede ser comprobado en la Tabla 3.2 (fila 4).

Un último aspecto que merece especial atención dentro del chequeo de los supuestos paramétricos es el análisis del grado de colinealidad exhibido entre las variables incluidas en los modelos. Variables altamente colineales pueden ser identificadas como aquellas que presenta un coeficiente de correlación por pares (R_{ij}) mayor que 0.7. Como puede comprobarse en la Tabla 3.2, (fila 5), las variables incluidas en los modelos exhiben un alto grado de colinealidad. En este caso la interpretación común de los coeficientes de regresión como la medida del cambio en el valor esperado de la variable respuesta, cuando una variable predictiva dada es aumentada en una unidad mientras que el resto son fijadas a un valor constante, no es totalmente aplicable. De aquí la importancia del cumplimiento de esta suposición si se pretende interpretar o asumir una relación causal entre los términos de la ecuación (215). No obstante, como el objetivo principal de estos MPs es de carácter predictivo, la violación de este supuesto no invalida su uso.

Los resultados obtenidos hasta el momento indican que la optimización simultánea de las propiedades antiretroviral frente a VIH-1 (EC_{50}) y tóxicas sobre células MT-4 del conjunto de 109 INNTIs puede ser llevado a cabo con un nivel adecuado de confiabilidad siempre considerando el dominio de aplicabilidad de los MPs desarrollados.

Tabla 3.3. Valores observados y predichos de las propiedades optimizadas, valores respectivos de deseabilidad individual y valores de deseabilidad global de los compuestos usados para el proceso de optimización multiobjetivos.

Comp. ID	$1/(1+EC_{50})$	Pred. $1/(1+EC_{50})$	$d(1/(1+EC_{50}))$	Pred. $d(1/(1+EC_{50}))$	$\sqrt{0.5+CC_{50}}$	Pred. $\sqrt{0.5+CC_{50}}$	$d\sqrt{0.5+CC_{50}}$	Pred. $d\sqrt{0.5+CC_{50}}$	$D_{CC50-EC50}$	Pred. $D_{CC50-EC50}$
1	0,685	0,593	0,681	0,587	6,332	8,927	0,175	0,303	0,345	0,422
2	0,847	0,853	0,847	0,852	7,28	6,322	0,222	0,174	0,433	0,385
3	0,667	0,63	0,662	0,625	6,267	6,341	0,172	0,175	0,337	0,331
4	0,869	0,637	0,869	0,632	6,237	7,094	0,17	0,213	0,385	0,367
5	0,676	0,911	0,671	0,911	5,897	5,46	0,153	0,132	0,321	0,346
6	0,826	0,573	0,825	0,567	5,406	5,209	0,129	0,119	0,326	0,26
7	0,662	0,602	0,658	0,596	6,237	5,083	0,17	0,113	0,335	0,26
8	0,946	0,67	0,947	0,665	5,805	5,685	0,149	0,143	0,375	0,308
9	0,606	0,454	0,6	0,445	6,007	6,496	0,159	0,183	0,309	0,285
10	0,725	0,531	0,721	0,524	5,766	6,851	0,147	0,201	0,326	0,324
11	0,941	0,715	0,942	0,712	5,983	6,31	0,158	0,174	0,385	0,352
12	0,8	0,556	0,798	0,549	5,486	4,774	0,133	0,098	0,326	0,232
13	0,35	0,335	0,339	0,324	6,133	3,862	0,165	0,053	0,237	0,131
14	0,546	0,719	0,54	0,716	5,75	6,611	0,146	0,189	0,281	0,367
15	0,119	0,115	0,105	0,1	6,051	5,565	0,161	0,137	0,13	0,117
16	0,253	0,308	0,24	0,296	5,75	5,695	0,146	0,143	0,187	0,206
17	0,826	0,883	0,825	0,882	5,654	5,417	0,141	0,13	0,342	0,338
18	0,916	0,883	0,916	0,882	5,609	5,419	0,139	0,13	0,357	0,338
19	0,074	0,287	0,058	0,275	13,131	10,272	0,511	0,37	0,172	0,319
20	0,901	0,738	0,901	0,735	5,772	7,224	0,147	0,219	0,364	0,401
21	0,937	0,703	0,938	0,699	5,543	5,191	0,136	0,118	0,357	0,288
22	0,265	0,775	0,253	0,773	5,648	9,966	0,141	0,355	0,189	0,524
23	0,813	0,649	0,811	0,644	8,807	11,147	0,297	0,413	0,491	0,516
24	0,043	0,188	0,027	0,174	14,512	12,763	0,58	0,493	0,125	0,293
25	0,229	0,2	0,216	0,187	6,534	11,073	0,185	0,409	0,2	0,276
26	0,606	0,574	0,6	0,568	12,45	12,409	0,478	0,476	0,536	0,52
27	0,176	0,371	0,162	0,36	7,751	8,27	0,245	0,271	0,199	0,312
28	0,274	0,504	0,262	0,497	6,066	4,849	0,162	0,102	0,206	0,225
29	0,546	0,341	0,54	0,33	13,705	10,051	0,54	0,359	0,54	0,344
30	0,68	0,767	0,676	0,765	7,115	8,634	0,214	0,289	0,38	0,47
$R^2_{D(CC50-EC50)} = 0,689$					$Adj.R^2_{D(CC50-EC50)} = 0,686$					

Tabla 3.3. (Continuación...)

Comp. ID	$1/(1+EC_{50})$	Pred. $1/(1+EC_{50})$	$d(1/(1+EC_{50}))$	Pred. $d(1/(1+EC_{50}))$	$\sqrt{0.5+CC_{50}}$	Pred. $\sqrt{0.5+CC_{50}}$	$d\sqrt{0.5+CC_{50}}$	Pred. $d\sqrt{0.5+CC_{50}}$	$D_{CC50-EC50}$	Pred. $D_{CC50-EC50}$
31	0,91	0,9	0,91	0,9	6,663	8,419	0,191	0,278	0,417	0,5
32	0,939	0,83	0,94	0,829	5,89	8,422	0,153	0,278	0,379	0,48
33	0,037	0,126	0,02	0,111	13,453	13,703	0,527	0,54	0,104	0,245
34	0,124	0,151	0,109	0,136	13,99	12,411	0,554	0,476	0,246	0,255
35	0,102	0,562	0,087	0,556	12,489	9,883	0,479	0,351	0,204	0,441
36	0,042	0,137	0,026	0,122	13,907	12,421	0,55	0,476	0,119	0,241
37	0,165	0,29	0,151	0,279	20,137	16,946	0,858	0,7	0,36	0,441
38	0,157	0,034	0,143	0,017	23,01	24,433	1	1,07	0,378	0,135
39	0,98	0,966	0,982	0,968	7,778	9,593	0,246	0,336	0,492	0,57
40	0,843	0,874	0,842	0,873	5,552	9,005	0,136	0,307	0,339	0,5179
41	0,903	0,89	0,904	0,89	6,29	9,441	0,173	0,329	0,395	0,541
42	0,27	0,37	0,258	0,36	5,474	6,843	0,132	0,2	0,185	0,268
43	0,915	0,685	0,915	0,681	14,05	12,894	0,557	0,5	0,714	0,583
44	0,942	0,874	0,943	0,873	15,411	14,596	0,624	0,584	0,767	0,714
45	0,577	0,647	0,571	0,642	14,429	13,77	0,575	0,543	0,573	0,591
46	0,38	0,321	0,37	0,31	13,494	12,885	0,529	0,499	0,443	0,393
47	0,794	0,963	0,792	0,964	15,827	12,46	0,645	0,478	0,714	0,679
48	0,917	1,153	0,918	1,158	6,537	4,407	0,185	0,08	0,412	0,304
49	0,865	1,038	0,864	1,041	8,654	9,83	0,29	0,348	0,501	0,602
50	0,796	0,635	0,794	0,63	11,616	12,958	0,436	0,503	0,589	0,563
51	0,484	0,544	0,476	0,537	6,484	6,873	0,182	0,202	0,295	0,329
52	0,779	0,802	0,777	0,8	13,943	12,298	0,551	0,47	0,654	0,613
53	0,991	0,971	0,993	0,972	11,987	13,481	0,455	0,529	0,672	0,717
54	0,991	0,946	0,993	0,947	11,554	9,629	0,433	0,338	0,656	0,566
55	0,898	0,887	0,898	0,887	13,727	13,053	0,541	0,507	0,697	0,671
56	0,982	0,922	0,984	0,923	9,661	11,469	0,34	0,429	0,578	0,629
57	0,981	0,908	0,983	0,908	15,636	16,33	0,635	0,67	0,79	0,78
58	0,988	0,77	0,99	0,768	13,081	13,654	0,509	0,537	0,71	0,642
59	0,973	0,874	0,974	0,874	5,714	6,032	0,144	0,16	0,375	0,374
60	0,925	0,912	0,926	0,912	15,604	14,761	0,634	0,592	0,766	0,735
61	0,943	1,073	0,944	1,076	16,026	13,571	0,654	0,533	0,786	0,757
$R^2_{D(CC50-EC50)} = 0,689$					$Adj.R^2_{D(CC50-EC50)} = 0,686$					

Tabla 3.3. (Continuación...)

Comp. ID	$1/(1+EC_{50})$	Pred. $1/(1+EC_{50})$	$d(1/(1+EC_{50}))$	Pred. $d(1/(1+EC_{50}))$	$\sqrt{0.5+CC_{50}}$	Pred. $\sqrt{0.5+CC_{50}}$	$d\sqrt{0.5+CC_{50}}$	Pred. $d\sqrt{0.5+CC_{50}}$	$D_{CC50-EC50}$	Pred. $D_{CC50-EC50}$
62	0,017	0,181	0	0,167	10,559	11,175	0,384	0,414	0	0,263
63	0,182	0,118	0,168	0,103	6,395	6,146	0,178	0,166	0,173	0,131
64	0,192	0,134	0,179	0,119	9,566	10,382	0,335	0,375	0,245	0,211
65	0,2	0,227	0,187	0,214	6,595	10,26	0,188	0,369	0,187	0,281
66	0,189	0,304	0,175	0,293	10,124	10,066	0,363	0,36	0,252	0,324
67	0,143	0,157	0,128	0,143	10,173	8,775	0,365	0,296	0,216	0,206
68	0,115	0,09	0,1	0,074	8,191	9,179	0,267	0,316	0,163	0,153
69	0,111	0,156	0,096	0,141	11,068	9,251	0,409	0,319	0,198	0,212
70	0,088	0,101	0,073	0,085	11,554	9,213	0,433	0,317	0,178	0,165
71	0,079	-0,065	0,064	-0,084	6,595	7,819	0,188	0,248	0,109	0
72	0,102	0,21	0,087	0,197	6,595	4,247	0,188	0,072	0,128	0,119
73	0,116	0,21	0,101	0,196	8,972	7,293	0,306	0,222	0,176	0,209
74	0,106	0,255	0,091	0,243	9,083	8,359	0,311	0,275	0,168	0,259
75	0,104	0,197	0,089	0,184	11,212	12,084	0,416	0,459	0,192	0,291
76	0,2	0,136	0,187	0,121	8,438	9,43	0,279	0,328	0,228	0,2
77	0,417	0,316	0,407	0,305	11,086	10,415	0,41	0,377	0,409	0,339
78	0,435	0,325	0,426	0,314	7,211	7,593	0,218	0,237	0,305	0,273
79	0,143	0,231	0,128	0,218	7,969	8,963	0,256	0,305	0,181	0,258
80	0,455	0,734	0,446	0,731	7,314	8,609	0,224	0,288	0,316	0,459
81	0,725	0,829	0,721	0,828	10,025	11,717	0,358	0,441	0,508	0,604
82	0,993	0,841	0,995	0,84	6,442	9,102	0,18	0,312	0,424	0,512
83	0,99	0,762	0,992	0,759	4,301	3,801	0,074	0,05	0,272	0,194
84	0,992	0,96	0,994	0,962	4,301	6,742	0,074	0,195	0,272	0,433
85	0,99	0,902	0,992	0,902	12,981	9,507	0,504	0,332	0,707	0,547
86	0,602	0,73	0,597	0,727	7,327	7,61	0,224	0,238	0,366	0,416
87	0,556	0,569	0,549	0,563	7,063	8,131	0,211	0,264	0,34	0,385
88	0,27	0,232	0,258	0,219	6,826	7,49	0,199	0,232	0,227	0,226
89	0,917	0,52	0,918	0,513	12,82	9,715	0,496	0,342	0,675	0,419
90	0,175	0,522	0,161	0,515	12,199	9,749	0,465	0,344	0,274	0,421
91	0,998	0,65	1	0,645	3,363	9,009	0,028	0,307	0,167	0,445
92	0,136	0,276	0,122	0,264	7,71	8,357	0,243	0,275	0,172	0,269

$$R^2_{D(CC50-EC50)} = 0,689$$

$$Adj.R^2_{D(CC50-EC50)} = 0,686$$

Tabla 3.3. (Continuación...)

Comp. ID	$1/(1+EC_{50})$	Pred. $1/(1+EC_{50})$	$d(1/(1+EC_{50}))$	Pred. $d(1/(1+EC_{50}))$	$\sqrt{0.5+CC_{50}}$	Pred. $\sqrt{0.5+CC_{50}}$	$d\sqrt{0.5+CC_{50}}$	Pred. $d\sqrt{0.5+CC_{50}}$	$D_{CC50-EC50}$	Pred. $D_{CC50-EC50}$
93	0,121	0,169	0,106	0,155	7,37	11,688	0,226	0,44	0,155	0,261
94	0,461	0,38	0,452	0,37	11,084	7,751	0,41	0,245	0,431	0,301
95	0,847	0,563	0,847	0,557	7,237	6,906	0,22	0,203	0,431	0,336
96	0,171	0,312	0,157	0,301	6,963	8,942	0,206	0,304	0,18	0,302
97	0,444	0,732	0,436	0,728	5,002	6,24	0,109	0,17	0,218	0,352
98	0,61	0,663	0,604	0,658	7,65	8,383	0,24	0,276	0,381	0,427
99	0,429	0,505	0,42	0,498	5,89	5,671	0,153	0,142	0,254	0,266
100	0,311	0,195	0,299	0,181	4,895	4,953	0,104	0,107	0,176	0,139
101	0,427	0,334	0,418	0,323	2,796	7,798	0	0,247	0	0,283
102	0,221	0,238	0,208	0,225	5,224	5,506	0,12	0,134	0,158	0,174
103	0,026	-0,027	0,009	-0,045	15,56	14,311	0,631	0,57	0,076	0
104	0,328	0,537	0,317	0,53	16,239	13,493	0,665	0,529	0,459	0,529
105	0,035	-0,089	0,018	-0,108	14,037	8,012	0,556	0,258	0,1	0
106	0,128	0,243	0,114	0,23	13,395	8,244	0,524	0,269	0,244	0,249
107	0,056	0,094	0,04	0,079	14,521	8,597	0,58	0,287	0,151	0,15
108	0,176	0,119	0,162	0,105	7,802	8,253	0,248	0,27	0,2	0,168
109	0,084	0,121	0,069	0,106	5,019	7,126	0,11	0,214	0,087	0,151
$R^2_{D(CC50-EC50)} = 0,689$					$Adj.R^2_{D(CC50-EC50)} = 0,686$					

Tabla 3.4. Valores predichos obtenidos durante el experimento de validación cruzada (dejando un caso fuera) de las propiedades optimizadas, valores respectivos de deseabilidad individual y valores de deseabilidad global para los compuestos usados para el proceso de optimización multiobjetivos.

Compuesto ID	VC_{L00} Pred. $1/1+EC_{50}$	VC_{L00} Pred. $d(1/(1+EC_{50}))$	VC_{L00} Pred. $\sqrt{0.5+CC_{50}}$	VC_{L00} Pred. $d\sqrt{0.5+CC_{50}}$	VC_{L00} Pred. $D_{CC50-EC50}$
1	0,588	0,582	9,098	0,312	0,426
2	0,853	0,852	6,284	0,173	0,383
3	0,627	0,622	6,345	0,176	0,33
4	0,622	0,617	7,142	0,215	0,364
5	0,929	0,929	5,446	0,131	0,349
6	0,555	0,548	5,201	0,119	0,255
7	0,597	0,591	5,022	0,11	0,255
8	0,649	0,645	5,677	0,142	0,303
9	0,442	0,433	6,522	0,184	0,282
10	0,518	0,511	6,914	0,204	0,323
11	0,699	0,695	6,342	0,175	0,349
12	0,534	0,527	4,689	0,094	0,222
13	0,334	0,323	3,356	0,028	0,095
14	0,734	0,73	6,657	0,191	0,374
15	0,115	0,1	5,543	0,136	0,116
16	0,312	0,301	5,692	0,143	0,208
17	0,886	0,886	5,405	0,129	0,338
18	0,881	0,88	5,41	0,129	0,337
19	0,308	0,297	10,065	0,36	0,327
20	0,727	0,723	7,362	0,226	0,404
21	0,687	0,683	5,16	0,117	0,283
22	0,818	0,817	10,389	0,376	0,554
23	0,639	0,635	11,419	0,427	0,520
24	0,205	0,191	12,648	0,487	0,305
25	0,197	0,183	11,23	0,417	0,276
26	0,571	0,565	12,406	0,475	0,518
27	0,385	0,375	8,309	0,273	0,32
28	0,518	0,51	4,658	0,092	0,217
29	0,322	0,311	9,732	0,343	0,327
30	0,782	0,78	8,741	0,294	0,479
31	0,897	0,897	8,468	0,281	0,502
32	0,797	0,795	8,727	0,293	0,483
33	0,134	0,119	13,729	0,541	0,254
34	0,153	0,138	12,321	0,471	0,255
35	0,621	0,616	9,537	0,333	0,453
36	0,147	0,133	12,15	0,463	0,248
37	0,309	0,298	15,913	0,649	0,439
38	-0,027	0	25,258	1	0
$Q^2_{D(CC50-EC50)} = 0,619$					

Tabla 3.4. (Continuación...)

Compuesto ID	VC_{Loo} Pred. $1/1+EC_{50}$	VC_{Loo} Pred. $d(1/(1+EC_{50}))$	VC_{Loo} Pred. $\sqrt{0.5+CC_{50}}$	VC_{Loo} Pred. $d\sqrt{0.5+CC_{50}}$	VC_{Loo} Pred. $D_{CC50-EC50}$
39	0,957	0,958	10,087	0,361	0,588
40	0,876	0,875	9,217	0,318	0,527
41	0,889	0,889	9,681	0,341	0,55
42	0,386	0,376	6,956	0,206	0,278
43	0,672	0,668	12,835	0,497	0,576
44	0,869	0,869	14,529	0,58	0,71
45	0,651	0,647	13,733	0,541	0,592
46	0,313	0,302	12,839	0,497	0,387
47	0,976	0,978	12,179	0,464	0,674
48	1,186	1	3,877	0,053	0,231
49	1,082	1	10,054	0,359	0,599
50	0,594	0,588	13,39	0,524	0,555
51	0,55	0,543	7,862	0,251	0,369
52	0,803	0,802	12,142	0,462	0,609
53	0,968	0,97	13,595	0,534	0,72
54	0,943	0,944	9,24	0,3189	0,548
55	0,886	0,886	12,985	0,504	0,668
56	0,915	0,915	11,676	0,439	0,634
57	0,899	0,899	16,988	0,702	0,795
58	0,754	0,751	13,697	0,539	0,636
59	0,864	0,863	6,128	0,165	0,377
60	0,91	0,91	14,243	0,566	0,718
61	1,099	1	13,464	0,528	0,726
62	0,193	0,18	11,198	0,416	0,273
63	0,111	0,096	6,124	0,165	0,125
64	0,13	0,115	10,421	0,377	0,209
65	0,23	0,217	10,459	0,379	0,287
66	0,313	0,301	10,061	0,359	0,329
67	0,159	0,145	8,671	0,291	0,205
68	0,087	0,071	9,228	0,318	0,151
69	0,158	0,144	9,161	0,315	0,213
70	0,101	0,086	9,109	0,312	0,164
71	-0,076	0	7,894	0,252	0
72	0,217	0,204	3,901	0,054	0,106
73	0,216	0,203	7,122	0,214	0,208
74	0,265	0,253	8,305	0,273	0,262
75	0,202	0,189	12,184	0,464	0,296
76	0,131	0,117	9,468	0,33	0,196
77	0,303	0,291	10,362	0,374	0,330
78	0,316	0,305	7,626	0,239	0,27

$$Q^2_{D(CC50-EC50)} = 0,619$$

Tabla 3.4. (Continuación...)

Compuesto ID	VC_{Loo} Pred. $1/1+EC_{50}$	VC_{Loo} Pred. $d(1/(1+EC_{50}))$	VC_{Loo} Pred. $\sqrt{0.5+CC_{50}}$	VC_{Loo} Pred. $d\sqrt{0.5+CC_{50}}$	VC_{Loo} Pred. $D_{EC_{50}-CC_{50}}$
79	0,243	0,231	9,018	0,308	0,267
80	0,755	0,753	8,784	0,296	0,472
81	0,843	0,842	11,859	0,448	0,614
82	0,821	0,82	9,34	0,324	0,515
83	0,742	0,739	3,333	0,027	0,14
84	0,955	0,956	6,98	0,207	0,445
85	0,885	0,885	9,217	0,318	0,53
86	0,778	0,776	7,626	0,239	0,431
87	0,57	0,564	8,191	0,267	0,388
88	0,226	0,213	7,568	0,236	0,224
89	0,486	0,478	9,569	0,335	0,4
90	0,552	0,545	9,635	0,338	0,429
91	0,572	0,566	9,542	0,334	0,435
92	0,29	0,279	8,403	0,277	0,278
93	0,171	0,157	11,889	0,45	0,266
94	0,379	0,369	7,662	0,241	0,298
95	0,548	0,542	6,893	0,203	0,331
96	0,329	0,318	9,24	0,319	0,318
97	0,749	0,746	6,316	0,174	0,361
98	0,673	0,669	8,442	0,279	0,432
99	0,509	0,502	5,644	0,141	0,266
100	0,174	0,16	4,967	0,107	0,131
101	0,325	0,314	8,061	0,26	0,286
102	0,239	0,226	5,519	0,135	0,175
103	-0,04	0	14,206	0,564	0
104	0,623	0,618	12,553	0,483	0,546
105	-0,103	0	7,654	0,24	0
106	0,26	0,248	7,9	0,252	0,25
107	0,102	0,086	8,309	0,273	0,154
108	0,115	0,1	8,305	0,273	0,165
109	0,124	0,109	7,229	0,219	0,155
$Q^2_{D(EC_{50}-CC_{50})} = 0,619$					

Una vez ajustado el modelo conteniendo todas las variables independientes se logra especificar la función de deseabilidad para los valores predichos de cada propiedad. Aquí, la actividad antiVIH-1 (EC_{50}) es maximizada mientras que la toxicidad sobre las células MT4 (CC_{50}) es minimizada con el objetivo de obtener luego del proceso de optimización un candidato que exhiba simultáneamente la mayor actividad antiVIH-1 a la vez que la menor toxicidad posibles. Los valores de deseabilidad individual de ambas respuestas se estimaron tal y como se detalla en la sección de “Materiales y Métodos”. En la Tabla 3.3 se muestran los valores calculados y

predichos de deseabilidad debido a la actividad antiVIH-1 ($1/(1+EC_{50})$) y la toxicidad sobre las células MT4 ($(\sqrt{CC_{50} + 1/2})$), así como la deseabilidad global para las 109 INNTIs usados en el estudio. Los resultados de la validación cruzada se muestran en la Tabla 3.4.

Tabla 3.5. Resultados del proceso de optimización simultanea multiobjetivos y ajuste de pesos.

<i>Niveles Óptimos de las Variables</i>	
piID = 216.197	BEHm8 = 2.496
MATS8e = -0.242	BELe4 = 1.227
GATS3m = 0.963	RDF030m = 9.978
EEig09x = 3.153	RDF135m = -2.061
EEig10x = 2.922	RDF150m = -0.451
BEHe2 = 3.901	G1e = 0.143
JGI4 = 0.045	Ks = 0.217
RDF130v = 2.594	H7v = 0.015
HTu = 24.156	R8v⁺ = 0.006
R7e = 0.614	R3p⁺ = 0.012
<i>Conjunto de pesos w_i que maximizan la concordancia entre D_i y Δ_i</i>	
$W_{piID} = 0.0284$	$W_{BEHm8} = 0.7286$
$W_{MATS8e} = -2.626$	$W_{BELe4} = -1.2573$
$W_{GATS3m} = -0.4767$	$W_{RDF030m} = 0.235$
$W_{EEig09x} = 1.7003$	$W_{RDF135m} = -0.2961$
$W_{EEig10x} = -3.2139$	$W_{RDF150m} = 0.8532$
$W_{BEHe2} = 4.686$	$W_{G1e} = 23.4602$
$W_{JGI4} = -107.318$	$W_{Ks} = -0.6094$
$W_{RDF130v} = 0.1213$	$W_{H7v} = 0.6381$
$W_{HTu} = 0.2282$	$W_{R8v+} = 57.8155$
$W_{R7e} = -2.5136$	$W_{R3p+} = -6.2499$
$1/1+EC_{50} = 0.908$	$\sqrt[2]{CC_{50} + 1/2} = 15.797$
$\Rightarrow EC_{50} = 0.101\mu M$	$\Rightarrow CC_{50} = 249.045\mu M$
$D_{CC50-EC50} = 0.819$	

R^2_D y $Adj.R^2_D$ son usados aquí como medidas globales de la calidad estadística de la función de deseabilidad global. En la Tabla 3.3 se muestra que estos valores están próximos a un valor de 0.7 lo que indica una calidad adecuada del proceso de optimización simultánea. Por otra parte, un valor de 0.62 de Q^2_D indica una capacidad predictiva adecuada del método para predecir los valores de D .

Seguidamente, la optimización de la deseabilidad global se condujo de acuerdo a lo descrito en la sección “Especificaciones de la funciones de deseabilidad”. Como resultado del proceso de optimización es posible

determinar los niveles de los descriptores moleculares empleados que simultáneamente producen la combinación más deseable entre la actividad antiVIH-1 y la toxicidad sobre las células MT4, expresadas a través de la deseabilidad global.

Los resultados del proceso de optimización simultánea se muestran en la Tabla 3.5 donde se indican los valores de las variables predictivas necesarios para obtener un candidato de las familias de las INNTIs con el mejor compromiso posible entre las propiedades antiVIH-1 y citotóxicas ($D = 1.00$).

Como se describió anteriormente, es posible el empleo de la metodología MOOP-DESIRE para el ordenamiento y filtrado de librerías de compuestos. Aquí usamos la librería de INNTIs empleadas también para demostrar la utilidad del método de ordenamiento propuesto en este trabajo.

Una vez determinados los valores de las variables predictivas necesarios para la obtención del candidato ideal estos son usados como patrón para ordenar la librería de 109 INNTIs usadas en este trabajo.

Tabla 3.7. Valores de Δ_i de los compuestos usados en el ordenamiento.

Compuesto	Δ_i	$^D\Delta_i$	Pred. $D_{(CC50-EC50)}$	Compuesto	Δ_i	$^D\Delta_i$	Pred. $D_{(CC50-EC50)}$
1	2,318	0,439	0,422	32	2,396	0,421	0,48
2	2,197	0,467	0,385	33	1,942	0,526	0,245
3	2,574	0,38	0,331	34	2,135	0,481	0,255
4	2,586	0,377	0,367	35	3,748	0,108	0,441
5	1,913	0,532	0,346	36	1,966	0,52	0,241
6	3,380	0,194	0,26	37	1,113	0,717	0,442
7	2,89	0,307	0,26	38	3,916	0,07	0,135
8	2,92	0,3	0,308	39	2,113	0,486	0,57
9	3,238	0,226	0,285	40	1,477	0,633	0,518
10	2,931	0,297	0,324	41	1,640	0,595	0,541
11	3,068	0,266	0,352	42	2,444	0,41	0,268
12	4,217	0,000	0,232	43	1,099	0,721	0,583
13	3,664	0,128	0,131	44	0,772	0,796	0,714
14	2,297	0,444	0,367	45	1,087	0,723	0,591
15	3,841	0,087	0,117	46	1,513	0,625	0,393
16	3,569	0,15	0,206	47	0,617	0,832	0,679
17	2,083	0,493	0,338	48	1,740	0,572	0,304
18	2,083	0,493	0,338	49	1,475	0,634	0,602
19	2,771	0,334	0,319	50	0,562	0,845	0,563
20	3,025	0,276	0,401	51	2,482	0,401	0,329
21	3,603	0,142	0,288	52	0,819	0,785	0,613
22	1,907	0,534	0,524	53	0,882	0,771	0,717
23	1,185	0,701	0,516	54	2,275	0,449	0,566
24	2,120	0,485	0,293	55	0,695	0,814	0,671
25	2,336	0,435	0,276	56	1,046	0,733	0,629
26	1,553	0,616	0,52	57	-0,111	1	0,78
27	2,937	0,296	0,312	58	0,493	0,861	0,642
28	3,196	0,236	0,225	59	1,663	0,59	0,374
29	3,262	0,221	0,344	60	0,160	0,937	0,735
30	1,55	0,616	0,47	61	0,631	0,829	0,757
31	0,993	0,745	0,5	62	2,581	0,378	0,263

Tabla 3.7. (Continuación...)

Compuesto	Δ_i	${}^D\Delta_i$	Pred. $D_{(CC50-EC50)}$	Compuesto	Δ_i	${}^D\Delta_i$	Pred. $D_{(CC50-EC50)}$
63	3,318	0,208	0,131	86	1,958	0,522	0,416
64	2,762	0,336	0,211	87	2,021	0,507	0,385
65	2,553	0,384	0,281	88	2,566	0,382	0,226
66	1,958	0,522	0,324	89	1,561	0,614	0,419
67	3,284	0,216	0,206	90	1,55	0,616	0,421
68	3,337	0,203	0,153	91	2,146	0,479	0,445
69	3,273	0,218	0,212	92	2,72	0,346	0,269
70	3,349	0,2	0,165	93	3,034	0,273	0,261
71	3,677	0,125	0	94	2,53	0,390	0,301
72	3,530	0,159	0,119	95	2,328	0,436	0,336
73	2,884	0,308	0,209	96	2,349	0,432	0,302
74	2,937	0,296	0,259	97	2,296	0,444	0,352
75	2,754	0,338	0,291	98	2,08	0,494	0,427
76	3,316	0,208	0,2	99	3,07	0,265	0,266
77	2,430	0,413	0,339	100	3,78	0,101	0,139
78	2,926	0,298	0,273	101	3,158	0,245	0,283
79	2,309	0,441	0,258	102	3,882	0,077	0,174
80	1,614	0,601	0,459	103	3,495	0,167	0
81	1,024	0,738	0,604	104	2,06	0,498	0,529
82	0,587	0,839	0,512	105	3,325	0,206	0
83	3,471	0,172	0,194	106	2,473	0,403	0,249
84	1,298	0,674	0,433	107	2,1	0,489	0,15
85	0,882	0,771	0,547	108	2,714	0,347	0,168
				109	3,438	0,18	0,151

Luego de la aplicación del proceso de optimización para estimar el juego óptimo de ponderaciones que conducen a minimizar la diferencia entre la descripción de las moléculas (Δ_i) y la solución (D_i) de las mismas, los valores de Δ_i de los 109 INNTIs es calculado. Este valor es usado para ordenar los compuestos con respecto al candidato óptimo determinado mediante el proceso de optimización simultánea de las propiedades antiVIH-1 y citotóxicas. Los valores de calculados de D_i y Δ_i así como estos últimos valores normalizados (${}^D\Delta_i$) mediante el empleo de funciones de deseabilidad de los 109 compuestos usados en el trabajo se muestran en la Tabla 3.7.

Basado en Δ_i es posible obtener un ordenamiento de la librería de 109 INNTIs con un índice de calidad corregido Ψ^* de 0.238 lo que representa un porcentaje de calidad del ordenamiento ($R\%$) igual al 76,162%. El ordenamiento obtenido comparado con el ordenamiento ideal se muestra en la Figura 3.3.

Como se ha podido apreciar, la calidad del ordenamiento obtenido ($R\% = 76,162$) es similar a los valores de predictividad obtenidos tanto para los modelos individuales como para el proceso global ($Q^2_{EC50} = 0.745$, $Q^2_{CC50} = 0.645$, $Q^2_{D(CC50-EC50)} = 0,619$), lo que confirma que la calidad tanto del proceso de optimización múltiple como del ordenamiento depende de la calidad de los modelos iniciales. Lo anterior indica que el proceso de ordenamiento refleja la calidad de los modelos y procesos en los que se basa.

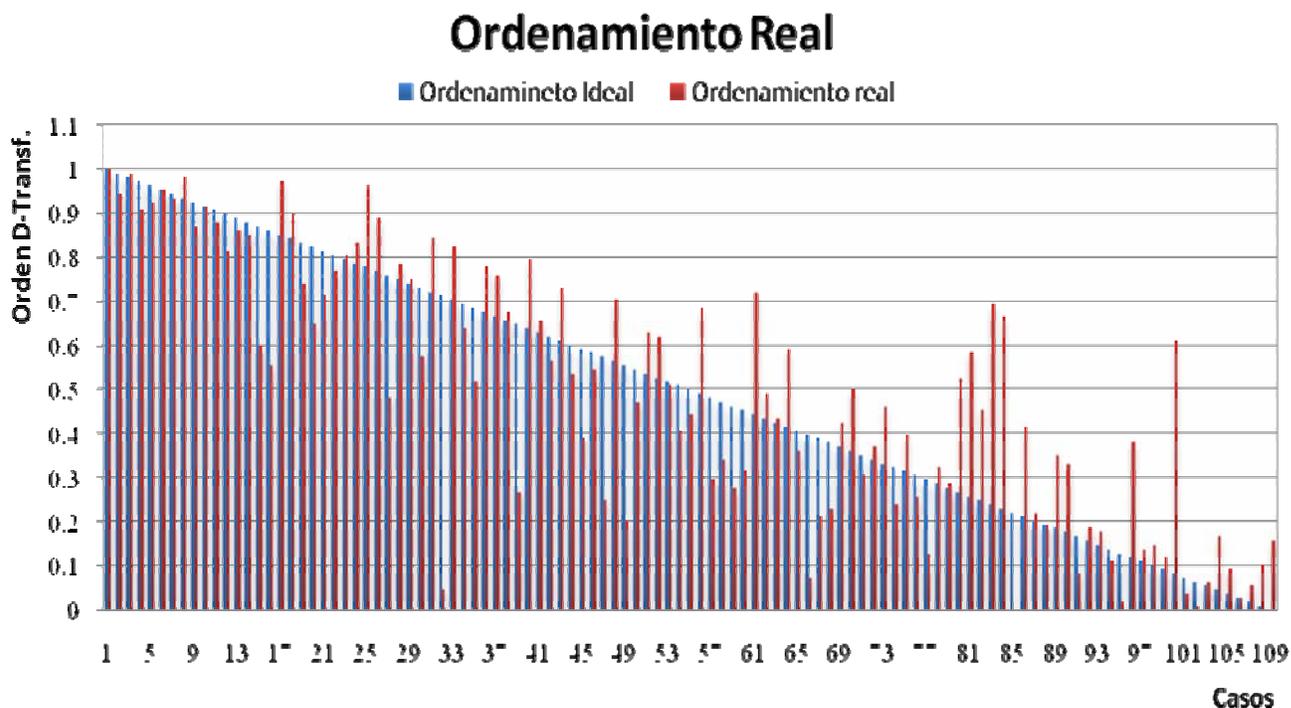


Figura 3.3. Ordenamiento de las 109 INNTIs en función de la su similitud con el candidato optimo (Δ_i).

Por otra parte, el objetivo de ordenar una librería de compuestos en función de su similitud con un candidato optimo es filtrar una parte de estos (los candidatos más cercanos y por tanto más semejantes al optimo), los que serian propuestos para ser sintetizados y probar de manera experimental su actividad inhibitoria sobre la TI del VIH-1 y su toxicidad sobre celulas MT4.

Consecuentemente, si tomáramos solamente el 10% (los mejores 11 compuestos) de nuestra base de datos para ser propuesta para su inclusión en el proceso de desarrollo de fármacos se incrementaría la probabilidad de encontrar candidatos prometedores ya que esta fracción presenta un índice de calidad corregido Ψ^* de 0.03 lo que representa un porcentaje de calidad del ordenamiento ($R_{\%}$) igual al 97.03%. El ordenamiento obtenido solo para esta fracción de la librería se muestra en la Figura 3.4.

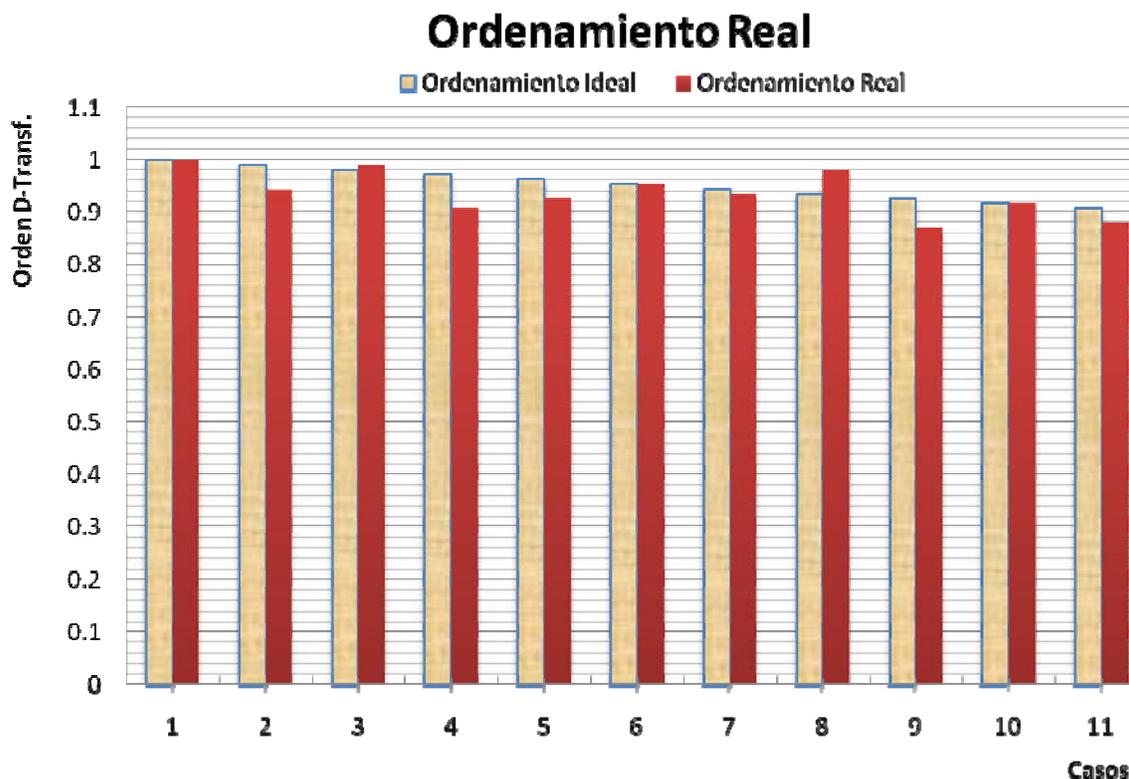


Figura 3.4. Ordenamiento obtenido para el 10% de la librería de compuestos.

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 y desarrollo de nuevos medicamentos 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 medicamento 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 (216). Igualmente, mediante el empleo de un jurado de modelos (MPs 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 más obra de la casualidad que fruto de una estrategia de diseño racional.

Como ha sido ilustrado, la metodología propuesta puede usarse como una estrategia racional para el filtrado a partir de librerías combinatorias de nuevos candidatos a fármaco; considerando únicamente aquellos candidatos incluidos dentro del dominio de aplicabilidad de los MPs en los que se basan los procesos de OMO y ordenamiento. En situaciones como esta, 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 en lugar de descriptores interpretables, 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 y desarrollo de nuevos 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 OMO. 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 farmacéutico global mejorado (ver referencia (213)). La Figura 3.5 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.

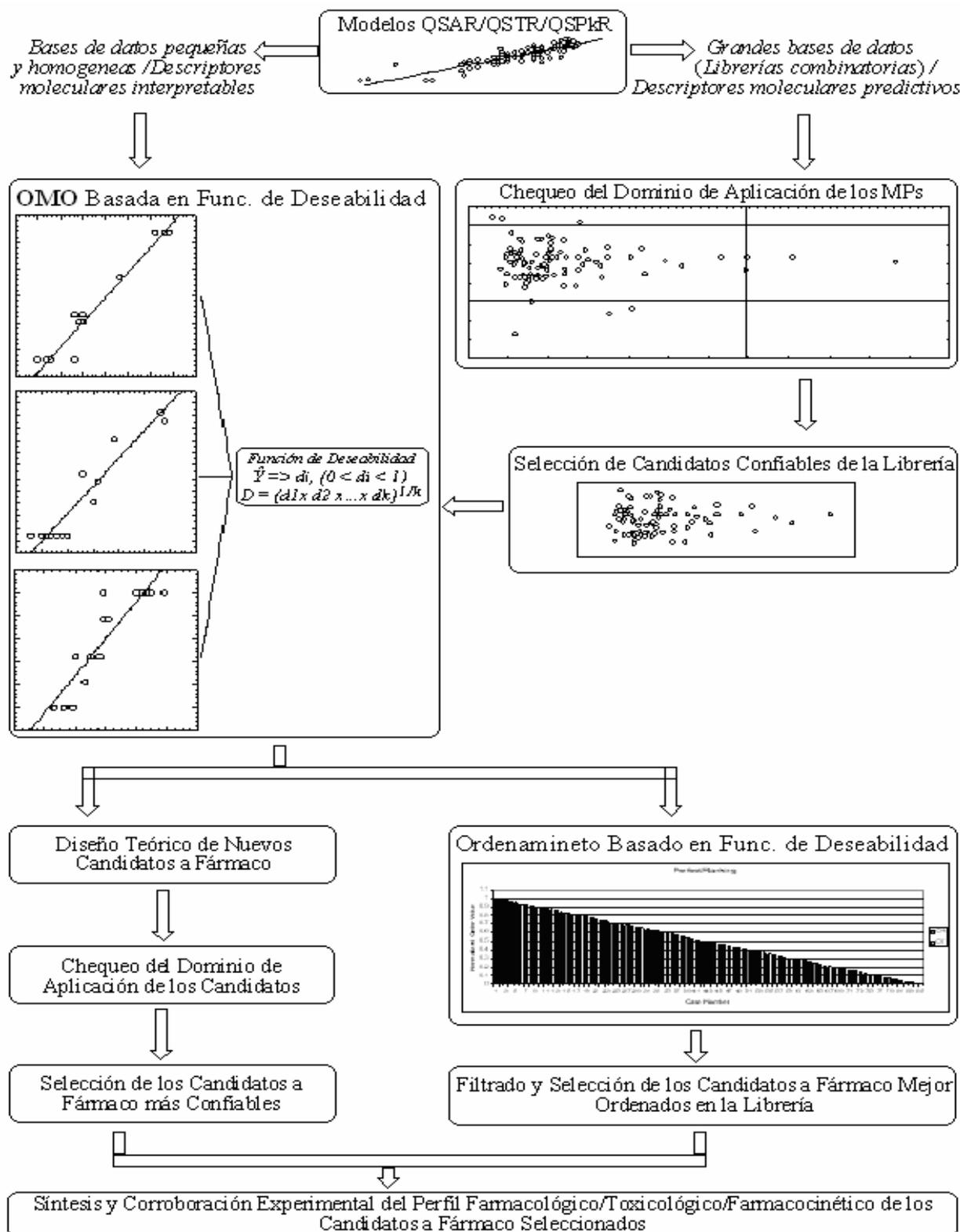


Figura 3.5. Descubrimiento y desarrollo racional de nuevos fármacos basado en la metodología MOOP-DESIRE.

CONCLUSIONES

CONCLUSIONES

“The important thing is not to stop questioning.”

Albert Einstein

Atendiendo a los resultados obtenidos se ha llegado a las siguientes conclusiones:

- Se Conformó una base de datos consistente, estructuralmente diversa y representativa de inhibidores no nucleosídicos de la TI del VIH-1 con información relativa a su actividad inhibitoria sobre la TI del VIH-1 y su toxicidad sobre células sanguíneas MT4.
- Se determinaron los niveles de las variables predictivas necesarios para obtener un análogo no nucleosídico con una combinación óptima entre su actividad inhibitoria sobre la TI del VIH-1 y su toxicidad sobre células sanguíneas MT4, mediante el empleo de un método de OMO basado en funciones de deseabilidad de Derringer.
- Se ordenaron los compuestos inhibidores no nucleosídicos de la TI del VIH-1 incluidos en la serie de entrenamiento con respecto a su similitud estructural con el candidato óptimo identificado durante el proceso de OMO mediante el empleo de un método de ordenamiento basado en funciones de deseabilidad.
- El criterio estructural de ordenamiento obtenido fue utilizado como esquema de ordenamiento y filtrado de nuevos candidatos tipo INNTI con perfiles de eficacia y seguridad simultáneamente optimizados.
- De manera general se propuso una estrategia de tamizaje virtual basada en estudios QSAR globales donde se consideró de forma simultánea la actividad inhibitoria sobre la TI del VIH-1 y la toxicidad sobre células sanguíneas MT4 de una familia estructuralmente diversa de análogos no nucleosídicos, lo que permite su uso en la identificación y selección racional de candidatos INNTI con perfiles de eficacia y seguridad simultáneamente optimizados.

RECOMENDACIONES

RECOMENDACIONES

- ✓ Emplear la metodología propuesta en la identificación y selección racional de candidatos INNTI con perfiles de eficacia y seguridad simultáneamente optimizados.
- ✓ Explorar el desempeño de técnicas no lineales de modelación, así como de métodos de optimización basados en algoritmos genéticos o evolutivos.
- ✓ Explorar el desempeño de otras medidas de similitud en el proceso de ordenamiento.

REFERENCIAS

BIBLIOGRÁFICAS

REFERENCIAS BIBLIOGRÁFICAS

1. Centers for Disease Control, (1981a). 1981 (Volltext: <http://hiv.net/link.php?id=144>).
2. Centers for Disease Control, (1981b). 1981.
3. Centers for Disease Control, (1981c). 1981.
4. Joan C. Effect of Thymoglobulin Induction on HIV-infected Renal Transplant Recipients: Differences between HIV-Positive and HIV-Negative Patients [serial on the Internet]. 2007; 23 (10).
5. Muñoz-Moreno JA, Carmina R, Fumaz MJ, Ferrer A. Assessing Self-Reported Adherence to HIV Therapy by Questionnaire: The SERAD (Self-Reported Adherence) Study [serial on the Internet]. 2007; 23 (10).
6. Barré-Sinoussi F, Cherman JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. . Isolation of a T-Lymphotropic Retrovirus from a patient at risk for Acquired Immune Deficiency Syndrome (AIDS) [serial on the Internet]. 1983; 220.
7. Broder S, Gallo RC. A pathogenic retrovirus (HTLV-III) linked to AIDS [serial on the Internet]. 1984; 311.
8. Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS [serial on the Internet]. 1984; 224.
9. Joint united nations. Programme on HIV/AIDS. <http://www.unaids.org>; AIDS epidemic update: 2007 [December].
10. UNAIDS/WHO. Epidemiological fact sheet on HIV/AIDS and sexually transmitted infections.: <http://www.who.ch/emc/diseases/hiv>; 2007 update.
11. Popovic MS, M. G.; Reed, E.; Gallo, R. C. . Detection, isolation and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and Pre-AIDS. Science1984(224):497-500.
12. Powell KL, Darby G. HIV reverse transcriptase as a target for antiviral drugs, in Design of Anti-AIDS drugs. Ed De Clercq, E Elsevier1990:123-40.
13. De Clercq E. Antiviral agents and human viral diseases. Philadelphia: 4th Ed. Lippincott Raven Publishers. ; 1997. p. 1-44.
14. De Clercq E. HIV inhibitors targeted at the reverse transcriptase [serial on the Internet]. 1992; 8.
15. Balzarini J, Karlsson A, Perez-perez M-J, Vrang L, Walbers J, Zhang H, et al. 1993; 192.
16. Ravindra K. Rawal, Rajkamal Tripathi, S. B. Katti, Christophe Pannecouque, and Erik De Clercq. Synthesis and evaluation of 2-(2,6-dihalophenyl)-3-pyrimidinyl-1,3-thiazolidin-4-one analogues as anti-HIV-1 agents [serial on the Internet]. 2007; 15.
17. Ekins S, Boulanger B, Swaan PW, Hupcey MA. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. J Comput Aided Mol Des2002;16(5-6):381-401.
18. Bleicher KH, Bohm HJ, Muller K, Alanine AI. Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discov2003 May;2(5):369-78.
19. Xu J, Hagler A. Chemoinformatics and Drug Discovery. Molecules2002;7:566-700.
20. Butina D, Segall MD, Frankcombe K. Predicting ADME properties in silico: methods and models. Drug Discov Today2002;7(11 Suppl):S83-S8.
21. Manly CJ, Louise-May S, Hammer JD. The impact of informatics and computational chemistry on synthesis and screening. Drug Discov Today2001 Nov 1;6(21):1101-10.
22. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ2003 Mar;22(2):151-85.
23. Nicolaou AC, Brown N, Pattichis CS. Molecular optimization using computational multi-objective methods. Curr Opin Drug Discov Devel2007;10(3):316-24.
24. Brown N, Lewis RA. Exploiting QSAR methods in lead optimization. Curr Opin Drug Discov Devel2006;9(4):419-24.
25. Hansch C. On the structure of medicinal chemistry. J Med Chem1976 Jan;19(1):1-6.
26. 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 Chem2007 Aug 1;15(15):5322-39.
27. 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 Model2006 Dec;25(4):410-22.
28. 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 Chem2006 Jun 15;14(12):4137-50.
29. Guo W, Hu X, Chu N, Yin C. Quantitative structure-activity relationship studies on HEPTs by supervised stochastic resonance. Bioorg Med Chem Lett2006 Mar 27.
30. Vasanathan 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.

31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.
45. Elías J. Cañas García-Otero, Francisco Javier García León, Gloria Andérica Frías. La infección por el VIH: Guía práctica 2003.
46. Alejandro Ruiz-Argüelles. El virus de la inmunodeficiencia humana. *Elementos* 49 2003: 31-7.
47. Gene Antonio. The AIDS cover-up? : the real and alarming facts about AIDS. 2nd ed., with supplement. ed1987. p. xii, 253 p. .
48. Streicher HZ, Reitz MS Jr, RC G. In: Mandell GL, Bennett JE, Dolin R. Principles and Practice of Infectious Diseases Human Immunodeficiency viruses. New York: Churchill Livingstone2000. p. 1874-87.
49. Quinones-Mateu ME, and E.J. Arts. Fitness of drug resistant HIV-1: methodology and clinical implications. *Drug Resist Update*2002;5(6): p. 224-33.
50. Oh SY CW, Raina J, et al. Identification of HIV-1 envelope glykoprotein in the serum of AIDS and ARC patients. *J Acquired Immune Defic Syndr*1992;(5): 251.
51. Sunila I VM, Pantaleo G, Fauci AS, Orenstein JM. Gp120 is present on the plasma membrane of apoptotic CD4 cells prepared from lymph nodes of HIV-1-infected individuals: an immunoelectron microscopic study. *AIDS* 1997: (11): 27-32.
52. Gelderblom HR GM, Scheidler A, Özel M, Pauli G. Zur Struktur und Funktion bei HIV. *AIFO*1993;(5): 231.
53. Barouch D. H., Powers J., Truitt D.M. ea. Dynamic immune responses maintain cytotoxic T lymphocyte epitope mutations in transmitted simian immunodeficiency virus variants [serial on the Internet]. 2005; 6 Available from: <http://amedeo.com/lit.php?id=15685174>.
54. Julia García Prado. Variaciones en el “fitness” del VIH-1 durante la Terapia [Memoria de tesis]. Barcelona: Universidad de Autónoma de Barcelona; 2005.
55. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time [serial on the Internet]. 1996; 217.
56. Drake JW. Rates of spontaneous mutation among RNA viruses [serial on the Internet]. 1993; 90.

57. Markowitz M, Mohri H, Mehandru S, et al. Infection with multidrug resistant, dual-tropic HIV-1 and rapid progression to AIDS: a case report. *Lancet*2005;(365): 1031-8. Abstract: (<http://amedeo.com/lit.php?id=15781098>).
58. Patel PH, Jacobo-Molina A, Ding J, Tantillo C, Clark ADJ, Raag R, et al. Insights into DNA polymerisation mechanisms from structure and function analysis of HIV-1 reverse transcriptase [serial on the Internet]. 1995; 34.
59. Spence R, Kati WM, Anderson KS, Johnson KA. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors [serial on the Internet]. 1995; 267.
60. De Clercq E. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) for the treatment of human immunodeficiency virus type 1 (HIV-1) infections: strategies to overcome drug resistance development. *Med Res Rev* 1996;16, 125-57.
61. De Clercq E. The role of non-nucleoside reverse transcriptase inhibitors (NRTIs) in the therapy of HIV-1 infection. *Antiviral Res* 1998; 38, 153-79.
62. Young SD. Non-nucleoside inhibitors of HIV-1 reverse transcriptase [serial on the Internet]. 1993; 1.
63. Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, et al. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *J Mol Biol*1994;243, 369-87.
64. Smerdon SJ, Jäger J, Wang J, Kohlstaedt LA, Chirino AJ, Friedman JM, et al. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1 [serial on the Internet]. 1994; 91.
65. De Clercq E. Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV infections [serial on the Internet]. 1995; 38.
66. Ilkay Küçükgül, Esra Tatar, Ş.GünizKüçükgül, Sevim Rollas, Clercq ED. Synthesis of some novel thiourea derivatives obtained from 5-[(4-aminophenoxy)methyl]-4-alkyl/aryl-2,4-dihydro-3H-1,2,4-triazole-3-thiones and evaluation as antiviral/anti-HIV and anti-tuberculosis agents [serial on the Internet]. 2007.
67. Ravindra K.Rawal, Rajkamal Tripathi, S.B. Katti, Pannecouque C, Clercq aED. Design, synthesis, and evaluation of 2-aryl-3-heteroaryl-1,3-thiazolidin-4-ones as anti-HIV agents [serial on the Internet]. 2007; 15.
68. Maria Letizia Barreca, Angela Rao, Laura De Luca, Maria Zappalà, Anna-Maria Monforte, Giovanni Maga, et al. Computational Strategies in Discovering Novel Non-nucleoside Inhibitors of HIV-1 RT [serial on the Internet]. 2005; 48(9).
69. Maria Letizia Barreca, Alba Chimirri, Erik De Clercq, Laura De Luca, Anna-Maria Monforte, Pietro Monforte, et al. Anti-HIV agents: design and discovery of new potent RT inhibitors [serial on the Internet]. 2002; 58.
70. Esnouf R, Ren JS, Ross C, Jones Y, Stammers D, Stuart D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors [serial on the Internet]. 1995; 2.
71. Andrea Rubbert GBaMO. Pathogenesis of HIV-1 Infection. In: Christian Hoffmann MD, Ph.D., Jürgen K. Rockstroh MD, Ph.D., Bernd Sebastian Kamps MD, editors. *HIV Medicine 2007*. Paris, Cagliari, Wuppertal: Flying; 2007. p. 59-72.
72. Lei Ji, Fen-Er Chen, Erik De Clercq, Jan Balzarini, Pannecouque aC. Synthesis and Anti-HIV-1 Activity Evaluation of 5-Alkyl-2-alkylthio-6-(arylcabonyl or α -cyanoarylmethyl)-3,4-dihydropyrimidin-4(3H)-ones as Novel Non-nucleoside HIV-1 Reverse Transcriptase Inhibitor. *Journal of Medicinal Chemistry*2007;Vol.50(No.8):1778-86.
73. Menendez-Arias L. Targeting HIV: antiretroviral therapy and development of drug resistance [serial on the Internet]. 2002; 23 (8).
74. Masquelier B, et al., . Primary infection with zidovudine-resistant HIV [serial on the Internet]. 1993; 329 (15).
75. Cao Y, Qin L, Zhang L, Safrit J, Ho DD. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection [serial on the Internet]. 1995; 332.
76. Pantaleo G, Menzo S, Vaccarezza M, Graziosi C, Cohen OJ, Demarest JF, et al. Studies in subjects with long-term nonprogressive human immunodeficiency virus infection [serial on the Internet]. 1995; 322.
77. Moyle GJ. Viral resistance patterns selected by antiretroviral drugs and their potential to guide treatment choice [serial on the Internet]. 1997; 6.
78. Vandamme AM, Van Vaerenberg K, De Clercq E. Anti-human immunodeficiency virus drug combination strategies [serial on the Internet]. 1998; 9.
79. Balzarini J. Suppression of resistance to drugs targeted to human immunodeficiency virus reverse transcriptase by combination therapy [serial on the Internet]. 1999; 58.
80. Vandamme AM, Van Laethem K, Van Vaerenbergh K, De Clercq E. *International Antiviral News* [serial on the Internet]. 1998; 6:10.
81. Balzarini J, Pelemans H, Pérez-Pérez MJ, San-Félix A, Camarasa MJ, De Clercq E, et al. Marked inhibitory activity of non-nucleoside reverse transcriptase inhibitors against human immunodeficiency virus type 1 when combined with (-)2',3'-dideoxy-3'-thiacytidine [serial on the Internet]. 1996; 49.
82. Iversen AK, Shafer RW, Wehrly K, Winters MA, Mullins JI, Chesebro B, et al. Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy [serial on the Internet]. 1996; 70.

83. Balter M. New HIV strain could pose health threat [serial on the Internet]. 1998; 218.
84. Yerly S, Kaiser L, Race E, Clavel F, Perrin L. Reverse transcriptase and protease gene analysis at the time of primary HIV-1 infection [serial on the Internet]. 24-27 June 1998.
85. Drews J. Drug discovery: a historical perspective. *Science*2000 Mar 17;287(5460):1960-4.
86. Kubinyi H. Strategies and recent technologies in drug discovery. *Pharmazie*1995 Oct;50(10):647-62.
87. Warne P, Page C. Is there a best strategy for drug discovery? *Drug News Perspect*2003 Apr;16(3):177-82.
88. Jorgensen WL. The many roles of computation in drug discovery. *Science*2004;303(5665):1813-8.
89. Armstrong JW. A review of high-throughput screening approaches for drug discovery. *Application note*. 1999.
90. Bajorath J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov*2002 Nov;1(11):882-94.
91. Lazo JS, Wipf P. Combinatorial chemistry and contemporary pharmacology. *J Pharmacol Exp Ther*2000 Jun;293(3):705-9.
92. Miertus S, Fassina G, Seneci PF. Concepts of Combinatorial Chemistry and Combinatorial Technologies. *Chem Listy*2000;94:1104-10.
93. Brown FK. Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery1998.
94. Leach AR, Gillet VJ. An introduction to chemoinformatics: Kluwer Academic Publishers; 2003.
95. Gasteiger J, Engel T. Chemoinformatics : a Textbook. Gasteiger J, Engel T, editors: Wiley-VCH; 2003.
96. Hann M, Green R. Chemoinformatics--a new name for an old problem? *Curr Opin Chem Biol*1999 Aug;3(4):379-83.
97. 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.
98. Apostolakis J, Caflisch A. Computational ligand design. *Comb Chem High Throughput Screen*1999 Apr;2(2):91-104.
99. Graves AP, Brenk R, Shoichet BK. Decoys for docking. *J Med Chem*2005 Jun 2;48(11):3714-28.
100. 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.
101. Wilton D, Willett P, Lawson K, Mullier G. Comparison of ranking methods for virtual screening in lead-discovery programs. *J Chem Inf Comput Sci*2003 Mar-Apr;43(2):469-74.
102. Finn P, Muggleton S, Page D, Srinivasan A. Pharmacophore discovery using the inductive logic programming language Progol. *Machine Learning*1998;30:241-70.
103. Warmuth MK, Liao J, Ratsch G, Mathieson M, Putta S, Lemmen C. Active learning with support vector machines in the drug discovery process. *J Chem Inf Comput Sci*2003 Mar-Apr;43(2):667-73.
104. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*2001 Mar 1;46(1-3):3-26.
105. 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.
106. 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.
107. Moreau G, P. B. Autocorrelation of molecular structures: Application to SAR studies. *Nouv J Chim*1980;757-64.
108. 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.
109. 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.
110. Lemmen C, Lengauer T. Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des*2000 Mar;14(3):215-32.
111. Dietterich TG, Lathrop RH, Lozano-Perez T. Solving the multiple instance problem with axis-parallel rectangles. *Artif Intell*1997;89(1):31-71.
112. Perry N, van Geerestein VJ. Database Searching on the basis of Three-Dimensional Similarity Using the sperm Program. *J Chem Inf Comput Sci*1992;32:607-16.
113. Kubinyi H. Comparative Molecular Field Analysis. In: Gasteiger J, editor. Handbook of Chemoinformatics From Data to Knowledge. Weinheim: Wiley-VCH; 2003. p. 1555-74.
114. Pepperrell CA, Willett P. Techniques for the calculation of three-dimensional structural similarity using interatomic distances. *J Comput Aided Mol Des*1991 Oct;5(5):455-74.
115. Wagener M, Sadowski J, Gasteiger J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic ah receptor activity by neural networks. *J Am Chem Soc*1995;117:7769-75.
116. Carhart RE, Smith DH, Venkataraghavan R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definitions and Applications. *J Chem Inf Comput Sci*1985;25:64-73.

117. Chen X, Russinko III A, Young SS. Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors. *J Chem Inf Comput Sci*1998;38:1054-62.
118. Swamidass SJ, Chen J, Bruand J, Phung P, Ralaivola L, Baldi P. Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics*2005 Jun;21 Suppl 1:i359-68.
119. Güner OF. Pharmacophore Perception, Development, and Use in Drug Design. volume 2 of IUL Biotechnology series2000.
120. Guner OF. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr Top Med Chem*2002 Dec;2(12):1321-32.
121. Pickett SD, Mason JS, McLay IM. Diversity profiling and design using 3D pharmacophores : Pharmacophores-Derived Queries (POD). *J Chem Inf Comput Sci*1996;36:1214-23.
122. McGregor MJ, Muskal SM. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J Chem Inf Comput Sci*1999 May-Jun;39(3):569-74.
123. McGregor MJ, Muskal SM. Pharmacophore fingerprinting. 2. Application to primary library design. *J Chem Inf Comput Sci*2000 Jan-Feb;40(1):117-25.
124. Brown RD, Martin YC. Use of Structure-Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J Chem Inf Comput Sci*1996;36:572-84.
125. Matter H, Pötter T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J Chem Inf Comput Sci*1999;39:1211-25.
126. Mason JS, Morize I, Menard PR, Cheney DL, Hulme C, Labaudiniere RF. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J Med Chem*1999 Aug 26;42(17):3251-64.
127. Abrahamian E, Fox PC, Naerum L, Christensen IT, Thogersen H, Clark RD. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J Chem Inf Comput Sci*2003 Mar-Apr;43(2):458-68.
128. Saeh JC, Lyne PD, Takasaki BK, Cosgrove DA. Lead hopping using SVM and 3D pharmacophore fingerprints. *J Chem Inf Model*2005 Jul-Aug;45(4):1122-33.
129. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. Gray J, editor. San Francisco: Morgan Kaufmann; 2005.
130. Duda RO, Hart PE, Stork DG. *Pattern Classification*. Second ed: Wiley-Interscience; 2001.
131. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*: Springer; 2001.
132. Van Waterbeemd H. *Chemometric methods in molecular design*. Manhnhold R, Krogsgaard-Larsen P, Timmerman H, Van Waterbeemd H, editors. New York: Wiley-VCH; 1995.
133. Jackson JE. *A User's Guide to Principal Components*. New York (NY): Wiley; 1991.
134. 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.
135. 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.
136. 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.
137. 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.
138. 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.
139. 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.
140. Venkatraman V, Dalby AR, Yang ZR. Evaluation of mutual information and genetic programming for feature selection in QSAR. *J Chem Inf Comput Sci*2004 Sep-Oct;44(5):1686-92.
141. 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.
142. 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.
143. 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.

144. 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.
145. Hansch C, Quinlan JE, Lawrence GL. Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J Org Chem*1968;33:347-50.
146. Aoyama T, Suzuki Y, Ichikawa H. Neural networks applied to quantitative structure-activity relationship analysis. *J Med Chem*1990 Sep;33(9):2583-90.
147. 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.
148. 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.
149. Zupan J, Gasteiger J. Neural networks in chemistry and drug design. 2nd ed. Weinheim: Wiley-VCH; 1999.
150. Devillers J. Neural Networks in QSAR and Drug Design. London: Academic Press; 1996.
151. Schneider G, Wrede P. Artificial neural networks for computer-based molecular design. *Prog Biophys Mol Biol*1998;70(3):175-222.
152. Brereton RG. Chemometrics. Horwood E, editor. Chichester (UK): Ellis Horwood; 1990.
153. 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.
154. Frank IE, Todechini R. The Data Analysis Handbook. Amsterdam (The Netherlands): Elsevier; 1994.
155. 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.
156. Massart DL, Vandeginste BGM, Buydens L, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics. Part A. Amsterdam (The Netherlands): Elsevier; 1997.
157. Massart DL, Vandeginste BGM, Buydens L, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics. Part B. Amsterdam (The Netherlands): Elsevier; 1998.
158. Legendre P, Legendre L. Numerical Ecology. Amsterdam (The Netherlands): Elsevier; 1998.
159. Draper N, Smith H. Applied Regression Analysis. New York (NY): Wiley; 1998.
160. Frank IE, Friedman JH. A Statistical View of Some Chemometrics Regression Tools. *Technometrics*1993;35:109-35.
161. Ryan TP. Modern Regression Methods. New York (NY): Wiley; 1997.
162. 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.
163. Atkinson AC. Plots, Transformations, and Regression. Oxford (UK): Clarendon Press; 1985.
164. Hendriks MMWB, de Boer JH, Smilde AK, Doornbos DA. Multicriteria Decision Making Chemom Intell Lab Syst1992;16:175-91.
165. Carlson R. Design and Optimization in Organic Synthesis. Amsterdam (The Netherlands): Elsevier; 1992.
166. Holland J. Adaptation in Artificial and Neural Systems. Ann Arbor (MI): University of Michigan Press; 1975.
167. 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.
168. Verbeeck K. Coordinated Exploration in Multi-agent Reinforcement Learning. Brussel, Belgium: Vrije Universiteit Brussel; 2004.
169. 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.
170. Yann C, Siarry P, editors. Multiobjective Optimization: Principles and Case Studies. Berlin, Germany: Springer-Verlag; 2004.
171. 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.
172. 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.
173. Shephird JK, Clark 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.
174. 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.
175. 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.

176. 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.
177. Lameijer EW, Kok JN, Back T, Ijerman AP. The molecule evaluator. An interactive evolutionary algorithm for the design of drug-like molecules. *J Chem Inf Model*2006;46(2):545-52.
178. Derringer G, Suich R. Simultaneous Optimization of Several Response Variables. *J Quality Technol*1980;12(4):214-9.
179. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. Fifth ed. New York: McGraw Hill; 2005.
180. Statsoft_Inc. *STATISTICA*. 6.0 for Windows ed2001.
181. Nelder JA, Mead R. A Simplex method for function minimization. *Computer Journal*1965;7:308-13.
182. Fletcher R, Reeves CM. Function minimization by conjugate gradients. *Computer Journal*1964;7:149-54.
183. Hooke R, Jeeves TA. Direct search solution of numerical and statistical problems. *J Assoc Comp Machin*1961;8:212-29.
184. Watson I, Marir F. *Case-based reasoning: a review*. The knowledge engineering review1994;9(4):Cambridge, UK: Cambridge University Press.
185. Coleman TF, Li Y. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J Optim*1996;6:418-45.
186. 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.
187. *MATLAB*. 7.2 ed: The MathWorks, Inc.; 2006.
188. Guang-Fu Sun, Xu-Xiang Chen, Fen-Er Chen, Yue-Ping Wang, Erik De Clercq, Jan Balzarini, et al. Nonnucleoside HIV-1 Reverse-Transcriptase Inhibitors, Part 5 Synthesis and Anti-HIV-1 Activity of Novel 6-Naphthylthio HEPT Analogues [serial on the Internet]. 2005; 53(8).
189. Yuan-ZhenXiong, Fen-Er Chen, Jan Balzarini, Erik De Clercq, Christophe Pannecouque. Non-nucleoside HIV-1 reverse transcriptase inhibitors Part 11:Structural modulations of diaryltriazines with potent anti-HIV activity [serial on the Internet]. 2007.
190. Angelo Ranise, Andrea Spallarossa, Silvia Schenone, Olga Bruno, Francesco Bondavalli, Laura Vargiu, et al. Design, Synthesis, SAR, and Molecular Modeling Studies of Acylthiocarbamates: A Novel Series of Potent Non-nucleoside HIV-1 Reverse transcriptase Inhibitors Structurally Related to Phenethylthiazolylthiourea Derivatives [serial on the Internet]. 2003; 46(5).
191. 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.
192. 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.
193. 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.
194. 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.
195. Fernandes PB. Mode of Action, In Vitro and In Vivo Activities of the Fluoroquinolones. *J Clin Pharmacol*1988;28:156-68.
196. 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.
197. 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.
198. Gracheck SJ, Mychalotka 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.
199. Pauwels R, Balzarini J, Baba M, Snoeck R, Schols D, Herdewijn P, et al. Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds. *J Virol Methods*1988;20:309-21.
200. Artico M, Silvestri R, Massa S, Loi AG, Corrias S, Piras G, et al. 2-Sulfonyl-4-chloroanilino Moiety: A Potent Pharmacophore for the anti-Human Immunodeficiency Virus Type 1 Activity of Pyrrolyl Aryl Sulfones. *J Med Chem*1996;39:522-30.
201. De Boor C. *A practical guide to splines*. New York: Springer-Verlag; 1978.
202. Gerald CF, Wheatley PO. *Applied numerical analysis*. 4th ed. Reading, MA: Addison Wesley; 1989.
203. CambridgeSoft. *ChemDraw Ultra*. 9.0 ed2004.

204. Todeschini R, Consonni V, Pavan M. DRAGON Software. 2.1 ed. Milano: Milano Chemometrics; 2002.
205. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J Chemom*1992;6:267-81.
206. 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.
207. 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.
208. 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.
209. Barbosa de Oliveira D, Gaudio AC. BuildQSAR. Vitória ES, Brasil: Physics Department-CCE, University of Espírito Santo; 2000.
210. Barbosa de Oliveira D, Gaudio AC. BuildQSAR: A new computer program for QSAR analysis. *Quant Struct-Act Relat*2000;19:599-601.
211. 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.
212. Atkinson AC. *Plots, Transformations and Regression*: Oxford:Clarendon Press; 1985.
213. Cruz-Monteaudo 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.
214. Stewart J, Gill L. *Econometrics*. 2nd edition ed. Allan P, editor. London: Prentice Hall; 1998.
215. Kutner MH, Nachtsheim CJ, Neter J, Li W. Multicollinearity and its effects. *Applied Linear Statistical Models*. Fifth ed. New York: McGraw Hill; 2005. p. 278-89.
216. Drews J. Innovation deficit revisited: Reflections on the productivity of pharmaceutical R&D. *Drug Discov Today*1998;3:491-4.

A NEXOS

Tabla 1 Compuestos seleccionados de literatura.

N ^o	Compuestos ^a	Ref. ^b	N ^o	Compuestos ^a	Ref. ^b
1	1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT)	A B	22	4-[4-Methylamino-6-(1-chloro-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9o)	C
2	2,3-dideoxyinosine (DDI)	C B	23	4-[4-Amino-6-(1-chloro-2-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (9p)	C
3	Trovirdine	D	24	4-[4-Amino-6-(6-bromo-2-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (9q)	C
4	4-[4-Methylamino-6-(1-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (8a)	C	25	4-[4-Azido-6-(1-chloro-2-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (9r)	C
5	4-[4-Amino-6-(1-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (8b)	C	26	O-(Benzyl) 2-furoyl (phenyl) thiocarbamate (13q)	D
6	4-[4-n-Propylamino-6-(1-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (8c)	C	27	O-(2-Phenethyl) (E)-cinnamoyl (phenyl) thiocarbamate (15b)	D
7	4-[4-Methylamino-6-(4-chloro-1-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (8d)	C	28	O-(2-Phenethyl) benzoyl (phenyl) thiocarbamate (15c)	D
8	4-[4-Amino-6-(4-chloro-1-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (8e)	C	29	O-(2-Phenethyl) 4-chlorobenzoyl (phenyl)thiocarbamate (15g)	D
9	4-[4-Ethylamino-6-(4-chloro-1-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (8f)	C	30	O-(2-Phenethyl) 2-furoyl (phenyl) thiocarbamate (15q)	D
10	4-[4-n-Propylamino-6-(4-chloro-1-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (8g)	C	31	O-(2-Phenoxyethyl) 2-phenoxyacetyl (phenyl)thiocarbamate (17a)	D
11	4-[4-Methylamino-6-(2-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (9a)	C	32	O-(2-Phenoxyethyl) (E)-Cinnamoyl (phenyl)thiocarbamate (17b)	D
12	4-[4-Amino-6-(2-naphthoxy)-1,3,5-triazine-2-yl]aminobenzonitrile (9b)	C	33	O-(2-Phenoxyethyl) benzoyl (phenyl) thiocarbamate (17c)	D
13	4-[4-Ethylamino-6-(1,6-dibromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9c)	C	34	O-(2-Phenoxyethyl) 4-chlorobenzoyl (phenyl)thiocarbamate (17g)	D
14	4-[4-n-Propylamino-6-(1,6-dibromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9d)	C	35	O-(2-Phenoxyethyl) 2,4-dichlorobenzoyl(phenyl)thiocarbamate (17k)	D
15	4-[4-i-Propylamino-6-(6-bromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9f)	C	36	O-(2-Phenoxyethyl) 3,5-dichlorobenzoyl (phenyl)thiocarbamate (17m)	D
16	4-[4-Methylamino-6-(6-bromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9g)	C	37	O-(2-Phenoxyethyl) 4-chloro-3-nitrobenzoyl(phenyl)thiocarbamate (17n)	D
17	4-[4-Methylamino-6-(1-bromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9h)	C	38	O-(2-Phenoxyethyl) 2-furoyl (phenyl) thiocarbamate (17q)	D
18	4-[4-Amino-6-(1-bromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9i)	C	39	O-(2-Phenoxyethyl) phenyl(thien-2-yl carbonyl)thiocarbamate (17r)	D
19	4-[4-Ethylamino-6-(1-chloro-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9k)	C	40	O-(2-Phenoxyethyl) benzoyl(4-fluorophenyl)thiocarbamate (18c)	D
20	4-[4-Methylamino-6-(1,6-dibromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9m)	C	41	(±) O-(1-Methyl-2-phenoxyethyl) phenoxyacetyl(phenyl)thiocarbamate (19a)	D
21	4-[4-Amino-6-(1,6-dibromo-2-naphthoxy)-1,3,5-triazine-2-yl] amino benzonitrile (9n)	C	42	(±) O-(1-Methyl-2-phenoxyethyl) 2-furoyl(phenyl)thiocarbamate (19q)	D

Tabla 1 (Continuación...)

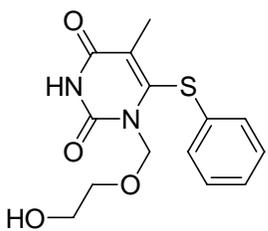
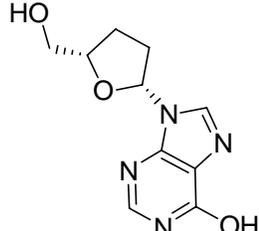
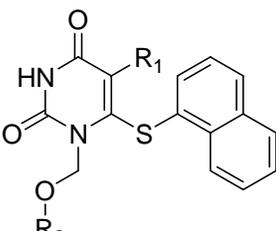
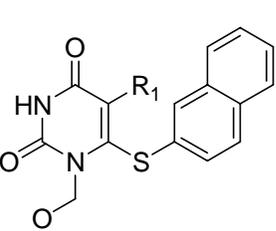
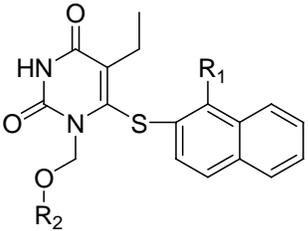
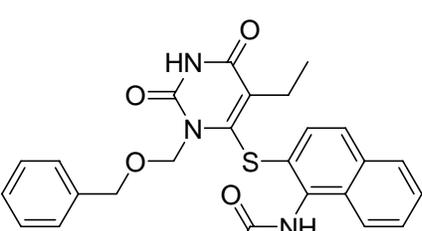
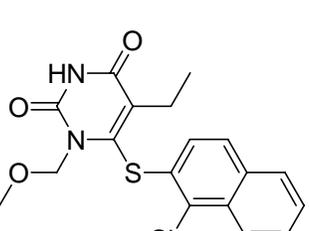
N ^o	Compuestos ^a	Ref. ^b	N ^o	Compuestos ^a	Ref. ^b
43	(±) O-(1-Methyl-2-phenoxyethyl) 4-nitrophenyl(thien-2-yl carbonyl) thiocarbamate (22r)	D	61	6-[α-Cyano-(1-naphthylmethyl)]-3,4-dihydro-5-ethyl-2-(4-nitrobenzylthio)pyrimidin-4(3H)-one (3o)	A
44	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl]3-bromophenyl (thien-2-ylcarbonyl)thiocarbamate (35r)	D	62	6-(α-Cyanobenzyl)-3,4-dihydro-5-ethyl-2-(4-nitrobenzylthio)-pyrimidin-4(3H)-one (3r)	A
45	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl] 3-nitrophenyl(thien-2-ylcarbonyl)thiocarbamate (36r)	D	63	6-(α-Cyano-2,6-dichlorobenzyl)-3,4-dihydro-5-ethyl-2-(4-nitrobenzylthio)pyrimidin-4(3H)-one (3s)	A
46	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl]4-chlorophenyl(2-furoyl)thiocarbamate (41q)	D	64	6-[α-Cyano-(1-naphthylmethyl)]-2-cyclopentylthio-3,4-dihydro-5-isopropylpyrimidin-4(3H)-one (3t)	A
47	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl]4-iodophenyl(thien-2-ylcarbonyl)thiocarbamate (43r)	D	65	6-(α-Cyanobenzyl)-3,4-dihydro-5-isopropyl-2-(4-methoxybenzylthio)pyrimidin-4(3H)-one (3v)	A
48	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl]2-furoyl(4-nitrophenyl)thiocarbamate (45q)	D	66	3,4-Dihydro-2-ethylthio-6-(1-naphthoyl)pyrimidin-4(3H)-one (4a)	A
49	O-[2-(1,3-Dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl] 4-nitrophenyl(thien-2-ylcarbonyl)thiocarbamate (45r)	D	67	6-Benzoyl-3,4-dihydro-2-isopropylthio-5-methylpyrimidin-4(3H)-one (4e)	A
50	6-[α-Cyano-(1-naphthylmethyl)]-3,4-dihydro-2-isopropylthiopyrimidin-4(3H)-one (3a)	A	68	3,4-Dihydro-5-ethyl-2-ethylthio-6-(1-naphthoyl)pyrimidin-4(3H)-one (4g)	A
51	6-[α-Cyano-(1-naphthylmethyl)]-2-cyclopentylthio-3,4-dihydro-5-methylpyrimidin-4(3H)-one (3c)	A	69	3,4-Dihydro-5-ethyl-2-isopropylthio-6-(1-naphthoyl)pyrimidin-4(3H)-one (4h)	A
52	6-[α-Cyano-(1-naphthylmethyl)]-3,4-dihydro-5-methyl-2-propynyl thiopyrimidin-4(3H)-one (3d)	A	70	2-Allylthio-3,4-dihydro-5-ethyl-6-(1-naphthoyl)pyrimidin-4(3H)-one (4i)	A
53	2-Benzoylmethylthio-6-[α-cyano-(1-naphthylmethyl)]-3,4-dihydro-5-methylpyrimidin-4(3H)-one (3e)	A	71	6-Benzoyl-2-cyclopentylthio-3,4-dihydro-5-ethylpyrimidin-4(3H)-one (4l)	A
54	2-(4-Chlorobenzoylmethylthio)-6-[α-cyano-(1-naphthylmethyl)]-3,4-dihydro-5-methylpyrimidin-4(3H)-one (3f)	A	72	2-Cyclopentylthio-3,4-dihydro-5-isopropyl-6-(1-naphthoyl)pyrimidin-4(3H)-one (4m)	A
55	6-(α-Cyanobenzyl)-3,4-dihydro-2-isopropylthio-5-methylpyrimidin-4(3H)-one (3g)	A	73	1-Ethoxymethyl-5-methyl-6-(1-naphthylthio)uracil (7a)	B
56	6-(α-Cyanobenzyl)-2-cyclopentylthio-3,4-dihydro-5-methylpyrimidin-4(3H)-one (3h)	A	74	1-[(Benzyloxy)methyl]-5-methyl-6-(1-naphthylthio)uracil (7b)	B
57	2-Benzoylmethylthio-6-(α-cyanobenzyl)-3,4-dihydro-5-methylpyrimidin-4(3H)-one (3i)	A	75	1-Methoxymethyl-5-ethyl-6-(1-naphthylthio)uracil (7c)	B
58	6-[α-Cyano-(1-naphthylmethyl)]-3,4-dihydro-5-ethyl-2-ethylthiopyrimidin-4(3H)-one (3j)	A	76	1-Ethoxymethyl-5-ethyl-6-(1-naphthylthio)uracil (7d)	B
59	6-[α-Cyano-(1-naphthylmethyl)]-2-cyclopentylthio-3,4-dihydro-5-ethyl pyrimidin-4(3H)-one (3l)	A	77	1-[(Benzyloxy)methyl]-5-ethyl-6-(1-naphthylthio)uracil (7e)	B
60	6-[α-Cyano-(1-naphthylmethyl)]-3,4-dihydro-5-ethyl-2-propynylthiopyrimidin-4(3H)-one (3n)	A	78	1-[3-Methyl(benzyloxy)methyl]-5-ethyl-6-(1-naphthylthio)uracil (7f)	B

Tabla 1 (Continuación...)

Nº	Compuestos ^a	Ref. ^b	Nº	Compuestos ^a	Ref. ^b
79	1-Methoxymethyl-5-isopropyl-6-(1-naphthylthio)uracil (7g)	B	95	1-[(2-Phenylethoxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7v)	B
80	1-Ethoxymethyl-5-isopropyl-6-(1-naphthylthio)uracil (7h)	B	96	1-(Hydroxyethoxymethyl)-6-(α -naphthalenethio)-5-isopropyluracil (7w)	B
81	1-[(2-Methoxyethyloxy)methyl]-5-isopropyl-6-(2-naphthylthio)uracil (7i)	B	97	1-Methoxymethyl-5-ethyl-6-(2-naphthylthio)uracil (8a)	B
83	1-[(Cyclopropylmethoxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7j)	B	98	1-Ethoxymethyl-5-ethyl-6-(2-naphthylthio)uracil (8b)	B
84	1-[(Benzyloxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7k)	B	99	1-[(Benzyloxy)methyl]-5-ethyl-6-(2-naphthylthio)uracil (8c)	B
85	1-[(3-Methyl-phenylmethyloxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7l)	B	100	1-Methoxymethyl-5-isopropyl-6-(2-naphthylthio)uracil (8d)	B
86	1-Ethoxymethyl-5-propyl-6-(1-naphthylthio)uracil (7m)	B	101	1-Ethoxymethyl-5-isopropyl-6-(2-naphthylthio)uracil (8e)	B
87	1-[(Benzyloxy)methyl]-5-propyl-6-(1-naphthylthio)uracil (7n)	B	102	1-[(Benzyloxy)methyl]-5-isopropyl-6-(2-naphthylthio)uracil (8f)	B
88	1-Ethoxymethyl-5-isobutyl-6-(1-naphthylthio)uracil (7o)	B	103	1-Methoxymethyl-5-ethyl-6-(1-nitro-2-naphthylthio)uracil (11a)	B
89	1-[(Benzyloxy)methyl]-5-isobutyl-6-(1-naphthylthio)uracil (7p)	B	104	1-Ethoxymethyl-5-ethyl-6-(1-nitro-2-naphthylthio)uracil (11b)	B
90	1-[(3-Fluorobenzyloxy)methyl]-5-ethyl-6-(1-naphthylthio)uracil (7q)	B	105	1-[(Benzyloxy)methyl]-5-ethyl-6-(1-nitro-2-naphthylthio)uracil (11c)	B
91	1-[(4-Fluorobenzyloxy)methyl]-5-ethyl-6-(1-naphthylthio)uracil (7r)	B	106	1-Methoxymethyl-5-ethyl-6-(1-amino-2-naphthylthio)uracil (12a)	B
92	1-[(Cyclohexylmethoxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7s)	B	107	1-Ethoxymethyl-5-ethyl-6-(1-amino-2-naphthylthio)uracil (12b)	B
93	1-[(3-Fluorobenzyloxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7t)	B	108	1-[(Benzyloxy)methyl]-5-ethyl-6-(1-acetamino-2-naphthylthio)uracil (13)	B
94	1-[(4-Fluorobenzyloxy)methyl]-5-isopropyl-6-(1-naphthylthio)uracil (7u)	B	109	1-Methoxymethyl-5-ethyl-6-(1-chloro-2-naphthylthio)uracil (15)	B

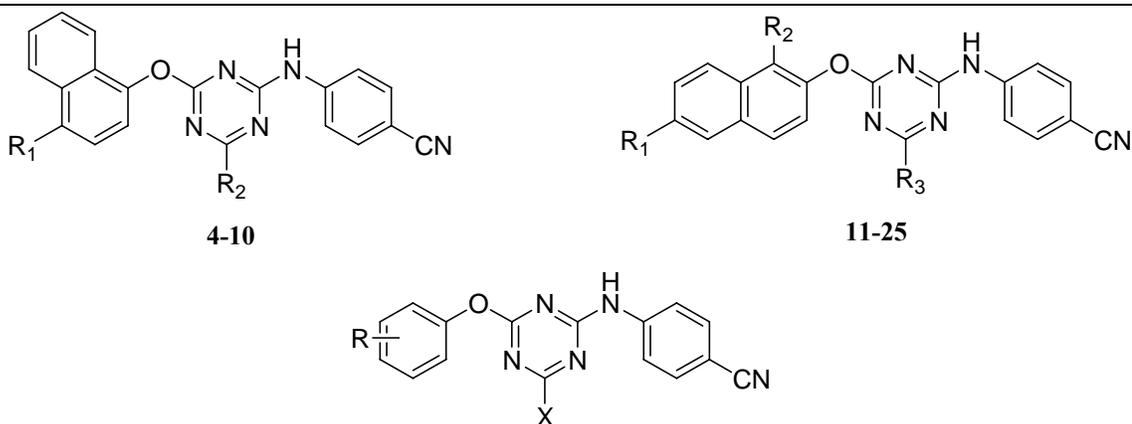
^aLa estructura molecular de estos inhibidores de la transcriptasa inversa se encuentran en la Tabla 2.2.^bReferencias tomadas de la literatura: A- LeiJi; Fen-ErChen; Erik De Clercq; JanBalzarini; Pannecouque, a. C. Journal of Medicinal Chemistry **2007**, Vol.50, 1778-1786; B- Guang-Fu Sun; Xu-Xiang Chen; Fen-Er Chen; Yue-Ping Wang; Erik De Clercq; Jan Balzarini; and Christophe Pannecouque. Chem. Pharm. Bull., **2005**; Vol. 53, 886 to 892; C- Yuan-ZhenXiong; Fen-Er Chen; Jan Balzarini; Erik De Clercq; Christophe Pannecouque. European Journal of Medicinal Chemistry, **2007**, 1 to 7; D- Angelo Ranise; Andrea Spallarossa; Silvia Schenone; Olga Bruno; Francesco Bondavalli; Laura Vargiu; Tiziana Marceddu; Massimo Mura; PaoloLaColla; and Alessandra Pani. Journal of Medicinal Chemistry, **2003**; Vol. 46, 768 to 781.

Tabla 2. Estructuras generales de los derivados de 1-[(2-Hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT)

 <p>1. HEPT *EC₅₀= 5.06; *CC₅₀=405</p>		 <p>2. DDI *EC₅₀= 5.37; *CC₅₀=529</p>		 <p>74-96</p>		 <p>97-102</p>			
 <p>103-107</p>		 <p>108. *EC₅₀= 8.78; *CC₅₀= 155.47</p>		 <p>109. *EC₅₀=22.66; *CC₅₀=192.9</p>					
Nº	R ₁	R ₂	*EC ₅₀	*CC ₅₀	Nº	R ₁	R ₂	*EC ₅₀	*CC ₅₀
74	Me	Et	0.46	39.6	91	<i>i</i> -Bu	<i>c</i> -Hex1-CH ₂	0.092	30.96
75	Me	Benzyl	0.18	52.5	92	<i>i</i> -Bu	3'-Fluorobenzyl	12.53	171.92
76	Et	Me	0.5	38.78	93	<i>i</i> -Bu	4'-Fluorobenzyl	0.11	32.82
77	Et	Et	0.15	38.4	94	<i>i</i> -Bu	PhCH ₂ CH ₂	0.067	30.23
78	Et	Benzyl	0.48	34.28	95	<i>i</i> -Bu	CH ₂ CH ₂ OAc	2.77	31.4
79	Et	3'-Methylbenzyl	0.21	28.73	96	<i>i</i> -Bu	CH ₂ CH ₂ OH	0.23	77.07
80	<i>i</i> -Pr	Me	0.51	38.4	97	Et	Me	22.02	210.09
81	<i>i</i> -Pr	Et	0.057	33.2	98	Et	Et	3.37	42.19
82	<i>i</i> -Pr	CH ₂ CH ₂ OCH ₃	0.65	35.58	99	Et	Benzyl	0.65	154.5
83	<i>i</i> -Pr	<i>c</i> -Pr-CH ₂	0.38	32.75	100	<i>i</i> -Pr	Me	4.69	59.58
84	<i>i</i> -Pr	Benzyl	0.063	35.3	101	<i>i</i> -Pr	Et	2.65	36.3
85	<i>i</i> -Pr	3'-Methylbenzyl	0.25	29.6	102	<i>i</i> -Pr	Benzyl	0.83	187.34
86	Pr	Et	1.86	37.11	103	Me	NO ₂	0.47	50.13
87	<i>i</i> -Bu	Et	0.83	32.57	104	Et	NO ₂	0.099	43.89
88	<i>i</i> -Bu	Benzyl	7.37	36.12	105	Benzyl	NO ₂	0.065	34.19
89	Et	3'-Fluorobenzyl	2.96	32.57	106	Me	NH ₂	25.99	180.48
90	Et	4'-Fluorobenzyl	0.21	31.47	107	Et	NH ₂	7.06	195.23

*Los valores EC₅₀ y CC₅₀ se determinan por los ensayos *in vitro*, en unidad µM

Tabla 3. Estructura general de serie de analogos de diaryltriazine (DATAs)

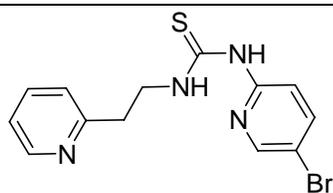


DATAs

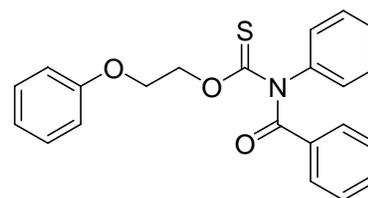
Nº	R ₁	R ₂	*EC ₅₀	*CC ₅₀	Nº	R ₁	R ₂	R ₃	*EC ₅₀	*CC ₅₀
4	H	NHMe	0.186	30.33	11	H	H	NHMe	0.26	250
5	H	NH ₂	0.107	39.07	12	H	H	NH ₂	0.09	42.23
6	H	<i>n</i> -PrNH	2.7	29.46	13	Br	Br	NHEt	0.156	74.4
7	Cl	NHMe	0.093	196.9	14	Br	Br	<i>n</i> -PrNH	0.256	134.44
8	Cl	NH ₂	0.062	237	15	Br	H	<i>i</i> -PrNH	1.065	41.54
9	Cl	NHEt	0.733	207.7	16	H	Br	NHMe	0.284	193.9
10	Cl	<i>n</i> -PrNH	1.63	181.6	17	Br	H	NHMe	0.0093	143.2
					18	Br	H	NH ₂	0.0094	133
					19	Cl	H	NHEt	0.114	187.93
					20	Br	Br	NHMe	0.0184	92.83
					21	Br	Br	NH ₂	0.0195	244
					22	Cl	H	NHMe	0.0118	170.6
					23	Cl	H	NH ₂	0.028	32.15
					24	H	Br	NH ₂	0.0808	243
					25	Cl	H	N ₃	0.06	256.32

*Los valores EC₅₀ y CC₅₀ se determinan por los ensayos *in vitro*, en unidad µM

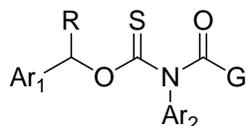
Tabla 4. Estructuras generales de serie O-sust-N-sust-N-acylthiocarbamates (ATCs)



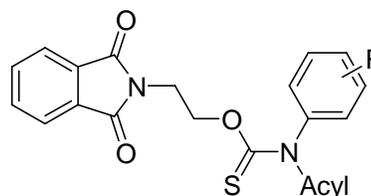
3. Troviridine

*EC₅₀= 0.02; *CC₅₀= 60

Compuesto líder



26-43



44-49

Nº	Ar ₁	R	Ar ₂	G-CO	*EC ₅₀	*CC ₅₀
26	Phenyl	H	C ₆ H ₅	2-furoyl	58	111
27	Benzyl	H	C ₆ H ₅	Trans-cinnamoyl	4.5	40.4
28	Benzyl	H	C ₆ H ₅	benzoyl	4.2	91
29	Benzyl	H	C ₆ H ₅	4-chlorobenzoyl	4	43
30	Benzyl	H	C ₆ H ₅	2-furoyl	4.3	102
31	phenoxymethyl	H	C ₆ H ₅	phenoxyacetyl	6	103
32	phenoxymethyl	H	C ₆ H ₅	trans-cinnamoyl	7.7	66.6
33	phenoxymethyl	H	C ₆ H ₅	benzoyl	8	122
34	phenoxymethyl	H	C ₆ H ₅	4-chlorobenzoyl	10.3	133
35	phenoxymethyl	H	C ₆ H ₅	2,4-dichlorobenzoyl	11.6	43
36	phenoxymethyl	H	C ₆ H ₅	3,5-dichlorobenzoyl	8.8	43
37	phenoxymethyl	H	C ₆ H ₅	4-chloro-3-nitrobenzoyl	7.6	80
38	phenoxymethyl	H	C ₆ H ₅	2-furoyl	8.4	82
39	phenoxymethyl	H	C ₆ H ₅	2-thenoyl	8.6	125.2
40	phenoxymethyl	H	4-F- C ₆ H ₅	benzoyl	4	70.7
41	phenoxymethyl	CH ₃	C ₆ H ₅	phenoxyacetyl	1.4	122.4
42	phenoxymethyl	CH ₃	C ₆ H ₅	2-furoyl	1.3	51.5
43	phenoxymethyl	CH ₃	4-NO ₂ -C ₆ H ₅	2-thenoyl	6	63
		R		Acyl		
44		3-Br		2-thenoyl	1.2	53
45		3-NO ₂		2-thenoyl	0.38	100
46		4-Cl		2-furoyl	0.007	41
47		4-I		2-thenoyl	0.01	18
48		4-NO ₂		2-furoyl	0.008	18
49		4-NO ₂		2-thenoyl	0.01	168

*Los valores EC₅₀ y CC₅₀ se determinan por los ensayos *in vitro*, en unidad µM

Tabla 5. Estructuras generales de nueva serie de análogos de S-DABO (S-DABOs)
2-alkoxy-3,4-dihydro-6-benzyl-4(3H)-pyrimidin-4-ones (DABOs)

Nº	R ₁	R ₂	R ₃	*EC ₅₀	*CC ₅₀
50	H	<i>i</i> -Pr	1-naphthyl	0.66	53.19
51	Me	cyclopentyl	1-naphthyl	0.8	49.39
52	Me	propynyl	1-naphthyl	2.7	46.09
53	Me	benzoylmethyl	1-naphthyl	0.09	163.86
54	Me	4-chlorobenzoylmethyl	1-naphthyl	4.71	148.32
55	Me	<i>i</i> -Pr	Ph	0.002	10.81
56	Me	cyclopentyl	Ph	6.34	58.95
57	Me	benzoylmethyl	Ph	7.28	53.81
58	Me	Et	1-naphthyl	1.17	122.35
59	Et	cyclopentyl	1-naphthyl	0.18	51.88
60	Et	propynyl	1-naphthyl	4.85	47.99
61	Et	4-nitrobenzyl	1-naphthyl	1.25	24.52
62	Et	<i>i</i> -Pr	Ph	0.64	58.02
63	Et	4-nitrobenzyl	Ph	1.33	34.19
64	Et	4-nitrobenzyl	2,6-Cl ₂ -Ph	2.22	23.46
65	<i>i</i> -Pr	cyclopentyl	1-naphthyl	1.34	7.32
66	<i>i</i> -Pr	4-methoxybenzyl	Ph	3.53	26.79
67	H	Et	1-naphthyl	37.58	241.61
68	Me	<i>i</i> -Pr	Ph	2.05	263.19
69	Et	Et	1-naphthyl	27.84	196.54
70	Et	<i>i</i> -Pr	1-naphthyl	6.79	178.92
71	Et	allyl	1-naphthyl	16.94	210.37
72	Et	cyclopentyl	Ph	4.69	60.37
73	<i>i</i> -Pr	cyclopentyl	1-naphthyl	10.87	24.69

*Los valores EC₅₀ y CC₅₀ se determinan por los ensayos *in vitro*, en unidad µM