



UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS
VERITATE SOLA NOBIS IMPONETUR VIRILISTOGA. 1948

FACULTAD QUIMICA-FARMACIA

Departamento de Licenciatura en Farmacia

Trabajo de Diploma

*Selección / identificación
asistida por computadora de nuevos compuestos líderes con
actividad anti-inflamatoria.*



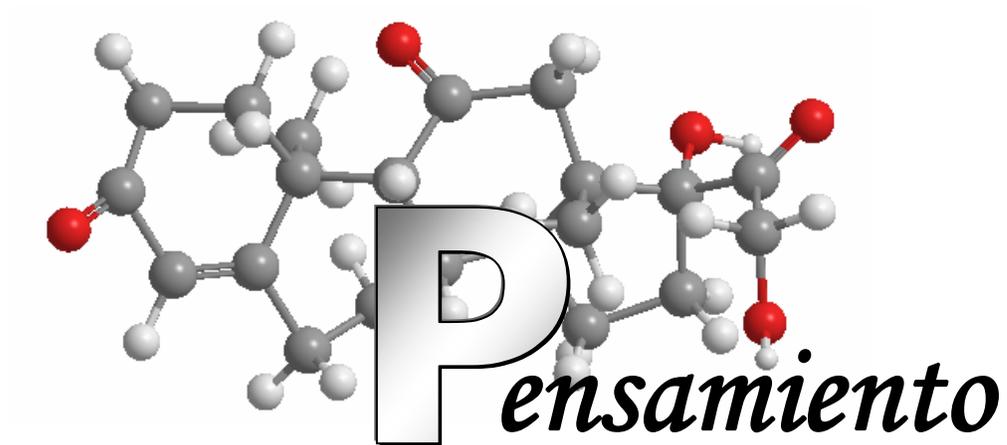
Autor: Carlos Andrés Cisneros Matos.

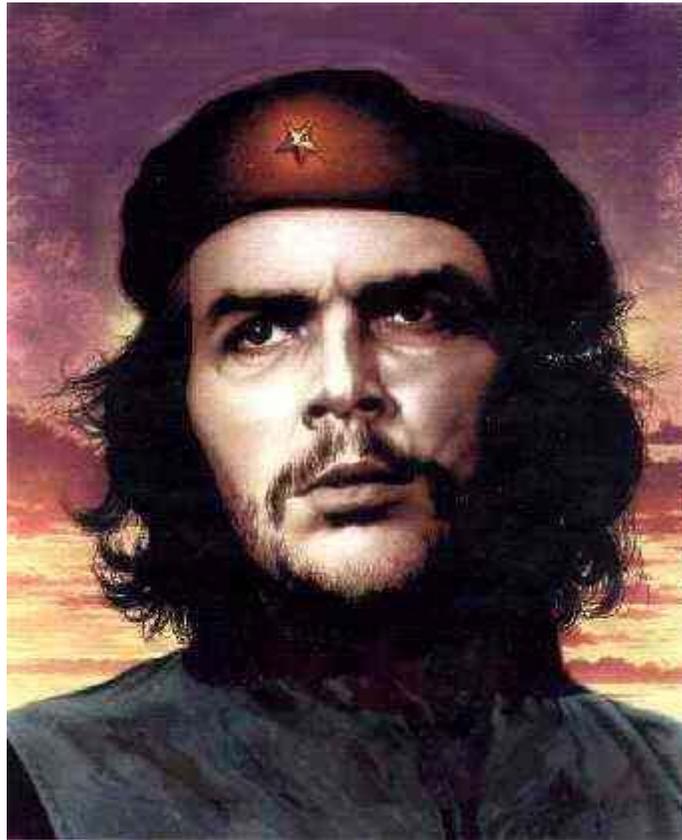
Tutores: Dr. Yovani Marrero Ponce. Profesor Asistente.

Lic. Dany Siverio Mota. Profesor Asistente.

Lic. Gerardo M. Casañola Martín. Profesor Instructor

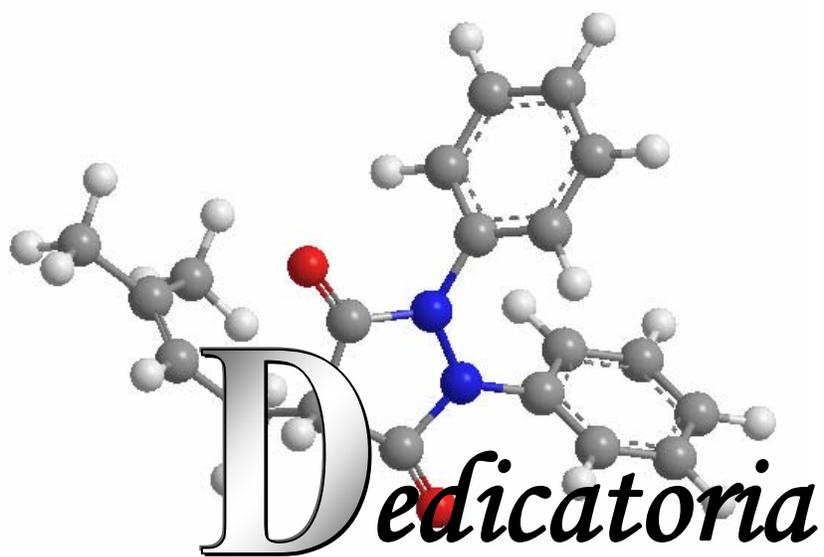






Lo que nosotros tenemos que practicar hoy, es la solidaridad. No debemos acercarnos al pueblo a decir: 'Aquí estamos. Venimos a darte la caridad de nuestra presencia, a enseñarte con nuestra presencia, a enseñarte con nuestra ciencia, a demostrarte tus errores, tu incultura, tu falta de conocimientos elementales'. Debemos ir con afán investigativo, y con espíritu humilde, a aprender en la gran fuente de sabiduría que es el pueblo".

Ernesto Guevara (Che).



Dedicataria

*Antes que todo a mi fiel amigo, al autor y consumidor de esta vida,
a aquel que me lleva en sus manos, a Dios.*

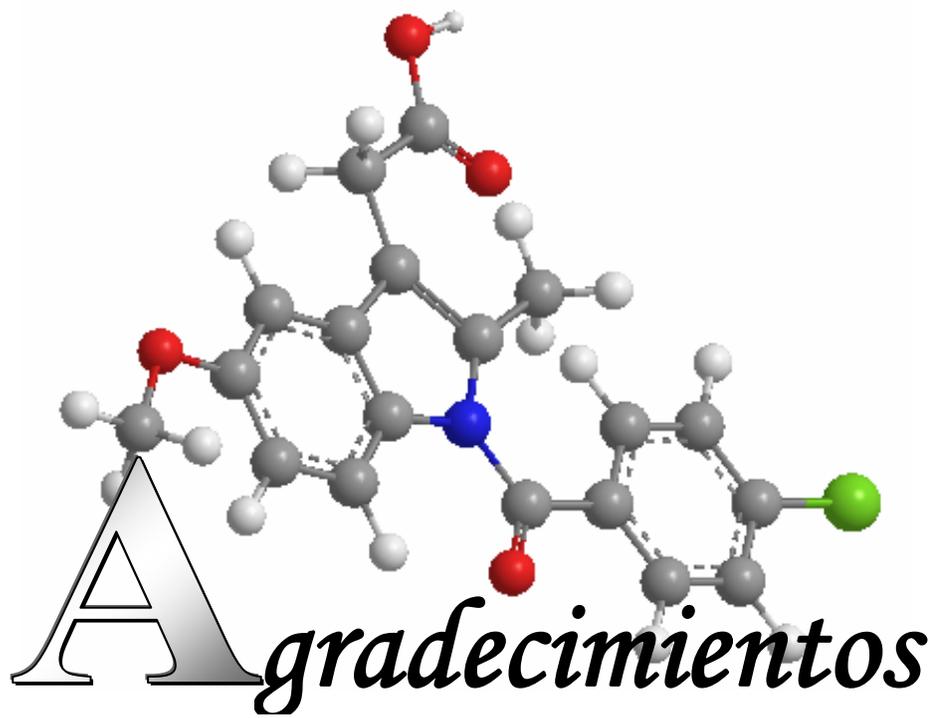
*La simple existencia de algunas personas trae felicidad y regocijo a
nuestras vidas. A ellas dedico este trabajo como muestra de su
permanencia en mi corazón.*

*» A mi primer y único amor, al complemento de mi vida, a mi esposa
Yaelis Cecilia.*

» A mis padres, mis buenos y abnegados padres.

*» A mi hermana y a toda la gente linda que conforman mi gran
familia.*

» A mis pastores Jesús y Lidia.



Siempre he confiado en un Dios de poder que no juega a los dados, que me acompaña cada día. Para ti Señor, mi primer agradecimiento. Sin Ti nunca hubiese llegado hasta aquí.

¿Qué puedo decirle a la vida al finalizar la carrera? Que resultó ser larga y trabajosa, es cierto; pero no por eso callará mi boca de agradecer a aquellos que en su diario andar fueron la pared donde pude recostarme:

Gracias Yaelita por tu constante apoyo, lo más lindo que ha pasado en mi vida fue conocerte, contigo he podido experimentar lo que es un verdadero amor, profundo y transparente.

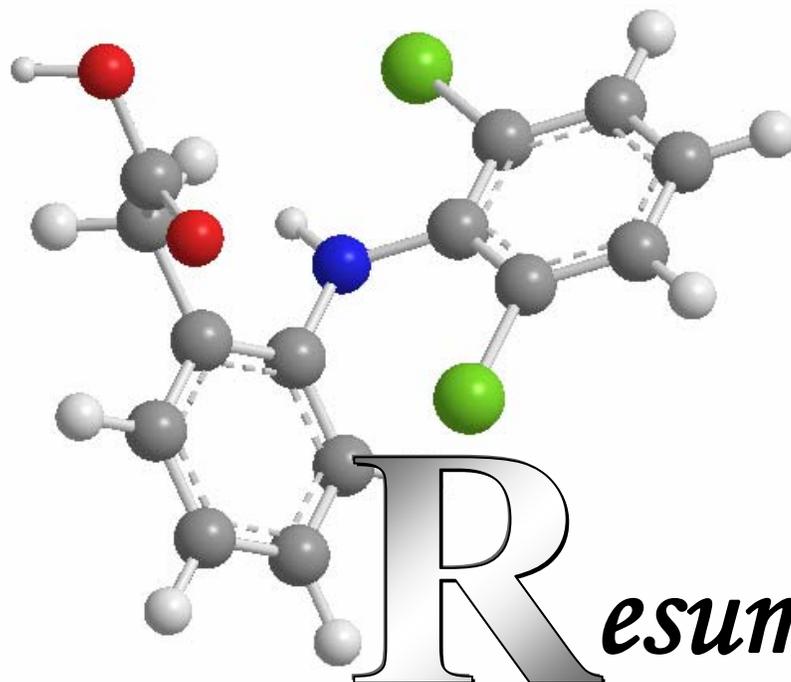
Gracias papi y mami por confiar siempre en mi, por darme lo mejor de ustedes, por haberme puesto en la senda correcta que ha hecho de mi lo que hoy con orgullo soy, por siempre gracias, siempre los querré.

Gracias a mi hermana, mi querida suegra y a toda mi gran familia por todo lo que han hecho por mi.

Gracias Jesús y china mis queridos pastores por los consejos y la ayuda que siempre me han dado.

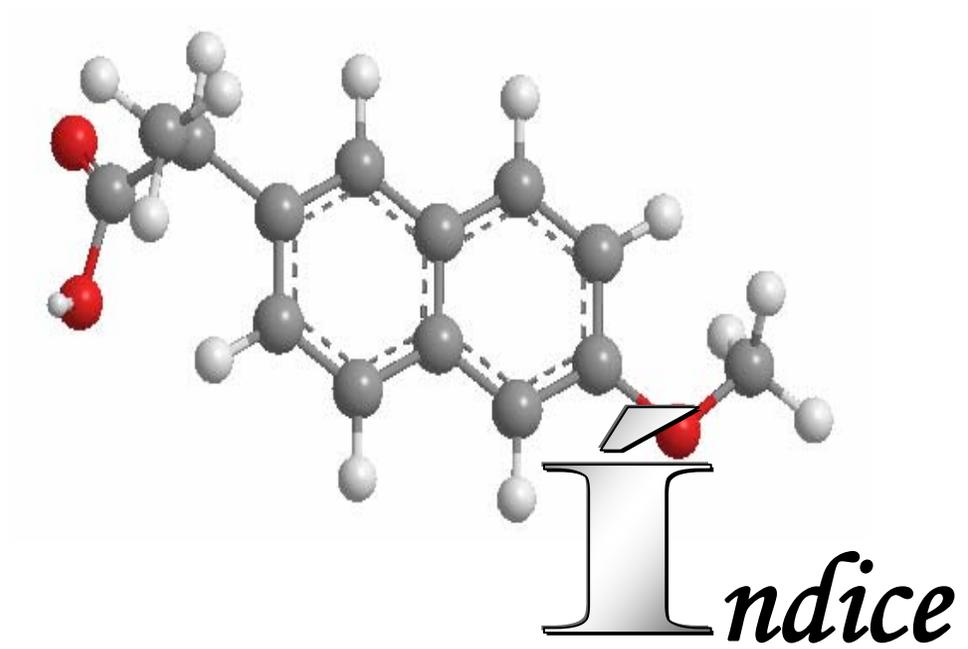
Gracias a mis compañeros de aula (en verdad me gustó compartir con ustedes), por el cariño que supieron darme; personas como ustedes (en especial a Yaima, Katia, Yaumary, Maidelys, Yusleidy, Enisleidy y Gastón) siempre estarán en mi corazón.

Gracias a mis tutores por el tiempo que me dedicaron. Gracias Dany, Yuly y Gerardo, sin ustedes se que esto no hubiese llegado al final.



Resumen

En este trabajo se desarrollan una serie de modelos de relaciones cuantitativas estructura-actividad (QSAR) para la modelación de la actividad antiinflamatoria que permitan realizar procesos de cribado virtual. Con este objetivo fue colectada de la literatura internacional una base de datos de 587 compuestos con actividad antiinflamatoria. Fueron empleados como descriptores moleculares los índices lineales de átomo implementados en el software TOMOCOMD-CARDD. Para identificar los descriptores que permitan la separación de la data en dos clases (compuestos antiinflamatorios y sin actividad antiinflamatoria) el análisis de pasos hacia delante fue usado como método de selección de variables y entonces los modelos fueron desarrollados usando el Análisis Discriminante Lineal como técnica estadística de clasificación. Para acceder al poder predictivo de los modelos, se realizó una validación con una serie de predicción externa. Los resultados de los análisis indican que los descriptores **TOMOCOMD-CARDD** totales y locales (heteroátomos e H-unidos a heteroátomos), proporcionan una excelente separación de la data (> 86%) en la serie de entrenamiento y en la serie de predicción. Finalmente, todos los modelos se utilizaron para estimar in-silico la actividad de compuestos con otros usos farmacológicos y nuevas entidades moleculares. Varios fármacos utilizados en la terapéutica actual y nuevas cabezas de serie fueron identificados como posibles antiinflamatorios pero se necesita de una posterior corroboración experimental. De forma general podemos concluir que el método TOMOCOMD-CARDD permite el desarrollo de modelos QSAR útiles en el descubrimiento biosilico de nuevos fármacos con actividad antiinflamatoria.



Indice.

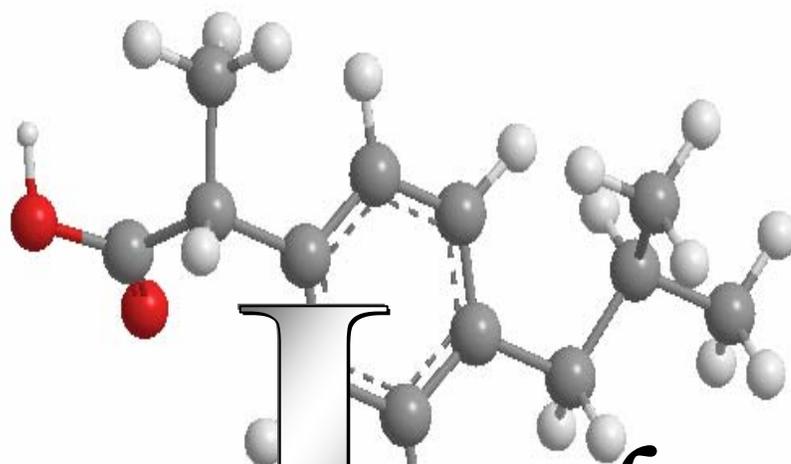
Introducción	1
Capítulo I: Revisión Bibliográfica	5
1.1. Inflamación. Consideraciones generales.....	5
1.2. Fármacos Antiinflamatorios.....	6
1.2.1. Fármacos Antiinflamatorios esteroideos.....	6
1.2.1.1. Mecanismo de acción.....	6
1.2.1.2. Efectos tóxicos.....	7
1.2.1.3. Usos terapéuticos.....	7
1.2.2. Fármacos Antiinflamatorios no esteroideos (AINEs).....	9
1.2.2.1. Orígenes históricos.....	9
1.2.2.2. Mecanismo de acción.....	9
1.2.2.3. Clasificación.....	10
1.2.2.3.1. Atendiendo a su Estructura química.	10
1.2.2.3.2. Desde el punto de vista clínico.....	12
1.2.2.4. Principales AINEs.....	13
1.2.2.4.1. Salicilatos.....	13
1.2.2.4.2. Pirazolonas y análogos.....	14
1.2.2.4.3. Derivados indolacéticos.....	14
1.2.2.4.4. Derivados arilacéticos.....	15
1.2.2.4.5. Derivados arilpropiónicos.....	15
1.2.2.4.6. Oxicamas y análogos.....	16
1.2.2.4.7. Fenamatos.....	16
1.2.2.5. Otras acciones farmacológicas de los AINEs.....	16
1.2.2.5.1. Acción analgésica.....	16
1.2.2.5.2. Acción antitérmica.....	17
1.2.2.5.3. Acción antiagregante plaquetaria.....	18
1.2.2.6. Efectos secundarios de los AINEs.	18
1.2.2.6.1. Gastrointestinal.	18
1.2.2.6.2. Renal.....	19
1.2.2.6.3. Reacciones alérgicas.....	20
1.2.2.6.4. Hematológicos.....	20
1.2.2.6.5. Inhibición de la motilidad uterina.....	20
1.2.2.6.6. Otros efectos adversos.....	20

Continuación

1.2.2.7.	Métodos tradicionales para la evaluación de la actividad antiinflamatoria	21
1.2.2.7.1.	Técnicas generales.....	21
1.2.2.7.1.1.	Modelos de inflamación aguda.	21
1.2.2.7.1.2.	Modelo de inflamación subcrónica o crónica.	22
1.2.2.7.2.	Técnicas específicas.	22
1.2.2.7.2.1.	Bolsa de aire en ratas.	22
1.2.2.7.2.2.	Peritonitis inducida por carragenina.	22
1.2.2.7.2.3.	Actividad ciclo – oxigenasa.	23
1.2.2.7.2.4.	Permeabilidad capilar incrementada por histamina.	23
1.2.2.7.2.5.	Método Embriones de peces cebrá.....	23
1.2.2.8.	Usos terapéuticos.Tratamiento del dolor.....	24
Capítulo II: Aspectos computacionales y quimiométricos.....		25
2.1.	Método computacional y análisis quimiométrico de los datos.....	25
2.1.1.	Generalidades.....	25
2.1.2.	Método computacional.....	27
2.1.2.1.	Descriptores <i>TOMOCOMD-CARDD</i> Basados en relaciones entre Núcleos Atómicos.....	28
2.1.2.1.1.	Información Química y Vector Molecular basado en Átomos.....	28
2.1.2.1.2.	“Background” sobre las Matrices de Densidad Electrónica Grafo-Teóricas No Estocásticas y Estocásticas.....	28
2.1.2.1.3.	Definición de los Índices Lineales basados.....	30
2.1.3.	Análisis quimiométrico de los datos.....	33
2.1.3.1.	Análisis Discriminante Lineal (ADL) para la predicción de la actividad.....	33
2.1.3.2.	Estimación de los Coeficientes.....	33
2.1.3.3.	Matriz de Clasificación.....	34
2.1.3.4.	Significación de la Función Discriminante (FD).....	34
2.1.3.5.	Criterios de Selección de Variables.....	35
2.1.3.5.1.	Selección del Número Óptimo de Predictores. Principio de la Parsimonia.....	36
2.1.3.5.2.	Tolerancia.....	36
2.1.3.5.	Multicolinealidad entre variables, redundancia de la Información y correlaciones lineales.....	37
2.1.3.7.	Validación Estadística de los Modelos QSAR.....	37
2.1.3.7.1.	División de la Data Original en Serie de Entrenamiento....	38

Continuación

Capítulo III: Materiales y Métodos	39
3.1. Base de datos de los compuestos usados en la obtención de los Modelos ADL-QSAR.....	39
3.2. Método Computacional. <i>TOMOCOMD-CARDD</i> software....	41
3.3. Desarrollo de las Funciones Discriminantes (FD) usando ADL.....	44
Capítulo IV: Resultado y discusión.....	46
4.1. Diseño de la base de datos.....	46
4.2. Desarrollo y Validación de las Funciones Discriminantes.....	49
4.3. Segunda Serie de Predicción Externa. Simulación de Screening' Virtual e Identificación de Nuevas Entidades Moleculares Como Fármacos Antiinflamatorio: Ensayos <i>In-Silico</i>	56
Conclusiones	58
Recomendaciones	59
Referencias bibliográficas.	60



I *ntroducción*

Introducción

La inflamación es una reacción defensiva perfectamente fisiológica y su importancia para el organismo es trascendental. Su sentido biológico consiste en intentar aislar un agente externo o interno que está produciendo una agresión física, para después tratar de destruirlo. En algunos casos suele tornarse complicado y pueden provocarse más daños de los que se pretendían evitar, debido a la falta de auto control del proceso, que provoca una cadena de eventos que escapan al control metabólico del tejido (1).

Hoy en día contamos con un buen arsenal de fármacos utilizados para tratar la inflamación, estos se encuentran divididos en dos grupos, por un lado los antiinflamatorios esteroides (glucocorticoides) y por el otro los antiinflamatorios no esteroides (AINEs).

Los glucocorticoides son antiinflamatorios inespecíficos, responden frente a cualquier causa de inflamación sea física, química, biológica e inmunológica. Actúan (como todas las hormonas esteroidales) en los distintos órganos efectores mediante la incorporación del esteroide al interior de la célula. En este orden se conoce la existencia de 2 tipos de receptores intracelular, siendo el número 2 el más específico para los glucocorticoides. Es importante conocer el uso de estos fármacos para lograr un buen efecto. Su toxicidad puede ser a nivel local y sistémico. Un uso crónico puede provocar atrofia de piel y alteraciones en el metabolismo, por el contrario el retiro de estos preparados puede producir un cuadro de pseudotumor cerebral que se manifiesta con edema de papila.

El grupo de los antiinflamatorios no esteroides (AINEs) ofrece una gama muy extensa de productos equiparables en eficacia y en incidencia de efectos adversos, y con grandes diferencias en la respuesta individual incluso entre miembros de una misma familia química. Los fármacos antiinflamatorios no esteroides (AINEs) son un grupo heterogéneo de sustancias con capacidad analgésica, antiinflamatoria y antipirética por su acción inhibiendo la ciclooxigenasa (COX) en sus distintos isómeros COX-1, COX- 2 y COX- 3; y de

esta forma la síntesis y liberación de prostaglandinas, mediando en la patogenia de la inflamación, la fiebre y el dolor.

A pesar de la importante utilidad de los AINEs no se puede desdeñar su efecto negativo potencial como consecuencia de su uso frecuente y sus efectos adversos, mayoritariamente debidos a la denominada gastroenteropatía inducida por AINEs que puede tener una expresión clínica muy diversa, en la que se incluyen desde síntomas dispépticos, hasta manifestaciones graves como la hemorragia o la perforación.

La obtención de nuevos compuestos cabezas de serie (o también conocidos como compuestos líderes) útiles en el tratamiento de la inflamación, constituye una tarea de primer orden para la industria farmacéutica actual. Tradicionalmente, el método que se emplea para la búsqueda de nuevos principios activos, es el basado en el sistema de “*prueba y error*” a través de ensayos masivos de gran número de sustancias químicas, el mismo es cada vez más ineficiente, pues se hace necesario ensayar más de 10.000 compuestos para encontrar el deseado. La *Boston Consulting Group* (BCG) realizó un estudio el cual involucró 50 compañías e instituciones académicas; los resultados obtenidos mostraron que el costo de desarrollo de un nuevo medicamento desde la identificación de su diana farmacológica, el descubrimiento y optimización de uno de los nuevos compuestos líderes, el desarrollo de los ensayos clínicos requeridos y su uso autorizado en terapéutica es como promedio de 880 millones de dólares (USD) y se necesita también como promedio un período de 15 años de investigación. Además, 7 de cada 10 medicamentos no recuperan su costo (2). Estas características desfavorables (inefectividad, alto costo y gran consumo de tiempo) hacen el método ‘tradicional’ de evaluación al *azar* inaplicable para los países en vías de desarrollo e incluso para las grandes transnacionales farmacéuticas.

Teniendo en cuenta que el paso limitante en el descubrimiento y desarrollo de nuevos fármacos, es la identificación y la optimización de nuevos compuestos líderes de un modo efectivo (en el menor tiempo posible y a un costo razonable), los métodos *in silico* vienen a ser la solución de ese problema, estas técnicas

han tenido y tienen potencial para mejorar significativamente el descubrimiento y posterior desarrollo de fármacos, por lo que el enfoque de diseño/descubrimiento de fármacos asistido por computadoras ofrece una alternativa para el mundo real de síntesis y evaluación.

Este procedimiento engloba todas las técnicas asistidas por ordenadores usadas en el diseño/descubrimiento y optimización de compuestos con propiedades deseadas y ha jugado un rol fundamental en el desarrollo de un número de fármacos que están ahora en el mercado. Este tipo de estudio se basa en el uso de un mundo “virtual” de hipótesis, generadas por computadoras y probadas en la práctica. Este tipo de procedimiento ‘*in silico*’ evita las tareas costosas para las actuales síntesis y bio-ensayos, que se hacen solamente después de la exploración de los conceptos iniciales con modelos QSAR.

El empleo de técnicas computacionales de diseño ‘*racional*’ constituye una útil herramienta para abaratar los costos y disminuir el tiempo de investigación requerido para la obtención de los objetivos propuestos. El sistema computacional *TOMOCOM-CARDD* nos permite la descripción de propiedades físicas, químico-físicas, químicas y biológicas (toxicológicas, farmacocinéticas y farmacológicas) así como el descubrimiento de nuevos compuestos bioactivos.

Por todo lo anterior se plantea el siguiente problema científico:

Los métodos tradicionales de ‘prueba y error’ para la búsqueda de nuevos fármacos han demostrado ser costosos y de baja efectividad. Aunque se han utilizado varios enfoques computacionales en la predicción y el diseño de fármacos antiinflamatorios no han sido probados aun los descriptores *TOMOCOMD-CARDD*.

Por lo que se formula la siguiente hipótesis:

Si es posible a través del uso de modelos matemáticos basados en el uso de los descriptores *TOMOCOMD-CARDD* obtener funciones que relacionen adecuadamente la estructura química con la actividad anti-inflamatoria se podrá

realizar la identificación/selección de nuevos compuestos líderes a través del tamizaje (cribado) virtual de grandes bases de datos.

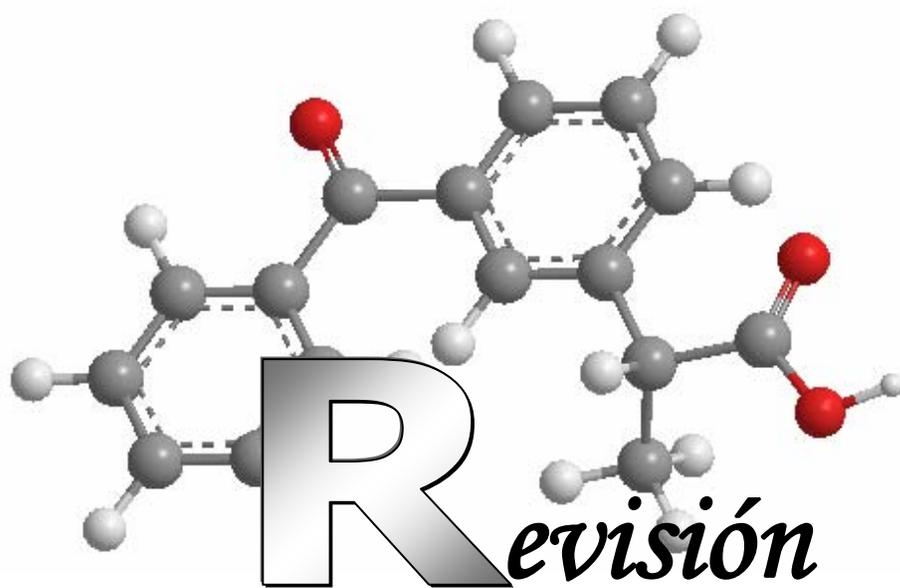
Siendo el objetivo general de este trabajo:

Desarrollar modelos matemáticos capaces de identificar/seleccionar nuevos compuestos líderes con actividad antiinflamatoria utilizando los descriptores *TOMOCOMD-CARDD*.

Objetivos específicos:

1. Recolectar una base de datos de la literatura de compuestos a los que se les ha reportado experimentalmente actividad antiinflamatoria para acceder al análisis y la modelación confiable de la data.
2. Obtener modelos QSAR utilizando descriptores *TOMOCOMD-CARDD* que permitan la descripción de la actividad antiinflamatoria,
3. Aplicar los modelos obtenidos al “screening” virtual de compuestos químicos para evidenciar las potencialidades del método *TOMOCOMD-CARDD* en el descubrimiento (selección/identificación y/o diseño) de nuevos compuestos líderes.

La **novedad científica** de este trabajo se fundamenta en la aplicación por vez primera y acorde con las últimas tendencias mundiales, de un método “*in silico*” (*TOMOCOMD-CARDD*) para la descripción en términos estructurales de la actividad antiinflamatoria. La utilización de esta metodología conlleva a un resultado final totalmente aplicable a la estrategia de trabajo de la unidad *CAMD-BIR* en su inmediato devenir, siendo también de posible utilización por otras instituciones científicas y académicas de nuestro país cuyo principal objetivo sea la racionalidad en la obtención de nuevos potenciales candidatos a fármacos con actividad antiinflamatoria.



Bibliográfica

Capítulo I

Capítulo I: Revisión bibliográfica.**1.1. Inflamación. Consideraciones generales.**

Los procesos inflamatorios se desarrollan como parte de los mecanismos de defensa que tienen lugar en el organismo en respuesta a estímulos de diversa naturaleza. Estos estímulos pueden abarcar desde lesiones provocadas por traumatismos mecánicos, procesos isquémicos, efecto de agentes patógenos como virus, parásitos o bacterias, interacciones antígeno-anticuerpo desarrollados por la actividad de los diferentes componentes del sistema inmunológico, hasta la acción de diversos productos químicos. Solo en los casos en que estos factores puedan ser eliminados por la activación de los sistemas de defensa del organismo, no se presentan visiblemente las consecuencias patológicas del proceso inflamatorio (3). Pero si ellos se encuentran en grandes cantidades, presentes en sitios inusuales, o son capaces de modificar algunas de las funciones del sistema inmunológico, pueden ocurrir daños tisulares como consecuencia de las reacciones inmunes y esto a su vez facilitar el desarrollo de diferentes enfermedades inflamatorias (4,5). La respuesta inflamatoria es coordinada por varias células como los macrófagos, linfocitos, leucocitos polimorfo nucleares, mastocitos entre otras, las cuales producen un gran número de mediadores que juegan un role importante en la fisiopatología del proceso inflamatorio, como los metabolitos del ácido araquidónico, las especies reactivas de oxígeno y nitrógeno, enzimas hidrolíticas, histamina, diversas citocinas, entre otros, que producidos en exceso, son responsables de severos daños tisulares (3). Los signos clásicos que macroscópicamente evidencian la inflamación son: calor (incremento de la temperatura en la zona inflamada), rubor (enrojecimiento, eritema), tumor (inflamación, edema), dolor (hiperalgesia) y pérdida de la función tisular, que son consecuencia de efectos inducidos por los mediadores en el influjo sanguíneo local, la permeabilidad vascular, la infiltración de leucocitos y la liberación de agentes inductores de dolor (1).

1.2. Fármacos Antiinflamatorios.

La inflamación es tan antigua como el ser humano y causante de diversos males, es por ello, que el hombre desde sus inicios se ha ocupado por aliviarla. Esto conllevó al surgimiento de los fármacos antiinflamatorios, los que se dividen en dos grandes grupos: Antiinflamatorios esteroideos y no esteroideos (AINEs).

1.2.1. Fármacos Antiinflamatorios esteroideos.

Los antiinflamatorios esteroideos están representados por los glucocorticoides (hormonas producidas por la glándula suprarrenal), los que ejercen una poderosa acción antiinflamatoria, sea cual fuere la causa de la inflamación (infecciosa, química, física o inmunológica), pudiendo inhibir tanto las manifestaciones inmediatas de la inflamación (rubor, dolor, etc.) como tardías, entendiéndose por tales ciertos procesos de cicatrización y proliferación celular. Inhiben la dilatación vascular, reducen la transudación líquida y la formación de edema, disminuyen el exudado celular y reducen el depósito de fibrina alrededor del área inflamada. Para que esta acción se manifieste, son necesarias dosis farmacológicas, pero la respuesta es tan intensa que los glucocorticoides están catalogados como los antiinflamatorios más eficaces (6). Dentro de este grupo de compuestos tenemos a la Cortisona, la Prednisolona, la Dexametasona, la Hidrocortisona y muchos otros más.

1.2.1.1. Mecanismo de acción.

El mecanismo fundamental de acción por donde actúan estos fármacos, es la estimulación de la síntesis de la proteína lipocortina. Esta enzima inhibe la fosfolipasa A2 que es la responsable de la obtención del ácido araquidónico a partir de fosfolípidos de membrana. El ácido araquidónico es el precursor de algunos mediadores celulares que intervienen en el proceso de inflamación: prostaglandinas (PG), tromboxanos y leucotrienos. Los glucocorticoides actúan a distintos niveles: desde la acción del macrófago hasta la amplificación de la respuesta humoral, celular y esto da cuenta de su acción antiinflamatoria e

inmunosupresora. A nivel del macrófago inhibe la liberación del factor de necrosis tumoral, impide la acción de cininas, también impide la acción de los interferones, inhibe la secreción de interleuquina 1, a nivel de las células T activadas, también interferirá en la secreción de interleuquina 2. Por tanto, a modo de resumen se puede decir que el mecanismo por el cual los glucocorticoides obtienen las propiedades antiinflamatorias e inmunosupresoras son:

1. Inhibe la síntesis de prostaglandinas y leucotrienos.
2. Inhibe la liberación de interleuquina 1.
3. Interfiere con la función del macrófago que es la célula donde se inicia la respuesta inmunológica.
4. Al interferir la secreción de interleuquina 2 va a frenar la amplificación de la respuesta inmune.
5. Inhibe otros factores que participan en el proceso inflamatorio e inmunitario como el factor activador de plaquetas (PAF) y el factor de necrosis tumoral (TNF).
6. Estabiliza las membranas lisosomales.
7. Mantiene el tono vascular al inhibir la oxido nítrico sintetasa (NOS).

1.2.1.2. Efectos tóxicos.

Se debe tener presente su uso en los niños por los efectos en el metabolismo del calcio, pues pueden producir un retardo en su crecimiento.

1.2.1.3 Usos terapéuticos.

Los glucocorticoides poseen dos usos terapéuticos fundamentales:

1. De sustitución a los glucocorticoides endógenos:
 - Insuficiencia suprarrenal aguda (dosis altas de corticoides).

- Insuficiencia suprarrenal crónica (tratamiento Standard, cortisol a dosis 20-25mg)
- Hiperplasia suprarrenal congénita.

2. Enfermedades no endocrinas:

- Alergias: deben administrarse por un período corto, si estamos frente a un shock anafiláctico usamos: adrenalina (es lo primero a utilizar y que le salvará la vida al paciente), posteriormente administración parenteral de hidrocortisona.
- Asma (debido a que en la patogenia del asma, los leucotrienos tienen gran participación en el proceso inflamatorio) no sólo se utilizan en crisis cuando no hay respuesta a broncodilatadores, sino también en tratamiento crónico por vía inhalatoria. Si el paciente a pesar de estar con corticoides inhalatorios entra en crisis estos pueden utilizarse por un período corto por vía oral.
- Mesenquimopatías: Artritis reumatoide.
- Enfermedades hematológicas: púrpura trombocitopénico idiopático, anemia hemolítica.
- Edema cerebral: tiene buena respuesta aquel producido por una masa tumoral.
- Hepatitis crónica activa.

Los glucocorticoides se utilizan en patologías con respuesta inmunológica exacerbada, como inmunosupresor en el caso de transplantes y en enfermedades autoinmunes y también en las terapias de cáncer (especialmente en los cánceres hematológicos en los cuales tiene efecto linfo lítico de ahí su efecto en los cánceres que dependen de la serie linfocítica, también para el manejo de complicaciones como la hipercalcemia) preferentemente la prednisona o la metilprednisolona (7).

1.2.2. Fármacos Antiinflamatorios no esteroideos (AINEs).

Los AINEs integran el grupo más amplio y heterogéneo de compuestos antiinflamatorios. A pesar de esta diversidad estructural, estas drogas comparten propiedades farmacológicas y en muchos casos efectos terapéuticos y colaterales.

1.2.2.1. Orígenes históricos.

La primera referencia a los AINEs se encuentra en el siglo V antes de Cristo, cuando el reconocido padre de la medicina griega, Hipócrates, empleó una pócima compuesta de extracto de corteza y hojas de sauce, *Salix latifolia*, para aliviar los dolores y controlar la fiebre. Se conoce que en la Edad Media, la corteza de sauce hervida era empleada por las curanderas para dar de beber a las personas que sentían dolores. El ingrediente activo de la corteza del sauce fue un glucósido amargo de color amarillento desarrollado en el año 1828 por el profesor de Farmacia Johan Andreas Buchner y quien la llamó salicina. Un año más tarde el químico francés Leroux logró convertir la salicina en cristales, aislándola así en su forma pura. En la hidrólisis, la salicina genera glucosa y alcohol salicílico, el mismo puede ser transformado en ácido salicílico in vivo o por manipulación química. Los excelentes resultados de este fármaco hicieron que Hoffman (un químico al servicio de la firma Bayer) preparara el ácido acetilsalicílico. No fue hasta 1899 que Dreser comenzó a utilizar dicho compuesto en medicina, una vez demostrado sus efectos antiinflamatorios.

1.2.2.2. Mecanismo de acción.

La acción básica es la inhibición de la ciclooxigenasa (COX), enzima que convierte el ácido araquidónico en endoperóxidos cíclicos, los cuales se transforman en PG y en tromboxanos. Esta acción inhibitoria de la enzima por los AINEs es también la responsable de los efectos indeseables de este grupo de fármacos. Los AINEs actuales no pueden inhibir la enzima lipooxigenasa, enzima responsable de la síntesis de leucotrienos, por lo que continúan formándose estos mediadores celulares y otros elementos activos. Este hecho

constituye una limitación en el empleo de éstos fármacos para controlar los procesos en los que intervengan numerosos mediadores. La inhibición de la acción de la ciclooxigenasa es de forma competitiva y reversible reduciendo de esta forma la síntesis de PG, excepto la aspirina que la inhibe irreversiblemente acetilando de manera covalente el sitio activo de la enzima. En las últimas décadas del siglo pasado, fueron identificadas dos tipos de ciclooxigenasa COX-1 y COX-2, la COX-1 es inducida por factores fisiológicos y tiene acciones protectoras, su activación produce tromboxanos A₂ a nivel plaquetario, Prostaciclina (PC) en la mucosa gástrica y PG E₂ (efecto vasodilatador) a nivel renal. La síntesis de COX-2 es inducida en presencia de patógenos o anomalías propias del organismo formándose los prostanoides. La inhibición de la COX1 es la causa de los efectos adversos más característicos de los AINEs aunque se han descrito efectos indeseables en el sistema cardiovascular para fármacos inhibidores selectivos de la COX2. En el 2002 Simmons y colaboradores, de la Universidad Brigham Young de Estados Unidos identificaron una nueva isoforma de ciclooxigenasa la cual fue denominada COX-3. Esta deriva del gen que codifica la COX-1, pero es estructural y funcionalmente distinta. Se estableció que su papel es a nivel de la producción de la fiebre, el dolor a nivel central y no a nivel de la inflamación. Al evaluar los medicamentos que producen inhibición farmacológica se identificó al acetaminofen como el más selectivo hacia ella (8).

1.2.2.3. Clasificación.

1.2.2.3.1. Atendiendo a su Estructura química.

a) Salicilatos:

ASA (ácido acetilsalicílico)

Acetilsalicilato de lisina

Benorilato

Diflunisal

Etersalato

Salsalato

Diflunisal

b) Derivados pirazolónicos:

Aminofenazona (dipirona o metamizol)

Fenilbutazona

Azaprofazona

c) Derivados del para-aminofenol:

Acetaminofen (paracetamol o tylenol)

d) Derivados del ácido acético:

Indometacina

Sulindaco

Glucametacina

e) Derivados carboxílicos y pirrolpirrólicos:

Etodolaco

Ketorolaco

f) Derivados del ácido fenilacético:

Diclofenaco (voltaren)

Aclofenaco

Tolmetina

Fenclofenaco

g) Derivados del ácido n-acetilantranílico:

Ácido mefenámico

Niflumico

Meclofenamico

Clonixinato de lisina

h) Derivados del ácido propiónico:

Ibuprofeno, Naproxeno, Ketoprofeno

Flurbiprofeno, Fenoprofeno, Oxaprozina

i) Derivados enólicos

Piroxican

Meloxican

Tenoxican

j) *Nimesulida, sulfonanilida*

k) *Grupo naftilcanonas:*

Nabumetona

L) Otros AINEs del grupo de los **COXIB**

-- *Productos que contienen como principio activo CELECOXIB:*

- CLOXIB - KLONAL
- COXEL - BETA

-- *Producto que contiene como principio activo VALDECOXIB:*

- BEXTRA – PFIZER

-- *Producto que contiene como principio activo ETORICOXIB:*

- ARCOXIA – MSD

-- *Producto que contiene como principio activo PARECOXIB:*

- BEXTRA IM/IV - PFIZER

1.2.2.3.2. Desde el punto de vista clínico.

Desde el punto de vista clínico se clasifican en cuatro grupos los que inhiben indistintamente la COX1 y la COX2 como el ibuprofeno, los que inhiben preferentemente la COX1 como la Indometacina y los que lo hacen sobre la COX2, nabumetona, meloxicam, rafecoxib, celecoxib, parecoxib (9).

Así el salsalato pertenece al grupo de los salicilatos, pero su poder de inhibición de la COX es distinto al de la aspirina y es comparable al del acetaminofen o paracetamol. Lo mismo sucede con el meloxicam que perteneciendo al grupo de las oxicamas su poder de inhibición es distinto al del piroxicam, es más inhibidor de la COX 2 que de la COX 1.

Interesa conocer la clasificación de los AINEs en relación a su poder de inhibición de la ciclooxigenasa (COX), tabla 1.1:

COX1 / COX2	↓COX1 / ↓COX2	↓COX1 / ↓↓COX2	Selectivo COX2	COX3
Piroxicam	Acetaminofen	Nimesulide	Colecoxib	Acetaminofen
Indometacina	Salsalato	Meloxicam	Rofecoxib	
Sulindac				
Tolmetin				
Ibuprofen				
Naproxen				
Fenoprofen				
Meclofenamato				
Diflunisal				
Ketoprofen				
Diclofenac				
Ketorolac				
Etodolac				
Nabumetona				
Oxaprozin				

Tabla 1.1. Clasificación de los AINEs en relación a su poder de inhibición de la COX.

1.2.2.4. Principales AINEs.

1.2.2.4.1. Salicilatos.

El ácido acetilsalicílico sigue teniendo acción antiinflamatoria equiparable a los medicamentos más potentes del grupo y, por tanto, es el medicamento de elección bajo un criterio estricto de costo/eficacia. Debe tenerse presente, sin embargo, que a las dosis utilizadas en enfermedades reumáticas (más de 3,5 g/día), un porcentaje alto de los pacientes no pueden tolerar los efectos secundarios gastrointestinales y neurológicos. El gran número de comprimidos de la toma diaria constituye también un obstáculo al cumplimiento de la prescripción.

Los preparados micronizados solubles o tamponados de ASA, como el acetilsalicilato de lisina, pueden mejorar la tolerancia gástrica. Benorilato y etersalato son ésteres de ASA con paracetamol y tienen también menos efectos gastrointestinales que el ASA, con la ventaja de poder administrarse cada 8-12 horas. Se transforman lentamente en el hígado en ácido acetilsalicílico y paracetamol. Parecidas ventajas tiene salsalato, que se absorben en el intestino delgado y se hidroliza lentamente a ácido salicílico. Diflunisal puede administrarse cada 12 horas y parece más eficaz y mejor tolerado que el ASA.

1.2.2.4.2. Pirazolonas y análogos.

Los derivados pirazolónicos tienen acción antiinflamatoria muy potente pero han sido sometidos a restricción legal por producir efectos secundarios graves, sobre todo discrasias sanguíneas. La incidencia es muy baja pero significativamente superior a la de otros grupos de antiinflamatorios. Por lo tanto, dentro de los AINEs, las pirazolonas deben considerarse medicamentos de último recurso, aunque muchos siguen estimando que la fenilbutazona es uno de los fármacos más eficaces en espondilitis anquilosante.

1.2.2.4.3. Derivados indolacéticos.

Indometacina sigue siendo uno de los antiinflamatorios más eficaces, pero tiene una incidencia elevada de efectos gastrointestinales y sobre el SNC (cefaleas, vértigo, etc). Los demás son análogos estructurales que buscan la misma acción con menos efectos secundarios, pero no hay que contar con que lo hayan conseguido. Si se busca un compuesto del grupo con menos efectos adversos que indometacina, tolmetin es posiblemente una buena elección ya que es el mejor documentado y el que estructuralmente se parece menos a indometacina, pero las diferencias de toxicidad no son muy grandes en cualquier caso.

1.2.2.4.4. Derivados arilacéticos.

Los fármacos de este grupo son parecidos a los del anterior. De hecho desde una perspectiva puramente química los derivados indolacéticos son un subconjunto de este grupo de los aril-acéticos. Se han clasificado en grupo aparte por la importancia que tiene la indometacina en la terapia antiinflamatoria. De todas formas diclofenac, en sus formas normal y retardada, es el AINE más prescrito en España, y, en la medida en que sean fiables las tablas comparativas de incidencia de efectos adversos, no puede decirse que sea mala elección.

Nabumetona es un profármaco que tras la absorción intestinal se transforma en un metabolito activo muy parecido al naproxeno (el radical ácido es acético, no propiónico, por ello se clasifica en este grupo y no en el siguiente). La administración es en dosis única diaria. En teoría es menos lesivo para la mucosa gástrica por ser un profármaco. Hasta que no haya más experiencia no puede decirse si esto se traduce en menor potencial ulcerógeno, pero la tolerancia gastrointestinal parece bastante buena.

1.2.2.4.5. Derivados arilpropiónicos.

Los derivados arilpropiónicos son generalmente menos potentes que los antiinflamatorios clásicos (ASA, fenilbutazona, indometacina), pero son mejor tolerados. Dentro de los AINEs, ibuprofeno tiene un excelente historial de seguridad, pero tal vez sea demasiado débil para ser útil en cuadros donde la inflamación sea muy predominante. Suele recomendarse de elección, como los analgésicos, en cuadros iniciales o predominantemente dolorosos.

Naproxeno es más potente, está en un lugar intermedio en las clasificaciones de tolerancia y tiene la ventaja de la toma cada 12 horas. Lo mismo ocurre con ketoprofeno. Por su parte, dexketoprofeno es el enantiómero activo del anterior y cabe esperar que tenga las mismas propiedades, aunque se administre a la mitad de dosis. El ácido tiaprofénico ha sido asociado a cuadros de cistitis y fenbufeno tiene una incidencia relativamente alta de reacciones dermatológicas.

1.2.2.4.6. Oxicamas y análogos.

Piroxicam tiene una semivida muy larga, que permite la administración una vez al día. Esta comodidad le ha convertido en uno de los antiinflamatorios más utilizados, aunque no es de los mejor tolerados. Meloxicam tiene selectividad alta hacia la ciclooxigenasa tipo 2 (inducible), pero no está claro que ello represente mayor seguridad en uso clínico. Parece equivalente en eficacia e incidencia de efectos adversos. Tenoxicam tiene una semivida aún más larga que el piroxicam (lo que significa que es el AINE de semivida más larga) pero es dudoso que esto suponga ya alguna ventaja frente al inconveniente teórico de que los efectos secundarios duran más tiempo. Nimesulida no es propiamente una oxicama, pero tiene una estructura química bastante parecida. También presenta una relativa selectividad hacia la COX-2. Lornoxicam es el más potente del grupo (entre 10 y 100 veces más que tenoxicam, 10 y 50 veces más que piroxicam y 20 veces más que diclofenac), y no presenta selectividad hacia la COX-2.

1.2.2.4.7. Fenamatos.

Tienen acción analgésica significativa pero efecto antiinflamatorio relativamente débil. El ácido meclofenámico es el antiinflamatorio más potente del grupo. Un inconveniente general es una incidencia relativamente alta de diarrea como efecto secundario (10).

1.2.2.5. Otras acciones farmacológicas de los AINEs.**1.2.2.5.1. Acción analgésica.**

La actividad antiálgica de los AINEs es de intensidad moderada o media, alcanzándose un techo analgésico claramente inferior al de los analgésicos opioides, pero frente a éstos presentan la ventaja de no alterar el sensorio o la percepción, lo cual redundo, en conjunto, en una utilización clínica menos comprometida.

Clásicamente se ha aceptado que la acción analgésica de los AINEs tiene lugar a nivel periférico, mediante la inhibición de la síntesis de las PG producidas en respuesta a una agresión o lesión tisular, impidiendo, por lo tanto, que los eicosanoides contribuyan, con su acción sensibilizadora sobre las terminaciones nerviosas nociceptivas, a aumentar la acción estimulante del dolor de otros mediadores allí liberados (histamina, bradicinina, etc.). Sin embargo, no existe ninguna correlación precisa entre la actividad anticiclooxigenasa *in vitro* y el efecto analgésico en todos los modelos de dolor experimental o clínico.

1.2.2.5.2. Acción antitérmica

La fiebre es una respuesta autónoma, neuroendocrina y conductual compleja y coordinada que se desencadena ante la existencia de una infección, lesión tisular, inflamación, rechazo de tejidos, tumores, etc., y sirve a una doble finalidad: alertar acerca de una situación anómala y potencialmente lesiva, y poner en marcha una serie de mecanismos fisiológicos para la defensa del organismo. Su manifestación cardinal es la elevación de la temperatura corporal del orden de 1 a 4 °C. El desencadenamiento de una reacción febril implica la existencia de *pirógenos endógenos* (citocinas, como IL-1, IL-6 e interferones) o tóxicos (endotoxinas o lipopolisacáridos liberados de bacterias gramnegativas) en áreas del SNC relacionadas con el control de la temperatura. Como áreas centrales importantes en el reconocimiento de los pirógenos endógenos circulantes destacan los órganos circunventriculares, que carecen de barrera hematoencefálica y, dentro de éstos, especialmente el órgano vascular de la *lamina terminalis*, donde se expresa la COX-1. El reconocimiento de los pirógenos origina la síntesis de prostaglandinas, principalmente E₂, como primer paso en la patogénesis de la fiebre, que actuarían (mediante el incremento en el AMPc) como mediadores paracrinos locales en el propio órgano vascular y áreas adyacentes (PO/HA), a partir de las cuales se coordinaría la respuesta febril.

1.2.2.5.3. Acción antiagregante plaquetaria

Es una acción que no comparten en la misma medida todos los AINE, aunque sea consecuencia de su efecto inhibitor de la COX-1. Reviste especial interés terapéutico en el caso del ASA debido, probablemente, al hecho de que su efecto inhibitor de la ciclooxigenasa es irreversible. Esta inhibición, que en la mayoría de las células del organismo se solventa con la síntesis de nuevas moléculas de COX, cobra un especial protagonismo, terapéutico o indeseable, en las plaquetas. Éstas son incapaces de sintetizar nuevas proteínas y una vez acetilada su COX, en el paso por la circulación portal previamente a la desacetilación hepática del ASA, resulta inhibida durante toda la vida de la plaqueta (8-11 días). Como consecuencia de esta acción se produce un marcado descenso de los niveles de TXA2 plaquetario (responsable de parte de los mecanismos que provocan la agregación plaquetaria). Esta acción, que es utilizada terapéuticamente en la prevención a largo plazo de accidentes tromboembólicos coronarios y cerebrales puede, asimismo, devenir en reacción adversa facilitando la aparición de hemorragias, especialmente en tratamientos o situaciones concurrentes que afecten la coagulación sanguínea (11).

1.2.2.6. Efectos secundarios de los AINEs.

1.2.2.6.1. Gastrointestinal.

El efecto secundario de los AINEs que con mayor frecuencia se observa es sobre el tracto gastrointestinal, provoca diarrea, estreñimiento, gastritis y dolor gástrico (15-25%), lesión de la mucosa gástrica o duodenal (ya que la integridad de la misma depende de la síntesis de PGs, los AINEs al disminuir esta síntesis modifica esta protección). Esta lesión se observan en el 40% de los pacientes sometidos a tratamientos con dosis terapéuticas bajas durante 3 meses aproximadamente. Este efecto sobre la mucosa tiene 2 componentes:

- 1 Efecto local agudo: la irritación local de las sustancias ingeridas permite la difusión retrógrada de ácido al interior de la mucosa gástrica y la inducción de daño tisular.
- 2 Efecto sistémico: descrito cuando se administraba un AINEs por otra vía distinta a la oral (ej. Intramuscular) en la cual no se produce la irritación directa, sin embargo observamos el daño a la mucosa, por inhibición de la biosíntesis de las PGs en estómago y, en particular, PGI₂ y PGE₂ que actúan como agentes citoprotectores de la mucosa estomacal.

En EE.UU. se calcula que por cada 10.000 prescripciones escritas ocurren 2 casos de toxicidad gastrointestinal (12). Se calcula que entre el 30 y el 50% de los pacientes que mueren de complicaciones relacionadas con enfermedad úlcero-péptica han tomado recientemente AINEs (13).

1.2.2.6.2. Renal

Las prostaglandinas ejercen efecto vasodilatador a nivel del riñón, la inhibición de éstas produce disminución en el flujo renal y la filtración glomerular(14). El uso de AINEs está asociado a toxicidad renal, los efectos más comunes son moderados y reversibles, sin embargo las complicaciones relativamente raras de nefritis intersticial y necrosis papilar son por lo general irreversibles (15). Se ha sugerido que Sulindac es el AINEs con menor toxicidad renal (16), sin embargo recientes trabajos ponen en duda esta aseveración (16,17). Al igual que en las complicaciones gastrointestinales, el uso de misoprostol parece ejercer un efecto protector a nivel renal (18). La nueva generación de AINEs selectiva hacia la COX-2 como el Meloxicam, promete reducir la toxicidad renal asociada a estos medicamentos (19).

1.2.2.6.3. Reacciones alérgicas

Las reacciones alérgicas debidas a AINEs son más frecuentes de lo que comúnmente se cree. En una revisión de 266 casos de anafilaxia, Kemp y cols. encontraron que el 20% fueron causados por medicamentos, y de este porcentaje más de la mitad correspondió a AINEs, superando inclusive a los (B-lactámicos (20). Es importante el interrogatorio de antecedentes alérgicos antes de iniciar una terapia con AINEs.

1.2.2.6.4. Hematológicos

Todos los AINEs inhiben la agregación plaquetaria aumentando el tiempo de sangrado. El uso de éstos debe efectuarse con precaución en pacientes con alteraciones hematológicas y en el caso de la aspirina ser suspendido una semana antes de un procedimiento quirúrgico. Han ocurrido casos de anemia aplásica con el uso de piroxican, y el uso prolongado de sulindac y ácido mefenámico se ha vinculado con anemia hemolítica (21).

1.2.2.6.5. Inhibición de la motilidad uterina

Se ha demostrado que los AINEs prolongan la gestación. Las PGs de las series E y F son potentes uterotrópicos, y su biosíntesis por parte del útero aumenta en forma extraordinaria horas antes del parto. Por tal razón, se ha planteado la hipótesis de que tal vez intervengan de modo importante en el desencadenamiento y progresión del trabajo de parto, y la expulsión (14).

1.2.2.6.6. Otros efectos adversos

El uso de AINEs puede elevar transitoriamente las enzimas hepáticas; también se han observado casos de hepatotoxicidad asociados a su uso. (21,22). En pacientes sensibles se han observado casos severos de asma desencadenados por el uso de AINEs. En el área dermatológica ha descrito reacciones como exantema y prurito.

1.2.2.7. Métodos tradicionales para la evaluación de la actividad antiinflamatoria.

Los métodos tradicionales de 'prueba y error' para la búsqueda de nuevos compuestos con actividad antiinflamatoria se llevan a cabo a través de dos tipos de técnicas: las técnicas generales, dentro de las cuales se encuentran los modelos de inflamación aguda y los modelos de inflamación subcrónica o crónica, y las técnicas específicas donde se incluye la determinación de la actividad antiinflamatoria mediante la bolsa de aire en ratas, la peritonitis inducida por carragenina, la actividad ciclo-oxigenasa y la permeabilidad capilar incrementada por histamina (23).

1.2.2.7.1. Técnicas generales

1.2.2.7.1.1. Modelos de inflamación aguda.

A Edema plantar por carragenia: una hora después de la administración del extracto o principio puro, del grupo control negativo y del grupo control positivo se administra una solución de carragenina en solución salina fisiológica en la aponeurosis plantar derecha del ratón. El grupo control recibe solamente el vehículo y otro grupo recibe Indometacina como agente antiinflamatorio. El volumen de la pata inyectada es medido antes y después de la inyección de carragenina, en un pletismómetro de agua o por un pie de rey. El porcentaje de inhibición de la reacción inflamatoria de la carragenina, se calcula en la fase aguda, a las 1, 2, 3,5 y 7 horas de la inyección de la misma. Existen reportes donde se emplean otros agentes irritantes, tal es el caso del Dextrano (23,24).

B Edema auricular: se basa en la aplicación de 12 – Tetradecanoil Forbol – 13 Acetato (TPA) en unión con el producto en estudio, en caso de que la solubilidad

lo permita, en la oreja del ratón. Transcurrida 4 h, los animales son sacrificados por dislocación cervical y se le corta una porción de la oreja inflamada e idénticamente otra de la oreja no inflamada. La media aritmética de las

diferencias entre el peso de ambas porciones da la magnitud del edema de cada lote. El fármaco de referencia es la Indometacina (23).

1.2.2.7.1.2. Modelo de inflamación subcrónica o crónica.

A Granuloma inducido por Discos de Algodón: el producto en estudio se administra 7 días consecutivos. Los animales se sacrifican al día siguiente a la última administración. Se extraen los granulomas formados, el timo y las glándulas suprarrenales. Se desecan los granulomas y se determina la diferencia de peso con respecto al disco de algodón implantado al comienzo del experimento (23).

1.2.2.7.2. Técnicas específicas.

1.2.2.7.2.1. Bolsa de aire en ratas.

A Técnica de Edwards: consiste en la formación de una bolsa de aire en el dorso del animal, al cabo de 6 días de la inyección de un agente irritante; recogiendo posteriormente el exudado y realizando con él las pruebas de interés (23).

B Técnica de Sedwick: los animales son anestesiados y rasurados en la zona dorsal y la nuca, procediendo a la inyección de aire en la misma. Al tercer día se inyectan nuevamente con aire. Al cuarto día se les inyecta carragenina y 6 h después se sacrifican procediendo a la extracción del exudado inflamatorio. (23).

1.2.2.7.2.2. Peritonitis inducida por carragenina.

Se comienza con la administración del producto en estudio. Después de 1 h se les inyecta (i.p.) carragenina como agente irritante y 5 h mas tarde se sacrifican los animales. Se les inyectan en la cavidad peritoneal PBS y después de un masaje se recogen los fluidos peritoneales y se procede a la determinación de los leucocitos presentes en una cámara de Neubauer (23).

1.2.2.7.2.3. Actividad ciclo – oxigenasa.**Actividad ciclo – oxigenasa y lipoxigenasa en leucocitos peritoneales de ratas:**

se basa en la estimulación de leucocitos peritoneales de ratas por ionóforo A23187, que determina la entrada de calcio a la célula, la activación de fosfolipasa A2 y la producción y liberación de metabolitos del ácido araquidónico producido por ambas vías (23).

B Actividad ciclo - oxigenasa en vesículas seminales: se emplea como sustrato ácido araquidónico radioactivo para separar después de los metabolitos formados y el sustrato sin reaccionar, por cromatografía en capa fina, y cuantificar los diferentes compuestos por centelleo líquido (23).

1.2.2.7.2.4. Permeabilidad capilar incrementada por histamina.

A Técnica de Ukada: consiste en la administración i.v. de azul de Evans, el cual se une a las proteínas plasmáticas, en particular a la albúmina, y por tanto colorear el área de la piel en la que tiene lugar la extravasación de proteínas debida a la histamina que previamente es inyectada intradérmicamente, lo que provoca un incremento de la permeabilidad capilar a las proteínas plasmáticas. (23).

B Técnica de Williams: permite cuantificar el edema producido por diversos mediadores de la inflamación por los compuestos en estudio, mediante la radioactividad presente en el plasma extravasado (23).

1.2.2.7.2.5. Método Embriones de peces cebra

Un método que en la actualidad es utilizado por cientos de grupos de investigación es el conocido por Embriones de peces cebra. El pez cebra se ha convertido en un modelo inigualable para investigar diferentes procesos biológicos y ahora, sus cualidades genéticas y embrionarias se aprovechan para buscar nuevos medicamentos que permitan controlar diversas enfermedades (25).

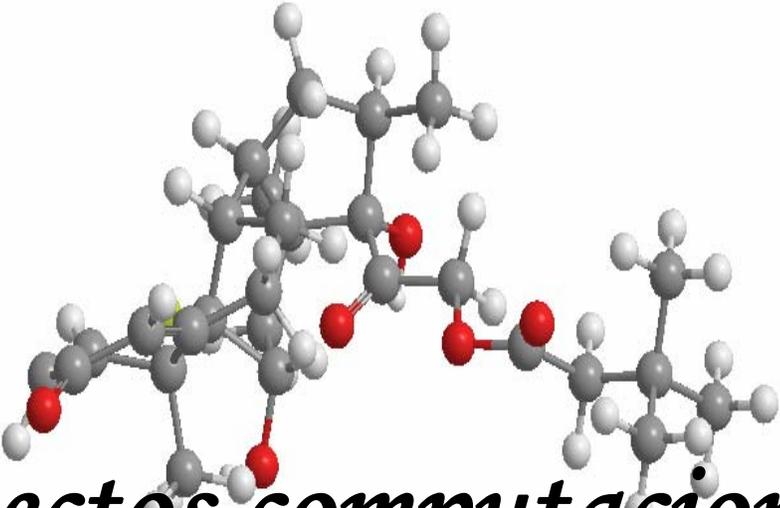
Los embriones de pez cebra facilitan el hallazgo de los genes que dirigen la construcción del cuerpo de los vertebrados desde que es una única célula hasta que se forman los órganos y extremidades (25).

Los embriones de pez cebra se desarrollan externamente y pueden ser observados y manipulados en todas sus etapas, con la ventaja de que su desarrollo es muy veloz (una hembra puede poner 200 huevos a la semana) y su organización es simple. Además, los embriones son transparentes, lo que permite visualizar su formación. Su genoma es aproximadamente la mitad de grande que el del ser humano. Estudios realizados han revelado numerosas similitudes genéticas entre el ser humano y ese pez tropical (26).

El 75% de sus genes son similares a los nuestros, lo que convierte a este animal acuático en un modelo para analizar procesos biológicos de relevancia médica, como los procesos inflamatorios, la creación de nuevos vasos sanguíneos, el crecimiento de las neuronas, la formación de hueso y músculo, o la resistencia a la insulina (25).

1.2.2.8. Usos terapéuticos. Tratamiento del dolor

- Neuralgias
- Cefaleas (cefaleas de tensión). Son útiles en la jaqueca (migraña) cuando la frecuencia del episodio es menor de 2 por mes y de moderada intensidad.
- Dolores de diverso tipo y origen (dolores radicales, dolores dentarios, por infecciones: otitis, sinusitis, etc.)
- Dolores postoperatorios o postparto (pequeña moderada intensidad).
- Dismenorreas. Las prostaglandinas contribuyen a la patogenia.
- Dolor por cáncer: Fundamentalmente en los primeros estadios. Se utilizan salicilatos u otros AINEs, posteriormente se emplean en asociación con opiáceos menores y por último opiáceos mayores (27).



A *spectos computacionales
y quimiométricos*

Capítulo II

2.3. Método computacional y análisis quimiométrico de los datos.

2.3.1. Generalidades.

Los estudios QSPR/QSAR constituyen un enfoque que permite entender como la variación estructural afecta la propiedad/actividad biológica de un conjunto de compuestos. En estos estudios, los *descriptores moleculares* (X) (*los descriptores moleculares son 'términos que caracterizan un aspecto específico de una molécula' y contienen información derivada de la representación estructural de las moléculas bajo estudio* (28,29) se correlacionan con una variable respuesta (Y). Es decir, este análisis puede definirse como una aplicación de métodos matemáticos y estadísticos al problema de encontrar una ecuación empírica de la forma $Y_i = f_i(X_1, X_2, \dots, X_n)$, donde Y_i son las propiedades y/o actividades biológicas de la molécula, y X_1, X_2, \dots, X_n son propiedades estructurales experimentales o calculadas (*descriptores moleculares*) de los compuestos. En este sentido, cada compuesto puede representarse como un punto en un espacio multidimensional, en los cuales los descriptores X_1, X_2, \dots, X_n son coordenadas independientes del compuesto. El objetivo más usual de este análisis es incrementar el entendimiento del sistema biológico bajo investigación o predecir la propiedad estudiada a un objeto (compuesto) no utilizado en la obtención del modelo.

En la actualidad, existe un gran número de *descriptores moleculares* (DMs) que pueden ser usados en estudios QSAR (30). La naturaleza de los DMs, depende de cual haya sido el proceder utilizado para la definición de los mismos, pudiendo tener en cuenta rasgos topológicos (que aquí llamaremos bidimensionales o 2D), geométricos (3D), y electrónicos de las moléculas. En este trabajo se utilizaron para la parametrización de la estructura molecular los índices lineales no estocásticos y estocásticos basados en las relaciones entre núcleos atómicos,

En una parte de este acápite se describirán de forma resumida los aspectos fundamentales de las técnicas quimiométricas empleadas para el desarrollo de los modelos QSAR. El término *quimiometría*, surgió en la década del 70 y se define como la disciplina química que utiliza métodos estadísticos y

matemáticos para seleccionar y optimizar los métodos analíticos y preparativos, así como procedimientos para el análisis e interpretación de los datos.

Antes de comenzar a describir el método computacional utilizado en el presente trabajo, daremos una breve introducción a las generalidades de la metodología QSAR, los que serán utilizados en el desarrollo de esta tesis y que son imprescindibles para la comprensión de los resultados tanto de la literatura como los alcanzados en el presente trabajo.

Los principios de la metodología QSAR pueden describirse mediante los siguientes pasos comunes (31).

- 1) Formulación del problema, se determina el objeto de análisis y el nivel de información requerido.
- 2) Parametrización cuantitativa de la estructura molecular de los compuestos químicos orgánicos/secuencia de biopolímeros.
- 3) Medición de la propiedad de interés ('efectos biológicos').
- 4) Escoger el tipo de modelo QSAR 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 tamizajes virtuales. Sin embargo, el desarrollo de cualquier QSAR es un ciclo interactivo.

Utilizando este procedimiento se han desarrollado un gran número de investigaciones en todo el mundo. No obstante, las aplicaciones de los DMS han estado dirigidas fundamentalmente hacia la predicción cuantitativa de propiedades físico-químicas y biológicas de compuestos orgánicos, en estudios que se han denominado QSPR y QSAR, respectivamente. Esta división, no es solo formal, porque aunque el método en ambos tipos de estudio es similar, por lo general, la actividad biológica es una propiedad mucho más compleja que las propiedades físico-químicas, debido a la gran cantidad de factores que influyen en la bio-actividad de un compuesto químico.

La aplicación de los DMs al diseño y selección de nuevas entidades químicas es probablemente una de las áreas más activas de investigación en la aplicación de tales DMs a problemas biológicos. Uno de los primeros ejemplos del diseño de nuevos compuestos en el uso de índices moleculares, fue descrito por la Upjohn & Pharmacy en 1993. Ellos fueron capaces de diseñar una nueva clase de compuestos de la familia de las heteropiperazinas con actividad contra la HIV-retrotransferasa (32). Más recientemente, Grassy y col. fueron capaces de diseñar y sintetizar un péptido que mostró una actividad inmunosupresora aproximadamente 100 veces mayor que los compuestos líderes ensayados (33).

2.3.2. Método computacional.

En esta sección describiremos de forma resumida los aspectos fundamentales de los DMs 2D (bi-dimensional), implementados en el programa **TOMOCOMD-CARDD**, que fueron utilizados en esta investigación. Una discusión más extensa sobre el formalismo matemático de estos DMs puede ser revisada en la literatura internacional que se refiere (34, 35,36).

Este método codifica la estructura molecular a través de aplicaciones matemáticas denominadas formas lineales. Con el propósito de calcular esta función algebraicas basadas en relaciones de núcleos atómicos, el *vector molecular basado en átomos*, \bar{x} (representación vectorial de la estructura molecular) y las *k-ésimas matrices de densidad electrónica grafo-teóricas no estocásticas y estocásticas*, \mathbf{M}^k y \mathbf{S}^k , respectivamente (representación matricial de la estructura molecular) deben ser calculadas(37,38). Tales relaciones de adyacencia entre núcleos atómicos (compartimiento de electrones de la capa de valencia) y codificación de la información química son utilizados en la obtención de la familia de DMs utilizados en el presente trabajo tal y como describiremos a continuación.

2.3.2.1. Descriptores TOMOCOMD-CARDD Basados en Relaciones entre Núcleos Atómicos.

2.3.2.1.1. Información Química y Vector Molecular basado en Átomos.

El vector molecular basado en átomo (\bar{x}), usado para representar las moléculas químicas, ha sido explicado en detalle en diferentes publicaciones internacionales (39). Los componentes (x) de \bar{x} son valores numéricos de una propiedad que caracteriza cada tipo de átomo (núcleo atómico) presente en la molécula. Es decir, estos valores (pesos o etiquetas de átomo) corresponden a diferentes propiedades atómicas. Por tanto, una molécula constituida por 5, 10, 15, ..., n núcleos atómicos puede ser representada por medio de vectores con 5, 10, 15, ..., n componente, perteneciendo a los espacios vectoriales \mathfrak{R}^5 , \mathfrak{R}^{10} , \mathfrak{R}^{15} , ..., \mathfrak{R}^n , respectivamente; donde n es la dimensión del conjunto de los reales. O sea, \bar{x} es un vector n -dimensional.

Este enfoque permite codificar moléculas orgánicas tales como el 3-mercapto-piridina-4-carbaldehído a través del vector molecular $\bar{x} = [x_{N1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}, x_{C6}, x_{C7}, x_{O8}, x_{S9}]$. Este vector pertenece a \mathfrak{R}^9

Diversos tipos de “pesos atómicos” (x) pueden ser utilizados para codificar información relacionada con cada núcleo atómico en la molécula. Estas etiquetas de átomo son números con un significado o interpretación química, tales como el Log P atómico (40), la contribución a la superficie polar de los átomos, (41), la refractividad molar atómica (42), las polarizabilidades atómica (P) (43,44,45), las cargas atómicas de Gasteiger-Marsilli (46), las masas atómicas (M) (47) los volúmenes de van der Waals (V), la electronegatividad en la escala de Pauling (E) (48), entre otras.

2.3.2.1.2. “Background” sobre las Matrices de Densidad Electrónica Grafo-Teóricas No Estocásticas y Estocásticas.

En topología molecular, la estructura química es expresada, generalmente, por un grafo molecular con hidrógenos suprimidos. Informalmente, un grafo molecular \mathbf{G} esta constituido por una colección de vértices (puntos) y otra de aristas (líneas o enlaces) conectando estos vértices (49,50). En términos matemáticos, un grafo simple \mathbf{G} es definido como $\mathbf{G} = (V, E)$ (51,52). El número de vértices en un grafo es designado como n y el número de aristas por m .

Otra de las formas de definir un grafo es la siguiente: consideremos un conjunto no vacío $V = \{v_i / i = 1, 2, \dots, n\}$, un conjunto no vacío $E = \{e_i / i = 1, 2, \dots, m\}$ y una aplicación θ , la cual asocia a cada elemento de E con un par no ordenado de elementos de V . Esta aplicación es denominada la aplicación de incidencia asociada con un grafo y le da singularidad al grafo para un conjunto V dado. Por tanto, para cada e_i existe un par no ordenado $[v_i, v_j]$ tal que $\theta(e) = [v_i, v_j]$. En este caso los dos conjuntos E y V junto con la aplicación θ forman un grafo.

Si en G hay pares repetidos (aristas múltiples, o sea vértices que están unidos por más de una arista), entonces el grafo G se llama *grafo con aristas múltiples* o *multigrafo*. Las aristas de la forma $\{v_i, v_j\}$, se denominan *lazos* o *bucles*. Si en G hay lazos (pueden también existir aristas múltiples), entonces el grafo G se llama *grafo con lazos* o *pseudografo* (ver Figura 2.1) (53).

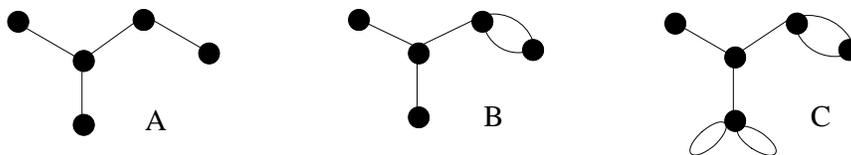


Figura 2.1. A) Grafo simple; B) Multigrafo; y C) Pseudografo.

A continuación relacionaremos una serie de definiciones sobre los términos más utilizados en los grafos moleculares, los cuales son muy útiles para describir varias características estructurales de estos: (54,55).

Si $a_k = \{v_i, v_j\}$ es una arista, entonces los vértices v_i, v_j , se llaman extremos de la arista a_k .

Si los vértices v_i y v_j son los puntos finales de e_i se denota como $e_i \sim [v_i, v_j]$, lo que se lee como ' e_i es *incidente* con v_i y v_j ' (enlace que los une).

Los vértices v_i, v_j se llaman *adyacentes* si existe una arista a_k tal que $a_k = \{v_i, v_j\} \in V$, (o sea, si existe una arista que los une). Dos aristas se denominan *adyacentes* si ellas tienen un vértice en común.

En un multigrafo, el *grado del vértice* v_i [$\delta(v_i)$] es el número de aristas del multigrafo que son incidentes al vértice v_i . En un pseudografo, el grado del vértice v_i es igual al número total de aristas (que no sean lazos) incidentes a este vértice, más el número de lazos incidentes a él.

Un camino (P) es una sucesión de aristas con vértices comunes. La *longitud* (l) de un camino es el número de aristas del mismo. Así por ejemplo, *camino de longitud cero* (P^0) es una sucesión de vértices que contiene solo un vértice.

Sea $G = (V, E)$ un pseudografo. El pseudografo $G' = (V', E')$ es un subgrafo de G si $V' \subseteq V$ y $E' \subseteq E$, o sea, si todos sus vértices y aristas también lo son en G . Los grafos II y III de la Figura 3.2 son subgrafos del grafo I, ya que todos los vértices y aristas de II y III están contenidos en I. El subgrafo $G - v_i$ se obtiene eliminando del grafo G el vértice v_i y todas sus aristas incidentes. Así por ejemplo, el grafo II es un subgrafo de I, obtenido al eliminar en I el vértice v_8 junto con sus aristas incidentes, e_{7-8} y e_{8-3} . El subgrafo $G - e_{ij}$ puede obtenerse eliminando del grafo G la arista e_{ij} . El grafo III es un subgrafo de I, obtenido al eliminar en I la arista e_{8-3} (ver Figura 1.3).

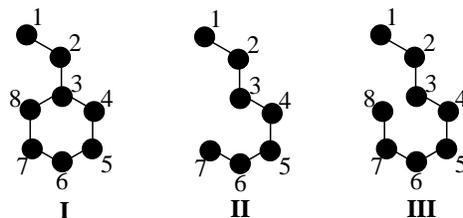


Figura 2.2. Representación esquemática de subgrafos

Recientemente, Marrero-Ponce y col. han propuesto nuevas matrices moleculares que describen los cambios en el tiempo de la distribución electrónica a través del esqueleto molecular (55,56). En este sentido, en este trabajo serán usados los índices lineales no-estocásticos y estocásticos. Algunos aspectos relacionados con estos descriptores moleculares serán tratados en este capítulo.

2.3.2.1.3. Definición de los Índices Lineales basados en Relaciones de Átomos.

Definición de los Índices Lineales Totales y locales para átomos, grupos y fragmentos moleculares: Los k -ésimas índices lineales locales (atómicos) para un átomo i en una molécula, son calculados como una aplicación lineal sobre \mathfrak{R}^n [$f: \mathfrak{R}^n \rightarrow \mathfrak{R}^n$; entonces f : Endomorfismo sobre \mathfrak{R}^n] en las bases canónicas (57,58). Específicamente, los k -ésimos índices lineales atómicos no

estocásticos y estocásticos, $f_k(\bar{x}_i)$ y ${}^s f_k(\bar{x}_i)$, son calculados a partir de las matrices \mathbf{M}^k y \mathbf{S}^k , como se muestra a continuación:

$$f_k(\bar{x}_i) = \sum_{j=1}^n m_{ij}^k x^j = [\mathbf{X}']^k = \mathbf{M}^k[\mathbf{X}] \quad (2.1)$$

$${}^s f_k(\bar{x}_i) = \sum_{j=1}^n s_{ij}^k x^j = [\mathbf{X}\mathbf{S}']^k = \mathbf{S}^k[\mathbf{X}] \quad (2.2)$$

donde n es el número de átomos de la moléculas y x^j son las coordenadas del vector molecular (\bar{x}) en el sistema de bases canónicas de \mathfrak{R}^n . Los coeficientes ${}^k m_{ij}$ y ${}^k s_{ij}$ son los elementos de \mathbf{M}^k y \mathbf{S}^k , respectivamente. O sea, que estas matrices denotan las matrices de $f_k(\bar{x}_i)$ y ${}^s f_k(\bar{x}_i)$, con respecto a la bases canónicas.

Nótese, que los índice lineales atómicos son definidos como una transformación lineal $f_k(\bar{x}_i)$ sobre un espacio vectorial molecular, \mathfrak{R}^n . Esta aplicación, es una correspondencia que asigna a cualquier vector X en \mathfrak{R}^n un vector $f(x)$ de forma tal que:

$$f(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 f(X_1) + \lambda_2 f(X_2) \quad (2.3)$$

para todo λ_1, λ_2 número reales y cualquier vector X_1, X_2 en \mathfrak{R}^n . En otras palabras, $f_k(\bar{x}_i)$ es una aplicación lineal dado que la imagen de la combinación lineal de dos vectores X_1 y X_2 , $\lambda_1 X_1 + \lambda_2 X_2$; es igual a la combinación lineal de las imágenes $f(X_1)$ y $f(X_2)$, $\lambda_1 f(X_1) + \lambda_2 f(X_2)$. Esta condición se denomina *condición de linealidad*. Las ecuaciones de definición para estos índices también puede ser escrita como una simple ecuación matricial (ver ecuaciones 2.1 y 2.2), donde $[\mathbf{X}]$ es el vector columna (una matriz de $n \times 1$) de coordenadas de \bar{x} en la base canónica de \mathfrak{R}^n , $[\mathbf{X}]^t$ es la matriz transpuesta de $[\mathbf{X}]$ (una matriz de $1 \times n$).

Este enfoque es similar al método **LCAO-MO** (siglas acrónimas de "Linear Combinations of Atomic Orbitals-Molecular Orbitals"). Realmente nuestro enfoque (para $k = 1$) es una aproximación muy similar al método de Hückel extendido, dado que nuestra matriz considera tanto electrones σ como π . La idea fundamental del método **LCAO-MO** es que los electrones en la molécula

están acomodados en orbitales moleculares definidos justamente como los átomos los acomodan en orbitales atómicos. Es decir, los orbitales moleculares son formados por combinaciones lineales de átomos que componen el sistema, lo cual puede ser escrito de la siguiente forma:

$$\psi_i = \sum_{j=1}^n c_{ij} \varphi_j \quad (2.4)$$

donde i es el número de MO ψ [en nuestro caso, $f_k(\bar{x}_i)$ o ${}^s f_k(\bar{x}_i)$]; j es el número de orbitales atómicos, φ -orbitals (en nuestro caso, x^j); c_{ij} (en nuestro caso, ${}^1 m_{ij}$ o ${}^1 s_{ij}$) son los coeficientes numéricos que definen la contribución individual de los AOs en un MO dado.

Los índices lineales totales constituyen funciones lineales (Algunos matemáticos usan el término *formas lineales*) sobre \Re^n (59). Es decir, los índices lineales totales constituyen aplicaciones lineales de \Re^n a escalares \Re [f: $\Re^n \rightarrow \Re$]. La definición matemática de este descriptor molecular es la siguiente (60,61).

$$f_k(\bar{x}) = \sum_{i=1}^n f_k(\bar{x}_i) = [\mathbf{u}]^t [\mathbf{X}']^k = [\mathbf{u}]^t \mathbf{M}^k [\mathbf{X}] \quad (2.5)$$

$${}^s f_k(\bar{x}) = \sum_{i=1}^n {}^s f_k(\bar{x}_i) = [\mathbf{u}]^t [\mathbf{X}\mathbf{S}']^k = [\mathbf{u}]^t \mathbf{S}^k [\mathbf{X}] \quad (2.6)$$

donde n es el número de átomos y $f_k(\bar{x}_i)$ y ${}^s f_k(\bar{x}_i)$ son los índices lineales atómicos no estocásticos y estocásticos obtenidos por las ecuaciones 2.1 y 2.2, respectivamente. En estas ecuaciones también se muestran las definiciones matriciales de estos índices totales, donde $[\mathbf{u}]^t$ es un vector fila (matriz fila) unitario de dimensión n .

Como puede observarse, los k^{th} índices lineales totales son calculados sumando todos los índices locales (átomos) de todos los átomos en la molécula. Además, si una molécula es particionada en Z fragmentos moleculares, los índices lineales totales no estocásticos [estocásticos] pueden ser particionados en Z índices lineales locales no estocásticos [estocásticos] $f_k(\bar{x}_i)$ [${}^s f_k(\bar{x}_i)$], $L = 1, \dots, Z$. Es decir, los índices lineales totales de orden k

pueden ser expresados como la suma de los índices locales de los Z fragmentos moleculares:

$$f_k(\bar{x}) = \sum_{L=1}^Z f_{kL}(\bar{x}) \quad (2.7)$$

$${}^s f_k(\bar{x}) = \sum_{L=1}^Z {}^s f_{kL}(\bar{x}) \quad (2.8)$$

Cada tipo de índice lineal local puede ser clasificado según el tipo de átomo que compone el fragmento. Así por ejemplo, se pueden calcular sobre heteroátomos, H-unido a heteroátomos (O, N y S), halógenos, cadenas alifáticas o aromáticas, entre otros.

2.3.3. Análisis quimiométrico de los datos.

2.3.3.1. Análisis Discriminante Lineal (ADL) para la predicción de la actividad.

La calidad y tipo de datos biológicos es un factor importante para seleccionar el método estadístico apropiado para desarrollar los modelos QSAR. Un gran número de pruebas biológicas producen resultados discretos, por ejemplo, activo o inactivo, o ++, +, 0, --, -. El ADL es una *técnica de clasificación y asignación de un individuo (compuesto) a un grupo (activo o inactivo) conocidas sus características* (30,62). En esta se dispone de una serie de grupos definidos *a priori*, con una serie de observaciones para cada individuo referidas a un conjunto de variables relevantes. En base a esta información se llega a calcular una *función discriminante* (FD). La FD es una ecuación lineal con una variable dependiente que representa la pertenencia a un grupo. *Combinaciones lineales* de variables independientes, sirven de base para clasificar a los individuos entre los grupos.

2.3.3.2. Estimación de los Coeficientes.

La información que contienen todas las variables independientes se analiza conjuntamente para obtener los coeficientes. Se trata de conseguir un promedio ponderado de las variables independientes para obtener una puntuación que permita distinguir entre grupos. En realidad, dados dos grupos

de compuestos, uno activo y otro inactivo dos funciones de clasificación, D_1 y D_2 son obtenidas (31,63).

$$D_1 = b_1X_1 + b_2X_2 + \dots + b_3X_3 + \dots \quad (2.9)$$

$$D_2 = a_1X_1 + a_2X_2 + \dots + a_3X_3 + \dots \quad (2.10)$$

Los coeficientes, a_i y b_i , son los llamados *pesos discriminantes* y se obtienen por el procedimiento de *regresión múltiple*. Esta función describe una línea, un plano o en general, una superficie (hiperplano) entre los grupos. La diferencia de las dos *funciones de clasificación* de cada uno de los grupos es la llamada FD, D_{12} .

$$D_{12} = D_1 - D_2 \quad (2.11)$$

1.3.3.3. Matriz de Clasificación.

La tabla que muestra los resultados de la *clasificación* con la FD se denomina *matriz de clasificación* o *de confusión*. Esta tabla provee el *porcentaje de casos bien clasificados* para cada grupo y de forma total. La evaluación de la función para todos los individuos que sirven de base para el análisis es también un criterio a considerar en la validación de la FD. Usualmente se exige que el porcentaje de casos bien clasificados no sea inferior a un 75 %, para que el criterio de clasificación sea considerado como aceptable (29,63).

2.3.3.4. Significación de la Función Discriminante (FD).

La hipótesis nula en el ADL puede formularse así: *no existen diferencias significativas entre las medias de las puntuaciones discriminantes de los grupos*. Una de las pruebas para comprobar la hipótesis estadística anterior se basa en la lambda (λ) de Wilks. En el caso de solo dos grupos, la λ de Wilks es (64).

$$\lambda = \frac{SC_{\text{int ragrupos}}}{SC_{\text{total}}} \quad (2.12)$$

donde SC es la *variabilidad*, por ejemplo, $SC_{\text{intragrupos}}$ es la *variabilidad intragrupos*.

Este valor representa la porción de la varianza total de las puntuaciones discriminantes que no ha sido explicada por la diferencia entre grupos. La λ de

Wilks toma valores entre 0 y 1. Mientras menor es la λ de Wilks, mayor es la diferencia entre las medias de las puntuaciones discriminantes de los grupos y esto indica el rechazo de la hipótesis nula. Es importante recalcar que aunque una λ sea significativa no puede interpretarse como una indicación de la eficacia de la FD. Lo único que prueba es que existen diferencias entre las medias. Pero diferencias pequeñas entre los grupos pueden ser estadísticamente significativas y en cambio no permitir una buena *discriminación* entre grupos. Evidentemente, si la lambda no es significativa la *discriminación* no será posible (65). La existencia de los grupos de poblacionales se comprueba a través de diferentes ensayos entre los que se encuentra el *Cuadrado de la Distancia de Mahalanobis*, D^2 , el cual mide la distancia al cuadrado entre los centroides de dos poblaciones. Sean p poblaciones de n_1, n_2, \dots, n_p , individuos cada una. En cada población se conocen v variables, x_1, x_2, \dots, x_v . A cada población le corresponde una matriz de observaciones. Se dispone por tanto de p matrices de $n \times v$. A partir de estos datos, y en notación matricial, Mahalanobis define la distancia entre los centroides de los grupos p y q por:

$$D_{pq}^2 = (\mu_p - \mu_q)' \Sigma^{-1} (\mu_p - \mu_q) \quad (2.13)$$

siendo μ_p y μ_q los vectores columna que contienen las medias de las variables de los grupos respectivos. Σ^{-1} es la inversa de la matriz de varianza-covarianzas intragrupos de los dos grupos conjuntamente. La prima (') indica la matriz transpuesta. A partir de la D^2 se puede estimar la F de Fischer y utilizarla como prueba de contraste:

$$F = D^2 \frac{n_p n_q (n_p + n_q - v - 1)}{(n_p + n_q)(n_p + n_q - 2)v} \quad (2.14)$$

2.3.3.5. Criterios de Selección de Variables.

Existen diversos criterios de selección de variables, los principales son los siguientes:

- a) Todas las variables se entran simultáneamente siempre que *satisfagan el criterio de tolerancia*.
- b) Se selecciona la variable que *minimice la lambda de Wilks*.

- c) Se selecciona la variable que *maximice la D^2 de Mahalanobis entre los grupos más próximos.*
- d) Se selecciona la variable que *maximice la menor F entre pares de grupos.*
- e) Se selecciona la variable que *minimice la suma de la variación no explicada entre grupos.*

2.3.3.5.1. Selección del Número Óptimo de Predictores. Principio de la Parsimonia.

La exactitud de un modelo de clasificación aumenta en la medida en que se añaden variables a la ecuación; pero a partir de cierto punto el incremento de esta para cada nueva variable que se añade, es insignificante. Un buen modelo no debe presentar ni demasiadas variables, ni debe olvidar las que sean verdaderamente relevantes. Es decir, debe cumplir el principio de la *parsimonia*, según el cual un fenómeno debe ser descrito con el número mínimo de elementos posibles.

Diversos procedimientos se han propuesto para seleccionar el número óptimo de variables a incluir en la ecuación, como por ejemplo la 'forward selection', 'backward elimination; y 'stepwise selection' (66). Este último método es el más utilizado (es una combinación de los dos anteriores) y sigue un proceso de selección de variables paso a paso.

2.3.3.5.2. Tolerancia.

La *tolerancia* es una medida del grado de asociación lineal entre las variables independientes. Para la variable i , la *tolerancia* es igual a $1 - R_i^2$, donde R_i^2 es la correlación múltiple al cuadrado entre la variable i considerada como variable dependiente y las demás variables independientes. Valores bajos en la *tolerancia*, indican que la variable i puede ser considerada como una combinación lineal de las otras variables independientes. Por tanto, la *tolerancia* de una variable, en un paso cualquiera del análisis 'stepwise', es la proporción de su varianza intra-grupo no explicada por otras variables del análisis.

2.3.3.6. Multicolinealidad entre variables, redundancia de la Información y correlaciones casuales.

El término 'multicolinealidad' se utiliza para describir la situación en que un gran número de descriptores moleculares están altamente intercorrelacionados. Las variables que se aproximan a ser una *combinación lineal* de las otras, se denominan *multicolineales* o *colineales* (67). Una 'multicolinealidad' alta, produce errores estándares altos en los coeficientes de regresión y dificulta estimar la importancia relativa de los descriptores en el modelo, lo cual afecta la interpretación de las actividades modeladas en términos estructurales. La importancia relativa puede determinarse al valorar el incremento en la R, cuando se añade una variable a la ecuación que ya contiene las demás variables (R_i^2). El método más utilizado para detectar la existencia de variables *colineales* es obtener una matriz de correlaciones entre los descriptores moleculares. Uno de los métodos más utilizados para detectar la interdependencia entre variables, es la *tolerancia*. Problemas con la *redundancia de la información* y la *colinealidad*, han sido ilustrados con el uso de DMs, tales como los índices de conectividad molecular (67,68). El nivel aceptable de *colinealidad* es algo subjetivo y en ese sentido se ha reportado que coeficientes de correlación entre las variables aceptables están en el rango de 0.4-0.9 (69).

2.3.3.7. Validación Estadística de los Modelos QSAR.

La significación estadística de la FD obtenida con el ADL debe ser probada analizando la λ de Wilks y la D^2 de Malahanobis, aunque según Kier, la calidad de la FD puede evaluarse de tres formas diferentes (70).

- 1) Comparación del valor de F con el valor tabulado.
- 2) Determinación de casos bien clasificados en la serie de entrenamiento (SE).
- 3) Validación externa.

Además, los métodos de *validación cruzada* también pueden aplicarse a este tipo de modelos. Ogino y col. han propuesto otro enfoque para seleccionar la mejor FD, la cual se selecciona teniendo en cuenta el análisis de la combinación de dos criterios (71)].

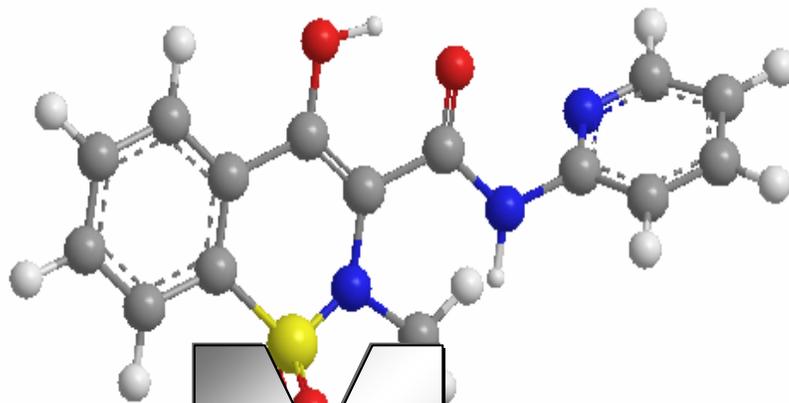
- 1) Una combinación de variables que minimice el número de compuestos mal clasificados.
- 2) El empleo del número menor de variables.
- 3) La colinealidad entre las variables independientes es minimizada.

Cuatro herramientas pueden ser utilizadas para acceder a la validación de los modelos QSAR obtenidos por RLM y la mayoría de estas pueden también extrapolarse a la validación de los modelos obtenidos con el ADL (61).

- 1) Aleatorización de la variable respuesta (Y- Randomización).
- 2) Validaciones cruzadas.
- 3) División de la data de compuestos en serie de entrenamiento (SE) y en serie de predicción (SP).
- 4) Confirmación del poder predictivo utilizando SP 'externas'.

2.3.3.7.1. División de la Data Original en Serie de Entrenamiento y Serie de Predicción.

Usualmente el procedimiento de VC es denominado *validación interna*, porque todos los compuestos que considera pertenecen a los mismos datos originales. Sin embargo, cuando el número de compuestos es grande, estos pueden dividirse en dos conjuntos separados de *entrenamiento o calibración* y otro conjunto de *validación o predicción (validación externa)*. Muchos investigadores consideran a los altos valores de q^2 ($q^2 > 0.5$) como un indicador del poder predictivo de un modelo QSAR. En contraste con estas especulaciones, varios investigadores han demostrado que la "única" condición necesaria y suficiente para poder estimar el verdadero poder predictivo de un modelo es comparar los valores predichos y observados de una extensa (suficientemente larga) SP externa (72,73).



Materiales y *Métodos*

Capítulo III

Capítulo III: Materiales y métodos.**3.1. Base de datos de los compuestos usados en la obtención de los Modelos ADL-QSAR.**

La aplicabilidad y la representatividad del presente método dependen de forma crítica de la selección de los compuestos que son utilizados como serie de entrenamiento para construir el modelo de clasificación. El aspecto más crítico en la construcción de la serie de entrenamiento es garantizar la gran diversidad molecular de la data. En orden de asegurar esta diversidad molecular nosotros hemos seleccionado una data compuesta por una gran cantidad de entidades moleculares, algunos reportados como antiinflamatorios y el resto conformada por compuestos con otros usos farmacológicos. La data de compuestos activos fue seleccionada considerando toda la representatividad de los diferentes núcleos estructurales (cabeza de series o compuestos líderes) y diferentes mecanismos de acción antiinflamatorios, un ejemplo de ello se muestra en la Figura 3.1. Las estructuras de todos los compuestos antiinflamatorios utilizados en este trabajo están en la Tabla 1 de los Anexos.

El conjunto de compuestos inactivos fue construido de la siguiente forma. Se seleccionó aleatoriamente una base de datos de fármacos con diferentes usos farmacológicos. Estos fármacos incluyen antivirales, hipnótico y sedante, anticonvulsivante y antiepiléptico, Diuréticos, radioprotector, Hipoglucemiante, antihipertensivo, anestésicos, anticoagulantes, antimaláricos, antihelmínticos, antidepresivos, antiprotozoos, opioides, antihistaminicos, antipsicóticos, antifúngicos, antibacteriales, etc. Esta claro que la declaración de estos compuestos como 'inactivos' (sin actividad antiinflamatoria) no garantiza que alguno de estos compuestos presente alguna actividad antiinflamatoria aún no detectada. Este problema puede verse reflejado en los resultados de la clasificación de la serie de compuestos inactivos. Sin embargo, alguno de estos compuestos puede ser detectado (clasificado) por la función de clasificación como antiinflamatorios. En este sentido, estas moléculas serían 'erróneamente'

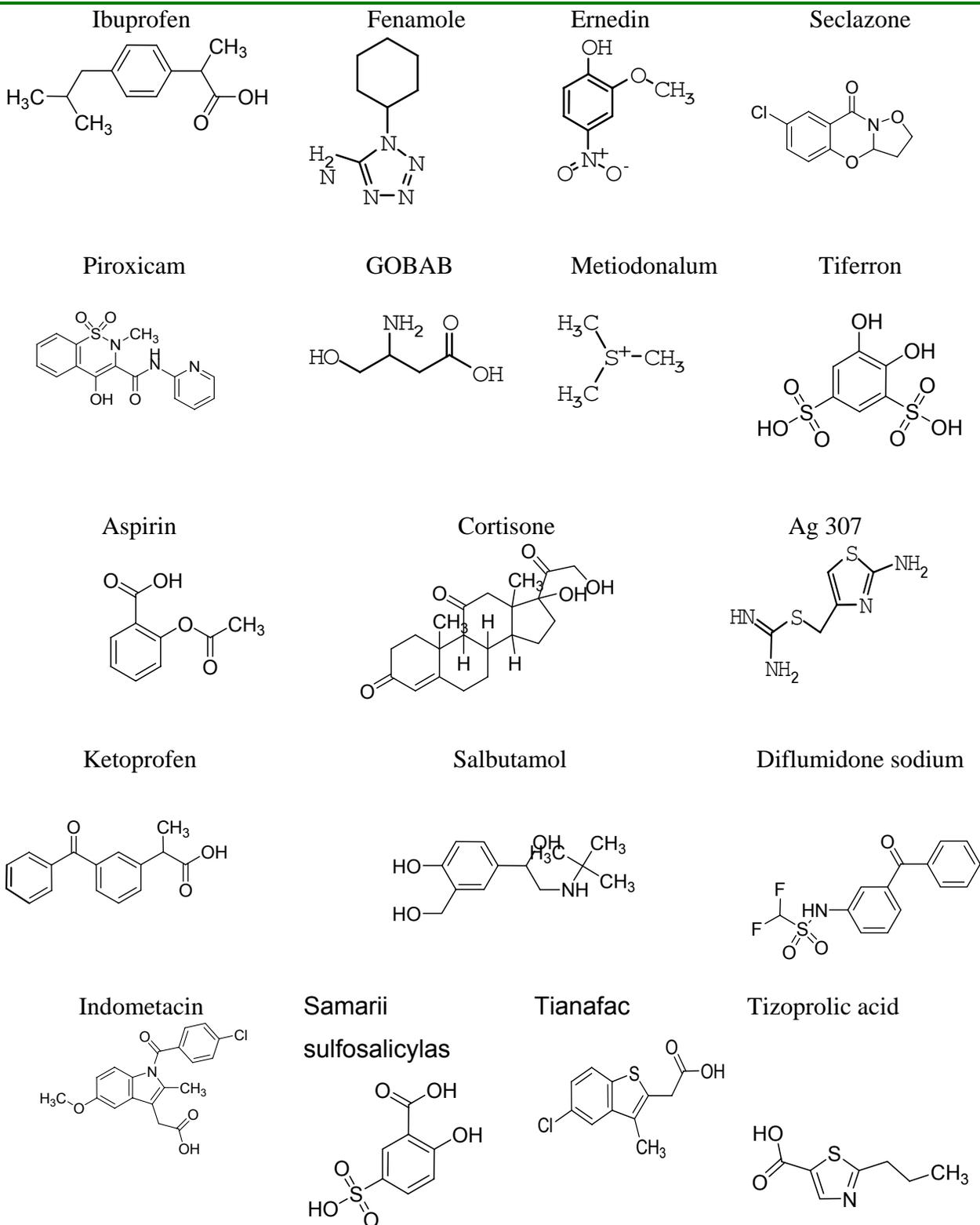


Figura 3.1. Ejemplo de la diversidad estructural de los compuestos con actividad antiinflamatoria utilizados en este estudio.

clasificadas como inactivas en un principio y pudieran ser escogidas para ser evaluadas experimentalmente.

El conjunto de la data de compuestos (activos e inactivos) fue dividido aleatoriamente en dos subconjuntos. Uno de estos conjuntos es usado como serie de entrenamiento para el desarrollo de la función de clasificación y el otro es usado como serie de predicción externa. Los compuestos incluidos en la serie de predicción externa nunca fueron usados en el desarrollo de los modelos cuantitativos. Todos los compuestos utilizados en este estudio fueron tomados del (74) y de la base de datos 'Merck Index' (75) donde los nombres, sinónimos y fórmulas estructurales para cada uno de los compuestos pueden ser encontrados.

3.2. Método Computacional. *TOMOCOMD-CARDD* software.

TOMOCOMD (76) es un programa interactivo para el diseño molecular e investigaciones bioinformáticas. El programa está compuesto por cuatro módulos o subprogramas. Cada uno de estos módulos consta de una interfaz gráfica que facilita al investigador la representación de las moléculas (drawing mode) y el cálculo de varias familias de descriptores moleculares (calculation mode). Los módulos han sido denominados con las siguientes siglas acrónimas: CARDD (Computed-Aided 'Rational' Drug Design), CAMPS (Computed-Aided Modeling in Protein Science), CANAR (Computed-Aided Nucleic Acid Research) y CABPD (Computed-Aided Bio-Polymers Docking). En este trabajo, nosotros usaremos el subprograma CARDD. Este módulo fue desarrollado basado en una filosofía amigable para el usuario, el cual no tiene que tener *a priori* ningún conocimiento de programación.

Los principales pasos para desarrollar un estudio QSPR/QSAR utilizando el enfoque topológico molecular ***TOMOCOMD-CARDD***, son resumidos brevemente a continuación:

1. Representar el pseudografo molecular de cada una de las moléculas de la base de datos a analizar, usando el módulo de dibujo del software. Este

procedimiento es llevado a cabo seleccionando el átomo deseado perteneciente a diferentes grupos de la tabla periódica en el momento de representar las moléculas,

2. Usar un 'peso' apropiado (etiqueta) de átomo, en orden de diferenciar cada tipo de átomo en la molécula. En este estudio hemos utilizado un esquema de ponderaciones constituido por 5 propiedades atómicas, las cuales son: masa atómica (M), volumen de van der Waals (V), polarizabilidad (P), electronegatividad en la escala de Mulliken (K) y electronegatividad en la escala de Pauling (G).

3. Computar los índices lineales totales y locales (tanto estocásticos como no-estocásticos). Este proceder es llevado a cabo en el módulo de cálculo del programa, el cual permite seleccionar la propiedad atómica y la familia que se desea calcular. Este programa genera una tabla en la cual las filas corresponden a los compuestos (casos) y las columnas a los índices moleculares calculados (variables).

4. Encontrar una o varias ecuaciones QSPR/QSAR usando técnicas estadísticas o de inteligencia artificial adecuada, tales como Regresión Lineal Múltiple (RLM), Análisis Discriminante Lineal (ADL), Redes Neuronales (RN), entre otras. Es decir, nosotros encontramos una relación cuantitativa entre una propiedad **P** y la estructura química codificada con los descriptores calculados. Así por ejemplo, la ecuación obtenida con los índices lineales totales tomaría la siguiente apariencia:

$$P = a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + \dots + a_k f_k(x) + c \quad (3.1)$$

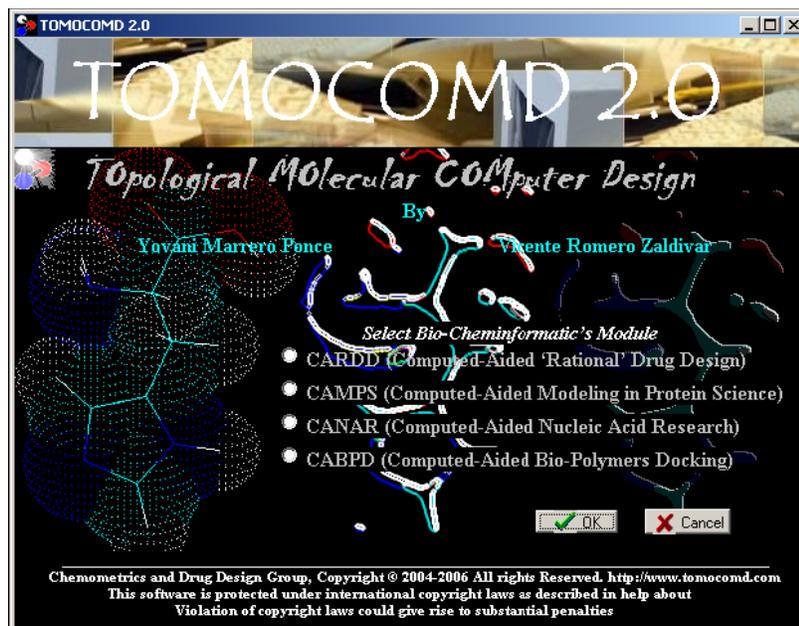
donde **P** es la medida de la propiedad (o actividad), $f_k(x)$ [o $f_{kL}(x)$] es el k^{th} índice lineal total [o local], y los términos a_k 's son los coeficientes obtenidos por el análisis estadístico multivariado.

5. Probar la robustez y demostrar el poder predictivo de las ecuaciones QSPR/QSAR obtenidas usando procedimientos de validación interna y externa.

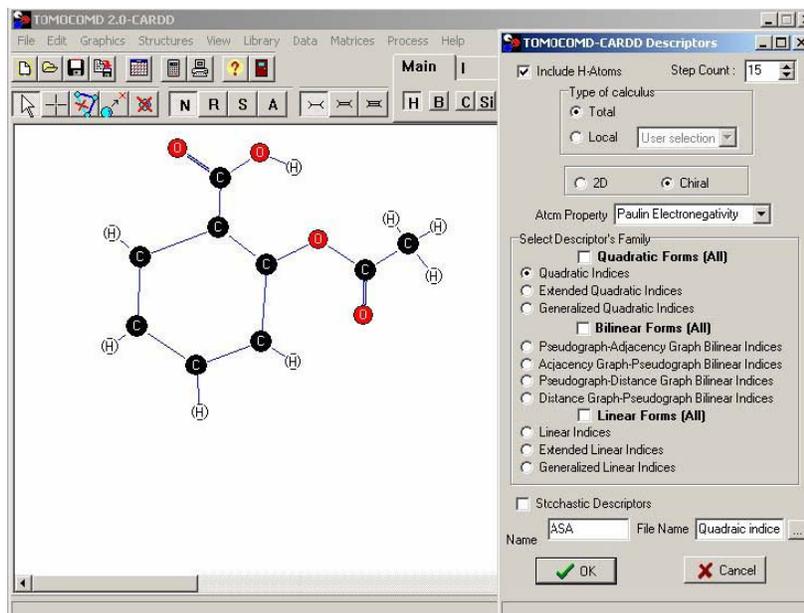
6. Desarrollar una interpretación estructural de los modelos QSAR/QSPR obtenidos, que permita interpretar la propiedad **P** estudiada.

Figura 3.2. TOMOCOMD-CARDD Software: A, Ventana para seleccionar el modulo de trabajo. B, Interfaz gráfica del subprograma de diseño de fármacos.

A



B



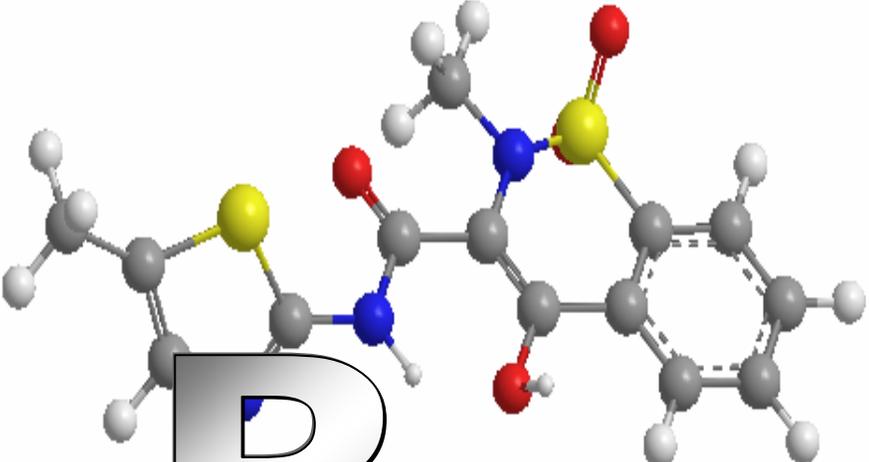
3.3. Desarrollo de las Funciones Discriminantes (FDs) usando ADL.

A pesar de que existen varias *técnicas quimiométricas* para generar FDs, tales como “soft independent modeling of class analogy” (SIMCA) o redes de neuronas (RNs) artificiales, el ADL fue seleccionado para obtener las *funciones de clasificaciones* sobre la base de la simplicidad del método (30). Los modelos QSAR-ADL se obtuvieron con el paquete de programas estadísticos STATISTICA (77). El método de selección de variables utilizado fue el de “pasos hacia delante (‘forward stepwise’). En todos los casos el estadístico F y la *tolerancia* se usaron para el control del proceder de selección. En este sentido, la *colinealidad* entre variables fue examinada utilizando las *matrices de correlaciones* entre las variables incluidas en el modelo. Siempre se utilizó, por defecto, como mínimo aceptable de *tolerancia* el valor de 0.01.

Con el objetivo de ensayar la *calidad y robustez* de los modelos obtenidos, siempre evaluamos varios parámetros estadísticos, tales como la λ de Wilks, el valor de F y el cuadrado de la distancia de Mahalanobis (D^2). Otro factor que se tuvo en consideración para evaluar la habilidad (poder discriminante) de las FDs obtenidas, fueron los *porcentajes de buena clasificación* en cada uno de los grupos y del modelo en general en la serie de entrenamiento. La clasificación de los casos se desarrolló por medio de la *probabilidad de clasificación posterior*. Usando la D^2 para la clasificación, se pueden obtener probabilidades. La probabilidad de que un caso pertenezca a un grupo particular es básicamente ‘proporcional’ a la D^2 del caso al centroide del grupo. En resumen, la *probabilidad posterior* es la probabilidad, basada en el conocimiento de los valores de otras variables, de que el respectivo caso pertenece a un grupo en particular. Además, se utilizaron varios parámetros estadísticos para evaluar la calidad de los modelos obtenidos (coeficiente de correlación de Matthews, sensibilidad, especificidad, etc), todos calculados a partir de la matriz de confusión (78).

Los resultados de las probabilidades de clasificación para cada compuesto se muestran utilizando el siguiente parámetro: $\Delta P\% = [P(\text{actv}) - P(\text{inactv})] \times 100$, donde $P(\text{actv})$ es la probabilidad con que la ecuación clasifica un compuesto como activo. Contrariamente, $P(\text{inactv})$ es la probabilidad con que el modelo clasifica un compuesto como inactivo. Cuando $\Delta P\%$ está en el rango $-5 < \Delta P\% < 5$ los compuestos fueron considerados como *no clasificados*.

Para acceder al *poder predictivo* de los modelos QSAR-ADL obtenidos, se desarrollaron procedimientos de *validación externa*. Finalmente fueron calculados en la serie de entrenamiento y predicción; el porcentaje de buena clasificación global (exactitud), especificidad, sensibilidad, la razón de falsos positivos (RFP) y el coeficiente de correlación de Mathews (MCC). Estos parámetros permiten evaluar la calidad de los modelos obtenidos. (79).



R *esultados y*
Discusión

Capítulo IV

Capítulo IV: Resultados y discusión.**4.1. Diseño de la base de datos.**

Las series activas e inactivas empleadas para obtener las funciones de clasificación son divididas en series de entrenamiento y predicción. En este sentido, primeramente debe ser probada la diversidad estructural de la base de datos, para ello se realizó un Análisis de Conglomerados Jerárquicos (CAs) a las series activas e inactivas de compuestos (80,81). Para desarrollar estos análisis de conglomerados fue utilizado el paquete estadístico STATISTICA (77) Los dendrogramas fueron obtenidos usando la distancia euclídeana (abcisa X) y al vincula completa (abcisa Y) como se muestra en las Figura 4.1 y 4.2, para la serie de compuestos activos e inactivos, respectivamente.

En ambos árboles binarios puede observarse un gran número de patrones estructurales diferentes, lo que prueba la variabilidad estructural de los compuestos seleccionados para conformar la base de datos.

Mediante este procedimiento es posible escoger de forma racional las series de entrenamiento y predicción aunque debido a la dificultad de evaluar los dendrogramas obtenidos, es necesario realizar otro tipo de análisis de cluster. Se realizaron dos análisis de cluster de partición (no-jerárquicos) llamados también *k*-MCA (análisis de cluster de *k*-medias) para dividir la base de datos en dos series: entrenamiento y predicción. La idea principal de este procedimiento consiste en hacer una partición de las series activas e inactivas en diferentes subconjuntos de compuestos estadísticamente representativos. Este procedimiento asegura que cualquiera de estos subconjuntos (determinado por los conglomerados derivados del *k*-MCA) estará representado en ambas series (entrenamiento y predicción).

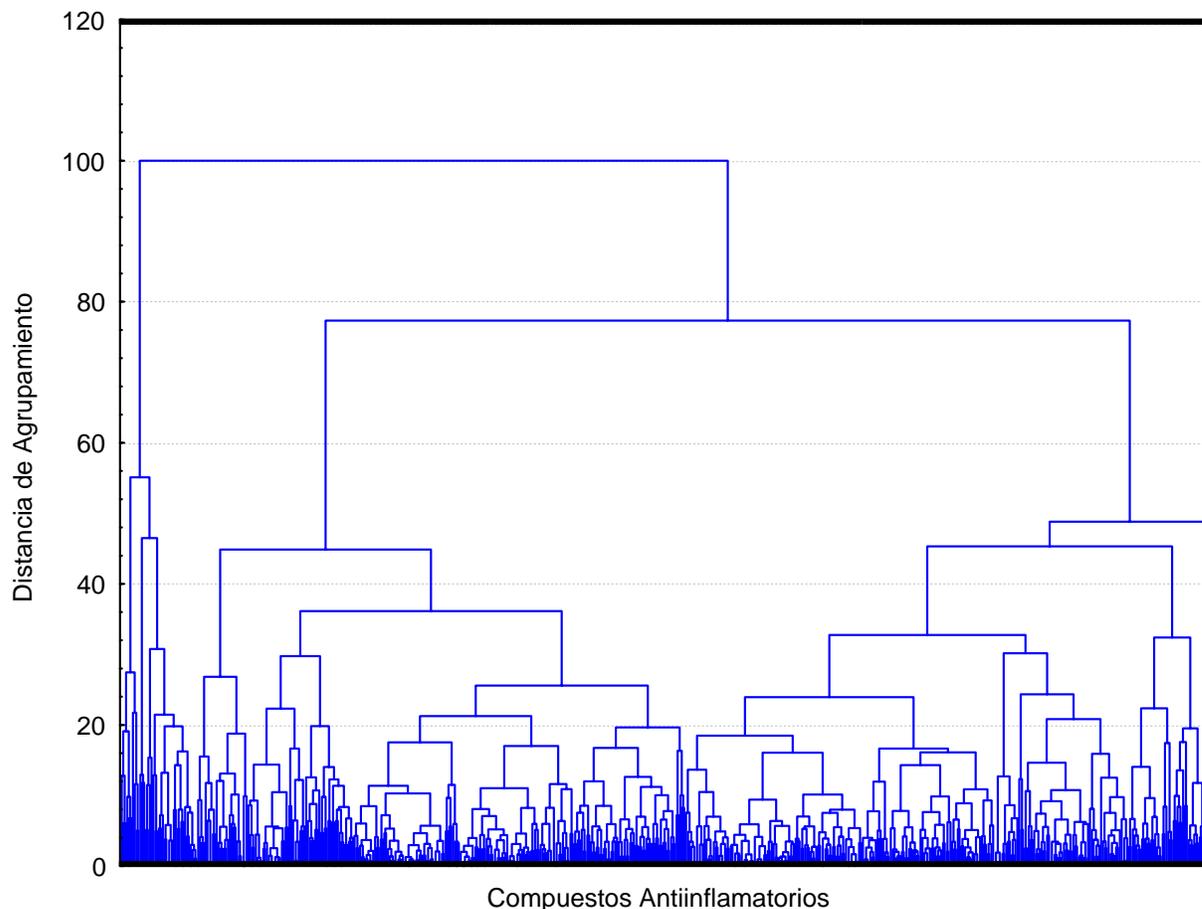


Figura 4.1. Dendrograma que muestra los resultados del k -NNCA jerárquico del conjunto de compuestos antiinflamatorios usados en la serie de entrenamiento y predicción del presente trabajo.

En este sentido, se realiza un primer k -MCA para la base de datos de compuestos activos, y luego otro k -MCA para el subconjunto de compuestos inactivos. En este primer k -MCA (k -MCA I) la data de compuestos antiinflamatorios fue dividida en 9 conglomerados. Además se efectúa un segundo k -MCA (k -MCA II) para la base de datos de inactivos, dividiendo la data en 10 conglomerados. Para estos análisis de cluster se emplean los índices lineales no-estocásticos y todas las variables empleadas muestran valores de significación de $p < 0.005$ para la prueba de Fisher. Estos resultados son mostrados en la Tabla 4.1.

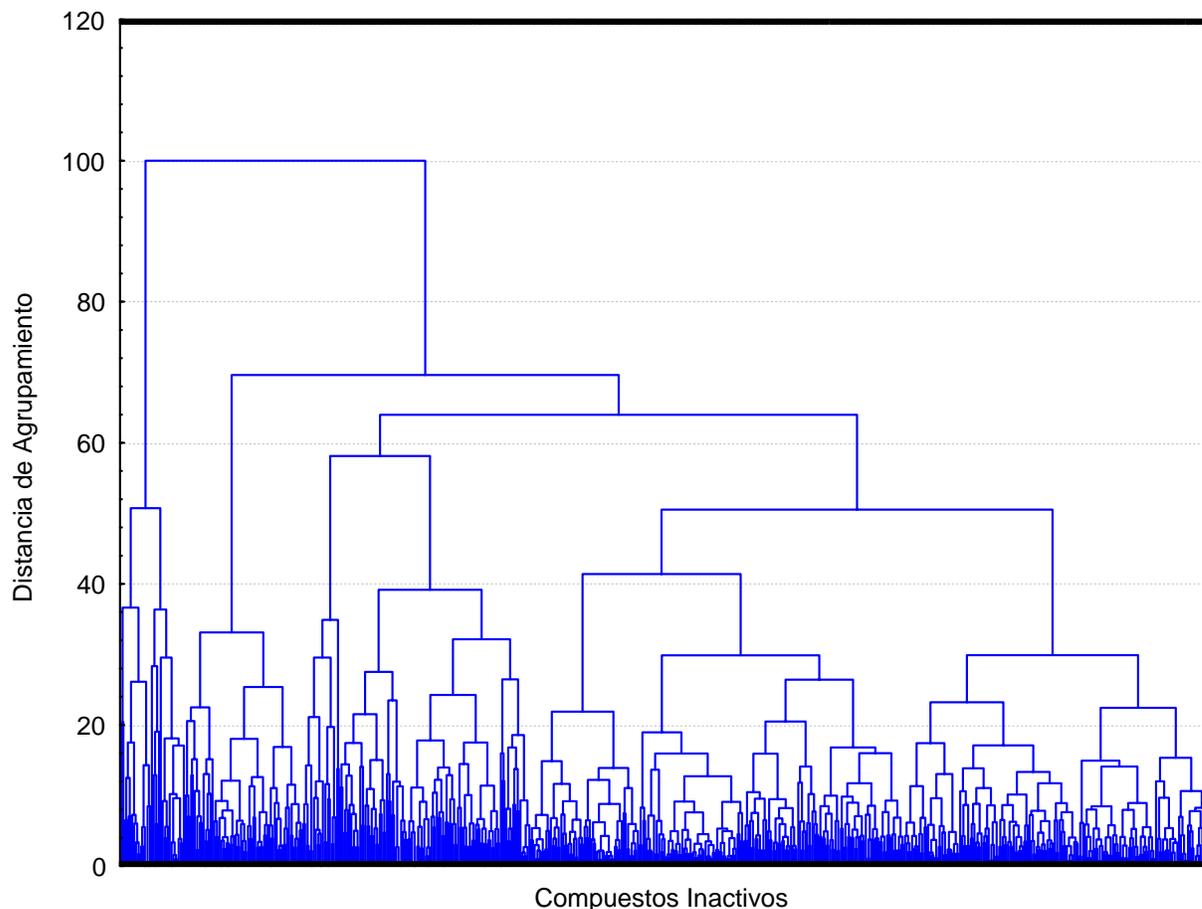


Figura 4.2. Dendrograma que muestra los resultados del k -NNCA jerárquico para el conjunto de compuestos inactivos usados en la serie de entrenamiento y predicción del presente trabajo.

A continuación se realiza la selección de las series de entrenamiento y predicción tomando compuestos de forma aleatoria de cada uno de los conglomerados formados. Como fue señalado anteriormente, estos 1213 compuestos fueron divididos en una serie de entrenamiento con 919 compuestos escogidos aleatoriamente, siendo 443 de ellos activos y 476 inactivos. El subconjunto restante conformado por 294 compuestos, con 144 antiinflamatorios y 150 con otros usos farmacológicos, se usa como serie de predicción para la validación externa de los modelos de clasificación.

Tabla 4.1. Principales resultados de los k -MCAs, para los compuestos antiinflamatorios y los compuestos inactivos.

Análisis de Varianza				
Variabes	Between SS ^a	Within SS ^b	Fisher(F)	p-level ^c
Análisis de Conglomerados de los Compuestos Antiinflamatorios (k - MCA I)				
$Vf_3(\bar{x})$	240.52	38.39	401.65	0.00
$Vf_{1L}^H(\bar{x} \mathbf{E})$	213.35	59.42	230.19	0.00
$Vf_7^H(\bar{x})$	209.32	48.90	274.45	0.00
$Gf_{2L}(\bar{x} \mathbf{E})$	174.14	71.72	155.67	0.00
$Mf_{0L}^H(\bar{x} \mathbf{E})$	180.77	85.47	135.59	0.00
$Pf_{1L}^H(\bar{x} \mathbf{E})$	197.43	47.40	267.05	0.00
$Kf_0(\bar{x})$	311.21	72.67	274.56	0.00
$Pf_4^H(\bar{x})$	234.51	40.56	370.71	0.00
$Mf_{1L}^H(\bar{x} \mathbf{E})$	207.44	53.82	247.11	0.00
$Gf_1(\bar{x})$	279.88	39.16	458.22	0.00
Análisis de Conglomerados de los Compuestos Inactivos (k – MCA II)				
$Vf_3(\bar{x})$	253.38	72.80	263.22	0.00
$Vf_{1L}^H(\bar{x} \mathbf{E})$	388.93	62.22	472.76	0.00
$Vf_7^H(\bar{x})$	260.13	86.26	228.05	0.00
$Gf_{2L}(\bar{x} \mathbf{E})$	458.04	72.83	475.62	0.00
$Mf_{0L}^H(\bar{x} \mathbf{E})$	425.55	136.08	236.51	0.00
$Pf_{1L}^H(\bar{x} \mathbf{E})$	405.26	62.62	489.45	0.00
$Kf_0(\bar{x})$	192.71	66.92	217.77	0.00
$Pf_4^H(\bar{x})$	281.11	69.04	307.91	0.00
$Mf_{1L}^H(\bar{x} \mathbf{E})$	527.46	85.78	465.04	0.00
$Gf_1(\bar{x})$	235.24	53.72	331.16	0.00

^aVariabilidad entre los grupos.^bVariabilidad dentro de los grupos.^cNivel de significación.

Estos compuestos de la serie de predicción no fueron usados para la construcción de los modelos de clasificación. En la figura 4.3 se muestra el procedimiento descrito para la selección de los subconjuntos representativos en las series de entrenamiento y predicción con el uso de los análisis de cluster correspondientes.

4.2. Desarrollo y Validación de las Funciones Discriminantes.

Con el objetivo de identificar compuestos con actividad antiinflamatoria fueron obtenidas funciones discriminantes sobre la base de una familia de DMs

implementados en el programa *TOMOCOMD-CARDD*. En este sentido, las estructuras moleculares de cada compuesto en la base de datos fueron

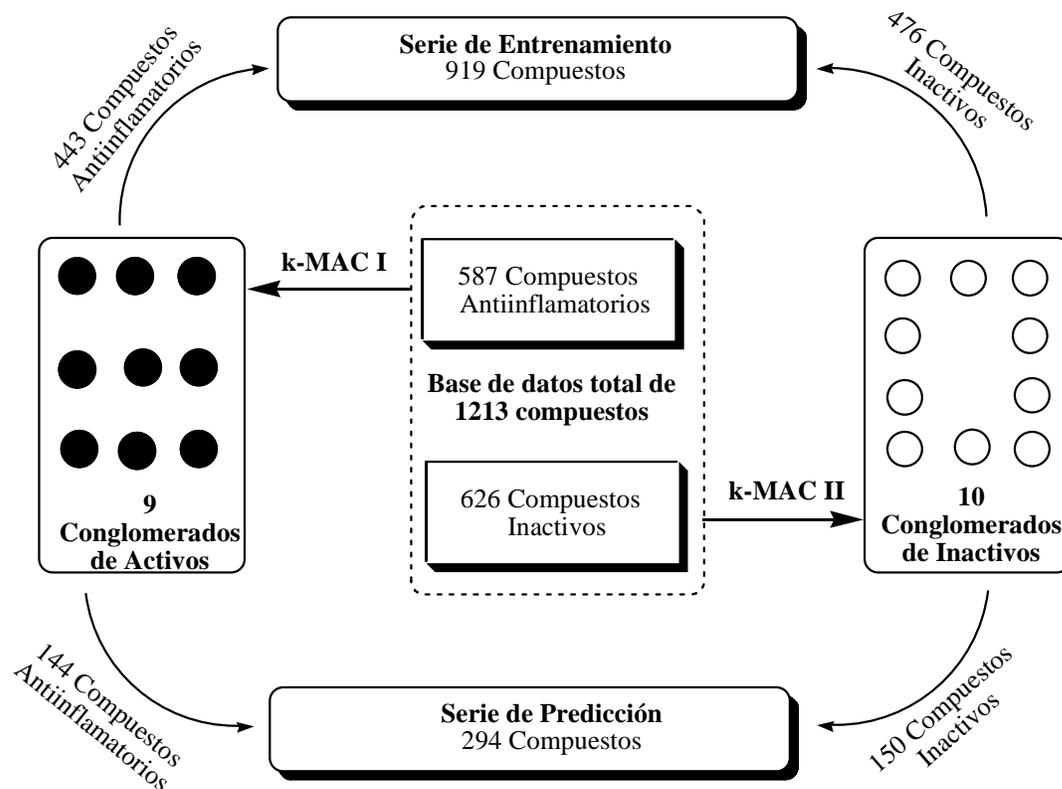


Figure 4.3. Algoritmo para el diseño de las series de entrenamiento y predicción.

parametrizadas haciendo uso de los índices lineales no-estocásticos y estocásticos basadas en relaciones entre núcleos atómicos. Dichos descriptores fueron calculados para todos los compuestos incluidos tanto en la base de datos [Serie de entrenamiento (SE) y serie de predicción (SP)] extraída de la literatura. Una vez calculados los índices lineales no-estocásticos y estocásticos se procedió a la obtención de los modelos mediante análisis discriminante lineal (ADL). Esta técnica permite encontrar una función discriminante con la habilidad de distinguir entre dos grupos de poblaciones (30). Para obtener la función discriminante que permitió la clasificación de compuestos como posibles activos (antiinflamatorios), fueron empleados como variables independientes los k -ésimos ($k = 15$) índices lineales totales y locales estocásticos y no estocásticos basados en relaciones de átomos no considerando y considerando átomos de

hidrógenos (H) en el pseudo-grafo molecular (G), [$f_k(\bar{x})$, $f_k^H(\bar{x})$, $f_{kL}(\bar{x}_E)$, $f_{kL}^H(\bar{x}_E)$ y $f_{kL}^H(\bar{x}_{E-H})$, respectivamente]. Estos índices fueron calculados con el programa **TOMOCOMD-CARDD** (76), y los modelos obtenidos se pueden observar en la Tabla 4.2.

Tabla 4.2. Modelos Discriminantes obtenidos usando los índices lineales totales y locales (no estocástico y estocástico).

Modelos QSAR-ADL obtenidos con los Índices Lineales no estocásticos

$$\begin{aligned} \text{Class} = & -2.960 - 8.662 \times 10^{-4} M f_4^H(\bar{x}) - 2.326 \times 10^{-2} M f_{3L}^H(\bar{x}_E) + 2.206 \times 10^{-2} M f_{3L}(\bar{x}_E) \\ & + 5.156 \times 10^{-2} M f_{2L}^H(\bar{x}_{E-H}) - 1.684 \times 10^{-4} M f_{7L}^H(\bar{x}_{E-H}) + 3.780 \times 10^{-3} M f_3(\bar{x}) + 1.123 \times 10^{-2} M f_1^H(\bar{x}) \\ & - 2.365 \times 10^{-2} M f_{0L}^H(\bar{x}_E) + 1.372 \times 10^{-7} M f_{12L}^H(\bar{x}_{E-H}) \end{aligned} \quad (4.1)$$

$$\begin{aligned} \text{Class} = & -2.881 + 2.356 \times 10^{-3} V f_3(\bar{x}) - 6.424 \times 10^{-3} V f_{3L}^H(\bar{x}_E) + 6.292 \times 10^{-3} V f_{3L}(\bar{x}_E) - 1.006 \times 10^{-5} V f_7^H(\bar{x}) \\ & + 4.579 \times 10^{-11} V f_{15}^H(\bar{x}) - 3.802 \times 10^{-2} V f_{0L}^H(\bar{x}_E) + 3.255 \times 10^{-2} V f_{0L}^H(\bar{x}_{E-H}) \end{aligned} \quad (4.2)$$

$$\begin{aligned} \text{Class} = & -2.843 + 7.474 \times 10^{-2} P f_2(\bar{x}) - 0.462 P f_{2L}^H(\bar{x}_E) + 0.425 P f_{2L}(\bar{x}_E) - 0.102 P f_{1L}(\bar{x}_E) \\ & - 3.497 \times 10^{-3} P f_4(\bar{x}) - 0.365 P f_{0L}(\bar{x}_E) + 1.612 P f_{0L}^H(\bar{x}_{E-H}) - 0.132 P f_{3L}^H(\bar{x}_{E-H}) \\ & + 2.625 \times 10^{-8} P f_{14L}^H(\bar{x}_{E-H}) + 3.830 \times 10^{-2} P f_0^H(\bar{x}) \end{aligned} \quad (4.3)$$

$$\begin{aligned} \text{Class} = & -3.401 + 6.997 \times 10^{-3} K f_3(\bar{x}) - 7.499 \times 10^{-2} K f_{1L}^H(\bar{x}_E) + 3.641 \times 10^{-6} K f_8^H(\bar{x}) + 0.101 K f_{2L}(\bar{x}_E) \\ & - 9.426 \times 10^{-2} K f_{2L}^H(\bar{x}_E) + 1.679 K f_{1L}^H(\bar{x}_{E-H}) - 2.088 K f_{0L}^H(\bar{x}_{E-H}) + 3.202 \times 10^{-7} K f_{12L}^H(\bar{x}_{E-H}) \end{aligned} \quad (4.4)$$

$$\begin{aligned} \text{Class} = & -3.433 + 7.523 \times 10^{-3} G f_3(\bar{x}) + 0.793 G f_{1L}^H(\bar{x}_E) - 1.310 \times 10^{-5} G f_7^H(\bar{x}) + 0.120 G f_{2L}(\bar{x}_E) \\ & - 0.105 G f_{2L}^H(\bar{x}_E) - 3.115 \times 10^{-4} G f_{5L}(\bar{x}_E) - 0.879 G f_{1L}(\bar{x}_E) - 0.266 G f_{2L}^H(\bar{x}_{E-H}) \\ & + 1.476 \times 10^{-8} G f_{15L}^H(\bar{x}_{E-H}) + 9.337 \times 10^{-7} G f_{12L}^H(\bar{x}_{E-H}) \end{aligned} \quad (4.5)$$

$$\begin{aligned} \text{Class} = & -2.955 + 2.430 \times 10^{-3} V f_3(\bar{x}) - 9.958 \times 10^{-3} V f_{3L}^H(\bar{x}_E) + 4.939 \times 10^{-3} V f_{3L}(\bar{x}_E) \\ & - 5.935 \times 10^{-6} M f_8^H(\bar{x}) + 3.249 \times 10^{-3} P f_{5L}^H(\bar{x}_E) + 2.091 \times 10^{-9} P f_{15}(\bar{x}) + 6.052 \times 10^{-3} G f_{4L}(\bar{x}_E) \\ & - 9.036 \times 10^{-3} K f_{3L}(\bar{x}_E) \end{aligned} \quad (4.6)$$

Tabla 4.2. Continuación

Modelos QSAR-ADL obtenidos con los Índices Lineales estocásticos

$$\begin{aligned} \text{Class} = & -3.543 - 9.025 \times 10^{-2} \text{Ms} f_0^H(\bar{x}) - 0.101 \text{Ms} f_{0L}^H(\bar{x}_E) - 0.196 \text{Ms} f_{15}^H(\bar{x}) + 0.137 \text{Ms} f_7(\bar{x}) \\ & - 0.114 \text{Ms} f_2^H(\bar{x}) + 0.173 \text{Ms} f_{13}^H(\bar{x}) + 9.044 \times 10^{-2} \text{Ms} f_{2L}^H(\bar{x}_E) - 8.142 \times 10^{-2} \text{Ms} f_{1L}^H(\bar{x}_E) \\ & - 7.664 \times 10^{-2} \text{Ms} f_{13L}^H(\bar{x}_E) + 5.715 \times 10^{-2} \text{Ms} f_{1L}(\bar{x}_E) \end{aligned} \quad (4.7)$$

$$\begin{aligned} \text{Class} = & -2.665 - 5.581 \times 10^{-2} \text{Vs} f_{14}(\bar{x}) + 2.432 \times 10^{-2} \text{Vs} f_0^H(\bar{x}) - 7.012 \times 10^{-2} \text{Vs} f_{0L}^H(\bar{x}_E) \\ & + 9.637 \times 10^{-2} \text{Vs} f_1(\bar{x}) - 6.526 \times 10^{-2} \text{Vs} f_2^H(\bar{x}) - 6.508 \times 10^{-2} \text{Vs} f_{13}(\bar{x}) - 7.853 \times 10^{-2} \text{Vs} f_{1L}(\bar{x}_E) \\ & + 0.216 \text{Vs} f_{14L}(\bar{x}_E) - 0.154 \text{Vs} f_{14L}^H(\bar{x}_E) + 0.11 \text{Vs} f_3(\bar{x}) \end{aligned} \quad (4.8)$$

$$\begin{aligned} \text{Class} = & -3.048 - 0.230 \text{Ps} f_{10}(\bar{x}) - 0.447 \text{Ps} f_L^H(\bar{x}_E) - 1.796 \text{Ps} f_{15}^H(\bar{x}) + 1.100 \text{Ps} f_1(\bar{x}) + 1.854 \text{Ps} f_{11}^H(\bar{x}) \\ & - 0.521 \text{Ps} f_{1L}^H(\bar{x}_{E-H}) - 1.202 \text{Ps} f_{1L}^H(\bar{x}_E) + 0.969 \text{Ps} f_{14L}(\bar{x}_E) - 0.461 \text{Ps} f_2^H(\bar{x}) \end{aligned} \quad (4.9)$$

$$\begin{aligned} \text{Class} = & -3.241 - 0.906 \text{Ks} f_0(\bar{x}) - 0.910 \text{Ks} f_{14L}(\bar{x}_E) - 1.912 \text{Ks} f_2^H(\bar{x}) + 1.845 \text{Ks} f_0^H(\bar{x}) + 1.174 \text{Ks} f_2(\bar{x}) \\ & + 0.622 \text{Ks} f_{2L}^H(\bar{x}_E) \end{aligned} \quad (4.10)$$

$$\begin{aligned} \text{Class} = & -3.212 - 1.013 \text{Gs} f_0(\bar{x}) - 0.355 \text{Gs} f_{14L}(\bar{x}_E) - 2.055 \text{Gs} f_2^H(\bar{x}) + 1.999 \text{Gs} f_8^H(\bar{x}) + 1.349 \text{Gs} f_2(\bar{x}) \end{aligned} \quad (4.11)$$

$$\begin{aligned} \text{Class} = & -2.797 - 8.387 \times 10^{-2} \text{Vs} f_{15}(\bar{x}) - 4.214 \times 10^{-2} \text{Ms} f_2^H(\bar{x}) + 8.159 \times 10^{-2} \text{Vs} f_1(\bar{x}) - 0.156 \text{Vs} f_{1L}^H(\bar{x}_{E-H}) \\ & + 0.655 \text{Gs} f_{3L}^H(\bar{x}_{E-H}) - 5.698 \times 10^{-2} \text{Ms} f_1^H(\bar{x}) + 0.269 \text{Ps} f_3^H(\bar{x}) - 6.667 \times 10^{-2} \text{Vs} f_3(\bar{x}) \end{aligned} \quad (4.12)$$

Modelos QSAR-ADL obtenidos con los Índices Lineales

$$\begin{aligned} \text{Class} = & -2.773 + 2.634 \times 10^{-2} \text{P} f_2(\bar{x}) - 0.582 \text{P} f_{2L}^H(\bar{x}_E) + 0.531 \text{P} f_{2L}(\bar{x}_E) + 5.951 \times 10^{-2} \text{Vs} f_{1L}(\bar{x}_E) \\ & - 3.623 \times 10^{-2} \text{Ms} f_{2L}^H(\bar{x}_E) - 9.326 \times 10^{-3} \text{Ms} f_0^H(\bar{x}) - 2.302 \times 10^{-2} \text{P} f_{3L}^H(\bar{x}_{E-H}) \end{aligned} \quad (4.13)$$

En la Tabla 4.3 se resumen los porcentajes de buena clasificación obtenidos en el ajuste de cada modelo utilizando la misma serie de aprendizaje. Esta Tabla también muestra los parámetros comúnmente usados en estadística médica

[Sensibilidad, especificidad, falsos positivos y coeficiente de correlación de Matthews (**C**)] (78).

En este contexto la sensibilidad no es más que la probabilidad de predecir un ejemplo positivo correctamente, mientras que la especificidad es la probabilidad de que una predicción positiva es correcta. A su vez, **C** cuantifica la fuerza de la relación lineal entre los descriptores moleculares y las clasificaciones, y puede proporcionar a menudo una evaluación mucho más equilibrada de la predicción que, por ejemplo, los porcentajes globales de clasificación correcta. Además se muestra otros estadísticos utilizados a la hora de determinar la calidad de los modelos, λ es el estadístico de Wilks, D^2 es el cuadrado de la distancia de Mahalanobis y F es el estadístico de Fischer. En todos los casos el valor del nivel de significación fue adecuado, $p < 0,05$.

Tabla. 4.3. Parámetros estadísticos de los modelos QSAR-ADL en la serie de entrenamiento.

Modelos	Coeficiente de correlación de Matthews (C)	Exactitud 'Q _{Total} ' (%)	Especificidad d (%)	Sensibilidad (%)	Razón falso: positivos (%)	Wilks ' λ	F	D ²
Modelos QSAR\ADL utilizando los índices lineales no estocásticos								
[Eq.4.1 (9)]	0.74	86.83	85.10	87.27	11.5	0.46	119.9	4.74
[Eq.4.2 (7)]	0.75	87.38	86.68	87.07	11.9	0.47	147.4	4.52
[Eq.4.3 (10)]	0.76	87.81	86.00	88.40	10.5	0.44	113.7	5.00
[Eq.4.4 (8)]	0.75	87.49	86.23	87.61	11.3	0.44	144.4	5.07
[Eq.4.5 (10)]	0.74	86.94	86.00	86.79	12.1	0.45	111.8	4.92
[Eq.4.6(8)]	0.77	88.57	86.68	89.30	9.66	0.44	145.7	5.12
Modelos QSAR\ADL utilizando los índices lineales estocásticos								
[Eq.4.7 (10)]	0.75	87.38	85.33	88.11	10.7	0.45	110.4	4.86
[Eq.4.8 (10)]	0.74	87.16	85.10	87.88	10.9	0.44	113.8	5.01
[Eq.4.9 (9)]	0.72	86.18	85.33	85.91	13.0	0.45	125.2	4.95
[Eq.4.10 (6)]	0.73	86.62	85.33	86.70	12.2	0.47	172.6	4.54
[Eq.4.11 (5)]	0.74	87.05	85.33	87.50	11.3	0.46	211.5	4.62
[Eq.4.12 (8)]	0.77	88.47	86.68	89.10	9.87	0.44	146.0	5.13
Modelo general QSAR\ADL utilizando los índices lineales no estocásticos y estocásticos.								
[Eq.4.13 (7)]	0.76	87.92	86.68	88.07	10.9	0.45	160.7	4.93

Tabla 4.4. Parámetros estadísticos para los modelos QSAR-ADL en la serie de predicción.

Modelos	Coefficiente de correlación de Matthews (C)	Exactitud 'Q _{Total} ' (%)	Especificidad d (%)	Sensibilidad (%)	Razón falsos positivos (%)
Modelos QSAR\ADL utilizando los índices lineales no estocásticos					
[Eq.4.1]	0.71	85.37	86.11	84.35	15.3
[Eq.4.2]	0.74	87.07	88.19	85.81	14.0
[Eq.4.3]	0.77	88.44	88.89	87.67	12.0
[Eq.4.4]	0.73	86.39	87.50	85.14	14.6
[Eq.4.5]	0.73	86.73	87.50	85.71	14.0
[Eq.4.6]	0.73	86.73	87.50	85.71	14.0
Modelos QSAR\ADL utilizando los índices lineales estocásticos					
[Eq.4.7]	0.73	86.39	85.42	86.62	12.6
[Eq.4.8]	0.78	88.78	87.50	89.36	10.0
[Eq.4.9]	0.76	88.10	85.42	89.78	9.33
[Eq.4.10]	0.75	87.41	88.19	86.39	13.3
[Eq.4.11]	0.75	87.41	86.11	87.94	11.3
[Eq.4.12]	0.80	89.80	90.28	89.04	10.6
Modelo general QSAR\ADL utilizando los índices lineales no estocásticos y estocásticos.					
[Eq.4.13]	0.76	88.10	86.81	88.65	10.67

Los modelos obtenidos con los índices lineales no-estocásticos y estocásticos (**Ecs 4.6 y 4.12**) resultantes de las combinación de las diferentes ponderaciones describen de una forma adecuada la actividad antiinflamatoria con valores de Q = 88.57% (C=0.77); y Q= 88.47% (C=0.77), respectivamente. En todos los casos se evidencian adecuados porcentajes de falsos activos (ver Tabla 4.3). Los falsos activos son compuestos inactivos que los modelos los clasifica como antiinflamatorio y los falsos inactivos son los compuestos activos clasificados como inactivos por los modelos. Los valores de $\Delta P\%$ (probabilidades posteriores) para estos compuestos en la SE, se muestran en la Tablas 2 y 3 de los Anexos.

Una de las aplicaciones más importantes de un modelo QSAR es su uso en la predicción de la actividad de nuevas entidades químicas, que incluso no tienen que tener existencia física aún. En este sentido, uno de los criterios para la aceptación o no de un modelo QSAR como los que se han desarrollado en este trabajo, están basados en los estadísticos de la serie de predicción externa (79). En esta serie los dos mejores modelos (Ecs 4.6 y 4.12) clasificaron

correctamente el 87.50% y el 90.27% de los compuestos activos y el 86.00% y el 89.33% de los compuestos inactivos, para un porcentaje de buena clasificación global de 86.73% y 89.80%, respectivamente. Como puede observarse, en esta serie el modelo ajustado usando los índices estocásticos (Ec. 4.12) describió más adecuadamente la actividad biológica de estos compuestos. Estos resultados se suman en la Tabla 4.4, donde también muestran los valores de otros parámetros estadísticos obtenidos en la serie de validación externa (exactitud, sensibilidad, especificidad, etc.) Los resultados de la clasificación (valores de las probabilidades posteriores) para cada uno de los compuestos activos e inactivos en la SP se pueden observar en las Tablas 4 y 5 de los Anexos.

De forma general podemos plantear que los dos modelos presentan parámetros estadísticos significativos; pero teniendo en cuenta los resultados en la clasificación de los compuestos en la SP externa, consideramos que la (Ec. 4.12) es superior a la (Ec. 4.6). No obstante, estas diferencias no son significativas y en la práctica todos los modelos obtenidos deben ser utilizados en paralelo en el momento de seleccionar compuestos como posibles antiinflamatorios. Este enfoque será evidenciado más adelante a través de un ejemplo de tamizaje virtual para el descubrimiento *biosilico* de nuevos antiinflamatorios. Uno de los principales parámetros que se debe tener en cuenta cuando los modelos son usados en la práctica de la química médica es *la razón de falsa alarma positiva (FAR)*. Este parámetro revela la probabilidad de seleccionar un compuesto negativo (no antiinflamatorios) como positivo (antiinflamatorios), o sea un falso activo.

El cálculo de este parámetro estadístico se realizó tanto para las SE como para las SP por cada uno de los modelos obtenidos. Teniendo en cuenta la cantidad de compuestos químicos que existen en la actualidad a los cuales no se les han realizado ningún tipo de ensayos, es preferible obtener una **FAR** lo más baja posible, pues de esta manera no estaremos evaluando compuestos que los modelos lo describen como positivo, siendo realmente negativos.

4.3. Segunda Serie de Predicción Externa. Simulación de Screening' Virtual e Identificación de Nuevas Entidades Moleculares Como Fármacos Antiinflamatorio: Ensayos *In-Silico*.

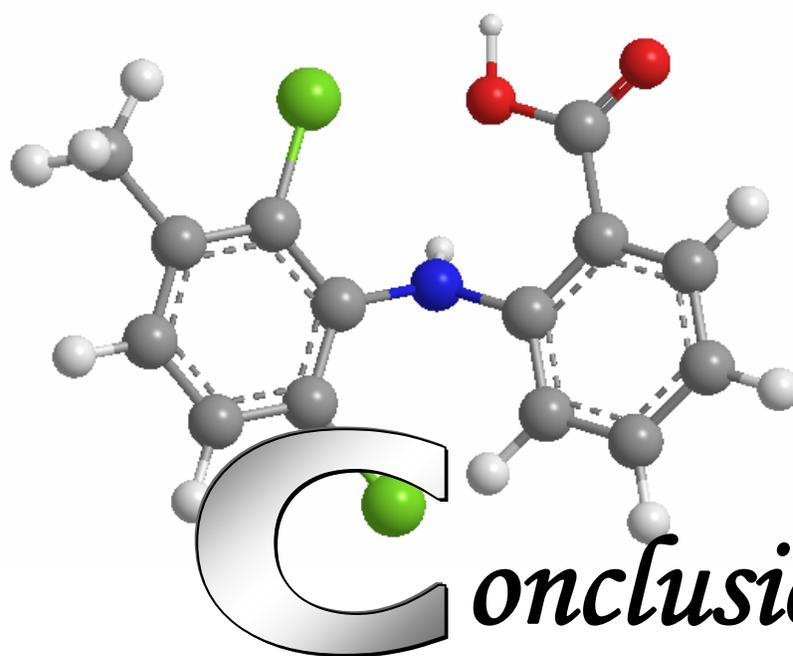
El 'screening' virtual ha emergido como una interesante alternativa para la evaluación masiva de compuestos químicos (32). Este proceso consiste en utilizar los modelos obtenidos en la evaluación virtual de grandes bases de datos en orden de encontrar un conjunto reducido de compuestos predichos con la actividad biológica deseada. Por tal motivo, conjuntamente con el desarrollo y corroboración de los modelos de predicción, se desarrollo un procedimiento de evaluación in silico de fármacos con diversos usos farmacológicos con el objetivo de identificar compuestos antiinflamatorios.

En este sentido, podemos plantear que existen 2 enfoques diferentes para encontrar un nuevo compuesto antiinflamatorios: 1) encontrar compuestos conocidos, con otras actividades o usos, en bases de datos químicas y probar su actividad antiinflamatoria experimentalmente o 2) diseño de nuevos compuestos para ser sintetizados en un laboratorio químico y posteriormente evaluados experimentalmente. Ambos enfoques son importantes y muy utilizados en la práctica farmacéutica, el primero permite la selección de compuestos con métodos de síntesis bien establecidos y en muchos casos su comportamiento toxicológico y farmacodinámico es bien conocido, sobre todo para el caso de compuestos comercializados como fármacos. El segundo enfoque permite el diseño de nuevas cabezas de serie (compuestos líderes) con la actividad deseada, pero ellos necesitan ser primeramente sintetizados, evaluados farmacológicamente y finalmente tienen que pasar a través de rigurosos 'test' toxicológicos y farmacodinámicos.

Por esta razón seleccionamos ambos métodos para la búsqueda de nuevos compuestos antiinflamatorios. En el primer caso desarrollamos una exhaustiva búsqueda en el Martin Negwer Handbook (74) de compuestos para ser evaluados en los modelos. Un reducido grupo fue identificado por las funciones discriminantes como posibles antiinflamatorios. Así por ejemplo, en la Tabla 6 de

los anexos se muestra una lista de compuestos que fueron identificados (seleccionados) por todos los modelos de clasificación como promisorios antiinflamatorios, es de destacar que todos estos compuestos muestran actividad con una probabilidad mayor que el 80%, para todas las ecuaciones. Entre estos, nosotros podemos encontrar conocidos fármacos con otras propiedades farmacológicas. Estos resultados son unos de los más importantes criterios de validación de los modelos desarrollados en este trabajo, pues estos son capaces de detectar una serie de compuestos como activos a partir de miles de compuestos incluidos en las bases de datos. Por supuesto que esto es un resultado preliminar y la actividad de los compuestos seleccionados como antiinflamatorios tienen que ser corroborada experimentalmente.

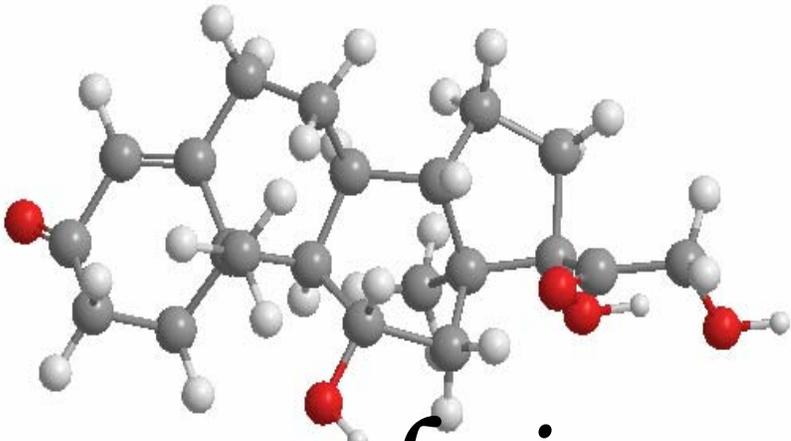
Tomando en consideración los resultados obtenidos, en la simulación de cribado virtual, que da un criterio de la fiabilidad de nuestros modelos para la discriminación de la actividad antiinflamatoria, se evaluaron nuevos compuestos de síntesis química a través de los modelos QSAR obtenidos y varios de ellos fueron identificados como posibles compuestos con actividad antiinflamatoria. Los resultados de la clasificación de estas nuevas entidades moleculares se muestran en la Tabla 7 de los Anexos. A partir de estos resultados, es posible contar con un grupo de moléculas como potenciales antiinflamatorios, sin embargo se hace necesario corroborar la actividad predicha a través de las pruebas experimentales.



onclusiones

Conclusiones.

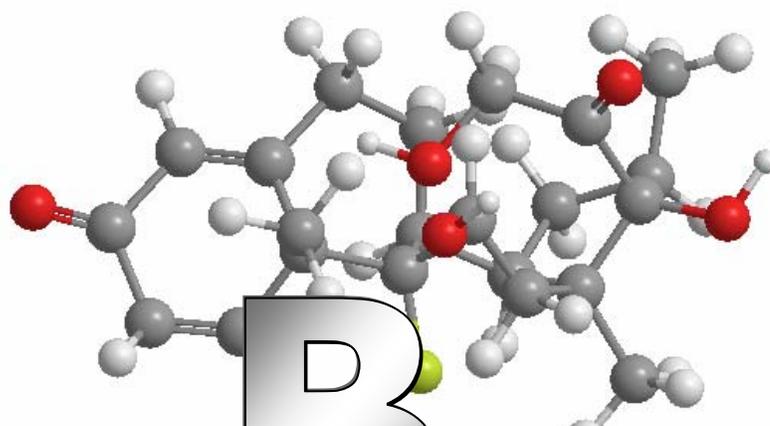
1. Se recolectó una base de datos de la literatura de compuestos a los que se les ha reportado experimentalmente actividad antiinflamatoria para acceder al análisis y la modelación confiable de la data.
2. Se obtuvieron modelos QSAR utilizando descriptores **TOMOCOMD-CARDD** que permitieron la descripción de la actividad antiinflamatoria.
3. Se desarrollaron procesos de validación externa de los modelos encontrados, para demostrar la robustez y el poder predictivo de los mismos.
4. Se aplicaron todos los modelos obtenidos al “screening” virtual de compuestos químicos para evidenciar las potencialidades del método **TOMOCOMD-CARDD** en el descubrimiento (selección/identificación y/o diseño) de nuevos compuestos líderes.



R *ecomendaciones*

Recomendaciones.

1. Desarrollar modelos QSAR utilizando otros descriptores **TOMOCOMD-CARDD** que permitan la descripción de la actividad antiinflamatoria
2. Realizar los ensayos experimentales de los compuestos seleccionados como posibles antiinflamatorios.



Bibliografía

Referencias bibliográficas.

- 1- Rocha E, Silva M. A brief survey of the history of inflammation. *Agents and Actions* 43:86-90, 1994.
- 2- PhaRMA Industry Profiles, Pharmaceutical Research and Manufacturers of America, 2000.
- 3- Navarro A, de las Heras B, Villar A. Anti-inflammatory and immunomodulating properties of a sterol fraction from *Sideritis foetens* Clem. *Biol. Pharm. Bull.* 24 (5) 470-473. 2001.
- 4- Ali H, Haribabu B, Richardson RM, Snyderman R. Mechanisms of inflammation and leukocyte activation. *Adv Rheumatol* 81:1-28, 1997.
- 5- Signore A, Annovazzi A, Corsetti F, Capriotti G, Chianelli M, De Winter F, Scopinaro F. Biological imaging for the diagnosis of inflammatory conditions. *BioDrugs* 16:241-259, 2002.
- 6- Florez de Qca Ftica) pag 906
- 7- <http://rubenbrizuela.americas.tripod.com/teoria/teoria8.htm>
- 8- Pradilla OE. Ciclooxygenasa 3: la nueva iso-enzima de la familia. *MedUNAB* 2004; 7:181-4
- 9- http://es.wikipedia.org/wiki/Antiinflamatorio_no_esteroideo
- 10-[http://html.rincondelvago.com/antiinflamatorio no esteroideos.html](http://html.rincondelvago.com/antiinflamatorio_no_esteroideos.html)
- 11-Florez de Qca Ftica) 4ta edición
- 12-Goodman TA, Simon LS. Minimizing the complication of NSAID therapy. *J Musculoskel Med*, 1994; 11:33-46
- 13-Hollander D, Gastrointestinal complications of nonsteroidal anti-inflammatory drugs: Prophylactic and therapeutic strategies. *Am J Med* 1994; 96:274-281.
- 14-Delmas PD, Non-steroidal anti-inflammatory drugs and renal function. *Br J Rheumatol* 1995; 34 (suppl.1): 25-28.
- 15-Ciabattoni G, Boss AH, Patragnani P, y cols. Effects on Sulindac on renal and extrarenal eicosanoid synthesis. *Clin Pharmacol Ther* 1987; 41:380-3.

- 16-Freed MI, Audet PR y cols. Comparative effects of nabumetone, sulindac and indomethacin on urinary prostaglandin excretion and platelet function in volunteers. *J Clin Pharmacol* 1994; 3 (11): 1098-1108.
- 17-Murray MD, Brater DC, Renal toxicity of the non-steroidal anti-inflammatory drugs. *Annu Rev Pharmacol Toxicol* 1993; 33:435-65.
- 18-Weir MR, Klassen DK, Hall PS y cols. Minimization of indometacin-induced reduction in renal function by misoprostol. *J Clin Pharmacol* 1991;31 729-35.
- 19-Engelhardt G. Pharmacology of Meloxicam, a new non-steroidal anti-inflammatory drug with an improved safety profile through preferential inhibition of COX-2. *Br J Rheumatol* 1996; 35(supl.1):4-12.
- 20- Kemp SF, Lockey RF, Wolf BL, Lieberman P. Anaphylaxis. A review of 266 cases. *Arch Intern Med.* 1995;155:1749-54.
- 21-Tintinall JE, Ruiz E, Krome RL. *Emergency Medicine. A comprehensive study guide.* Editorial McGraw Hill. Cuarta edición. 1996. 793.
- 22-Zaragoza A, Alfonso V, Roig E. Hepatotoxicidad inducida por AINES: aceclofenac y diclofenac. *Rev Esp Enferm Dig* 1995; 87(6): 472-5.
- 23-Sánchez, C.; Pinzon, R.: *Manual de técnicas de investigación. Actividad antiinflamatoria.* Marzo. Pág. 81-82. 1995.
- 24-Silva, J.; Abebe, W.: Analgesic and anti-inflammatory effects of essential oils of Eucalyptus. *Januaral of Ethnopharmacologic.* Pág. 277-283. 2003.
- 25- <http://revista.consumer.es/>
- 26- Centro Sanger de Gran Bretaña, junio 2007
- 27-Bases farmacológicas de la terapéutica, Goodman and Gilman, 6ta Edición.
- 28-Van de Waterbeemd, H.; Carter, R., E.; Grassy, G.; Kubinyi, H.; Martin, Y., C.; Tute, M., S.; Willett, P. *Annu. Rep. Med. Chem.* 1998, 33, 397.
- 29-Todeschini, R.; Consonni, V. (Eds.) *Handbook of Molecular Descriptors;* Wiley-VCH: Weinheim (Germany), 2000.

- 30-Van de Waterbeemd, H. (Ed.); *Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry, Vol. 2)*, John Wiley & Sons: New York. 1995.
- 31-Basak, S., C.; Magnuson, V., R.; Niemi, G., J.; Regal, R., R. *Discrete Appl. Math.* 1988, 19, 17.
- 32-Lajiness, M., S. In *Computational Chemical Graph Theory*; Rouvray, D. H. (Ed.); Nova Sci., New York. 299, 1990.
- 33-Basak, S., C.; Grunwald, G., D. *J. Chem. Inf. Comput. Sci.* 1995, 35, 366.
- 34-Basak, S., C.; Grunwald, G., D. *SAR QSAR Environ. Res.* 1994, 2, 289.
- 35-Basak, S., C.; Grunwald, G., D. *SAR QSAR Environ. Res.* 1995, 3, 265.
- 36-Marrero-Ponce, Y. *Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. Molecules* 2003, 8, 687-726.
- 37-Basak, S., C.; Gute, B., D. *SAR QSAR Environ. Res.* 1997, 7, 1.
- 38-Casañola-Martin, G. M.; Khan, M. T. H.; Marrero-Ponce, Y.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F. *New Tyrosinase Inhibitors Selected by Atomic Linear Indices-Based Classification Models. Bioorg. Med. Chem. Letter.* 2006, 16, 324-330.
- 39-Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc, R.; Buelow, R. *Nature Biotech.* 1998, 16, 748.
- 40-Marrero-Ponce, Y. *Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Definition, Significance-Interpretation and Application to QSAR Analysis of Flavone Derivatives as HIV-1 Integrase Inhibitors. J. Chem. Inf. Comput. Sci.* 2004, 44, 2010-2026.
- 41-Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. *Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix" and Their Stochastic Forms: A Novel Approach for Virtual Screening and in silico Discovery of New Lead Paramphistomicide Drugs-like Compounds. J. Mol. Struct. (Theochem).* 2005, 717, 67-79.
- 42-Marrero-Ponce, Y.; Castillo-Garit, J., A. *3D-Chiral Atom, Atom-type, and Total Non-Stochastic and Stochastic Molecular Linear Indices and their*

- Applications to Central Chirality Codification. J. Comput.-Aided Mol. Design. In Press (DOI: DO00017575)
- 43-Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón Pérez, M.; García Sánchez, R. Non-Stochastic and Stochastic Linear Indices of the “Molecular Pseudograph’s Atom Adjacency Matrix”: Application to “in silico” Studies for the Rational Discovery of New Antimalarial Compounds. *Bioorg. Med. Chem.* 2005, 13, 1293-1304.
- 44-Marrero-Ponce, Y.; Iyarreta-Veitía, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Ávila, P. E.; Kirchgatter, K.; Machado, Y. Ligand-Based Virtual Screening and in silico Design of New Antimalarial Compounds Using Non-Stochastic and Stochastic Total and Atom-type Quadratic Maps. *J. Chem. Inf. Comput. Sci.* 2005, 45, 1082-1100.
- 45-Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; Sánchez, A. M.; Torrens, F.; Castro, E. A. Atom, Atom-Type and Total Molecular Linear Indices as a Promising Approach for Bioorganic & Medicinal Chemistry: Theoretical and Experimental Assessment of a Novel Method for Virtual Screening and Rational Design of New Lead Anthelmintic. *Bioorg. Med. Chem.* 2005, 13, 1005-1020.
- 46-Marrero-Ponce, Y.; Medina-Marrero, R.; Martinez, Y.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. A. Non-Stochastic and Stochastic Linear Indices of the Molecular Pseudograph’s Atom Adjacency Matrix: A Novel Approach for Computational –in silico- Screening and “Rational” Selection of New Lead Antibacterial Agents. *J. Mol. Mod. J. Mol. Mod.* 2006, 12, 255–271.
- 47-Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a Novel Approach for Computer-Aided “Rational” Drug Design: I. Theoretical and Experimental Assessment of a Promising Method for Computational

- Screening and in silico Design of New Anthelmintic Compounds. *J. Comput.-Aided Mol. Des.* 2004, 18, 615-633.
- 48-Marrero-Ponce, Y. Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Molecular Descriptors and QSPR/QSAR Applications. *Bioorg. Med. Chem.* 2004, 12, 6351-6369.
- 49-Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro E. Atom, Atom-Type and Total Linear Indices of the “Molecular Pseudograph’s Atom Adjacency Matrix”: Application to QSPR/QSAR Studies of Organic Compounds. *Molecules.* 2004, 9, 1100-1123.
- 50-Marrero-Ponce, Y.; Cabrera, M., A.; Romero, V.; Ofori, E.; Montero, L. A. Total and Local Quadratic Indices of the “Molecular Pseudograph’s Atom Adjacency Matrix”. Application to Prediction of Caco-2 Permeability of Drugs. *Int. J. Mol. Sci.* 2003, 4, 512-536.
- 51-Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. A New Topological Descriptors Based Model for Predicting Intestinal Epithelial Transport of Drugs in Caco-2 Cell Culture. *J. Pharm. Pharm. Sci.* 2004, 7, 186-199.
- 52-Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. 3D-Chiral Quadratic Indices of the “Molecular Pseudograph’s Atom Adjacency Matrix” and their Application to Central Chirality Codification: Classification of ACE Inhibitors and Prediction of σ -Receptor Antagonist Activities. *Bioorg. Med. Chem.* 2004, 12, 5331-5342.
- 53-Marrero-Ponce, Y.; Cabrera, M. A.; Romero-Zaldivar, V.; Bermejo, M.; Siverio, D.; Torrens, F. Prediction of Intestinal Epithelial Transport of Drug in (Caco-2) Cell Culture from Molecular Structure using ‘in silico’ Approaches During Early Drug Discovery. *Internet Electronic J. Mol. Des.* 2005, 4, 124-150.
- 54-Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. Nucleic Acid Quadratic Indices of the “Macromolecular Graph’s Nucleotides Adjacency Matrix”.

- Modeling of Footprints after the Interaction of Paromomycin with the HIV-1 Ψ -RNA Packaging Region. *Int. J. Mol. Sci.* 2004, 5, 276-293.
- 55-Marrero-Ponce, Y.; Medina-Marrero, R.; Castillo-Garit, J. A.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. Protein Linear Indices of the "Macromolecular Pseudograph's α -Carbon Atom Adjacency Matrix" in Bioinformatics. 1. Prediction of Protein Stability Effects of a Complete Set of Alanine Substitutions in Arc Repressor. *Bioorg. Med. Chem.* 2005, 13, 3003-3015.
- 56-Marrero-Ponce, Y.; Torrens, F. Bond-Extended non-Stochastic and Stochastic Bilinear Indices. 1. QSPR Applications to the Description of Physicochemical Properties of Octane Isomers. *J. Theor. Chem. (THEOCHEM)*. Submitted for publication.
- 57-Marković, S., Gutman, I., *J. Mol. Struct. (Theochem)*, 235 (1991) 81.
- 58-Potapov, V.M. Stereochemistry, Mir, Moscow, 1978.
- 59-Alzina, R. B. Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL Y SPAD; PPU, SA: Barcelona, 1989; Chapter 8-10, Vol. 1, p. 202-264.
- 60-STATISTICA ver. 5.5, Statsoft, Inc. 1999.
- 61-Fan, Y.; Shi, L., M.; Kohn, K., W.; Pommier, Y.; Weinstein, J., N. *J. Med. Chem.* 2001, 44, 3254
- 62-Belsey, D., A.; Kuh, E.; Welsch, R., E. *Regression Diagnostics*; Wiley: New York, 1980.
- 63-Basak, S., C.; Balaban, A., T.; Grunwald, G., D.; Gute, B., D. *J. Chem. Inf. Comput. Sci.* 2000, 40, 891.
- 64-Patel, H.; Cronin, M., T., D. *J. Chem. Inf. Comput. Sci.* 2001, 41, 1228.
- 65-Cronin, M., T., D.; Schultz, T., W. *J. Mol. Struct. (Theochem)*. 2003, 622, 39.
- 66-Egan, W., J.; Morgan, S., L. *Anal. Chem.* 1998, 70, 2372.
- 67-Kier, L., B. *J. Pharm. Sci.* 1980, 69, 416.
- 68-Cativiela, C.; Garcia, J., I.; Fernandez-Alvarez, E.; Elorriaga, C. *Acta Cim. Hung.* 1992, 130, 129.

- 69-Ogino, A.; Matsumura, S.; Fujita, T. J. Med. Chem. 1980, 23, 437.
- 70-Rose, K.; Hall, L., H.; Kier, L., B. J. Chem. Inf. Comput. Sci. 2002, 42, 651.
- 71-Gironés, X.; Gallegos, A.; Ramon, C-D. J. Chem. Inf. Comput. Sci. 2000, 46, 1400.
- 72-Suzuki, T.; Ide, K.; Ishida, M.; Shapiro, S. J. Chem. Inf. Comput. Sci. 2001, 41, 718.
- 73-Golbraikh, A.; Tropsha, A. J. Mol. Graph. Modell. 2002, 20, 269.
- 74-M. Negwer, (Ed), Organic-Chemical Drugs and their Synonyms; Akademie-Verlag, Berlin, (1987)
- 75-Chapman & Hall. The Merck Index. Twelfth Edition. (1996)
- 76-Marrero-Ponce, Y. and Romero, V. TOMOCOMD software. Central University of Las Villas. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a Preliminary Experimental Version; in Future a Professional Version will be Obtained upon Request to Y. Marrero., (2002) yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es
- 77-STATISTICA (data analysis software system), vs 6.0; StatSoft Inc., (2001)
- 78-Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. Bioinformatic (2000), 16, 412–424.
- 79-Wold, S. and Erikson, L. Statistical Validation of QSAR Results. Validation Tools; In: Manhnhold, R., Krogsgaard-Larsen, L. Timmerman, H. (Eds.). Methods and Principles in Medicinal Chemistry. Vol. 2, Chemometric Methods in Molecular Design, Van Waterbeemd, H. (Ed.), VCH, Weinhiem. 309-318 (1995)
- 80-Mc Farland, J. W; Gans, D. J. Cluster Singnificance Analysis. In Chemometric Methods in Molecular Design, Waterbeemd, H., VCH Publishers: New York, 1995; pp 295-307.
- 81-Johnson, R. A.; Wichern, D. W. Applied Multivariate Statistical Analysis. Prentice-Hall: New Jersey, 1988.

