



UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS

Trabajo de Diploma

*Los Índices Lineales Moleculares: Un
Novedoso Enfoque para la Modelación de
la Actividad carcinogénica de compuestos
químicos*

Autora: Yuleimi Díaz Pérez

Tutores: Lic. Juan Alberto Castillo Garit

Dr. Yovani Marrero Ponce

Dr. Antonio Pérez Donato

Departamento Farmacia, Facultad de Química y Farmacia

Santa- Clara 2006

ABSTRACT

A novel approach has been used to predict the carcinogenicity activity, non-stochastic and stochastic atom-based linear indices. Two QSAR models were developed to predict the carcinogenic and non-carcinogenic activity on a data set of 189 compounds. The accuracy of these models was 78.31% and 79.37% for non-stochastic and stochastic model respectively. The predicted power of the models was assessed by the use of two external prediction sets, the first and second exercises of the National Toxicology Program. The results achieved with **TOMOCOMD-CARDD** molecular descriptors were similar to other previously reported. We can conclude that non-stochastic and stochastic atom based linear indices could be an interesting tool to predict the carcinogenicity potential of chemical compounds.

RESUMEN

Se utilizó un enfoque computacional conocido como ***TOMOCOMD-CARDD*** para modelar la actividad carcinógena de compuestos químicos. Específicamente fueron empleados los índices lineales tanto estocásticos como no estocásticos para modelar el potencial carcinogénico. Se obtuvieron dos funciones discriminantes para predecir la actividad carcinógena y no carcinogénica de 189 compuestos. El porcentaje de la buena clasificación de dichos modelos fue de 78.31% y 79.37% respectivamente. Los modelos obtenidos fueron validados haciendo uso de dos series de predicción externas; las mismas son el primer y segundo ejercicio del Programa Nacional de Toxicología de Estados Unidos. (Nacional Toxicology Program, NTP por sus siglas en inglés), conformados por 43 compuesto del primer ejercicio y 26 compuestos del segundo ejercicio. Nuestros resultados fueron comparados con otros reportados en la literatura internacional observándose correspondencia entre nuestros resultados y los obtenidos por otros métodos. Podemos concluir que los índices lineales estocásticos y no-estocásticos pueden ser empleados como una herramienta alternativa para predecir el potencial carcinogénico de compuestos químicos.

TABLA DE CONTENIDO

	Resumen	
	Abstract	
	Glosario de Términos	
	Tabla de Contenido	
	Capítulo 1. INTRODUCCIÓN	1
	Capítulo 2. REVISIÓN BIBLIOGRÁFICA	4
2.1	Conceptos Básicos y Definiciones	4
2.2	El Cáncer, Origen de la Enfermedad	8
2.2.1	Posibles causas	9
2.2.2	Apariencia microscópica de las células cancerosas	10
2.3	La Carcinogénesis como Proceso	10
2.3.1	Mecanismos moleculares en el desarrollo de la carcinogénesis	11
2.3.1.1	<i>Mecanismos Moleculares de Defensa</i>	11
2.3.1.2	<i>Etapas de la Carcinogénesis</i>	11
2.4	Tipos de Carcinógenos	15
2.4.1	Carcinógenos genotóxicos	16
2.4.2.1	<i>Carcinógenos Endógenos</i>	16
2.4.2.2	<i>Carcinógenos Exógenos</i>	17
2.4.2	Carcinógenos no genotóxicos o epigenéticos	17
2.4.3	Clasificación de los carcinógenos	18
2.5	Los Agentes Genotóxicos y el Daño Genético	20
2.5.1	Metabolismo de los agentes genotóxicos	21
2.5.2	La inhibición del metabolismo	22
2.5.3	Interacciones con el ADN	23
2.6	Métodos para Determinar la Acción Carcinogénica	24
2.6.1	Determinación de carcinógenos genotóxicos	24
2.6.2	Bioensayo in vivo. Determinación de carcinogenicidad en animales	27
2.7	Métodos Alternativos para Valorar la Carcinogénesis	28
2.7.1	Modificaciones propuestas al bioensayo clásico	28
2.7.2	Alternativas sin animales	30
	Capítulo 3. MÉTODO COMPUTACIONAL Y ANÁLISIS QUIMIOMÉTRICO DE LOS DATOS	34
3.1	Método Computacional	35
3.1.1	Descriptores <i>TOMOCOMD-CARDD</i> basados en relaciones entre núcleos atómicos	36
3.1.1.1	<i>Información Química y Vector Molar Basados en Átomos</i>	36
3.1.1.2	<i>Background" Sobre las Matrices de Densidad Electrónica Grafo-Teóricas No Estocásticas y Estocásticas</i>	37
3.1.1.3	<i>Definición de los Índices Lineales basados en Relaciones de Átomos</i>	42
3.2	Análisis Quimiométrico de los Datos	44
3.2.1	Análisis Discriminante Lineal (ADL) para la predicción de la actividad	44
3.2.2	Estimación de coeficientes	44
3.2.3	Matriz de clasificación	45
3.2.4	Significación de la Función Discriminante (FD)	45

3.2.5	Criterios de selección de variables	46
3.2.5.1	<i>Selección del Número Óptimo de Predoectores. Principio de la Parsimonia</i>	46
3.2.5.2	<i>Tolerancia</i>	47
3.2.6	Multicolinialidad entre variables, redundancia de la información y correlaciones casuales	47
3.2.7	Validación estadística de los modelos QSAR	47
3.2.7.1	<i>Procesos de validación a los modelos QSAR-ADL obtenidos.</i>	48
3.2.8	Conclusiones parciales de capítulo	48
	Capítulo 4. PARTE EXPERIMENTAL	50
4.1	Base de Datos de los Compuestos Usados en la Obtención de los Modelos ADL-QSAR	50
4.2	Método Computacional. TOMOCOMD CARDD software	51
4.3	Desarrollo de las Funciones Discriminates (FDs) usando ADL	53
	Capítulo 5. RESULTADOS Y DISCUSIÓN	55
5.1	Obtención de los Modelos de Clasificación con Índices Moleculares.	55
5.1.1	Desarrollo de Modelos QSAR-ADL con Índices Lineales no-estocásticos	55
5.1.2	Desarrollo de Modelos QSAR-ADL con Índices Lineales estocásticos	61
5.1.3	Comparación entre los Índices Lineales Moleculares y otros Enfoques Teóricos	66
	Capítulo 6. CONCLUSIONES	68
	Capítulo 7. RECOMENDACIONES	69
	Capítulo 8. REFERENCIAS BIBLIOGRÁFICAS	70
	Anexos	85

GLOSARIO DE TÉRMINOS

2D Bidimensional

3D Tridimensional

ADME Absorción, Distribución, Metabolismo y Excreción

ADN Ácido desoxiribonucleico

AIH Absorción Intestinal en Humanos

ant.: antónimo, opuesto

ARN Ácido ribonucleico

CoMFA Comparative Molecular field Analysis

cruzada LOO

D^2 Distancia de Mahalanobis al cuadrado

DL₅₀ dosis letal media.

EC₅₀ Concentración Efectiva 50

F Biodisponibilidad

F_{exp} Razón de Fisher experimental

IC₅₀ Concentración Inhibitoria 50

ITs Índices Topológicos

LGO Procedimiento de validación interna cruzada ‘leave-group-out’

LNO Procedimiento de validación interna cruzada ‘leave-n-out’

Log (1/EC₅₀) Logaritmo del inverso de la Concentración Efectiva 50

Log (1/IC₉₀) Logaritmo del inverso de la Concentración Inhibitoria 50

Log PCaco-2 Logaritmo del coeficiente de permeabilidad intestinal “*in vitro*” en células de carcinoma de colon

LOO Procedimiento de validación interna cruzada ‘leave-one-out’

m. est.: más estricto

m. gral.: más general

MAE Error medio absoluto

NEM Nueva Entidad Molecular

q^2 Coeficiente de correlación al cuadrado del procedimiento de validación

$qk(x)$ k -ésimo Índice cuadrático total de la matriz de adyacencia entre vértices de un

seudografo molecular

QSAR Quantitative Structure Activity Relationships

QSPeR Quantitative Structure Permeability Relationships

QSPR Quantitative Structure Property Relationships

R Coeficiente de correlación

R^2 Coeficiente de determinación o coeficiente de correlación al cuadrado

s Desviación estándar

scv Desviación estándar del procedimiento de validación cruzada LOO

SE Serie de entrenamiento

sin.: sinónimo

SP Serie de predicción

t. rel.: término relacionado...

Te Temperatura de ebullición

TOMOCOMD TOPological MOlecular COMputer Design

TOPS MODE TOPological Substructural Molecular DEsign

VC Validación cruzada

λ Lambda de Wilks

μ_k k -ésimo momento espectral

1

INTRODUCCIÓN

El cáncer es considerado una de las enfermedades más importantes a nivel mundial, cada año se presentan más de seis millones de muertes por algún tipo de neoplasia maligna en el ámbito mundial. Asimismo, se estima que cerca de veinte millones de personas presentarán algún tipo de cáncer cada año, de las cuales cerca de nueve millones corresponderán a casos incidentes.

Constituye esta la segunda causa de muerte en los países desarrollados, después de las enfermedades cardiovasculares, aunque, si se consideran separadamente la enfermedad cardiovascular y la cardiopatía isquémica, el cáncer (incluyendo todas las localizaciones) ocuparía la primera posición.

En la actualidad la enfermedad es contraída en muchas ocasiones por la exposición de sustancias cancerígenas en horario laboral trayendo consigo la muerte de unas 5.000 personas al año en España.

La inducción de cáncer provocado por la exposición crónica a sustancias químicas fue originalmente descrita por Percival Pott en 1775, quien descubrió la aparición de cáncer de escroto en algunos limpiadores de chimeneas. El médico inglés estableció la inducción de tumores por exposición a agentes cancerígenos (hollín), propuso la prevención por medio de la reducción a la exposición y comprobó la sensibilidad individual, ya que no todos los deshollinadores desarrollaban cáncer de escroto. A principios del siglo XX se hicieron experimentos con animales de laboratorio, los cuales demostraron que el alquitrán, que contiene grandes cantidades de hidrocarburos aromáticos policíclicos, genera tumores. Asimismo se demostró que otro grupo de compuestos, las aminas aromáticas, producen cáncer de vejiga. Otros compuestos con gran potencia carcinogénica fueron descubiertos durante los experimentos realizados para provocar cáncer experimentalmente, como ocurrió con las diferentes nitrosaminas (Rodríguez, 1995).

El cáncer puede originarse a partir de cualquier tipo de célula en cualquier tejido corporal, no es una enfermedad única sino un conjunto de enfermedades que se clasifican en función del tejido

y célula de origen. Existen varios cientos de formas distintas, siendo tres los principales subtipos: los *sarcomas* proceden del tejido conectivo como huesos, cartílagos, nervios, vasos sanguíneos, músculos y tejido adiposo. Los *carcinomas* proceden de tejidos epiteliales como la piel o los epitelios que tapizan las cavidades y órganos corporales, y los tejidos glandulares de la mama y próstata. En el tercer subtipo se encuentran las *leucemias* y *linfomas* que incluyen los cánceres de los tejidos formadores de las células sanguíneas. Producen inflamación de los ganglios linfáticos, invasión del bazo y médula ósea, y sobreproducción de células blancas inmaduras.(Romina, 1997)

La carcinogénesis química ha sido el objeto fundamental en el intento de crear modelos capaces de predecir, debido a que la determinación experimental, para un solo producto químico es aproximadamente de 2 años en el análisis de roedor, lo que implica mucho tiempo y recursos. También el aumento de presiones sociales y económicas de reducir el uso de las pruebas de animales y la existencia de bases de datos relativamente grandes aparece como razones importantes para desarrollar modelos “en sílico” en carcinogénesis.

Sin embargo dentro de los principales problemas para la predicción “in sílico” encontramos: la complejidad biológica del mecanismo bioquímico involucrado en la carcinogénesis química que debe ser modelado por técnicas estadísticas, los descriptores para las estructura y propiedades químicas que son inadecuados para predecir la carcinogenicidad, los modelos de relación de estructura actividad (SAR) los cuales ignoran la importancia de las variables biológicas y la clasificación de carcinógenos en roedores los cuales son demasiados imprecisas para sacar modelos exactos a partir de ellos. La mayoría de los esfuerzos para modelar las relaciones cuantitativas de estructura actividad (QSAR) se han concentrado en el desarrollo de predecir la carcinogenicidad en modelos.

Sistema de objetivos

Objetivo general:

Establecer los criterios científicos que permitan estimar la carcinogénesis con vistas a establecer un novedoso enfoque para modelar esta actividad en compuestos químicos.

Objetivos específicos:

- ✓ *Identificar cuales son los métodos (in vitro, in vivo, etc.) que permitan predecir la actividad carcinogénica de compuesto químicos.*
- ✓ *Identificar las bases teóricas de interés para los debates de la comunidad científica relacionado con la carcinogénesis.*
- ✓ *Comprobar la interrelación existente entre el efecto carcinogénico y el diseño experimental preclínico que se le aplica a los compuestos químicos.*
- ✓ *Desarrollar modelos de clasificación para discriminar la actividad carcinogénica de compuestos químicos utilizando los índices lineales moleculares.*
- ✓ *Validar la calidad y robustez de los modelos desarrollados*
- ✓ *Comparar con otros enfoque previamente descritos en la literatura.*

2

REVISIÓN BIBLIOGRÁFICA: CARCINOGENESIS

2.1 Conceptos Básicos y Definiciones

Adenocarcinoma. Tumor maligno originado en el epitelio glandular o que forma estructuras de tipo glandular. t. rel. **cáncer, adenoma.**

Adenoma. Tumor benigno desarrollado en el epitelio glandular o que forma estructuras de tipo glandular. t. rel. adenocarcinoma.

Benigno. 1. Enfermedad que no produce efectos nocivos persistentes. 2. Tumor que no invade otros tejidos (metástasis); aunque hay pérdida del control de crecimiento no se pierde el de situación. ant. **maligno.**

Bioensayo. Procedimiento para evaluar la actividad biológica, la presencia o la cantidad de una sustancia (tóxico, toxina, hormona, antibiótico, etc.) mediante la medida de sus efectos sobre un organismo o cultivo celular en comparación con una preparación estandar apropiada. m. gral. **ensayo.** Nagel, 1991.

Cáncer. Denominación de las tumoraciones malignas. Los **carcinomas** se originan en las células epiteliales; los **sarcomas** en el tejido conjuntivo. m. est. **carcinoma, epitelioma, sarcoma.** t. rel. **carcinógeno, carcinogénesis,**

carcinogenicidad, neoplasia, tumor.

Carcinogénesis. Proceso de inducción de neoplasias malignas por agentes físicos, químicos o biológicos. t. rel. **carcinógeno, cancerígeno.**(WHO, 1989a)

Carcinogenicidad. Capacidad para inducir cáncer; según la evidencia que sobre ello se tenga de cada agente (sin atender a la potencia) se distingue:

1. Evidencia suficiente. Se ha establecido relación de causalidad con confianza razonable entre la exposición al agente y la aparición de cáncer en humanos o en animales. 2. Evidencia limitada. Se ha observado asociación positiva entre exposición y cáncer, pero la evidencia no es suficiente por no poder descartarse con suficiente seguridad la participación de casualidad, sesgo o de factores que induzcan a confusión. 3. Evidencia inadecuada. Los estudios disponibles son de insuficiente calidad, consistencia o

de escasa representatividad estadística para permitir una conclusión acerca de la existencia o ausencia de una relación de causalidad. 4. Falta de evidencia. Se dispone de estudios adecuados, que abarcan el rango de dosis a los que los seres humanos pueden estar expuestos y que se confirman mutuamente, que no demuestran la existencia de la inducción de cáncer a ningún nivel de exposición. La designación de «falta de carcinogenicidad» está inevitablemente limitada a los tipos de cáncer, dosis, tiempo de exposición y demás circunstancias consideradas en los estudios. Además, nunca puede excluirse la posibilidad de algún riesgo a los niveles de exposición estudiados. 5. Evaluación global. A partir de todos los datos disponibles, cada agente se clasifica mediante un criterio científico que tiene en cuenta la evidencia obtenida en los estudios sobre humanos y en la experimentación animal y cualquier otro dato relevante, en:

Grupo 1. El agente es carcinógeno en humanos; se utiliza esta calificación sólo cuando existe evidencia de ello. Grupo 2. Cuando la evidencia de carcinogénesis en humanos es casi suficiente, pero no hay suficiente evidencia experimental, atendiendo a los datos epidemiológicos experimentales y cualquier otro relevante, se distingue:

Grupo 2A: Agente probablemente carcinógeno en humanos; existe limitada evidencia sobre humanos pero suficiente con animales. Grupo 2B: Agente posiblemente carcinógeno; la evidencia en humanos es limitada y tampoco hay suficiente evidencia con animales de experimentación. Grupo 3. El agente no es clasificable como carcinógeno para humanos, y no puede incluirse en otro grupo. Grupo 4. El agente probablemente no es carcinógeno en humanos; la evidencia disponible, tanto de humanos como de experimentación animal así lo sugiere (IARC, 1987)

Carcinógeno. Agente físico, químico o biológico capaz de incrementar la incidencia de neoplasias malignas. sin. **Cancerígeno** (IARC, 1987)

Carcinoma. Tumor maligno de células epiteliales. sin. **epitelioma**. t. rel. **sarcoma**.

Cocarcinógeno. Factor físico, químico o biológico que intensifica el efecto de un carcinógeno.

Dosis tóxica. Proporción de una sustancia que produce intoxicación sin que llegue a ser letal.

Ensayo de carcinogenicidad. Estudio a largo plazo (crónico) diseñado para cualquier posible efecto carcinógeno de una sustancia.

Ensayo de toxicidad. Estudio experimental de los efectos adversos de una sustancia sobre un organismo vivo, durante un tiempo determinado y condiciones definidas. t. rel. **ensayos de toxicidad aguda, ensayos de toxicidad crónica, ensayos de carcinogenicidad**.

Ensayo de toxicidad aguda. Estudio experimental para determinar los efectos adversos que pueden aparecer en un corto tiempo (usualmente dos semanas) después de una dosis única de una sustancia, o de varias dosis administradas en 24 h. t. rel. **ensayo límite, dosis letal media (DL50)**.

Ensayo de toxicidad crónica. Estudio en el cual se observan organismos a lo largo de una gran parte de su vida, durante y después de la exposición a la sustancia que se ensaya. sin. **ensayo a largo plazo**. ant. **ensayo de toxicidad aguda** (WHO, 1978a)

Epigenesis, epigenético. Cambios en un organismo a causa de alteraciones en la expresión de la información genética, sin alteración del genoma; se afecta el fenotipo pero no el genotipo. t. rel. **mutación, fenotipo, tumor.**

Epitelioma. Tumor originado en el epitelio. m. est. **carcinoma.**

Genotoxicidad. Capacidad para causar daño al material genético; el daño puede ser de tipo mutágeno o carcinógeno.

In vitro. Literalmente, en vidrio. Estudio de laboratorio realizado sobre células, tejidos u órganos aislados o con sistemas subcelulares o bioquímicos (enzimas). ant. **in vivo.**

In vivo. Estudio realizado sobre individuo vivo. ant. **in vitro**

Maligno. 1. Tendencia a empeorar progresivamente hasta la muerte, si no se trata. 2. En relación con el cáncer, células que crecen de forma incontrolada y con tendencia a invadir y destruir otros tejidos. t. rel. **cáncer, metástasis, tumor, hipertermia.** ant. **benigno.**

Metástasis. 1. Movimiento de bacterias u otras células, especialmente las cancerosas, de una parte del cuerpo a otra, dando lugar a modificación en la localización espacial de una enfermedad o de sus síntomas. 2. Crecimiento de microorganismos patógenos o de células anormales lejos de su lugar de origen en el cuerpo.

Mutación. Cualquier cambio heredable, relativamente estable, del material genético que puede ser una transformación química de un gen individual (mutación génica o puntual) que altera su función, o un reordenamiento, ganancia o pérdida de un cromosoma, visible al microscopio (mutación cromosómica). Las mutaciones pueden ocurrir en células germinales y transmitirse a la descendencia o en células somáticas y pasar de una célula a otra al dividirse éstas. t. rel. **cromosoma, gen, clastogénesis, genotoxicidad.**

Mutagenicidad. Capacidad de un agente biológico, químico o físico para inducir cambios heredables (**mutaciones**).

Mutágeno. Cualquier sustancia que puede inducir cambios heredables (mutaciones) en el genotipo de una célula como consecuencia de alteraciones o de pérdida de genes o de cromosomas o de parte de los mismos.

Oncogen. Gen que puede causar transformación neoplásica de las células: los oncogenes son genes con ligeros cambios respecto a genes normales llamados proto-oncogenes.

Oncogénesis. Producción de tumores.

Procarcinógeno. Sustancia que ha de ser metabolizada para inducir tumores malignos.

Sarcoma. Tumor maligno que crece en el tejido conectivo y compuesto primariamente de células anaplásicas que simulan tejido de sostén.

Seguridad química. Garantía práctica de que los organismos no están expuestos a cantidades tóxicas de sustancias químicas; esto implica conseguir un riesgo aceptablemente bajo de exposición a sustancias potencialmente tóxicas (Duffus, 1986)

Toxicidad. Capacidad para producir daño a un organismo vivo, en relación con la cantidad o dosis de sustancia administrada o absorbida, la vía de administración y su distribución en el tiempo (dosis única o repetidas), tipo y severidad del daño, tiempo necesario para producir éste, la naturaleza del organismo afectado y otras condiciones intervinientes.

Toxicidad aguda. Capacidad de una sustancia para producir efectos adversos dentro de un corto plazo de tiempo (usualmente hasta 14 d) después de la administración de una dosis única (o una exposición dada) o tras dosis o exposiciones múltiples en 24 h. t. rel. **efecto agudo.** ant. **toxicidad crónica.**

Toxicidad crónica. Capacidad de una sustancia para producir efectos adversos consecuentes a una exposición prolongada; éstos pueden aparecer durante o después de interrumpida la exposición. t. rel. **ensayo de toxicidad crónica.** ant. **toxicidad aguda** (IRIS, 1986)

Toxicidad subcrónica. 1. Efectos adversos ocasionados por administración o exposición repetida de una sustancia durante un corto período de tiempo, usualmente el 10% de la vida (al menos 90 días en animales). 2. Capacidad para producir efectos adversos tras exposición subcrónica. t. rel. **ensayos de toxicidad subcrónica.**

Toxicidad subcrónica, ensayo de. Experimentación animal para estudiar los efectos producidos por una sustancia cuando se administra repetida o continuamente durante 90 días (WHO, 1979)

Tóxico. Cualquier agente químico o físico capaz de producir un efecto adverso para la salud. Todos los agentes físicos y químicos son tóxicos potenciales, ya que su acción depende de la dosis y de las circunstancias individuales y ambientales (Repetto, 1988)

Toxicología. 1. Disciplina científica dedicada al estudio del peligro actual o potencial presentado por los efectos nocivos de las sustancias químicas sobre los organismos vivos y ecosistemas, de las relaciones de tales efectos nocivos con la exposición, y de los mecanismos de acción, diagnóstico, prevención y tratamiento de las intoxicaciones. 2. Ciencia que estudia las sustancias químicas y los agentes físicos en cuanto que son capaces de producir alteraciones patológicas a los seres vivos, a la par que estudia los mecanismos de producción de tales alteraciones y los medios para contrarrestarlas, así como los procedimientos para detectar, identificar y determinar tales agentes y valorar su grado de toxicidad (Repetto, 1987)

Transgénico. Animal portador de un gen que fue introducido por microinyección de ADN en el núcleo del huevo fertilizado.

Tumor. 1. Inflamación (bulto) o crecimiento anormal de un tejido, ya sea benigno o maligno. 2. Crecimiento anormal, en velocidad y estructura a partir del tejido normal, sin utilidad fisiológica. sin. **neoplasia**.

Xenobiótico. En sentido estricto, cualquier sustancia que interactúa con un organismo y que no es uno de sus componentes naturales. sin. **sustancia exógena, sustancia extraña.** (Repetto y Sans, 1995)

2.2 El Cáncer, Origen de la Enfermedad

El vocablo cáncer viene del griego y significa cangrejo. El médico griego Galeno lo describía como "un tumor que se extiende por los lados mediante unas prolongaciones anormales que invaden los tejidos adyacentes. Esto se parece a las patas de un cangrejo...". Durante siglos se siguió la doctrina de *Hipócrates* con respecto a esta dolencia: "el tratamiento mata a estos enfermos".

El cáncer, que puede originarse a partir de cualquier tipo de célula en cualquier tejido corporal, no es una enfermedad única sino un conjunto de al menos 100 enfermedades distintas aunque relacionadas, a menudo con causas diferentes.

Se origina cuando las células normales se transforman en cancerígenas, es decir, adquieren la capacidad de multiplicarse descontroladamente e invadir tejidos y otros órganos. Este proceso se denomina *carcinogénesis*.

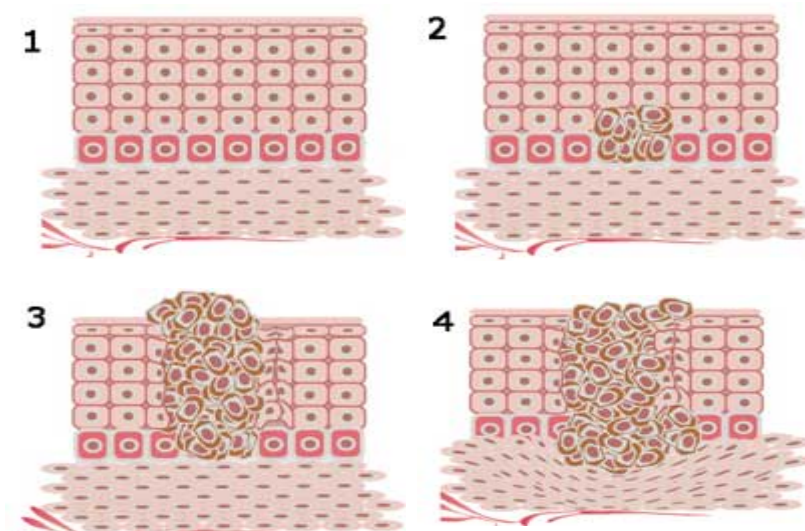


Figura 1. Crecimiento de las células carcinogénicas

La carcinogénesis dura años y pasa por diferentes fases. Las sustancias responsables de producir esta transformación se llaman *agentes carcinógenos*. Un ejemplo de ellos son las radiaciones ultravioleta del sol, el asbesto o el virus del papiloma humano. (Schneider BL y Kulezz-Martin, 2004; Hunter y col, 1997; Nurse y Bissett, 1981)

2.2.1 Posibles causas

Los *hábitos de vida* son fundamentales a la hora de desarrollar determinadas enfermedades como la obesidad, las enfermedades las enfermedades cardíacas y el cáncer. El 75-80% de los cánceres se deben a la acción de agentes externos que actúan sobre el organismo, causando alteraciones en las células. Por el hecho de ser externos, son modificables. En el otro 20-25% de los casos, no se conocen con exactitud los mecanismos por los que se produce y desarrolla el cáncer. Se piensa que puede ser debido a mutaciones espontáneas de los genes o por la acción de algún factor externo que aún no se ha identificado. Actualmente es uno de los principales temas de investigación.

Para que el cáncer se origine debe producirse de cuatro a seis mutaciones o alteraciones genéticas celulares, por lo que todo apunta a que los factores de riesgo deben estar en contacto con el organismo durante un considerable periodo de tiempo (años). Esto también explicaría que el riesgo de padecimiento de un cáncer aumente con los años.

A los agentes externos se les denomina factores de riesgo o agentes carcinógenos. Son las sustancias, que en contacto con un organismo son capaces de generar en él enfermedades cancerosas. Su naturaleza es variada, habiéndose encontrado factores físicos, químicos y biológicos.

La mayoría de los *carcinógenos químicos* están relacionados con actividades industriales, por lo que gran parte de los cánceres producidos por ellos, se dan en los países desarrollados. De los 7 millones de compuestos químicos conocidos, en unos 2000 se ha descrito algún tipo de actividad carcinogénica y muy pocos están en contacto directa o indirectamente con el ser humano. Además, independientemente de su composición, la capacidad de que una sustancia produzca cáncer va a depender de la cantidad de dosis recibida y del tiempo de exposición a la sustancia. El amianto, arsénico, benceno, cadmio, mercurio, níquel, plomo, hidrocarburos clorados, naftilamina, son algunos de los agentes con actividad carcinogénica más usuales.

Entre los *agentes físicos* destacan las radiaciones ionizantes (Rayos X), las radiaciones no ionizantes (rayos ultravioleta del sol), y las radiaciones que emite la propia corteza terrestre (radón). Otra fuente de agentes físicos cancerígenos es la provocada por accidentes nucleares como es el caso de las fugas producidas en centrales nucleares.

En los últimos, años los *agentes biológicos* están tomando cada vez más protagonismo en la carcinogénesis humana. Hoy día sabemos que el 18% de los cánceres son atribuibles a infecciones persistentes provocadas por virus, bacterias o parásitos, entre los que destacan el virus del papiloma humano (cáncer de cuello uterino), el virus de la hepatitis B (cáncer de hígado), el helicobacter pylori (cáncer de estómago).

2.2.2 Apreciación Microscópica de las Células Cancerosas

Bajo el microscopio el tejido canceroso se caracteriza por su distintiva apariencia, apreciándose en la figura 2. Entre las características que se deben buscar están: un gran número de células de forma irregular, dividiéndose, la variación en el tamaño y la forma nuclear, la variación en el tamaño y la forma de la célula, la pérdida de las características especializadas de la célula, la pérdida de la organización normal del tejido y un límite externo mal definido del tumor.

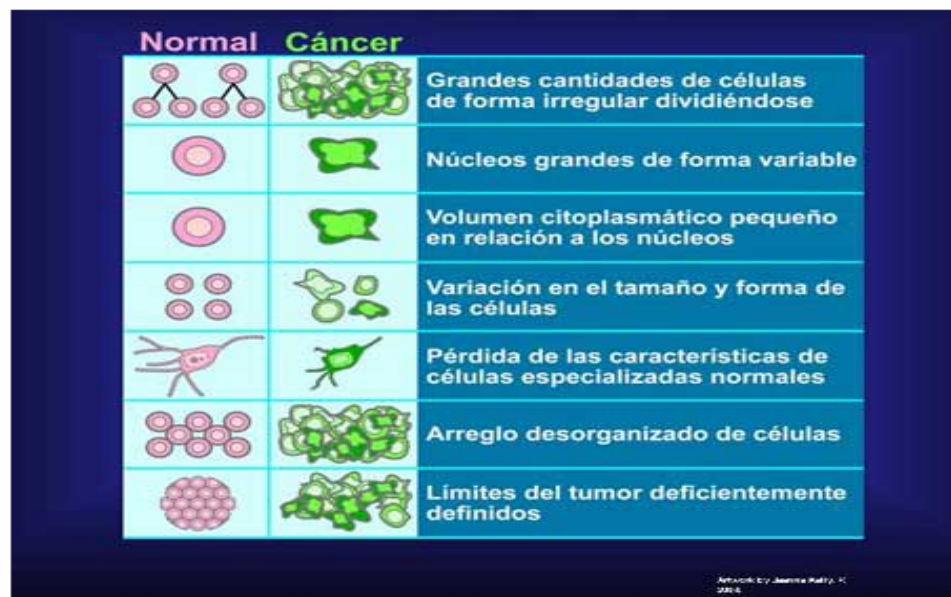


Figura 2. Apariencia microscópica de las células cancerosas

2.3 La Carcinogénesis como Proceso.

En la dieta diaria se ingieren diferentes tóxicos y mutágenos, los organismos han desarrollado un sistema de adaptación relacionado a su capacidad individual de detoxificación;

Cabe señalar que existen dos mecanismos por los cuales los genes pueden alterarse:

a) **GENÉTICO** donde se producen alteraciones estructurales del genoma por **cambios** en la disposición de los propios genes o de sus bases, como son **las mutaciones, translocaciones o deleciones**

b) **EPIGENÉTICO** en acciones moleculares por alteraciones de las enzimas o de los sustratos de las mismas, tal el caso de la metilación de las bases. (Laird, 1997) Este mecanismo generalmente compromete simultáneamente los dos alelos y la **hipometilación** conduce a la mayor expresión de los genes, por lo tanto, una mayor cantidad de la enzima **metiltransferasa** que inhibe la metilación, puede conducir a la mayor expresión de **oncogenes**. Esta enzima se encuentra elevada en los tejidos tumorales.

2.3.1 Mecanismos moleculares en el desarrollo de la carcinogénesis.

2.3.1.1 Mecanismos Moleculares de Defensa

La célula expuesta a tantos factores que pueden dañar el genoma, cuenta sin embargo con mecanismos de defensa, de los cuáles depende la normal replicación celular. Ya hemos mencionado algunos de ellos pero consideramos de interés insistir sobre los mismos:

- La apoptosis, o muerte celular programada.
- Las proteínas *anticiclinas* que “enlentecen” el ciclo celular y dan tiempo a actuar a los mecanismos reparadores del genoma.
- Las proteínas del complejo NER (nucleotide-excisión-repair) también denominadas MMR (mismatch repair genes) que localizan los sectores dañados, devanan la hélice, excluyen el segmento de ADN afectado e incorporan las secuencias correctas.
- El acortamiento fisiológico de los telómeros. Los telómeros, son secuencias del genoma que se encuentran en los extremos de los cromosomas y no se ha constatado hasta ahora, que codifiquen señales de proliferación. Impiden la pérdida y alteración espontánea de las secuencias de ADN y al mismo tiempo son marcadores del envejecimiento celular, puesto que se van acortando con cada división y llega un punto que ellos mismos inducen el “suicidio” de la célula. Cuando los telómeros son “repuestos” por acción de una telomerasa, no llega ese final conveniente para el porvenir del “clon”, de ahí, que la presencia de telomerasas, constituya un signo desfavorable. Las telomerasas se sobreexpresan en los tejidos neoplásicos. Sin embargo, esta enzima puede estar aumentada también en células normales que requieren reproducción activa útil, como son las germinales y algunas hematopoyéticas. (Robbins y col, 1995; Romero y col, 1993; Rosemberg y col, 1990; Rosemberg, 1990; Sidransky, 1996)

3.1.1.2 Etapas de la Carcinogénesis

Cuando el clínico se encuentra ante un tumor, no observa más que un pequeño momento de la vida del proceso canceroso, el denominado periodo clínico. Este periodo, a su vez, puede ser subdividido en dos fases, una fase local, en la que el tumor se encuentra todavía localizado en las estructuras primarias

afectadas, y una fase de generalización, en la que en ocasiones se produce la diseminación del tumor (aparición de metástasis), configurando lo que hemos venido a denominar "enfermedad cancerosa".

Sea con una sola o con las dos subfases, este periodo puede ser asimilado a lo que representa la parte visible de un iceberg, que mantiene fuera de nuestra vista el periodo más largo, el que va desde la actuación de la causa hasta la emergencia clínica y que encierra el periodo precanceroso (iniciación, promoción y una parte más o menos importante de la progresión) (Figura 3).

La primera etapa del proceso de la carcinogénesis, absolutamente preclínica y en una primera etapa aún no cancerosa (precancerosa) consta de tres etapas principales:



Iniciación:

Ocurre a nivel del genoma y las alteraciones pueden darse en los tumores benignos y malignos al igual que la segunda etapa, pero la tercera, o sea la de progresión, es exclusiva de la transformación maligna.

Los agentes que actúan en la primer etapa pueden ser: **FÍSICOS - QUÍMICOS o VIRALES.**

Los carcinógenos físicos están constituidos por las **radiaciones** que dañan, ionizando las bases, deprimen el gen de la proteína p53, pueden estimular citoquinas como la IL 1 y 6, que actúan como verdaderos factores de crecimiento, facilitan la formación de radicales libres y pueden lesionar el gen que codifica para el Complejo Mayor de Histocompatibilidad (CMH) el cual se encuentra en el Cr 6.

Las fuentes radiantes pueden surgir de la metodología diagnóstica o terapéutica como así también por exposición a los rayos solares en forma persistente o por emanaciones de **radón** de los suelos. (Leon, J. y col, 1988)

Los carcinógenos químicos tienen como blanco preferencial al nitrógeno de la **guanina** (alquilantes, aminas aromáticas, nitrosaminas y grasas poliinsaturadas) produciendo mutaciones irreversibles.

La aflatoxina (aislada de alimentos contaminados con un tipo de hongo) se considera oncogénica para la célula hepática. Se atribuyen efectos genotóxicos a los compuestos policlorados contenidos en insecticidas y plaguicidas, así como también productos de la manufactura de materiales eléctricos y plásticos formando parte de los contaminantes ambientales, que llegan a los seres vivos a través del aire, del agua y de los alimentos. (Leon, J. y col, 1988)

Los carcinógenos virales actúan introduciendo sus propias **oncoproteínas** al genoma de la célula afectada con lo que la misma cambiará su código normal, por el que le imponen los oncogenes virales. Tal es el caso del **papiloma virus humano, del Epstein Bar y de las hepatitis B y C.**

Los oncogenes virales se ubican generalmente en las proximidades de *proto-oncogenes* o de *oncogenes supresores*, activando a los primeros y desactivando a los segundos. (Gallo y col, 1977; Green y col, 1977; I,A y col, 1996; Teramoto y col, 1996)

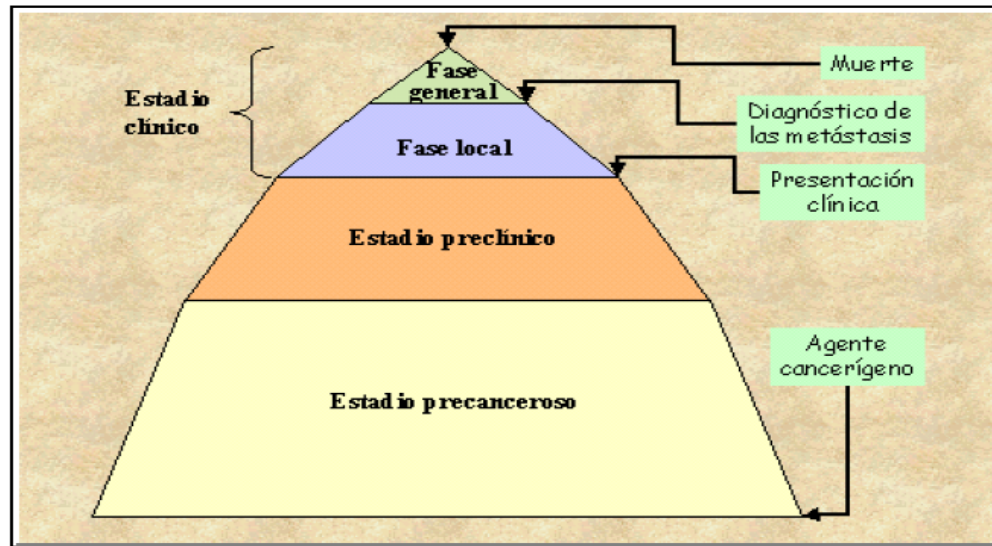


Figura 3. Representación esquemática de la evolución espontánea de un cáncer estándar



Promoción:

Representa la etapa de crecimiento tisular con la formación del tumor. Participan: los factores de crecimiento y los receptores a los factores de crecimiento, como así también la angiogénesis y degradación de las matrices extracelulares.

Los factores de crecimiento (FC), son péptidos producidos por las mismas células o por las vecinas y actúan como **facilitadores** de la mitosis incorporando en fase **S**, a algunas células que se encuentran en fase **G0** o **G1** prolongada. (Colville y col, 1997; Halachmi y col, 1994; Khan y col, 1994)

Los FC se sintetizan en una célula y luego migran al espacio intercelular, ejerciendo sus acciones sobre las células vecinas. (Liotta y col, 1988; Maciag, 1990; Masakazu y col, 1996) Los primeros FC descubiertos fueron el de crecimiento neuronal (**NGF**) y el epidérmico (**EGF**), a los que se sumaron muchos más, entre ellos el derivado de plaquetas (**PDGF**), el de hepatocitos (**HGF**), el de crecimiento de fibroblastos (**FGF**) el estimulante de crecimiento de colonias, el símil insulina (**IGF-1**).

Algunas hormonas ejercen acciones similares a estos factores peptídicos una vez que fueron captadas por los receptores de membrana o intracitoplasmáticos. Es reconocido el efecto proliferativo de los estrógenos sobre los epitelios mamarios y del tracto genital; las gonadotropinas hipofisarias, estimulan especialmente al epitelio ovárico; la prolactina ejerce su acción en el ámbito de la mama y también del ovario; la insulina de origen pancreático y el factor símil insulina, de origen hepático, son verdaderos factores de crecimiento. (Halachmi y col, 1994; Khan y col, 1994)

Merece un comentario aparte la acción de los **estrógenos**, tan vinculados a cánceres hormonodependientes y su uso como terapia hormonal de reemplazo, a originando controversias al respecto.

- a) estimulan algunos proto-oncogenes como el FOS
- b) estimulan la acción de otros factores de crecimiento (FC) y los receptores para FC.
- c) Facilitan la síntesis y liberación de prolactina
- d) Estimulan la síntesis de receptores de PG
- e) Aumentan el AMP cíclico que participa en la transducción de señales activando la replicación celular.
- f) Eleva los niveles de **cíclicas** (especialmente la E) posiblemente por mecanismos indirectos.(vía estimulación de proto-oncogenes por el AMPc responsive elements)

Los receptores de membrana, son compuestos gluco-proteicos, que se unen a los factores de crecimiento y transmiten los mensajes proliferativos por intermedio de sus conexiones transmembrana. Algunas veces, la sobre expresión de éstos receptores los hace **autoinducibles** es decir, que se encuentran en acción permanente aún en ausencia del factor de crecimiento.

Algunas **citocinas**, que son productos de distintos tipos de células, pueden ejercer efectos moduladores o inhibitorios de la proliferación; tal es el caso del factor **TGF beta, del interferón y del TNF o factor de necrosis tumoral** que antagonizan a los factores de crecimiento.

Progresión:

Implica la capacidad de **invadir** tejidos vecinos o a distancia, por parte de la célula tumoral maligna. Esa capacidad está codificada también en los genes de la misma con modificaciones estructurales y funcionales. (Robbins y col, 1995)

Las células normales, se encuentran “ancladas” en un habitat que le es propio. El contacto con las células vecinas controla su propia división celular y existen moléculas de adhesión que las mantienen próximas y permiten la transmisión de señales de una a otra; las células normales son incapaces de atravesar la membrana basal que las separa del tejido conjuntivo sub-basal de donde obtiene los materiales que la nutren; tampoco tienen capacidad de introducirse a los capilares sanguíneos o linfáticos, aunque los linfocitos, hacen excepción a esta particularidad.

En algunas ocasiones tras estas primeras etapas en el proceso carcinogénico es posible objetivar lesiones tisulares (evidenciables microscópicamente) denominadas precancerosas (situaciones patológicas que, de manera estadísticamente significativa, preceden o favorecen la aparición de procesos tumorales). (Domínguez, 2004)

Un resumen más comprensible se muestra en la Figura 4.

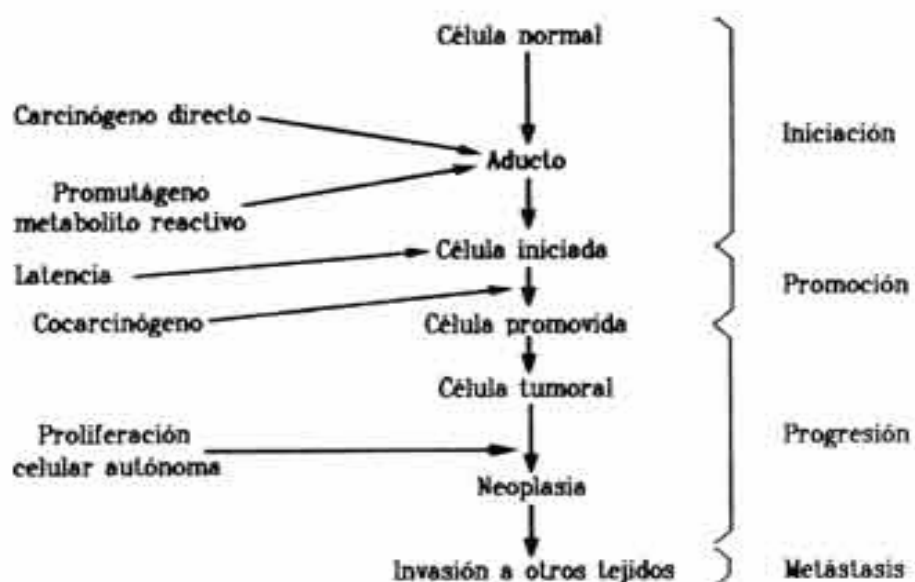


Figura 4. Resumen del proceso canceroso

2.4 Tipos de Carcinógenos

Algunos compuestos químicos de acción carcinogénica son genotóxicos, es decir, actúan a través de su interacción con los ácidos nucleicos. Otros carcinógenos presentan mecanismos de acción no genéticos, u epigenéticos, entre los que son bien conocidos los efectos de plásticos implantados en el organismo, del asbesto que destruye a los lisosomas, y de los medicamentos inmunosupresores como la azatropina, que actúan como promotores (Figura 5).

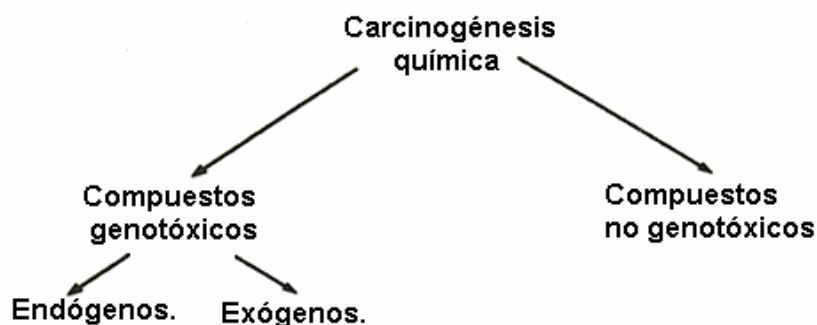


Figura 5. Carcinógenos químicos.

2.4.1 Carcinógenos genotóxicos:

Algunos agentes carcinógenos, en especial los agentes iniciadores y progresores se caracterizan por su capacidad de alterar la estructura del ADN y/o de los cromosomas. Estos efectos genotóxicos inducen directamente la aparición de células neoplásicas (transformadas o malignas). Tales efectos genotóxicos podríamos sumarizarlos en mutaciones, formación de aductos y aberraciones cromosómicas. Sin embargo, en la mayor parte de los casos, la acción carcinógena de estos agentes consiste en un aumento del potencial oxidativo de las células lo cual resulta en modificaciones en el ADN (oxidación del ADN) o formación de uniones covalentes de los agentes o sus metabolitos a las cadenas de ADN (aductos). En la acción de este tipo de sustancias juega un papel fundamental el metabolismo celular, a través del cual se produce la biotransformación de sustancias en principio inocuas a compuestos (generalmente reactivos) que presentan capacidad genotóxica y que son llamados (carcinógenos finales).

En todo caso, la acción de un agente carcinogénico debe acompañarse, para que ésta sea efectiva, de un desbalance en los mecanismos de reparación de ADN.

2.4.1.1 Carcinógenos Endógenos

Los mecanismos implicados en la carcinogénesis endógena son: oxidación por especies reactivas del oxígeno, reducción con antioxidantes, reacción con radicales libres e inhibición en la reparación de la oxidación del ADN. Los carcinógenos endógenos son especies reactivas del oxígeno. Las más importantes son los radicales hidroxilo ($\text{OH}\cdot$) y el oxígeno singlete (O_2), destacando también el anión superóxido (O_2^-), peróxido de hidrógeno (H_2O_2), especies peroxiladas ($\text{RO}\cdot$) y especies alcoxiladas ($\text{RO}\cdot$). Las especies reactivas del oxígeno se producen a partir de las reacciones celulares como la respiración celular (transporte electrónico mitocondrial), procesos de síntesis y degradación del metabolismo (metabolismo del ácido araquidónico, β -oxidación de los ácidos grasos de alto peso molecular, oxidación de aminoácidos, síntesis del ácido ascórbico, oxidación de poliaminas, esteroidogénesis, oxidación de purinas), biotransformación de xenobióticos (transporte electrónico microsomal, oxidación peroxidativa, funciones de las oxidasas) y activación de células fagocíticas (leucocitos periféricos, macrófagos, células de Kupffer del hígado y células claras del pulmón).

La oxidación endógena del ADN puede ser nuclear o mitocondrial. La oxidación del ADN mitocondrial es mucho más elevada debido a una mayor producción del anión superóxido durante la respiración mitocondrial. La mayor parte de los metabolitos responsables de la oxidación del ADN celular son el anión superóxido, el peróxido de hidrógeno y sus derivados. Las células poseen mecanismos eficaces de protección como la glutatión peroxidasa y la superóxido dismutasa, las cuales inactivan estos compuestos o antioxidantes específicos (metabolitos hidrofóbicos como el α -tocoferol e hidrofílicos como el ascorbato, ciertas proteínas, etc) los cuales disminuyen la cantidad de bases oxidadas del ADN.

También se pueden encontrar productos de gran actividad antioxidante en los vegetales y frutas. Sin embargo, no todos los productos antioxidantes producen efectos beneficiosos.

2.4.1.2 Carcinógenos Exógenos.

Los carcinógenos exógenos son aquellos que incrementan la oxidación del ADN como, por ejemplo, agentes de proliferación de peroxisomas, benceno, arsénico, estradiol, nitrosaminas, bromuro de potasio, radiaciones ultravioletas (UVA y UVB) e ionizantes (rayos X). Muchas sustancias inorgánicas, particularmente hierro, cromo, cobalto (II) y sales de níquel en presencia de H_2O_2 producen la oxidación de las bases del ADN. Por otro lado, la exposición a la polución de los ambientes urbanos produce también altos niveles de ADN oxidado. La exposición a altos niveles de agentes contaminantes como el ozono, materia particulada, aldehídos, metales y óxidos de nitrógeno incrementan también los niveles de 8-oxo-dG. La luz UVB (290-320 nm) produce mutaciones en el ADN y como consecuencia tumores de piel. La luz UVA (320-400 nm) es menos carcinógena y produce mayormente oxidación del ADN mediante la activación fotodinámica de especies reactivas del oxígeno. Hay compuestos químicos que pueden ser afectados por la radiación UV y producir la oxidación del ADN, principalmente mediante la generación de especies reactivas del oxígeno. Por ejemplo, el azul de metileno produce oxígeno singlete en presencia de UV. El oxígeno singlete formado reacciona con residuos de guanina produciéndose 8-oxo-dG. Otro ejemplo son ciertos antibióticos (tetraciclina, fluoroquinolonas) con capacidad fototóxica.

2.4.2 Carcinógenos no genotóxicos o epigenéticos.

Son aquellos compuestos químicos que actúan por mecanismos que no incluyen la modificación directa del ADN, dando lugar finalmente a células genéticamente inestables (tumor). Estos agentes parece ser que modulan el crecimiento y la muerte celular. En general, estos compuestos actúan modificando la fisiología normal de órganos y sistemas específicos produciéndose una sobreestimulación persistente cuyo resultado es una replicación celular intensificada (alteración del ciclo celular con efecto mitogénico). Esto puede dar lugar a un incremento de mutaciones espontáneas y de las probabilidades de alterar el ADN, tanto por factores endógenos como exógenos antes de que haya posibilidad de reparación. Normalmente se trata de compuestos exógenos, aunque en determinadas circunstancias compuestos endógenos (hormonas) podrían considerarse como carcinógenos epigenéticos. La acción carcinógena de los compuestos puede tener diferentes mecanismos pero todos ellos comparten las siguientes características principales:

✓ **Especificidad:**

Los compuestos epigenéticos, al contrario que los genotóxicos, pueden ser más específicos en su capacidad de inducir carcinogénesis ya que frecuentemente inducen la formación de tumores en una especie animal, un sexo determinado y en la mayor parte de los casos en uno o varios órganos determinados dentro de una especie. Esta especificidad puede ser explicada por diferencias fisiológicas, metabólicas y de sensibilidad inter-especies.

✓ **Existencia de un umbral en el desarrollo del tumor:**

En la mayoría de los casos el efecto carcinógeno se produce solamente cuando se administran altas dosis de los compuestos por lo que la carcinogénesis no aparecerá hasta que se alcance un determinado umbral. Según estos datos, se pueden construir curvas de dosis-respuesta para correlacionar qué dosis son perjudiciales. El análisis de estas curvas dosis-respuesta son de especial utilidad para determinar a qué niveles de un compuesto determinado no se produce efecto adverso y cuales constituyen factores de riesgo para el desarrollo del tumor en humanos.

✓ **Reversibilidad:**

Los carcinógenos epigenéticos actúan generalmente como promotores del tumor cuando son administrados continua y prolongadamente. Los efectos producidos pueden revertir parcialmente cuando se interrumpe la administración del compuesto.

✓ **Citotoxicidad:**

Los agentes epigenéticos son citotóxicos, produciendo un perjuicio crónico en las células que resulta en un aumento en la proliferación celular. Este incremento en la proliferación celular puede ser responsable del desarrollo neoplásico, ya que el ADN es cada vez más sensible a mutaciones a lo largo de sucesivas divisiones celulares. Por otro lado, la modificación producida en el ADN ya sea de forma endógena o exógena tiene posibilidades muy altas de convertirse en mutaciones heredables puesto que las posibilidades de reparación disminuyen.

En general, los agentes epigenéticos se pueden considerar como promotores en la expansión de células espontáneamente iniciadas. Algunos de estos agentes químicos son el benceno, cloroformo, tricloroetileno, furfural, metapirileno, lindano y bifenilos policlorinados. Un ejemplo clásico de carcinogénesis epigenética es la aparición de hepatomas o hepatocarcinomas inducidos en modelos. (Domínguez, 2004)

2.4.3 Clasificación de los carcinógenos

La evidencia más relevante para asegurar que cualquier sustancia química supone un riesgo de carcinogenicidad para el hombre deriva de los estudios epidemiológicos realizados en la especie humana, pero la asunción de riesgos que supondría el uso indiscriminado de nuevas sustancias químicas sin unos previos estudios del potencial carcinogénico de esas sustancias no podría ser socialmente aceptable. A pesar de los importantes avances en el desarrollo de pruebas para detectar precozmente la actividad

carcinogénica, sin necesidad de tener que recurrir al concurso de seres vivos complejos, que exigen estudios largos, pesados y muy costosos, una parte importante de los estudios debe realizarse todavía en animales porque, tal como señala la Organización Mundial de la Salud, la única prueba definitiva de actividad carcinogénica continúa siendo el desarrollo de un tumor histológicamente evidenciable en un animal. De todos modos, no se puede afirmar con precisión absoluta que una sustancia que ha sido comprobada como carcinogénica en animales lo vaya a ser también en los humanos; sin embargo, en la mayoría de los casos, los carcinógenos comprobados en los humanos también lo son al menos en una especie animal y, a menudo, en varias.

Tabla 2. Clasificación de las sustancias químicas cancerígenas (IARC)

Grupo	Descripción	Referencia a la carcinogenicidad en los animales
1	La sustancia es cancerígena para el hombre (o las condiciones de la exposición conllevan exposiciones cancerígenas para el hombre)	Aunque no existan pruebas suficientes de carcinogenicidad en humanos, excepcionalmente se puede incluir en esta categoría a una sustancia si las pruebas son suficientes en animales y los mecanismos implicados son significativos para el hombre
2A	La sustancia es posiblemente cancerígena para el hombre (o las condiciones de la exposición conllevan exposiciones posiblemente cancerígenas para el hombre)	Cuando existen pruebas limitadas de carcinogenicidad en humanos y pruebas suficientes en experimentación animal En algunos casos, se pueden incluir sustancias de las que existen pruebas inadecuadas de carcinogenicidad en humanos y pruebas suficientes en animales (siempre que haya evidencias claras de que en la patogenia están implicados mecanismos significativos para el hombre)
2B	La sustancia es posiblemente cancerígena para el hombre (o las condiciones de la exposición conllevan exposiciones posiblemente cancerígenas para el hombre)	Cuando existen pruebas limitadas de carcinogenicidad en humanos y pruebas insuficientes en experimentación animal En algunos casos, se pueden incluir sustancias de las que existen pruebas inadecuadas de carcinogenicidad en humanos y pruebas suficientes en animales. Ocasionalmente, si existen pruebas inadecuadas de carcinogenicidad en humanos, pruebas limitadas en animales y otros datos significativos que lo apoyen.
3	Los datos no permiten que la sustancia pueda ser clasificada respecto a su carcinogenicidad para el hombre	Cuando existen pruebas inadecuadas de carcinogenicidad en humanos y pruebas inadecuadas o limitadas en experimentación animal Excepcionalmente, si hay pruebas inadecuadas en humanos y suficientes en animales pero existen evidencias claras de que el mecanismo de carcinogenicidad en animales no funciona en humanos.

En este sentido, se puede asegurar que existe una clara correlación positiva entre las observaciones realizadas en el hombre y los índices de carcinogenicidad en los animales, que además, a menudo han precedido a las observaciones humanas. Por ello, la observación de carcinogenicidad de una determinada sustancia en una especie animal debería, al menos, ser interpretada como una señal de atención para estudiar la adopción de medidas preventivas.

La utilización de los animales en la evaluación de la carcinogenicidad de las sustancias químicas resulta obligatoria ya que la misma clasificación de las sustancias carcinogénicas, se basa, entre otros criterios, en la existencia o no de suficiente evidencia de carcinogenicidad en animales. A este respecto no citaremos aquí más que la que probablemente es la de mayor prestigio internacional actualmente, la de la Agencia Internacional para la Investigación sobre el Cáncer (International Agency for Research on Cancer, IARC) (Tabla 1), en la que (aunque domina el criterio de las pruebas en humano) la existencia de pruebas en los animales permite ubicar (con carácter habitual, ocasional o excepcional) a las sustancias en los distintos grupos de riesgo carcinogénico. (Pitot III y Dragan, 1996; Hasmall y Roberts, 1997; García y col, 2003; Domínguez, 2004)

Además, los productos químicos pueden ser clasificados en una de las cinco categorías de toxicidad basadas en la toxicidad aguda por vía oral, cutánea o por inhalación, con arreglo a los criterios numéricos expresados en valores aproximados de la DL50 (oral, cutánea) o CL50 (inhalación), tal como se indica en el anexo 1, al que acompañan unas notas explicativas.

2.5 Los Agentes Genotóxicos y el Daño Genético.

Durante el proceso de mutagenicidad el agente xenobiótico ingresa al organismo, se absorbe, se distribuye y atraviesa las membranas. Una vez dentro de la célula, el agente químico puede ser reactivo por sí mismo (de acción directa), o bien puede ser activado por las enzimas metabólicas, en cuyo caso es de acción indirecta y se llama promutágeno. Se produce entonces la interacción con el ADN, que puede ser reparada eficiente o ineficientemente de manera tal que el daño genético inicial se fijará o no, expresándose en las diferentes estirpes celulares, tal como se muestra en el esquema de la figura 6.

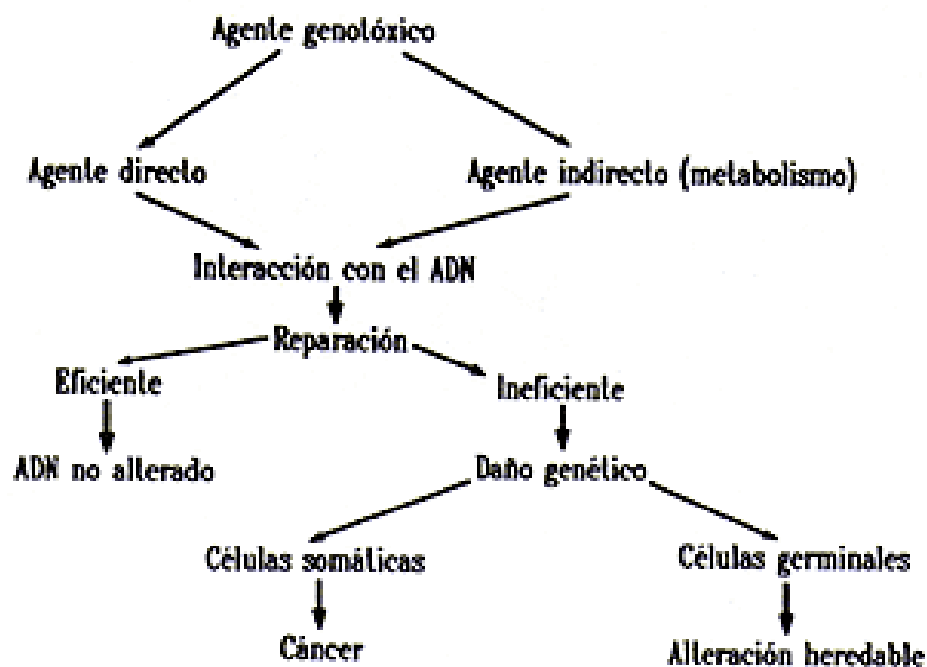


Figura 6. Los agentes genotóxicos y el daño genético inducido.

2.5.1 Metabolismo de los agentes genotóxicos

En realidad, la gran mayoría de los agentes genotóxicos son inertes en los seres vivos. Es a través de las enzimas metabólicas que las genotoxinas son biotransformadas a productos más reactivos, o electrofílicos, capaces de interactuar con diversas macromoléculas celulares, tales como las proteínas y los ácidos nucleicos. Los organismos procariontes son incapaces de bioactivar promutágenos, y entre los eucariontes existen diferencias importantes en cuanto a la capacidad metabólica; se conoce que cada especie desplegó durante la evolución un grupo particular de enzimas para neutralizar los efectos nocivos de las toxinas naturales de origen vegetal. De hecho, las enzimas metabólicas muestran diferencias considerables en los diferentes órganos del individuo, entre los individuos de la misma especie y entre las diferentes especies. La actividad enzimática varía en el individuo dependiendo de la edad, el sexo, factores nutricionales, niveles hormonales y otros factores biológicos.

En principio, el conjunto de enzimas de los eucariontes hidroliza, oxida y reduce compuestos extraños, reacciones que se llevan a cabo en el sistema enzimático del citocromo P-450 que se encuentra en el citoesqueleto y en el retículo endoplásmico de las células con núcleo. Los productos intermedios así generados en ocasiones se conjugan con proteínas, formándose compuestos altamente reactivos. Es decir,

en las células existen numerosas enzimas que activan a los promutágenos, pero también otras enzimas que desintoxican e inactivan a los productos intermedios: el equilibrio entre estas dos funciones celulares es el que en última instancia determina el potencial genotóxico del promutágeno (compuesto químico inerte que requiere ser metabolizado, transformándose así en un compuesto electrofílico y por lo tanto reactivo).

En la figura 7 se muestra la activación inicial de algunos promutágenos. Muchos de ellos pasan por diversos procesos metabólicos, generándose varios productos intermedios. El compuesto electrofílico terminal es el que va a interactuar con los átomos nucleofílicos del ADN (los sitios nucleofílicos de las bases nitrogenadas son los centros que pueden ser atacados por moléculas electrofílicas, por ejemplo el nitrógeno 7 y el oxígeno 6 de la guanina).

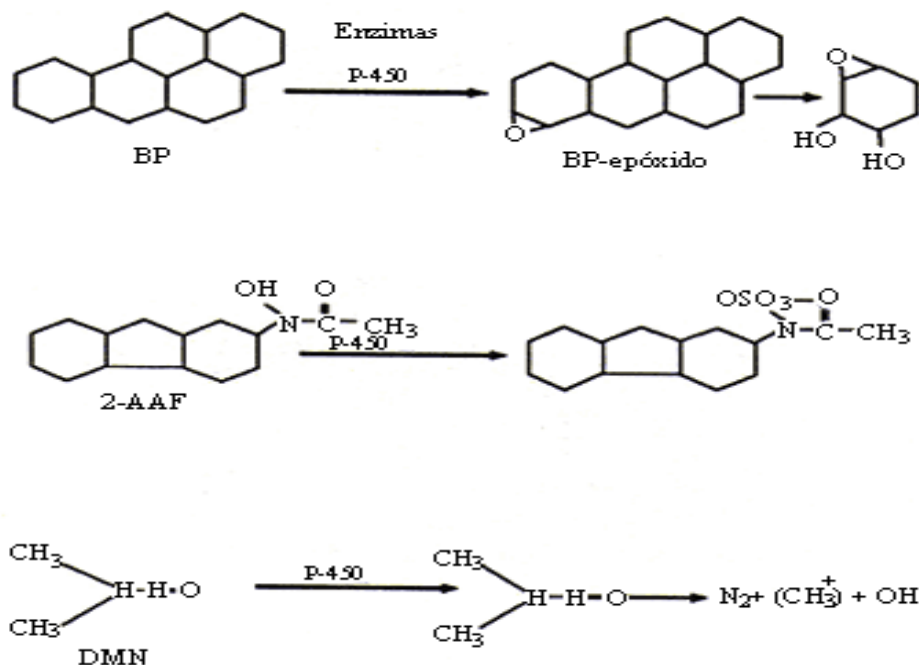


Figura 7. Activación metabólica de algunos promutágenos.

Es importante mencionar que existen también compuestos químicos que no son carcinógenos, pero que potencian el efecto de carcinógenos. Estos agentes químicos se llaman cocarcinógenos y suelen actuar en la etapa de promoción tumoral.

2.5.2 La inhibición del metabolismo

Durante el metabolismo normal de las células se generan radicales libres que suelen ser muy reactivos y, por lo tanto, potencialmente muy dañinos. Los organismos han desarrollado mecanismos que,

por cierto, están muy conservados evolutivamente, para atrapar a los radicales libres. Entre estos mecanismos están diversas enzimas que catalizan la conversión de oxígeno reducido (O_2^-) a peróxido de hidrógeno (H_2O_2) y de éste a agua y oxígeno ($H_2O + O_2$), y otros como el glutatión, que reacciona directamente con los compuestos electrofílicos de acción directa, o con los producidos durante el metabolismo.

Así tenemos que los radicales libres se forman como productos intermedios en los procesos bioquímicos naturales. Se piensa que las enfermedades degenerativas como la arterioesclerosis, el cáncer y el envejecimiento celular se deben en gran medida a la pérdida de la capacidad enzimática de las células para atrapar radicales libres.

En los alimentos que ingerimos normalmente existen mutágenos y antimutágenos, y durante el metabolismo se generan compuestos mutagénicos, como las nitrosaminas, que se producen en el estómago al reaccionar los nitritos que se emplean como aditivos de alimentos con las aminas presentes en la carne. Es un hecho conocido que la dieta y los hábitos diarios de la persona influyen notablemente en el tipo de cáncer que los individuos desarrollan. Evidencias experimentales han mostrado que la ingesta diaria de vitaminas como la A, C, E, y los betacarotenos, que son cofactores que atrapan radicales libres protegen a los individuos en contra de los efectos nocivos de los radicales libres. Los mecanismos de acción de estas vitaminas son variados, el tocoferol o vitamina E puede interferir durante la formación de nitrosaminas, atrapa-radicales libres, al igual que la vitamina C y los beta-carotenos, y la vitamina A suprime la fase de promoción tumoral.

2.5.3 Interacciones con el ADN

Los productos reactivos generados a través del metabolismo interactúan con el ácido desoxirribonucleico, produciéndose lesiones premutagénicas, o aductos, que en muchos casos se fijan y producen mutaciones puntuales en el ADN, tales como sustituciones de bases, transiciones y transversiones, o bien mutaciones de corrimiento del marco de lectura.

Sin embargo, en muchos casos, las lesiones premutagénicas son eficientemente reparadas por enzimas que funcionan en los organismos para mantener la integridad y fidelidad de los ácidos nucleicos. Se piensa que las enzimas que intervienen en los procesos de reparación aparecieron pronto en la evolución, ya que están presentes en las bacterias.

Los mecanismos de reparación pueden funcionar antes o después de la replicación del ADN. Su eficiencia varía, ya que pueden reparar eficientemente, es decir, sin errores, situación que se presenta

cuando la exposición a agentes genotóxicos es baja; o bien reparar de manera ineficiente, promoviendo errores en el ADN, lo que depende de la saturación del primer mecanismo y que generalmente ocurre cuando hay exposiciones altas (Figura 8).

Sin embargo, ambos mecanismos se ven afectados por numerosas variables además de la exposición. Dependen también de la estructura química del mutágeno, del tipo de aducto formado y de la cantidad de daño inducido.

Una vez establecidos estos principios generales de interacción de los agentes genotóxicos con las macromoléculas celulares, analizaremos los tipos de agentes tóxicos y los efectos biológicos y genéticos que producen en los seres vivos. (Rodríguez, 1995)

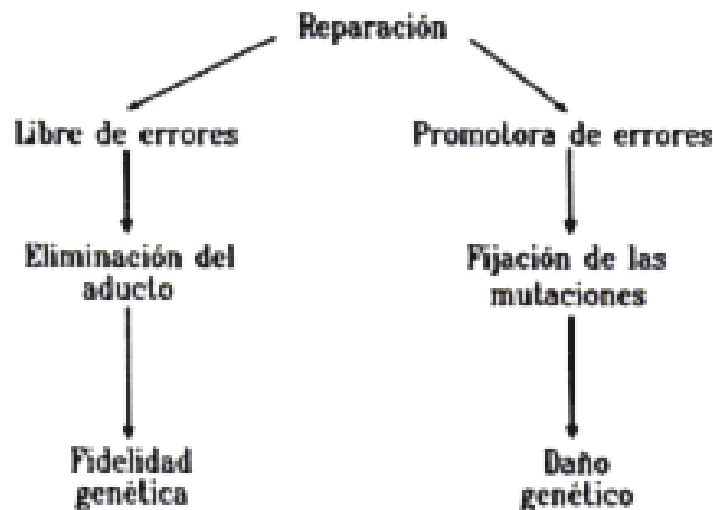


Figura 8. Mecanismos de reparación del ADN.

2.6 Métodos para Determinar la Acción Carcinogénica.

2.6.1 Determinación de carcinógenos genotóxicos:

Existe un alto nivel de redundancia en los tipos de ensayos incluidos en las evaluaciones genotoxicológicas por lo que los mismos se han agrupado en 4 clases que miden:

1. Mutación génica.
2. Aberraciones cromosómicas.

3. Daño primario al ADN (mide la respuesta secundaria producida como resultado de una lesión primaria, tales como estimulación de la reparación del ADN , recombinación somática entre homólogos o cromátidas hermanas, o la ruptura de cadena del ADN).

4. Transformación celular (no evalúa efectos directos sobre el ADN sino la habilidad de la sustancia de ensayo de transformar las células normales en neoplásicas, esta categoría de daño es exclusivamente una medida del potencial oncogénico sin ninguna o poca relevancia para la producción de daños heredables.

Existen numerosos ensayos para evaluar los posibles efectos genotóxicos de las sustancias. Estos se evalúan a su vez en múltiples biomodelos o sistemas biológicos, desde bacterias, hongos y levaduras y plantas hasta células aisladas y mamíferos íntegros. Pero para poder realizar estudios evaluativos más homogéneos y predictores de acciones genotóxicas se han realizado numerosos intentos y se han propuesto baterías de varios ensayos en orden lógico y en las que se incluyan las evaluaciones de las principales categorías de daño genético. Actualmente se realizan esfuerzos mancomunados para armonizar las guías para la evaluación genotoxicológica de las sustancias químicas y dentro de las Conferencias Internacionales de Harmonización (ICH) se ha propuesto un grupo de ensayos que son legislado por EUA, UE y Japón.

La batería propuesta está conformada por cuatro ensayos básicos:

1- Ensayo de detección de mutación génica en bacterias. Ensayo de mutación reversa en *Salmonella typhimurium*/E. Coli.

Este ensayo es el más ampliamente utilizado para propósitos de screening de mutágenos y carcinógenos (es el único validado). Combina una alta sensibilidad con una relativa facilidad técnica, rapidez y economía.

En este ensayo se utilizan diferentes cepas mutantes auxotróficas (incapaces de sintetizar histidina) de *Salmonella typhimurium*. Cada una de estas cepas tiene diferentes mutaciones que desactiva el gen que codifica para la enzima requerida en la síntesis de este aminoácido vital. De manera que no pueden crecer en un cultivo a no ser que el medio esté suplementado con este aminoácido. Si el gen afectado es mutado se produce una reversión al estado salvaje u original y entonces la bacteria será capaz de crecer en ausencia del aminoácido este fenómeno es conocido como reversión y las colonias como revertantes. Además de esa mutación en el operón histidina, presentan otras mutaciones que incrementan su sensibilidad a los mutágenos, ej:

- Mutación *rfa*, causa pérdida parcial de los lipopolisacáridos de la pared celular bacteriana, incrementándose la permeabilidad a macromoléculas.

- Mutación *uvrB*, incrementa la sensibilidad a los mutágenos por delección de los genes que codifican para los mecanismos de reparación por excisión.
- Plásmido pKM101 (factor R) causa un incremento en los mecanismos de reparación propensos a error aumentando la sensibilidad, también confiere resistencia al Ampicillin lo que es un buen marcador para detectar la presencia del plásmido.

Dos tipos de mutaciones son detectables por este ensayo: sustitución de pares de bases y corrimiento en el marco de lectura.

Esas líneas detectan cambios en los sitios guanina-citosina de los genes histidina, sin embargo, algunos carcinógenos también modifican los pares de bases adenina-timina, por lo que más recientemente se han introducido líneas que detecten mutaciones en estos sitios, ej. La cepa TA 102 de *Salmonella typhimurium* y *E. coli* WP2uvrA la cual detecta esta mutación en el gen triptófano.

2- Ensayo *in vitro* de Linfoma L5178Y tk +/- de ratón.

Este ensayo evalúa dual punto final de acción genotóxica (mutación génica y aberración cromosómica) ya que es capaz de detectar mutaciones puntuales que involucran sustituciones de base, delecciones, corrimientos y reordenamientos dentro del *locus TK* y además detecta lesiones clastogénicas que involucran delecciones multilocus y múltiples genes. Las células de mamífero generalmente tienen 2 copias de cada gen, pero esta línea tiene solamente una copia funcional del gen (células heterocigóticas para el locus timidina-kinasa derivadas de linfoma de ratón). La timidina-kinasa no es una enzima esencial, la misma incorpora timidina exógena por fosforilación de la timidina. La trifluorotimidina (TFT) también puede ser fosforilada por la enzima y por consiguiente las células que contienen la enzima son sensibles a los efectos citotóxicos y citostáticos de la TFT. Las células son expuestas al compuesto con y sin activación metabólica e incubadas por un período de tiempo que permita la expresión de cualquier mutación que conlleve a la transformación a células homocigóticas tk^{-/-} (enzima infuncional), las cuales son resistente a la TFT y permanecen viables en presencia de esta sustancia.

Se evalúan varias concentraciones, al menos 4 niveles de dosis, la máxima debe reducir la sobrevivencia en un 80 a un 90%, y para compuestos no tóxicos esta debe ser el estándar límite de 5000 µg/mL.

Se considera un resultado positivo si hay un incremento estadísticamente significativo en la frecuencia de mutaciones.

3- Ensayo citogenético *in vitro* con células de ovarios de hámster (CHO).

Estos ensayos son utilizados para confirmar si un mutágeno presuntivo es mutagénico para animales mayores. Las células de mamífero presentan un mayor grado de organización que las bacterias y su capacidad metabólica, y de reparación del ADN también es mucho más compleja. Una gran variedad

de líneas celulares y sistemas han sido utilizados para estudiar la actividad mutagénica, sin embargo la mayoría de los estudios se llevan a cabo en los sistemas siguientes:

- Líneas permanentes o inmortales: células de ovario de hámster (CHO), células de pulmón de hámster (V79) y células de linfoma de ratón (L5178Y).
- Cultivos primarios de linfocitos.

En contraste con los ensayos de mutación génica los ensayos citogenéticos se utilizan para determinar los cambios inducidos en la estructura o número de cromosomas, lo que es visto a través de un microscopio. Existen muchas técnicas para las evaluaciones citogenéticas tanto *in vitro* como *in vivo*, que se dividen en tres tipos diferentes:

- Las que detectan aberraciones cromosómicas.
- Las que detectan intercambios entre cromátidas hermanas.
- Las que detectan micronúcleos.

Los métodos se basan en la exposición de células proliferativas (en división) al agente de ensayo durante un tiempo que garantice que las células estén expuestas durante todas las etapas del ciclo celular, entonces son tratadas con un veneno del huso (colchicina) para acumular las células en metafase y que permita su visualización microscópica.

Los ensayos citogenéticos *in vitro* utilizan líneas celulares permanentes (inmortales) usualmente derivadas de Hámster Chino, ej. Células de ovario y de pulmón; y cultivos primarios, generalmente de origen humano como linfocitos humanos.

4- Ensayo *in vivo* de inducción de eritrocitos policromatófilos micronucleados en médula ósea de ratón.

2.6.2 Bioensayo *in vivo*. Determinación de carcinogenicidad en animales.

Objetivo. El objetivo de los ensayos de larga duración de carcinogenicidad es observar animales de ensayo por la mayor porción de su tiempo de vida para evaluar si desarrollan lesiones neoplásicas durante o después de exposición a varias dosis de una sustancia de ensayo y por una apropiada ruta de exposición.

Procedimiento. Los estudios deben ser desarrollados en dos especies de mamíferos, generalmente en ratas y ratones a causa de su período de vida relativamente corto, costos de mantención limitados, uso generalizado en estudios farmacológicos y toxicológicos, su inducción de susceptibilidad a tumores y la disponibilidad de cepas isogénicas y suficientemente caracterizadas. El estudio se inicia cuando estos animales son adultos jóvenes, en número por sexo equivalente, se utilizan al menos 100 roedores (50 de cada sexo) en cada nivel de dosis y un grupo control de vehículo o no tratado. Si se planifican sacrificios

intermedios durante el estudio el número de animales deberá ser aun mayor. Se mantienen bajo condiciones climáticas e iluminación estandarizadas y homogéneas; y alimentados bajo regímenes a voluntad y con componentes alimenticios controlados para evitar influencia de impurezas y contaminantes en estos. Se utilizan al menos tres niveles de dosis previamente seleccionadas y de forma racional, estas deben ser espaciadas para producir una gradación de los efectos. Las vías de administración más frecuentes son oral, dérmica e inhalatoria. Esta dependerá de la naturaleza físico-química de la sustancia de ensayo y del futuro uso en humano. Los animales se administran sobre la base de 7 días a la semana, aunque 5 días son aceptables. Los animales en ensayo son observados al menos una vez al día recogiendo todos los síntomas y alteraciones que muestren. Igualmente se evalúa la evolución del peso corporal y el consumo de agua y alimentos. Se realizan pruebas y análisis de patología clínica que incluyen determinaciones hematológicas y hemoquímicas; también screening inmunotoxicológicos y un análisis total de patología macroscópica e histopatología de todos los órganos, estructuras y sistemas de los animales en ensayo. Todos los hallazgos del estudio se analizan en conjunto con los de estudios previos. Dicha evaluación incluye la relación entre la dosis de la sustancia de ensayo y la presencia, incidencia y severidad de las anomalías (incluyendo las conductuales y clínicas), lesiones macroscópicas, órganos diana identificados, cambios en el peso corporal, efectos en la mortalidad y cualquier otro efecto tóxico general o específico. Al finalizar el análisis global se emite un criterio de decisión y se clasifica la sustancia.

2.7 Métodos Alternativos para Valorar la Carcinogénesis.

2.7.1 Modificaciones propuestas al bioensayo clásico:

Ratones:

Los ratones han demostrado ser particularmente pobres indicadores de potenciales carcinógenos humanos. Se propone la eliminación de su uso.

Ratones alterados genéticamente:

Se utilizan varias líneas transgénicas y knockout, incluso potencialmente para propósitos regulatorios. (Goodman, 2001; Thompson y Sistare, 2003; Cannon, 2003) Los transgénicos tienen activados oncogenes introducidos en su genoma (lo que facilitan las neoplasias); y los knockout tienen inactivados o eliminados genes supresores de tumores. No han sido aún validados, pero la FDA ha aceptado datos de transgénicos como parte de la determinación de seguridad de algunos fármacos. (Combes y col, 2002)

Ej. de líneas:

Transgénicos: TgAC, TgrasH2

Knockout: heterocigótico p53, homocigótico XPA, homocigótico XPA-heterocigótico p53.

En un estudio en que se evaluaron 21 productos (genotóxicos y no genotóxicos, carcinógenos y no) con estos se concluyó que estos modelos pueden ser útiles como screening y no ofrecen una especificidad definitiva para carcinógenos humanos, ni pueden distinguir definitivamente entre genotóxicos y no genotóxicos. (Goodman, (2001).

Ratones neonatales:

Muy sensibles para carcinógenos genotóxicos pero no para epigenéticos. (McClain y col, 2001)

Reducción del período del ensayo clásico:

Se está abogando por varios investigadores que el período de evaluación sea reducido a 12 a 18 meses porque en ese período es cuando en la mayoría de las evaluaciones (se analizaron 210 carcinógenos) se revelan los efectos tumorigénicos.

Modelos de iniciación-promoción:

La exposición a un carcinógeno durante la iniciación tumoral puede ser continuado por una exposición prolongada a un promotor no carcinogénico, o viceversa; así se acelera el proceso de carcinogénesis y se reduce el ensayo. (Conolly y col, 1988)

Las ICH acordaron que estos modelos pueden utilizarse.

Estudios farmacocinéticos, farmacodinámicos y de corta y media duración en roedores y humanos:

Se propone que en el caso de fármacos humanos la apreciación de la estructura química, (Ashby, 1996) la farmacocinética (ADME) y la farmacodinámica (mecanismo de acción), combinado con la evaluación de corta duración de las propiedades, como la genética, toxicidad en roedores y tejido-específica, pudiera permitir predicciones confiables de potenciales carcinógenos genotóxicos y no genotóxicos. Otros plantean la utilización de ensayos combinados de genotoxicidad in vivo y estudios toxicológicos en dos especies de 3 a 6 meses y ensayos clínicos fase I y II para identificar factores de riesgo carcinogénico (Monro, 1996). En general varios autores plantean que muchas veces, si se realizan dichos estudios y se realiza una interpretación correcta de los resultados, el ensayo convencional en roedores es redundante.

2.7.2 Alternativas sin animales:

SARs y QSARs

Las relaciones estructura-actividad (SARs) predicen actividades biológicas como la carcinogénesis, basado en la presencia de sub-estructuras moleculares u otros indicadores químicos los cuales les confieren actividad biológica a la molécula original. Las relaciones cuantitativas estructura-actividad (QSARs) son descripciones matemáticas de la relación entre las propiedades físico-químicas de las moléculas y sus actividades biológicas. (Comber y col, 2003)

En 1991, Sabih y Tennant (Ashby y Tennant, 1991), al analizar los resultados de 301 químicos evaluados, demostraron que la presencia de grupos amino aromáticos y grupos nitro, agentes alquilantes, y otros grupo químicos incrementan la probabilidad de carcinogénesis en roedores. Cronin y colaboradores (Cronin, 2003) describieron un número de sub-estructuras moleculares electrofílicas comunes a un rango de toxicidades potenciales en varias especies, incluyendo mutagenicidad y carcinogenicidad. A pesar de que han ocurrido fallos iniciales, los datos de QSAR más recientes han demostrado ser muy útiles para predecir la carcinogenicidad de compuestos. Matthews y Contrera (Matthews y Contrera, 1998) describieron la evaluación de un sistema computarizado QSAR que demostró un 97% de sensibilidad para carcinógenos roedores y 98% de especificidad para no carcinógenos. Estos análisis tienen la marcada ventaja de ser muy rápidos y relativamente baratos.

Ensayos in vitro:

Los ensayos *in vitro*, incluyendo bacterias, levaduras, protozoos, cultivos celulares de células de mamíferos y humanos, pueden contribuir junto con una caracterización del “peso de la evidencia” lo cual es suficiente para hacer innecesaria el bioensayo en roedores. Brusick (Brusick, 1977) encontró una correlación de aproximadamente 90% entre mutagénesis microbiana *in vitro* y propiedades carcinogénicas en mamíferos para una larga lista de químicos. Tennant y colaboradores (Tennant y col, 1990) predijeron exitosamente los hallazgos de 86% de 44 químicos a los que se evaluó carcinogénesis por el Programa Nacional de Toxicología (NTP), por medio del uso de la mutagenicidad en *Salmonella*, ensayos de toxicidad sub-aguda en roedores (90 días) combinado con información estructural química. Los ensayos de genotoxicidad de Ames, mutación reversa en *Samonella typhimurium* y aberraciones cromosómicas han sido aceptados por agencias regulatorias. (Dearfield, 1991)

Ensayos de transformación celular

Generalidades de transformación celular

Estos detectan cambios morfológicos que orientan a signos tempranos de identificación fenotípica de carcinogénesis.

El fenómeno de la transformación morfológica de la célula incluye cambios en el control del comportamiento y crecimiento de las células cultivadas, caracterizadas por uno o mas de las siguientes características (dependiendo del sistema celular):

- Alteraciones en la morfología celular.
- Perfiles desorganizados del crecimiento de las colonias.
- Adquisición de “crecimiento independiente de anclaje”. (Barrett y Fletcher, 1987; Yuspa y Poirier, 1988).

Tales efectos resultan de cambios en la expresión de oncogenes y genes supresores de tumores.

La transformación ha sido definida como la inducción de ciertas alteraciones fenotípicas en células cultivadas que constituyen características tumorigénicas. (Barrett y col, 1986)

Estas alteraciones fenotípicas pueden ser inducidas por exposición de células normales a procesos de carcinogénesis o por la expresión de oncogenes activos en estas células. Las células transformadas que han adquirido todas las características de células malignas tienen la habilidad de formar tumores invasivos en animales susceptibles. Tradicionalmente esto es determinado por inyección de células transformadas en ratones atímicos o animales recién nacidos de la misma especie de la línea celular, no obstante algunos sistemas de transformación tales como el ensayo de células SHE (Syrian Hamster Embryo) ha demostrado ser altamente predictivo. (LeBoeuf y col, 1996)

Desde su descubrimiento la transformación celular in vitro ha demostrado ser un proceso de múltiple pasos los cuales modelan los diferentes estadios de la carcinogénesis in vivo (Landolph, 1985; LeBoeuf, 1999), varias razones apoyan esta expresión.

a) La inducción de transformación celular resulta de la exposición a químicos los cuales son carcinogénicos en animales.

b) Las células transformadas malignamente y que son inyectadas en animales hospederos susceptibles pueden formar tumor a diferencia de las no transformadas que no lo pueden provocar.

c) La transformación celular in vitro ocurre en varias etapas donde adquieren varios fenotipos diferentes y donde las células progresan de un fenotipo normal a uno maligno en experimentos de transfección con oncogenes activados, y por el hecho que los agentes promotores de tumores pueden aumentar la frecuencia de transformación celular en células tratadas con bajas concentraciones de carcinogenicidad. (Heidelberger y col, 1983; Müller y col, 1993)

La transformación celular in vitro parece brindar algunas evidencias específicas y cruciales para determinar el potencial tumorigénico de un químico lo que no puede ser brindado por los ensayos genotóxicos a pesar de que varios ensayos de transformación in vitro se han desarrollado su relevancia y confiabilidad para identificar carcinógenos genotóxicos no ha sido aun ampliamente aceptada.

Indudablemente para determinar este tipo de carcinogenicidad se han utilizado los ensayos de genotoxicidad (Müller y col, 1993). Existe un interés en el uso de la transformación celular para detectar carcinogenicidad por el hecho de que algunos ensayos son sensibles a químicos que actúan por un diverso rango de mecanismos, incluyendo mecanismos genotóxicos y no genotóxicos (Landolph, 1985; Landolph y col, 1996).

Existen varias razones para el relativo bajo uso de bioensayos de transformación celular para predecir carcinogenicidad, esto incluye: a) Las habilidades técnicas requeridas, b) la carencia de estandarización y caracterización y c) el hecho de que la prevalencia y fiabilidad de los ensayos no han sido establecidos en estudios independientes hasta donde ellas pueden ser consideradas para ser validadas con aceptación y uso regulatorio.

Ensayos de transformación celular en roedores:

Estos incluyen las células de periodo de vida finito tales como ensayo de células SHE, así como las líneas celulares inmortalizadas, ejemplo Balb/c3T3 y C3H/10T $\frac{1}{2}$. La transformación morfológica de las colonias celulares y formación de focos de una monocapa celular son los puntos finales mas usados.

La transformación celular puede ser ensayada simultáneamente en los mismos cultivos con otro puntos finales, ejemplo mutación génica, formación de intercambio de cromátidas hermanas, daño cromosómico, inducción de micronúcleos y aneuploidias (Landolph, 1985; Tsutsui y col, 1997; Fritzenschaf y col, 1993). Estos estudios combinados pueden ser útiles en proveer información adicional que puede ser valiosa en la interpretación de los datos sobre transformación ejemplo: en decidir cuando la actividad transformante de un químico se debe a mecanismo genotóxico o no genotóxico.

Ensayo de transformación celular en células de embriones de Hamster Sirio (SHE). Ha sido descrito como el ensayo de corta duración más predictivo para carcinógenos en roedores. (Zhang y col, 2004) Detecta transformaciones celulares morfológicas, el estadio fenotípico de carcinogénesis más tempranamente identificable. (se han demostrado correlaciones de hasta 91%. (Pienta y col, 1977) Su principal ventaja con otros ensayos in vitro es la habilidad de detectar algunos químicos genotóxicos al igual que no genotóxicos. (Mauthe y col, 2001; Amacher y Zelljadt, 1983)

El sistema de transformación basado en células humanas

Indudablemente un ensayo de transformación ideal debe ser aquel que utilice células humanas debido a que la mayoría de los ensayos de transformación desarrollados para determinar el riesgo de transformación humana a potenciales carcinogénicos. Además la validación de los ensayos de

transformación celular humana tendrá que ser desarrollada tomando como referencia los datos en animales que posean suficiente relevancia para el riesgo en humanos.

Los dos sistemas de transformación celular humana que han sido desarrollados y estudiados son:

✓ *El modelo de transformación celular en queratinocitos (HaC)*. Esta fue derivada por inmortalización espontánea d ATe queratinocitos humanos normales muy probablemente debido a mutaciones en el gen P₅₃ y la consecuente pérdida de genes sensibles (Boukamp y col, 1988; Boukamp y col, 1997). Esta célula puede ser propagada indefinidamente, tienen un cariotipo aneuploide con cambios cromosómicos específicos, pero poseen enteramente muchas propiedades fenotípicas de los queratinocitos humanos (Breitkreutz, 1998).

✓ *Líneas celulares MSU-1*. Esta línea consiste en un modelo de transformación celular en fibroblastos humanos, en esta el encogen v-myc fue transfectado en fibroblastos humanos normales, las células clonadas que expresan la potencia v-myc fueron las seleccionadas e inmortalizadas, posteriormente se obtuvo una variante de la línea celular, la msu-1.1.

Ambas líneas celulares son no tumorigénicas en ratones atímicos.

Uso no regulatorio de ensayos de transformación celular. Estos han sido utilizados en estudios mecanísticos de carcinogenicidad y para los posibles mecanismos de acción de carcinogenicidad conocidos o sospechosos, han sido ampliamente empleados a la hora de tomar decisiones tempranas en la priorización de químicos para ser evaluados en ensayos de carcinogenicidad en animales. El análisis de las bases de datos disponibles sugieren que solo el ensayo de transformación SHE ha demostrado ser el que mejor cumple con los parámetros exigidos desde el punto de vista de propósitos de escrinig carcinogénico.

Uso regulatorio de los ensayos de transformación celular. La transformación celular como un punto final aparece en algunos esquemas y estrategias de ensayos para evaluación de la mutagenicidad/carcinogenicidad de químicos (Gatehouse, 1994), no obstante tales ensayos son recomendados solamente cuando se sospecha fundamentalmente que las bases moleculares del fenotipo transformado y su relación con el cáncer in vivo no están totalmente esclarecidos, dicha problemática se ha originado a causa de la carencia de comprensión de los mecanismos fundamentales incluidos en la transformación, la subjetividad alejada de los puntos finales in vivo que se obtienen, la carencia de datos de ensayos regulatorios en transformaciones celulares y la ausencia de adecuada validación de estudios.

Otros procesos y análisis que contribuyen a la determinación de la carcinogenicidad son: cDNA microarrays (de muy reciente utilización y con promisorio futuro) investigaciones epidemiológicas y evaluación de datos históricos.

3

MÉTODO COMPUTACIONAL Y ANÁLISIS QUIMIOMÉTRICO DE LOS DATOS

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) (van de Waterbeemd y col., 1998; Todeschini y Consonni, 2000) 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 (Karelson, 1995). 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 el presente trabajo se utilizaron para la parametrización de la estructura molecular los DMs **TOMOCOMD-CARDD** no estocásticos y estocásticos basados en relaciones entre núcleos atómicos. Estos índices moleculares serán descritos a continuación en la primera porción de la primera parte de este capítulo.

En la segunda porción de la primera parte de este capítulo 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.(van de Waterbeemd, 1995).

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 (van de Waterbeemd, 1995): 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)

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 (Basak y col., 1988; Lajiness, 1990; Basak y Grunwald, 1994, 1995a,b; Basak y Gute, 1997). 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 (Grassy y col., 1998).

3.1 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 (Marrero-Ponce, 2003, 2004, Marrero-Ponce y col, 2003, 2004a,b,c,d,e,f,g,h, 2005a,b,c,d,e,f,g,h,i, 2006a,b; Meneses y col, 2005; Montero y col, 2005, 2006; Caseñola y col, 2006; Castillo-Garit y col, 2006; Celeste y col, 2006; Marrero-Ponce y Torrens, 2004).

Este método codifica la estructura molecular a través de aplicaciones matemáticas denominadas formas cuadráticas, lineales y bilineales. Con el propósito de calcular estas funciones 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 (Marrero-Ponce, 2003, 2004, Marrero-Ponce y col, 2003, 2004a,b,c,d,e,f,g,h, 2005a,b,c,d,e,f,g,h,i, 2006a,b; Meneses y col, 2005; Montero y col, 2005, 2006; Caseñola y col, 2006; Castillo-Garit y col, 2006; Celeste y col, 2006; Marrero-Ponce y Torrens, 2004) 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 las 3 principales familias de DMs utilizados en el presente trabajo tal y como describiremos a continuación.

3.1.1 Descriptores TOMOCOMD-CARDD basados en relaciones entre núcleos atómicos.

3.1.1.1 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 (Marrero-Ponce, 2003, 2004, Marrero-Ponce y col, 2003, 2004a,b,c,d,e,f,g,h, 2005a,b,c,d,e,f,g,h,i, 2006a,b; Meneses y col, 2005; Montero y col, 2005, 2006; Caseñola y col, 2006; Castillo-Garit y col, 2006; Celeste y col, 2006; Marrero-Ponce y Torrens, 2004). 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 $\bar{x} = [x_{N1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}, x_{C6}, x_{C7}, x_{O8}, x_{S9}]$ (ver estructura molecular en la Tabla 3.1). 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 (Wang y col, 2000), la contribución a la superficie polar de los átomos (Ertl y col, 2000), la refractividad molar atómica (Ghose y Crippen, 1987), las polarizabilidades atómica (P) (Millar, 1990; Kier y Hall, 1986), las cargas atómicas de Gasteiger-Marsilli (Gasteiger y Marsilli, 1978), las masas atómicas (M)

(Kier y Hall, 1986), los volúmenes de van der Waals (V), la electronegatividad en la escala de Pauling (E) (Pauling, 1939), entre otras.

3.1.1.2 “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 **G** esta constituido por una colección de vértices (puntos) y otra de aristas (líneas o enlaces) conectando estos vértices.⁵³⁻⁵⁵ En términos matemáticos, un grafo simple **G** es definido como $G = (V, E)$ (Harary, 1976; Cvetkovic y col, 1995; Balaban, 1976). 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 $\{v_i, v_i\}$, 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 3.1) (Rouvray y Balaban, 1976; Trinajstić, 1983; Gutman,y Polansky, 1986; Harary, 1971; Cvetkovic y col, 1995; Balaban, 1976).

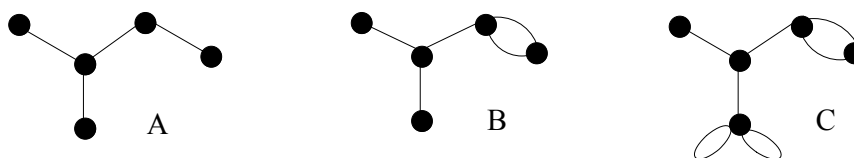


Figura 3.1. A) Grafo simple; B) Multigrafo; y C) Pseudografo.

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 (Rouvray y Balaban, 1976; Trinajstić, 1983; Gutman y Polansky, 1986; Harary, 1971; Cvetkovic y col, 1995; Balaban, 1976; Gorbátov, 1988; Busacker y Saaty, 1965; Wilson, 1972).

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 3.2).

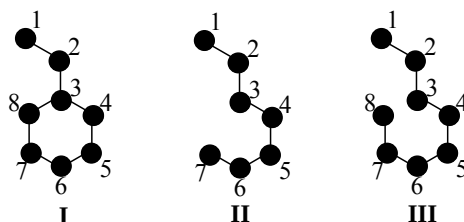


Figura 3.2. Representación esquemática de subgrafos

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 (Marrero-Ponce, 2003, 2004, Marrero-Ponce y col, 2003, 2004a,b,c,d,e,f,g,h, 2005a,b,c,d,e,f,g,h,i, 2006a,b; Meneses y col, 2005; Montero y col, 2005, 2006; Caseñola y col, 2006; Castillo-Garit y col, 2006; Celeste y col, 2006; Marrero-Ponce y Torrens, 2004). Las $n \times n$ k -ésimas (k^{th}) matrices de densidad electrónica grafo-teóricas no estocásticas del pseudografo molecular G , M^k , son matrices simétricas y cuadradas, donde n es el número de átomos en la molécula. Luego, $M(G) \equiv M = [m_{ij}]$, donde los elementos m_{ij} se definen como sigue:

$$m_{ij} = P_{ij} \text{ if } i \neq j \text{ and } \exists e_k \in E(G) \quad (3.1)$$

$$= L_{ij} \text{ if } i = j$$

$$= 0 \text{ otherwise}$$

donde $E(\mathbf{G})$ representa el conjunto de las aristas de $\mathbf{G}$. P_{ij} es el número de aristas (enlaces químicos) entre los vértices (núcleos atómicos) v_i y v_j . L_{ij} es el número de lazos en v_i .

Los elementos m_{ij} ($m_{ij} = P_{ij}$) de la matriz $\mathbf{M}$ representan los enlaces químicos entre un núcleo atómico i y otro j . La matriz $\mathbf{M}^k$ provee el número de *camino de longitud k* que une cada par de vértices v_i y v_j . Por esta razón, cada arista en $\mathbf{M}^1$ (para $k = 1$) representa 2 electrones perteneciente al enlace covalente entre los dos núcleos atómicos i y j . Así por ejemplo, los coeficientes m_{ij} y m_{ji} de $\mathbf{M}^1$ son iguales a 1, 2 y 3 cuando un simple, doble o triple enlace, respectivamente, aparece entre los núcleos atómicos (vértices) v_i y v_j . Las moléculas que contienen anillos aromáticos con más de una estructura canónica son representadas como pseudografo. Este es el caso de los ciclos de piridina, naftaleno, quinoleína, etc; donde la presencia de los electrones de los orbitales PI (π) son representados por medio de lazos sobre cada núcleo atómico del anillo aromático. En cambio, los anillos aromáticos con una sola estructura canónica, tales como el furano, tiofeno, pirrol, entre otros, son representados como *multigrafos*. Con el propósito de ilustrar el cálculo de estas matrices, tomemos como un simple ejemplo la molécula seleccionada en la sección anterior, 3-mercapto-piridina-4-carbaldehído. La Tabla 3.1 muestra la estructura química y el pseudografo molecular etiquetado de este compuesto.

Tabla 3.1. A) Estructura química de la molécula de 3-mercapto-piridina-4-carbaldehído y su pseudografo molecular etiquetado, G. B) y C) Cero ($k = 0$), primera ($k = 1$), segunda ($k = 2$) y tercera ($k = 3$) potencia de las matrices de densidad electrónica grafo-teóricas no estocásticas y estocásticas de G, respectivamente.

A)

Estructura molecular

Pseudografo molecular (pseudografo con atomos de H suprimido)

Vertices

Heteroatoms

C atoms

B) k -ésimas potencia de las matrices de densidad electrónica grafo-teóricas estocásticas, M^k ($k = 0-3$)

Orden cero ($k = 0$)

Orden uno ($k = 1$)

Orden dos ($k = 2$)

Orden tres ($k = 3$)

C) k -ésimas potencia de las matrices de densidad electrónica grafo-teóricas no-estocásticas, S^k ($k = 0-3$)^{a,b}

Orden uno ($k = 1$)

Orden dos ($k = 2$)

Orden tres ($k = 3$)

^ala potencia

cero ($k = 0$) de la matriz de densidad electrónica grafo-teórica estocástica, S^0 , coincide con la matriz de esa misma potencia no estocástica ($M^0 = S^0$)

^blos valores de los elementos de las k -ésimas matrices S^k (s_{ij}^k) han sido aproximados.

Nótese, que la matriz $\mathbf{M}^1$ es un modelo grafo-teórico de la estructura electrónica de la molécula y puede ser considerada como una matriz tipo-Hückel extendida. En cambio, la matriz $\mathbf{M}^k$ ($k = 0, 1, 2, 3, \dots$) provee el número de *camino de longitud k* que une a los vértices v_i y v_j . Es decir, $\mathbf{M}^1$ considera todos los electrones de la capa de valencia (electrones σ y π de la molécula) en un paso, y sus potencias ($k = 0, 1, 2, 3, \dots$) pueden ser consideradas como un modelo de la interacción electrónica en la “red molecular” en k pasos. Este modelo puede ser visto como un intermedio entre la ecuación de Schrödinger de la mecánica cuántica y las ideas clásicas del enlace químico (Klein, 2003).

Desde el punto de vista químico-físico, este enfoque puede ser entendido como un modelo simple del movimiento intramolecular de todos los electrones de la capa de valencia en la “red molecular” de enlace covalente. Consideremos un ejemplo hipotético en el cual un conjunto de átomos está libre en un espacio y en un tiempo inicial (t_0) arbitrario, a este tiempo los electrones están distribuidos alrededor de los núcleos atómicos ($\mathbf{M}^0$). Paulatinamente estos electrones pueden distribuirse alrededor de los núcleos en intervalos discretos de tiempo t_k . En este sentido, un electrón perteneciente a un átomo i puede moverse a otros átomos a diferentes periodos discretos de tiempo t_k ($k = 0, 1, 2, 3, \dots$) a través de la red del enlace químico. Este comportamiento se tiene en cuenta en las respectivas matrices $\mathbf{M}^k$. En la Tabla 3.1 (columna izquierda) se ejemplifica esta representación ($\mathbf{M}^0$ - $\mathbf{M}^3$) para la molécula del 3-mercapto-piridina-4-carbaldehído.

Las k -ésimas matrices de densidad electrónica grafo-teóricas estocásticas, $\mathbf{S}^k$, puede ser obtenida directamente de $\mathbf{M}^k$ (ver ecuación 3.2). Los elementos ${}^k s_{ij}$ se definen como se muestra en la ecuación 3.2:

$${}^k s_{ij} = \frac{{}^k m_{ij}}{{}^k \text{SUM}_i} = \frac{{}^k m_{ij}}{{}^k \delta_i} \quad (3.2)$$

donde ${}^k m_{ij}$ son los elementos de la k -ésima potencia de $\mathbf{M}$, y ${}^k \text{SUM}_i$ es la suma de la fila i -ésima de $\mathbf{M}^k$ o grado del vértice de orden k del átomo i , ${}^k \delta_i$. Es importante destacar que la matriz $\mathbf{S}^k$ tiene la propiedad de que la suma de los elementos en cada fila es 1. Una matriz con elementos enteros-positivos y que poseen esa propiedad es llamada una **matriz estocástica** (Edwards y Penney, 1988a).

Esta transformación propuesta en la ecuación 3.2 normaliza cada fila de la matriz original y por tanto, sus k -ésimos elementos constituyen las probabilidades de transición con las cuales un electrón se mueve de un átomo i a otro j en un periodo de tiempo discreto t_k . En la columna derecha de la Tabla 5.1 se muestra, a modo de ejemplo, las matrices estocásticas de orden 0-3 para la molécula del 3-mercapto-piridina-4-carbaldehído. Nótese que los k -ésimos elementos s_{ij}

toman en consideración la información de la topología molecular en k pasos a través de todo el esqueleto covalente (red química σ - y π -). Así por ejemplo, las probabilidades de auto-retorno de segundo orden ($^2s_{ii}$) [esta es la probabilidad con la que un electrón retorna a su átomo original en t_2] varían regularmente según los diferentes “estados híbridos” de un átomo i en la molécula. Es decir, un electrón debe tener una probabilidad de retorno a un átomo de C sp mayor que a un átomo de C sp^2 (o sp^3) en t_2 [$p(C_{sp}) > p(C_{sp}^2) > p(C_{sp}^2 \text{ arom}) > p(C_{sp}^3)$] (ver Tabla 3.1 para más detalles). Esto es un resultado lógico si tomamos en cuenta las propiedades electrónicas (por ejemplo su escala de electronegatividad) de cada una de las diferentes hibridaciones de estos átomos.

3.1.1.3 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 $\Re^n$ [$f: \Re^n \rightarrow \Re^n$; entonces f : Endomorfismo sobre $\Re^n$] en las bases canónicas (Gorbátov, 1988; Busacker y Saaty, 1965; Wilson, 1972; Klein, 2003; Edwards y Penney, 1988a; Edwards y Penney, 1988b; Browder, 1996; Axler, 1996). 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:

$$f_k(\bar{x}_i) = \sum_{j=1}^n m_{ij} x^j = [\mathbf{X}']^k = \mathbf{M}^k[\mathbf{X}] \quad (3.3)$$

$$^s f_k(\bar{x}_i) = \sum_{j=1}^n s_{ij} x^j = [\mathbf{XS}']^k = \mathbf{S}^k[\mathbf{X}] \quad (3.4)$$

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 $\Re^n$. 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, $\Re^n$. Esta aplicación, es una correspondencia que asigna a cualquier vector X en $\Re^n$ un vector $f(x)$ de forma tal que:

$$f(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 f(X_1) + \lambda_2 f(X_2) \quad (3.5)$$

para todo λ_1, λ_2 número reales y cualquier vector X_1, X_2 en $\Re^n$. 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 3.16 y 3.17), donde $[X]$ es el vector columna (una matriz de $n \times 1$) de coordenadas de $\bar{x}$ en la base canónica de $\mathfrak{R}^n$, $[X]^t$ es la matriz transpuesta de $[X]$ (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 sigma como pi. 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:

$$\psi_i = \sum_{j=1}^n c_{ij} \varphi_j \quad (3.6)$$

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 $\mathfrak{R}^n$ (Gorbátov, 1988; Busacker y Saaty, 1965; Wilson, 1972; Klein, 2003; Edwards y Penney, 1988a; Edwards y Penney, 1988b; Browder, 1996; Axler, 1996). Es decir, los índices lineales totales constituyen aplicaciones lineales de $\mathfrak{R}^n$ a escalares $\mathfrak{R}$ [$f: \mathfrak{R}^n \rightarrow \mathfrak{R}$]. La definición matemática de este descriptor molecular es la siguiente: (Rouvray, 1976; Trinajstić, 1992; Gutman y Polansky, 1986; Harary, 1971; Cvetkovic y col, 1995; Balaban, 1976)

$$f_k(\bar{x}) = \sum_{i=1}^n f_k(\bar{x}_i) = [\mathbf{u}]^t [\mathbf{X}']^k = [\mathbf{u}]^t \mathbf{M}^k [\mathbf{X}] \quad (3.7)$$

$${}^s f_k(\bar{x}) = \sum_{i=1}^n {}^s f_k(\bar{x}_i) = [\mathbf{u}]^t [\mathbf{X}\mathbf{S}']^k = [\mathbf{u}]^t \mathbf{S}^k [\mathbf{X}] \quad (3.8)$$

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 3.3 y 3.4, 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:

$$f_k(\bar{x}) = \sum_{L=1}^Z f_{kL}(\bar{x}) \quad (3.9)$$

$$^s f_k(\bar{x}) = \sum_{L=1}^Z ^s f_{kL}(\bar{x}) \quad (3.10)$$

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.

3.2 Análisis Quimiométrico de los Datos:

3.2.1 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* (van de Waterbeemd, 1995; Alzina, 1989). 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.

3.2.2 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 (van de Waterbeemd, 1995; Alzina, 1989):

$$D_1 = b_1 X_1 + b_2 X_2 + \dots + b_3 X_3 + \dots \quad (3.11)$$

$$D_2 = a_1 X_1 + a_2 X_2 + \dots + a_3 X_3 + \dots \quad (3.12)$$

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} .

$$D_{12} = D_1 - D_2 \quad (3.13)$$

3.2.3 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 (Todeschini y Consonni, 2000; STATISTICA, 1999).

3.2.4 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 λ de Wilks. En el caso de solo dos grupos, la λ de Wilks es (Alzina, 1989):

$$\lambda = \frac{SC_{\text{int ragrupos}}}{SC_{\text{total}}} \quad (3.14)$$

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 λ no es significativa la *discriminación* no será posible (Alzina, 1989).

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:

$$D_{pq}^2 = (\mu_p - \mu_q)' \Sigma^{-1} (\mu_p - \mu_q) \quad (3.15)$$

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 varianzas-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:

$$F = D^2 \frac{n_p n_q (n_p + n_q - v - 1)}{(n_p + n_q)(n_p + n_q - 2)v} \quad (3.16)$$

3.2.5 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*.

3.2.5.1 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' (Alzina, 1989; STATISTICA, 1999); también se han implementado combinaciones de estos métodos así han surgido el 'backward stepwise' y el 'forward stepwise'. Este último método es el más utilizado y sigue un proceso de selección de variables paso a paso.

3.2.5.2 Tolerancia.

La *tolerancia* es una medida del grado de asociación lineal entre las variables independientes (Belsey y col, 1980). 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.

3.2.6 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* (Alzina, 1989; STATISTICA, 1999; Belsey y col, 1980). Una 'multicolinealidad' alta, dificulta estimar la importancia relativa de los descriptores en el modelo, lo cual afecta la interpretación de las actividades modeladas en términos estructurales. En estos casos los modelos se considera que están sobreajustados y esto es una dificultad cuando el objetivo es interpretar el modelo obtenido para la actividad en términos estructurales. 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 (Basak y col, 2000; Patel y Cronin, 2001). El nivel aceptable de *colinealidad* es algo subjetivo y en ese sentido, según plantean Cronin y Schultz, la misma debe ser la menor posible (Cronin y Schultz, 2003). No obstante lo antes planteado no es objetivo de este trabajo interpretar los modelos en términos estructurales sino obtener modelos capaces de predecir la actividad carcinogénica con la mayor precisión posible.

3.2.7 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 (Kier, 1980): 1) comparación del valor de F con el valor tabulado, 2) determinación de casos bien clasificados en la serie de entrenamiento (SE) y 3) validación externa. Además, los métodos de *validación cruzada* también pueden aplicarse a este tipo de modelos (Cativela y col, 1992). 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

(Ogino y col, 1980): 1) una combinación de variables que minimice el número de compuestos mal clasificados, 2) El empleo del número menor de variables, y 3) la colinealidad entre las variables independientes es minimizada.

3.2.7.1 Procesos de validación a los modelos QSAR-ADL obtenidos.

Para acceder al *poder predictivo* del modelo QSAR-ADL obtenido, se pueden desarrollar procedimientos de *validación externa* y de *validación interna*. La *validación interna* se utiliza cuando todos los compuestos pertenecen a los mismos datos originales y son utilizados en la construcción del modelo. Este proceso consiste en remover sucesivamente compuestos de la data y obtener el modelo para los restantes; utilizando los que se extrajeron para predecir su actividad. No obstante, se ha demostrado que la “única” condición necesaria y suficiente para poder estimar el verdadero poder predictivo de un modelo es analizando su poder predictivo en una serie de predicción externa (es decir utilizando compuestos que no fueron empleados en el entrenamiento del modelo). En este sentido, en nuestro estudio utilizamos el procedimiento de *validación externa* (Wold y Erikson, 1995; van de Waterbeemd, 1995) para ellos empleamos como series de predicción externas, el primer y segundo ejercicio de la Programa Nacional de Toxicología de Estados Unidos. (Nacional Toxicology Program, NTP por sus siglas en inglés), conformados por 43 compuesto del primer ejercicio y 26 compuestos del segundo ejercicio. La calidad de los modelos también fue evaluada chequeando los principales parámetros estadísticos propuestos en la literatura para los modelos de clasificación: exactitud (Q_{total}), sensibilidad, especificidad y la proporción de falsos activos.

3.2.8 Conclusiones parciales del capítulo.

En el presente, el uso de los DMs abarca las principales áreas de investigación en el desarrollo de fármacos: Descubrimiento de compuestos líderes y optimización de las actividades biológicas de estos (no solo las farmacológicas, sino también las farmacocinéticas y toxicológicas). En el presente, la mayoría de los investigadores que trabajan en esta rama del saber, concentran sus esfuerzos en el desarrollo de mejores DMs y de mejores modelos matemáticos que permitan el descubrimiento *in silico* de nuevos compuestos líderes bioactivos.

Todos los modelos desarrollados están basados en datos (y teoría), por lo que la calidad y representatividad de estos datos es esencial para la credibilidad, la utilidad y la interpretabilidad de los modelos. Por la forma detallada en que han sido expuestos los pasos y estrategias de trabajo en el análisis multivariado de los datos puede dar la impresión que la obtención de modelos QSAR robustos y con un adecuado poder predictivo es más arte que ciencia, lo cual en parte es verdad. Ahora bien, ¿puede un modelo QSAR ser utilizado para hacer predicciones a compuestos fuera de la SE? Para confirmar esto un riguroso proceso de validación tiene que llevarse a cabo. Así por ejemplo, los modelos QSAR que son desarrollados en base solamente a

una SE, sin hacer referencia a una SP; las predicciones externas deben hacerse con mucha cautela y las conclusiones sobre el modelo y la propiedad modelada no pueden ser tan categóricas. En contraste, si un modelo es transparente, mecanística o estructuralmente interpretable y predice satisfactoriamente la actividad biológica de compuestos en una SP, entonces puede ser considerado que dentro de las limitaciones de un modelo, este es muy útil en el diseño/descubrimiento de nuevos materiales o compuestos bioactivos y que los errores potenciales han sido minimizados.

4

PARTE EXPERIMENTAL

4.1 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.

La data de carcinogenicidad usada fue obtenida de la base de datos de potencia carcinogénica (CPDB por sus sigla en ingles), establecida por Gold y Zeiger (Gold y Zeiger, 1997) y disponible en Internet en (<http://potency.berkeley.edu/cpdb.html>) (ver Anexo 2). La CPDB es un solo recurso estandarizado por muchos años de carcinogénesis crónica y término largo. Esta base de datos contiene una gran diversidad de estructuras químicas (más de 1300 sustancias reportadas), e incluyen datos tumorales reconocidos a partir de todos los reportes técnicos de bioensayos de NCI/NTP, así como datos adicionales extraídos de alrededor de 1200 fuentes de literaturas sujetas a una revisión extensiva (Gold y col, 1999).

La data establecida esta compuesta por 189 compuestos (119 carcinogénicos y 70 no carcinogénicos). Los modelos obtenidos fueron validados utilizando como series de predicción externas, el primer y segundo ejercicio del Programa Nacional de Toxicología de Estados Unidos. (Nacional Toxicology Program, NTP por sus siglas en inglés). Estos ejercicios fueron utilizados en la pasada década para probar el poder predictivo de los modelos teóricos que habían sido desarrollados por la comunidad científica. El primero de estos ejercicios esta compuesto por 43 compuestos químicos de diferentes clases. Muchos de los sistemas que participaron en este ejercicio fueron enfoque de tipo relación estructura-actividad (SAR por sus siglas en ingles) y de tipo relación cuantitativa estructura-actividad (QSAR); otros enfoques que también participaron fueron los que relacionan la carcinogenesis con otros evento biológico a mas corto plazo conocido como relaciones actividad-actividad (AAR por sus siglas en inglés). El segundo ejercicio esta compuesto de 26 compuestos químicos que en aquel momento estaban siendo ensayados. En este estuvieron involucradas predicciones basadas en sistemas de expertos humanos los cuales fueron los que alcanzaron los mejores resultados.

El problema fundamental en dichos ejercicios estuvo basado en que muchos de los compuestos no carcinógenos se predicen como carcinógenos en diversos sistemas. Los resultados de los dos ejercicios comparativos del NTP son muy importantes para juzgar la capacidad verdadera de predecir la carcinogénesis de productos químicos no comprobados. En ambos ejercicios, el mejor funcionamiento fue logrado por los métodos que confiaron en gran parte en el juicio experto humano. Sin importar esto, los métodos clásicos de QSAR pueden generar modelos absolutamente satisfactorios para las clases individuales de mutágenos y de agentes carcinógenos (Benigni, 2000; Franke, 2001; Zhang, 1992).

La función discriminante fue obtenida haciendo el análisis discriminante lineal implementado en el paquete estadístico versión 6.0. (StatSoft, 2002). Los parámetros predeterminados de este programa se usaron en el desarrollo del modelo. Las variables a ser incluidas en la ecuación se seleccionaron por el procedimiento de paso hacia adelante (forward stepwise) como estrategia en el método de selección de variable.

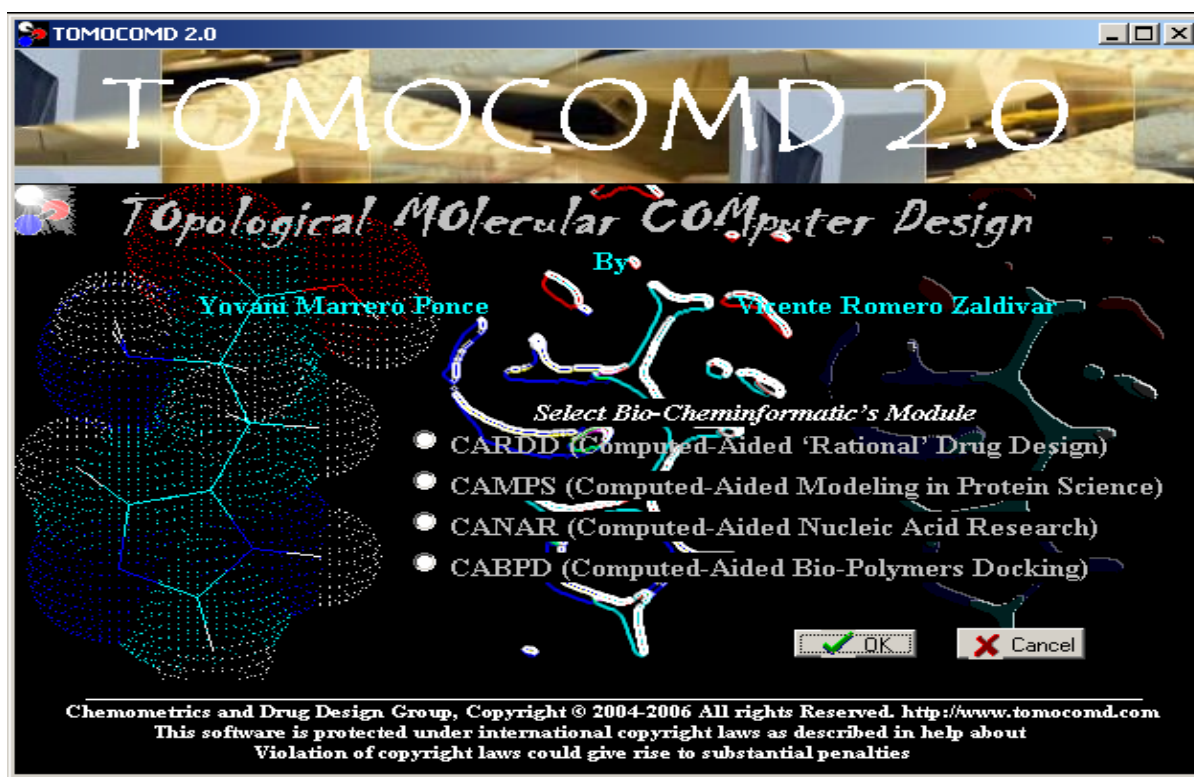
4.2 Método Computacional. TOMOCOMD-CARDD software

TOMOCOMD (Marrero-Ponce y Romero, 2002) es un programa interactivo para el diseño molecular e investigaciones bioinformáticas. El programa esta 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 siguiente 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 solo demostraremos los rasgos sobresalientes de uno de estos subprogramas: CARDD. Este modulo fue desarrollado basado en una filosofía amigable para el usuario, el cual no tiene que tener *a priori* ningún conocimiento de programación.

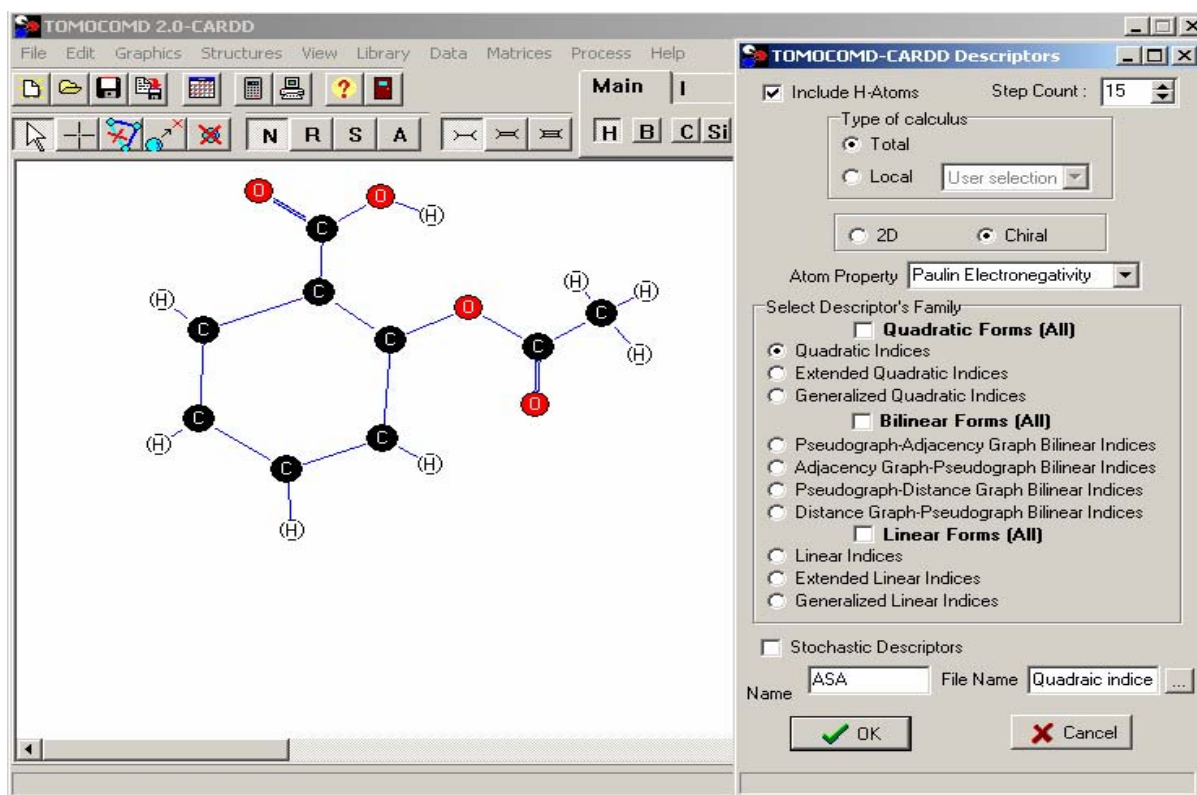
Los principales pasos para desarrollar un estudio QSPR/QSAR utilizando el enfoque topológico molecular **TOMOCOMD-CARDD**, son resumidos brevemente a continuación:

1. Representar el seudografo 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 4 propiedades atómicas,
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



A



B

Figura 4.1. TOMOCOMD-CARDD Software: A, Ventana para seleccionar el modulo de trabajo. B, Interfaz gráfica del subprograma de diseño de fármacos.

descriptores calculados. Así por ejemplo, la ecuación obtenida con los índices lineales totales tomaría la siguiente apariencia:

$$P = a_0b_0(x) + a_1b_1(x) + a_2b_2(x) + \dots + a_kb_k(x) + c \quad (4.1)$$

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

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.

4.3. 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 (RN) artificiales, el ADL fue seleccionado para obtener las *funciones de clasificaciones* sobre la base de la simplicidad del método (van de Waterbeemd, 1995). Los modelos QSAR-ADL se obtuvieron con el paquete de

programas estadísticos STATISTICA (StatSoft Inc., 2001). 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 (exactitud, sensibilidad, especificidad, etc), todos calculados a partir de la matriz de confusión (Baldi, 2000).

Los resultados de las probabilidades de clasificación para cada compuesto se muestran utilizando el siguiente parámetro: $\Delta P\% = [P(\text{activ}) - P(\text{inactiv})] \times 100$, donde $P(\text{activ})$ es la probabilidad con que la ecuación clasifica un compuesto como activo. Contrariamente, $P(\text{inactiv})$ es la probabilidad con que el modelo clasifica un compuesto como inactivo.

Para acceder al *poder predictivo* de los modelos QSAR-ADL obtenidos, se desarrollaron procedimientos de *validación interna* y *externa*. En la *validación interna*, se llevó a cabo en cada caso un proceder de validación cruzada (VC) ‘dejando *grupos de compuestos fuera*’ (leave-group-out). Esta metodología retira sistemáticamente un grupo de compuestos de la SE y construye el modelo con los datos reducidos y la función obtenida se utiliza para predecir los compuestos que han sido extraídos. Este proceder se repite hasta que todos los casos (en este caso grupos) han sido retirados una vez y predichos por el modelo generado con las respectivas bases de datos reducidas. Finalmente, para un exhaustivo ensayo del *poder predictivo* de los modelos obtenidos, en cada caso se utilizó una SP externa (Wold y Erikson, 1995).

5

RESULTADOS Y DISCUSIÓN

5.1 Obtención de los Modelos de Clasificación con Índices Moleculares.

5.1.1 Desarrollo de Modelos QSAR-ADL con Índices Lineales no-estocásticos

Con el objetivo de modelar la actividad carcinogénica, desarrollamos modelos discriminantes utilizando el ADL y los índices lineales no-estocásticos como técnica estadística y como descriptores moleculares, respectivamente. En este estudio utilizamos las ponderaciones de cuatro propiedades atómicas: 1) masa atómica (M), 2) polarizabilidad atómica (P), 3) electronegatividad atómica de Pauling (G) y 4) volumen de van der Waals atómico (V), para el cálculo de los descriptores moleculares. La mejor función de clasificación obtenida con el proceder ‘forward stepwise’ es dada a continuación junto con sus parámetros estadísticos:

$$\begin{aligned} \text{Class} = & 1.9644 - 7.414 \times 10^{-7} M f_{10}^H(x_E) + 1.726 \times 10^{-6} M f_{11}(x_{E-H}) - 4.472 \times 10^{-8} M f_{14}(x_{E-H}) \\ & + 5.704 \times 10^{-6} V f_7^H(x) + 1.955 \times 10^{-3} V f_4^H(x_E) - 3.688 \times 10^{-6} V f_{10}(x_E) - 3.087 \times 10^{-3} P f_4^H(x) \\ & + 9.771 \times 10^{-5} P f_9(x_E) + 4.581 \times 10^{-2} G f_3^H(x_E) - 2.397 \times 10^{-2} G f_4^H(x_E) \\ & + 6.049 \times 10^{-7} G f_{11}^H(x_E) + 8.812 \times 10^{-7} G f_{12}(x_E) \end{aligned} \quad (5.1)$$

$$N = 189 \quad \lambda = 0.66 \quad D^2 = 2.09 \quad F(12,176) = 7.2373 \quad p < 0.0001$$

donde N es el número de compuestos, λ es el estadístico de Wilks, D^2 es el cuadrado de la distancia de Mahalanobis y F es el estadístico de Fischer.

El objetivo fundamental de este proceder es obtener funciones a partir de espacios experimentales iguales, con el objetivo de llevar a cabo un estudio comparativo sobre el poder descriptivo de la actividad carcinogénica de los descriptores moleculares **TOMOCOMD-CARDD** (índices lineales basadas en átomo) utilizados para obtener los modelos QSAR-ADL.

En el desarrollo de las funciones discriminantes fueron utilizados los índices lineales totales y locales y no-estocásticos (Ec.5.1). La serie de entrenamiento estuvo conformada por 189 compuestos, de ellos 119 compuestos activos (carcinógenos) y 70 compuestos inactivos (no

carcinógenos). Los modelos obtenidos fueron validados haciendo uso de dos series de predicción externas; las mismas son el primer y segundo ejercicio de la Programa Nacional de Toxicología de Estados Unidos. (Nacional Toxicology Program, NTP por sus siglas en inglés), conformados por 43 compuesto del primer ejercicio y 26 compuestos del segundo ejercicio.

Tabla 5.1. Resultados de la Clasificación de los Compuestos Incluidos en las Series de Entrenamiento y de Predicción Obtenidos Utilizando los Índices Lineales no estocásticos

	Exactitud 'Q _{Total} ' (%)	Sensibilidad (%)	Especificidad (%)	"Hit" de Falsos Positivos (%)
índices cuadráticos no-estocásticos				
SE	78.31	82.35	83.05	28.57
SP: 1 ^{er} Ejercicio	53.49	69.57	55.17	65.00
SP: 2 ^{do} Ejercicio	57.69	68.75	64.71	60.00

En la Tabla 5.1 se suman los resultados de la clasificación obtenidos en el ajuste de el modelo, así como los resultados en los dos ejercicios de la NTP. En el modelo obtenido para el índice lineal no-estocástico se evidencia un por ciento de clasificación de 78,30% para la SE, donde se clasificaron incorrectamente 41 compuestos de un total de 189 compuestos; en la SP externa se demuestra que para el segundo ejercicio (Q = 74.41%) los índices lineales describen en una forma mas adecuada la actividad carcinogénica que para el primer ejercicio (Q = 53.48%). Los falsos activos son compuestos inactivos que el modelo los clasifica como carcinogénicos y los falsos inactivos son los compuestos activos clasificados como inactivos por el modelo.

En la Tabla 5.2 se da una explicación más detallada de clasificación de la serie de entrenamiento con relación a los falsos positivos y los falsos negativos, así como su probabilidad correspondiente.

Tabla 5.2 Resultados de clasificación de los compuestos de la serie de entrenamiento para Índices Lineales no-estocásticos.

Compuestos	ΔP%	Clas	Compuestos	ΔP%	Clas
<i>Compuestos carcinógenos</i>					
[4-Chloro-6-(2,3-xylidino)-2-pyrimidinylthio]acetic acid	32.22	+	Diethylstilbestrol	46.98	+
1-(2-Hydroxyethyl)-1-nitrosourea	37.25	+	Dihydrosafrole	-9.08	-
1,1-Dimethylhydrazine	79.20	+	dl-Ethionine	33.61	+
1,2-Dibromoethane	72.89	+	Doxylamine succinate	-18.27	-
1,2-Dimethylhydrazine	77.96	+	Ethylene thiourea	66.37	+
1'-Hydroxysafrole	2.73	+	Formaldehyde	35.15	+
1-Nitroso-3,4,5-trimethylpiperazine	57.16	+	Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	1.62	+
2,2,2-Trifluoro-N-[4-(5-nitro-2-furyl)-2-thiazolyl]acetamide	-20.90	-	Glu-P-1	29.18	+
2,3,7,8-Tetrachlorodibenzo-p-dioxin	62.02	+	Glu-P-2	19.75	+
2,4,5-Trimethylaniline	61.03	+	Hexachlorobenzene	70.00	+
2,4,6-Trimethylaniline	59.81	+	HYDRAZINE	87.43	+

Tabla 5.2 (Cont)

Compuestos	ΔP%	Clas	Compuestos	ΔP%	Clas
2,4-Diaminotoluene	75.03	+	HYDRAZINE SULFATE	82.23	+
2,4-Dinitrotoluene, practical grade	60.68	+	IQ	13.32	+
2,5-Dimethoxy-4'-aminostilbene	30.74	+	Isoniazid	60.45	+
2,5-Xylidine	64.89	+	Lead Acetate, basic	-15.12	-
2-Acetylaminofluorene	64.66	+	MeA-α-C acetate	49.95	+
2-Amino-4-(5-nitro-2-furyl)thiazole	20.39	+	MeIQx	-26.78	-
2-Hydrazino-4-(5-nitro-2-furyl)thiazole	62.08	+	Melphalan	11.39	+
2-Hydrazino-4-(p-aminophenyl)thiazole	54.71	+	Methyl tert-butyl ether	47.46	+
2-Hydrazino-4-(p-nitrophenyl) thiazole	44.37	+	Metronidazole	49.85	+
2-Naphthylamine	78.06	+	N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide	-21.25	-
3-(5-Nitro-2-furyl)-imidazo(1,2-α)pyridine	33.70	+	N-[4-(5-Nitro-2-furyl)-2-thiazolyl]formamide	-54.15	-
3,3',4,4'-Tetraaminobiphenyl	72.93	+	N-[5-(5-Nitro-2-furyl)-1,3,4-thiadiazol-2-yl]acetamide	24.17	+
3-Aminotriazole	49.55	+	N-Hydroxy-2-acetylaminofluorene	71.45	+
3-Nitro-3-hexene	54.08	+	Nitrobenzene	55.06	+
4-Chloro-4'-aminodiphenylether	64.99	+	Nitroso-2,3-dihydroxypropyl-2-oxopropylamine	74.76	+
5-Azacytidine	72.95	+	Nitroso-2-oxopropylethanolamine	73.46	+
Acetaldehyde	52.55	+	Nitrosodibutylamine	26.96	+
Acetamide	70.02	+	N-Nitroso-2,3-dihydroxypropyl-2-hydroxypropylamine	78.98	+
Acetaminophen	41.13	+	N-Nitrosoallyl-2-oxopropylamine	49.48	+
AF-2	32.74	+	N-Nitrosodimethylamine	71.90	+
Aldrin	51.20	+	N-Nitrosomethyl-2,3-dihydroxypropylamine	75.84	+
α-1,2,3,4,5,6-Hexachlorocyclohexane	84.03	+	N-Nitrosomorpholine	71.38	+
Aramite	12.86	+	N-Nitroso-N-methylurea	32.53	+
Auramine-O	-47.60	-	N-Nitrosopiperidine	59.02	+
Azobenzene	43.06	+	N-Nitrosopyrrolidine	67.18	+
Benzidine	77.29	+	o-Phenylenediamine	81.62	+
Benzo(a)pyrene	97.79	+	o-Toluidine	75.85	+
β-Propiolactone	-3.76	-	p,p'-DDE	27.07	+
Bis-(chloromethyl)ether	12.08	+	Phenacetin	21.04	+
Bis-2-hydroxyethylthiocarbamic acid,potassium	43.51	+	Phenobarbital, sodium	-86.26	-
Butylated hydroxyanisole	20.36	+	PhIP	-38.65	-
Caffeic acid	13.30	+	Piperonyl butoxide	-83.49	-
Captafol	38.00	+	Propylthiouracil	-29.63	-
Captan	-24.80	-	Safrole	-6.10	-
Carbon tetrachloride	54.11	+	Sesamol	24.91	+
Catechol	68.42	+	Sterigmatocystin	-66.29	-
Chlorambucil	-16.92	-	Streptozotocin	-15.41	-
Chloroform	58.70	+	Styrene oxide	45.96	+
Chloromethyl methyl ether	27.66	+	Tamoxifen citrate	5.69	+
Chrysazin	-22.75	-	Thioacetamide	90.78	+
Ciprofibrate	9.95	+	Thiouracil	34.18	+
Cyclophosphamide	55.61	+	Trichloroethylene	84.44	+

Tabla 5.2 (Cont)

Compuestos	ΔP%	Clas	Compuestos	ΔP%	Clas
D & C red no	79.21	+	Trp-P-1 acetate	40.04	+
DDT	59.45	+	Trp-P-2 acetate	56.95	+
Diallylnitrosamine	49.57	+	Uracil	-22.32	-
Dibromodulcitol	78.93	+	Urethane	-25.92	-
Dibromomannitol	78.93	+	Vinyl acetate	8.39	+
Dichloroacetylene	73.70	+	Vinyl chloride	66.89	+
Dieldrin	53.64	+			
<i>Compuestos no carcinógenos</i>					
1,1,1-Trichloroethane(grado tec	74.08	+	FD & C blue no1	-97.43	-
1,2-Dichloroethane	72.24	+	FD & C blue no 2	-94.28	-
1,4-Dichlorobenzene	63.96	+	FD & C green no 3	-97.21	-
1-[(5-Nitrofurfurylidene)amino]hydantoin	-30.69	-	FD & C red no 3	-92.94	-
1-Chloro-2,4-dinitrobenzene	39.45	+	FD & C yellow no 5	-86.70	-
2,3,4,5,6-Pentachlorophenol (Dowicide EC-7)	72.64	+	FD & C yellow no 6	-57.06	-
2,4,5-Trichlorophenoxyacetic acid	52.03	+	Fenvalerate	-96.81	-
3-Nitro-4-Hydroxyphenylarsonic Acid	-22.74	-	Flecainide acetate	-90.13	-
5,5-Diphenylhydantoin	-74.79	-	Fluoxetine	-76.62	-
6-Dimethylamino-4,4-diphenyl-3-heptanone	-12.78	-	Hexamethylenetetramine	-94.07	-
Adipamide	25.41	+	Isopropyl-N-(3-chlorophenyl) carbamate	-63.70	-
Aspirin	-70.75	-	Maleic hydrazide	32.19	+
Barium Acetate	-37.26	-	Methotrexate	-94.07	-
Benzalazine	-77.78	-	Misoprostol	-92.59	-
Benzoate, sodium	-17.14	-	Monochloroacetic acid	75.06	+
Benzoguanamine	62.54	+	Nefiracetam	-56.20	-
Black PN	-99.37	-	N-Nitrosocimetidine	40.72	+
C.I. pigment yellow 12	-82.95	-	Oxamyl	-75.86	-
C.I. pigment yellow 16	-85.88	-	Oxprenolol	10.92	+
C.I. pigment yellow 83	-74.64	-	Phenyl isothiocyanate	-51.16	-
Cadmium Acetate	-25.30	-	Prazepam	-67.75	-
Caffeine	-95.65	-	Praziquantel	-92.18	-
Chlorodifluoromethane	20.29	+	Propranolol	65.85	+
Chromium (Iii) Acetate	-91.17	-	Quercetin dihydrate	23.64	+
Compound 50-892	-80.88	-	Rutin trihydrate	-23.26	-
Cyclohexylamine	57.82	+	Saccharin	-56.56	-
Deltamethrin	-98.59	-	Sorbic acid	-13.11	-
Diazepam	-48.63	-	Sotalol	-29.14	-
Dichlorodifluoromethane	-0.39	-	Tetramethylthiuram disulfide	-14.29	-
Dimethylformamide	-46.51	-	Tilidine fumarate	-87.97	-
Diphenyl-p-phenylenediamine	43.16	+	Toluene Diisocyanate	-93.86	-
Erythorbate, sodium	42.02	+	Trichlorofluoromethane	29.22	+
Estazolam	-2.82	-	Trifluralin, technical grade	-74.66	-
Ethynodiol diacetate	-98.18	-	Triprolidine	-2.83	-
Etodolac	-49.31	-	Urea	56.73	+

El modelo obtenido fue validado utilizando como series de predicción externas, el primer y segundo ejercicio del Programa Nacional de Toxicología de Estados Unidos. (NTP por sus siglas en inglés). El problema fundamental en dichos ejercicios ha estado basado en que muchos de los

compuestos no carcinógenos se predicen como carcinógenos por diversos sistemas incluso cuando se utilizan predicciones basadas en sistemas expertos humanos.

Los resultados de la predicción, de los modelos obtenidos por métodos clásicos de QSAR, en estos ejercicios comparativos del NTP brindan un criterio suficiente para juzgar la capacidad verdadera de dichos modelos de predecir la carcinogénesis a productos químicos no comprobados.

En la Tabla 5.3 se muestran los resultados de clasificación global del primer ejercicio así como su respectiva probabilidad.

Tabla 5.3 Resultados de la clasificación en la serie de predicción para el primer ejercicio de la NTP, con los Índices Lineales no-estocásticos.

Compuestos	Actividad	AP%	Clas
naphthalene	+	80.94	+
o-benzyl-p-chlorophenol	+	44.39	+
2.2-bis(bromomethyl)-1.3-propanediol	+	4.06	+
t-butyl alcohol	+	71.69	+
3.4-dihydrocoumarin	+	11.75	+
mercuric chloride	+	83.73	+
Methylphenidate Hydrochloride	+	-70.87	-
Triamterene	+	-9.63	-
diphenylhydantoin	+	-84.73	-
pentachloroanisole	+	11.90	+
p-nitrobenzoic acid	+	-10.69	-
tris(2-chloroethyl)phosphate	+	40.87	+
direct blue 218	+	100.00	+
C.I. Pigment Red 3	+	25.49	+
2.4-diaminophenol dihydrochloride	+	85.31	+
salicylazosulfonylpyridine	+	44.61	+
C.I. Acid Red 114	+	-85.90	-
C.I. direct blue 15	+	-99.93	-
coumarin	+	-8.67	-
2.3-dibromo-1-propanol	+	75.13	+
3.3'-dimethylbenzidine	+	67.95	+
o-nitroanisole	+	33.31	+
1.2.3-trichloropropane	+	73.08	+
dl-amphetamine sulfate	-	56.01	+
promethazine hydrochloride	-	-6.92	-
resorcinol	-	68.88	+
gamma-butyrolactone	-	-11.74	-
manganese sulfate monohydrate	-	85.96	+
monochloroacetic acid	-	57.47	+
p-nitrophenol	-	65.00	+
tricresyl phosphate	-	-65.94	-
ethylene glycol	-	70.02	+
theophyllin	-	-93.59	-

Tabla 5.3 (Cont)

Compuestos	Actividad	$\Delta P\%$	Clas
4,4'-thiobis(6-t-butyl-m-cresol)	-	-23.16	-
chloramine	-	86.20	+
4,4'-diamino-2,2'-stilbene disulfonic acid	-	-87.04	-
methyl bromide	-	66.86	+
sodium azide	-	91.97	+
C.I. pigment red 23	-	-66.60	-
4'-Hydroxyacetanilide	-	44.05	+
titanocene dichloride	-	100.00	+
HC yellow 4	-	51.42	+
p-nitroaniline	-	71.16	+

En la Tabla 5.4 se muestran los resultados de clasificación global para el segundo ejercicio así como su respectiva probabilidad, demostrando un buen porcentaje de clasificación.

Tabla 5.4 Resultados de la clasificación de los compuestos de la serie de predicción en el segundo ejercicio para Índices Lineales no-estocásticos.

Compuestos	Actividad	$\Delta P\%$	Clas
oxymethalone	+	-81.23	-
primaclone	+	-75.76	-
phenolphthalein	+	-35.89	-
D&C yellow no	+	-12.33	-
anthraquinone	+	0.57	+
1,2-dihydro-2,2,4-trimethylquinoline	+	62.45	+
chloroprene	+	50.15	+
isobutene	+	44.32	+
methyleugenol	+	-1.46	-
nitromethane	+	80.96	+
pyridine	+	68.98	+
diethanolamine	+	64.85	+
furfuryl alcohol	+	56.08	+
tetrahydrofuran	+	35.93	+
ethylene glycol monobutyl ether	+	25.00	+
ethylbenzene	+	39.47	+
scopolaminehydrobromide trihydrate	-	-75.50	-
codeine	-	-22.56	-
emodin	-	-28.98	-
t-butylhydroquinone	-	45.61	+
iso-butyraldehyde	-	17.40	+
1-chloro-2-propanol	-	70.54	+
xylene sulfonic acid, sodium salt	-	4.70	+
cinnamaldehyde	-	19.34	+
sodium nitrite	-	87.36	+
citral	-	-45.45	-

5.1.2 Desarrollo de Modelos QSAR-ADL con Índices Lineales estocásticos

Posteriormente, otra FD fue obtenida usando una combinación del ADL y los índices lineales estocásticos. Esta ecuación se muestra a continuación:

$$\begin{aligned} \text{Class} = & 2.2673 - 0.167 {}^P f_0^H(x) + 0.171 {}^P f_3(x) - 6.641 {}^P f_4(x_{E-H}) + 17.058 {}^P f_8(x_{E-H}) \\ & - 9.479 {}^P f_{12}(x_{E-H}) - 1.214 {}^P f_{15}(x_{E-H}) - 3.805 {}^G f_{15}^H(x_E) - 1.050 {}^G f_2(x_E) \\ & - 2.658 {}^G f_3(x_E) + 7.426 {}^G f_5(x_E) \end{aligned} \quad (5.2)$$

$$N = 189 \quad \lambda = 0.63 \quad D^2 = 2.42 \quad F(10,178) = 10.1740 \quad p < 0.0001$$

En este estudio también utilizamos las ponderaciones de cuatro propiedades atómicas: 1) masa atómica (M), 2) polarizabilidad atómica (P), 3) electronegatividad atómica de Pauling (G) y 4) volumen de van der Waals atómico (V) para el cálculo de los descriptores moleculares. El modelo desarrollado con los índices lineales estocásticos (Ec. 5.2) clasifica correctamente el 79.37% de los compuestos en la SE, clasificando incorrectamente solo 39 entidades de un total de 189. En este modelo se obtuvo una exactitud global para las SP externas de 60.46%, para el primer ejercicio (17/43), y 74.42% para el segundo ejercicio se obtuvo (11/26). Estos resultados muestran la estabilidad y el poder predictivo de la función de clasificación obtenida con los índices lineales estocásticos.

Los resultados obtenidos en el desarrollo de este modelo se presentan de forma resumida en la Tabla 5.5. El conjunto completo de compuestos incluidos en las series de entrenamiento y de predicción.

Tabla 5.5 Resultados de la Clasificación de los Compuestos Incluidos en las Series de Entrenamiento y de Predicción Obtenidos Utilizando los Índices Lineales Estocásticos.

	Exactitud 'Q _{Total} ' (%)	Sensibilidad (%)	Especificidad (%)	"Hit" de Falsos Positivos (%)
índices cuadráticos estocásticos				
SE	79.37	81.51	85.09	24.29
SP: 1 ^{er} Ejercicio	60.47	73.91	60.71	55.00
SP: 2 ^{do} Ejercicio	57.69	81.25	61.91	80.00

En la Tabla 5.2 se da una explicación mas detallada de clasificación de la serie de entrenamiento con relación a los falsos posito y los falsos negativos, así como su probabilidad correspondiente.

Tabla 5.6 Resultados de la clasificación para los compuestos de la serie de entrenamiento utilizando los Índices Lineales no-estocásticos.

Compuestos	$\Delta P\%$	Clas	Compuestos	$\Delta P\%$	Clas
<i>Compuestos carcinógenos</i>					
[4-Chloro-6-(2,3-xylidino)-2-pyrimidinylthio]acetic acid	68.68	-	Diethylstilbestrol	44.71	+
1-(2-Hydroxyethyl)-1-nitrosourea	6.45	+	Dihydrosafrole	28.51	+
1,1-Dimethylhydrazine	32.27	+	dl-Ethionine	33.23	+
1,2-Dibromoethane	5.51	+	Doxylamine succinate	51.63	-
1,2-Dimethylhydrazine	7.20	+	Ethylene thiourea	57.42	-
1'-Hydroxysafrole	29.29	+	Formaldehyde	28.94	+
1-Nitroso-3,4,5-trimethylpiperazine	4.36	+	Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	46.61	+
2,2,2-Trifluoro-N-[4-(5-nitro-2-furyl)-2-thiazolyl]acetamide	14.95	+	Glu-P-1	22.65	+
2,3,7,8-Tetrachlorodibenzo-p-dioxin	18.16	+	Glu-P-2	17.00	+
2,4,5-Trimethylaniline	15.35	+	Hexachlorobenzene	33.75	+
2,4,6-Trimethylaniline	18.68	+	hydrazine	1.86	+
2,4-Diaminotoluene	5.70	+	hydrazine sulfate	34.07	+
2,4-Dinitrotoluene, practical grade	15.44	+	IQ	21.27	+
2,5-Dimethoxy-4'-aminostilbene	23.61	+	Isoniazid	29.04	+
2,5-Xylidine	14.19	+	Lead Acetate, basic	37.24	+
2-Acetylaminofluorene	24.50	+	MeA-alpha-C acetate	19.75	+
2-Amino-4-(5-nitro-2-furyl)thiazole	25.87	+	MeIQx	19.73	+
2-Hydrazino-4-(5-nitro-2-furyl)thiazole	3.39	+	Melphalan	28.38	+
2-Hydrazino-4-(p-aminophenyl) thiazole	3.27	+	Methyl tert-butyl ether	22.04	+
2-Hydrazino-4-(p-nitrophenyl) thiazole	4.48	+	Metronidazole	43.96	+
2-Naphthylamine	8.80	+	N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide	65.82	-
3-(5-Nitro-2-furyl)-imidazo(1,2-alpha)pyridine	16.01	+	N-[4-(5-Nitro-2-furyl)-2-thiazolyl]formamide	60.44	-
3,3',4,4'-Tetraaminobiphenyl	4.14	+	N-[5-(5-Nitro-2-furyl)-1,3,4-thiadiazol-2-yl]acetamide	67.29	-
3-Aminotriazole	16.39	+	N-Hydroxy-2-acetylaminofluorene	19.10	+
3-Nitro-3-hexene	17.49	+	Nitrobenzene	9.83	+
4-Chloro-4'-aminodiphenylether	12.19	+	Nitroso-2,3-dihydroxypropyl-2-oxopropylamine	9.11	+
5-Azacytidine	49.81	+	Nitroso-2-oxopropylethanolamine	4.92	+
Acetaldehyde	16.29	+	Nitrosodibutylamine	5.92	+
Acetamide	23.35	+	N-Nitroso-2,3-dihydroxypropyl-2-hydroxypropylamine	8.24	+
Acetaminophen	53.96	-	N-Nitrosoallyl-2-oxopropylamine	4.23	+
AF-2	8.01	+	N-Nitrosodimethylamine	4.27	+
Aldrin	22.16	+	N-Nitrosomethyl-2,3-dihydroxypropylamine	5.19	+
alpha-1,2,3,4,5,6-Hexachlorocyclohexane	5.62	+	N-Nitrosomorpholine	2.22	+
Aramite	19.53	+	N-Nitroso-N-methylurea	4.96	+
Auramine-O	72.11	-	N-Nitrosopiperidine	2.91	+
Azobenzene	20.49	+	N-Nitrosopyrrolidine	2.85	+
Benzidine	5.30	+	o-Phenylenediamine	5.32	+
Benzo(a)pyrene	25.52	+	o-Toluidine	4.00	+
beta-Propiolactone	29.76	+	p,p'-DDE	47.54	+

Tabla 5.6 (Cont)

Compuestos	$\Delta P\%$	Clas	Compuestos	$\Delta P\%$	Clas
Bis-(chloromethyl)ether	77.61	-	Phenacetin	64.17	-
Bis-2-hydroxyethylthiocarbamic acid,potassium	81.37	-	Phenobarbital, sodium	75.70	-
Butylated hydroxyanisole	47.53	+	PhIP	19.70	+
Caffeic acid	47.44	+	Piperonyl butoxide	58.56	-
Captafol	15.22	+	Propylthiouracil	60.27	-
Captan	39.80	+	Safrole	24.09	+
Carbon tetrachloride	79.49	-	Sesamol	20.46	+
Catechol	17.72	+	Sterigmatocystin	67.36	-
Chlorambucil	51.84	-	Streptozotocin	75.13	-
Chloroform	63.21	-	Styrene oxide	18.22	+
Chloromethyl methyl ether	59.74	-	Tamoxifen citrate	64.44	-
Chrysazin	41.44	+	Thioacetamide	51.76	-
Ciprofibrate	38.19	+	Thiouracil	45.34	+
Cyclophosphamide	30.27	+	Trichloroethylene	11.12	+
D & C red no	34.88	+	Trp-P-1 acetate	30.96	+
DDT	22.05	+	Trp-P-2 acetate	19.20	+
Diallylnitrosamine	4.44	+	Uracil	48.42	+
Dibromodulcitol	7.50	+	Urethane	11.56	+
Dibromomannitol	7.50	+	Vinyl acetate	24.21	+
Dichloroacetylene	14.98	+	Vinyl chloride	64.17	+
Dieldrin	32.20	+			
<i>Compuestos no carcinógenos</i>					
1,1,1-Trichloroethane(grado tec	11.55	+	FD & C blue no	98.67	-
1,2-Dichloroethane	12.29	+	FD & C blue no	77.88	-
1,4-Dichlorobenzene	4.01	+	FD & C green no	99.21	-
1-[(5-Nitrofurfurylidene)amino]hydantoin	13.19	-	FD & C red no	65.62	-
1-Chloro-2,4-dinitrobenzene	60.55	+	FD & C yellow no	94.32	-
2,3,4,5,6-Pentachlorophenol (Dowicide EC-7)	10.18	+	FD & C yellow no	79.10	-
2,4,5-Trichlorophenoxyacetic acid	29.07	+	Fenvalerate	92.94	-
3-nitro-4-hydroxyphenylarsonic acid	13.58	-	Flecainide acetate	98.77	-
5,5-Diphenylhydantoin	65.36	-	Fluoxetine	96.41	-
6-Dimethylamino-4,4-diphenyl-3-heptanone	74.06	-	Hexamethylenetetramine	89.86	-
Adipamide	71.26	+	Isopropyl-N-(3-chlorophenyl) carbamate	79.76	-
Aspirin	42.24	-	Maleic hydrazide	43.97	+
Barium Acetate	73.95	-	Methotrexate	80.75	-
Benzalazine	80.78	-	Misoprostol	96.97	-
Benzoate, sodium	91.94	+	Monochloroacetic acid	8.70	+
Benzoguanamine	41.40	+	Nefiracetam	71.92	-
Black PN	12.95	-	N-Nitrosocimetidine	77.56	-
C.I. pigment yellow 12	99.88	-	Oxamyl	76.05	-
C.I. pigment yellow 16	85.27	-	Oxprenolol	81.01	-
C.I. pigment yellow 83	95.65	-	Phenyl Isothiocyanate	26.32	+
Cadmium Acetate	95.70	-	Prazepam	63.42	-
Caffeine	58.61	-	Praziquantel	78.19	-
Chlorodifluoromethane	94.21	-	Propranolol	83.61	-
Chromium (Iii) Acetate	87.04	-	Quercetin dihydrate	47.49	+
Compound 50-892	95.43	-	Rutin trihydrate	88.08	-

Tabla 5.6 (Cont)

Compuestos	$\Delta P\%$	Clas	Compuestos	$\Delta P\%$	Clas
Cyclohexylamine	27.83	+	Saccharin	58.83	-
Deltamethrin	96.22	-	Sorbic acid	33.13	+
Diazepam	54.15	-	Sotalol	83.05	-
Dichlorodifluoromethane	93.54	-	Tetramethylthiuram disulfide	69.35	-
Dimethylformamide	63.13	-	Tilidine fumarate	87.16	-
Diphenyl-p-phenylenediamine	45.06	+	toluene diisocyanate	84.26	-
Erythorbate, sodium	78.77	-	Trichlorofluoromethane	88.22	-
Estazolam	50.58	-	Trifluralin, technical grade	99.34	-
Ethynodiol diacetate	97.76	-	Triprolidine	57.07	-
Etodolac	92.27	-	Urea	8.91	+

La clasificación de los dos ejercicio se ve mejorada en los índices lineales estocásticos, pues los resultados muestran mejores clasificaciones en este modelo que los expuestos anteriormente, mostrándose de mejor forma en la Tablas 5.7 y 5.8 el primer y segundo ejercicio respectivamente.

Tabla 5.7 Resultados de la clasificación para los compuestos de la serie de predicción con los Índices Lineales estocásticos.

Compuestos	Actividad	$\Delta P\%$	Clas
naphthalene	+	19.06	+
o-benzyl-p-chlorophenol	+	27.74	+
2,2-bis(bromomethyl)-1,3-propanediol	+	34.88	+
t-butyl alcohol	+	36.71	+
3,4-dihydrocoumarin	+	12.07	+
mercuric chloride	+	48.23	+
Methylphenidate Hydrochloride	+	86.83	-
Triamterene	+	9.17	+
diphenylhydantoin	+	81.51	-
pentachloroanisole	+	73.09	-
p-nitrobenzoic acid	+	47.09	+
tris(2-chloroethyl)phosphate	+	28.85	+
direct blue 218	+	99.60	-
C.I. Pigment Red 3	+	25.13	+
2,4-diaminophenol dihydrochloride	+	3.86	+
salicylazosulfonylpyridine	+	42.67	+
C.I. Acid Red 114	+	99.45	-
C.I. direct blue 15	+	99.26	-
coumarin	+	37.34	+
2,3-dibromo-1-propanol	+	3.46	+
3,3'-dimethylbenzidine	+	9.63	+
o-nitroanisole	+	14.95	+
1,2,3-trichloropropane	+	5.85	+
dl-amphetamine sulfate	-	62.92	-
promethazine hydrochloride	-	50.16	-
resorcinol	-	15.25	+
gamma-butyrolactone	-	42.19	+

Tabla 5.7 (Cont)

Compuestos	Actividad	$\Delta P\%$	Clas
manganese sulfate monohydrate	-	53.48	-
monochloroacetic acid	-	11.86	+
p-nitrophenol	-	7.85	+
tricresyl phosphate	-	78.42	-
ethylene glycol	-	11.58	+
theophyllin	-	96.65	-
4,4'-thiobis(6-t-butyl-m-cresol)	-	84.17	-
chloramine	-	2.99	+
4,4'-diamino-2,2'-stilbene disulfonic acid	-	36.75	+
methyl bromide	-	21.42	+
sodium azide	-	10.52	+
C.I. pigment red 23	-	80.75	-
4'-Hydroxyacetanilide	-	50.31	-
titanocene dichloride	-	29.91	+
HC yellow 4	-	79.84	-
p-nitroaniline	-	4.64	+

Tabla 5.8 Resultados de la clasificación para los compuestos de la serie de predicción utilizando los Índices Lineales estocásticos.

Compuestos	Actividad	$\Delta P\%$	Clas
oxymethalone	+	81.28	-
primaclone	+	78.49	-
phenolphthalein	+	61.85	-
D&C yellow no	+	46.12	+
anthraquinone	+	32.66	+
1,2-dihydro-2,2,4-trimethylquinoline	+	41.77	+
chloroprene	+	19.66	+
isobutene	+	19.73	+
methyleugenol	+	30.26	+
nitromethane	+	1.51	+
pyridine	+	13.67	+
diethanolamine	+	31.86	+
furfuryl alcohol	+	14.41	+
tetrahydrofuran	+	20.23	+
ethylene glycol monobutyl ether	+	22.91	+
ethylbenzene	+	23.00	+
scopolaminehydrobromide trihydrate	-	65.41	-
codeine	-	60.65	-
emodin	-	49.27	+
t-butylhydroquinone	-	33.18	+
iso-butyraldehyde	-	33.70	+
1-chloro-2-propanol	-	11.73	+
xylene sulfonic acid, sodium salt	-	12.94	+
cinnamaldehyde	-	29.62	+
sodium nitrite	-	11.21	+
citral	-	43.08	+

5.1.3 Comparación entre los Índices Lineales Moleculares y otros Enfoques Teóricos

Los resultados de predicción de modelos teóricos en los dos ejercicios comparativos del NTP son muy importantes para juzgar la verdadera capacidad de predecir, el potencial carcinogénico a productos químicos que aun no han sido ensayados, de dichos modelos. Los modelos predictivos obtenidos para series de productos químicos no congenéricos, lograron una exactitud máxima de alrededor 65-70%; sin embargo, este máximo no fue logrado fácilmente, y muchos enfoques, incluyendo varios disponibles comercialmente, demostraron que su capacidad predictiva era limitada al ser enfrentadas a estos dos ejercicios. En ambos ejercicios, el mejor resultado fue logrado por los enfoques que confiaron en gran parte en el juicio experto humano. No obstante, los métodos clásicos de QSAR pueden generar modelos satisfactorios para clases individuales de agentes mutágenos y carcinógenos, con una exactitud en la gama de 80-90%. Las razones de las dificultades encontradas por los diferentes enfoques generales han sido discutidas en varias publicaciones (Benigni 2000; Richard 1994; Richard y Benigni 1998). Resumiendo, algunas de dichas razones son las siguientes: a) mientras que los modelos clásicos de QSAR describen una clase de carcinogenesis y un mecanismo principal a la vez, los acercamientos generales apunta modelar varios mecanismos al mismo tiempo, que es una tarea más difícil; b) la base de datos de experimentos en roedores (alrededor 1.500 productos químicos bioensayados) no es suficientemente representativo en términos de número de productos químicos por mecanismo/clase química; c) la evolución y los cambios de los tipos químicos de interés (drogas nuevas, pesticidas nuevos, etc...) plantea continuamente nuevos desafíos a los que no puede hacer frente basándose en la información proporcionada por la base de datos histórica de carcinogénesis. Hasta cierto punto, los enfoques basados en sistemas de expertos humanos podían superar estas dificultades por su capacidad de tener acceso a varias fuentes de la información de una manera flexible.

En este sentido varias investigaciones han sido realizadas con varios descriptores moleculares y el ADL para predecir la actividad carcinógena. En la Tabla 5.9 y 5.10 se muestran los principales resultados obtenidos con estos enfoques computacionales y los archivados por los modelos desarrollados en este trabajo utilizando los descriptores **TOMOCOMD-CARDD**. Como se puede observar, los modelos reportados en esta tesis son similares a muchos de los modelos obtenidos previamente. Teniendo en cuenta que muchos de estos involucraron la utilización de sistemas basados en el conocimiento de expertos humanos podemos decir que los índices lineales estocásticos y no-estocásticos pueden ser utilizados como una herramienta alternativa para este tipo de estudios.

Tabla 5.9 Comparación entre los Modelos Desarrollados en este Estudio con Otros Enfoque Computacionales, (primer ejercicio de la NTP)

Método de Predicción	Predicción en %
Ten	75
RAS	68
Wei	65
Ke	65
Índices Lineales estocásticos	60
DER	59
TOP	57
Ben	57
Sty	57
Deh	56
TOPS-MODE	55
Lij	55
Índices Lineales no-estocásticos	53
COM	53
CAS	49

Tabla 5.10 Comparación entre los Modelos Desarrollados en este Estudio con Otros Enfoque Computacionales, (segundo ejercicio de la NTP).

Método de Predicción	Predicción en %
Oncologic	65
SHE	65
R1	64
TOPS-MODE	63
Huff y otros	62
R2	61
Benigni y otros	61
Tennant y otros	60
Índices Lineales estocásticos	58
Índices Lineales no-estocásticos	58
Ashby	57
Bootman	53
FALS	50
Benigni	48
RASH	45
Ke(estimado)	44
COMPACT	43
DERECK	43
Salmonella	33
Purdy	32
PROGOL	29
CAS	25

6

CONCLUSIONES

- ✓ *Se Identificaron los métodos (in vitro, in vivo, etc) que permiten predecir la actividad carcinogénica de compuestos.*
- ✓ *Se Identificaron las bases teóricas de interés para los debates de la comunidad científica relacionado con carcinogenesis.*
- ✓ *Se comprobó la interrelación existente entre el efecto carcinogénico y el diseño experimental preclínico que se aplica a los compuestos químicos.*
- ✓ *Se desarrollaron modelos de clasificación para discriminar la actividad carcinogénica de compuestos químicos utilizando los índices lineales moleculares.*
- ✓ *Se validó la calidad y robustez de los modelos desarrollados utilizando dos series de predicción externas.*
- ✓ *Se compararon nuestros resultados con otros reportados en la literatura obteniéndose resultados promisorios.*

7

RECOMENDACIONES

- ✓ *Continuar el desarrollo de este trabajo con los índices cuadráticos y bilineales basados en átomos así como a los índices cuadráticos, lineales y bilineales pero en esta ocasión basados en enlace.*
- ✓ *Comparar estos modelos con otros métodos donde se empleen otras técnicas estadísticas de clasificación diferentes del análisis discriminante lineal.*

8

REFERENCIAS BIBLIOGRÁFICAS

A

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.

Amacher, D.E. & Zelljadt, I. The morphological transformation of Syrian hamster embryo cells by chemicals reportedly nonmutagenic to *Salmonella typhimurium*. *Carcinogenesis* 4. (1983) 291–295.

Ashby, J. Alternatives to the 2-species bioassay for the identification of potential human carcinogens. *Human and Experimental Toxicology* 15. 1996, 183–202.

Axler, S. *Linear Algebra Done Right*. Springer-Verlag: New York, 1996, pp 37-70.

B

Balaban, A., T. (Ed.), *Chemical Applications of Graph Theory*; Academic Press, London, 1976.

Barrett, J.C. & Fletcher, W.F. Cellular and molecular mechanisms of multi-step carcinogenesis in cell culture models. In *Mechanisms of Environmental Carcinogenesis*, Vol. 2, Multistep Models of Carcinogenesis. 1987, pp. 73-116. Boca Raton, FL, USA: CRC Press. 764 R.

Barrett, J.C., Kakanuga, T., Kuroki, T., Neubert, D., Trosko, J.E., Vasilieu, J.M., Williams, G.M. & Yamasaki, H. Mammalian cell transformation in culture. In *Long and Short Term Test Assays for Carcinogens: A Critical Appraisal. IARC Scientific Publication No. 83* (ed. R. Monsanto, H. Bartsch, H. Vaimo, J. Wilbourn & H. Yamasaki). 1986, pp. 267-286. Lyon, France: IARC.

Basak, S., C.; Balaban, A., T.; Grunwald, G., D.; Gute, B., D. *J. Chem. Inf. Comput. Sci.* 2000, 40, 891.

Basak, S., C.; Grunwald, G., D. *J. Chem. Inf. Comput. Sci.* 1995, 35, 366.

Basak, S., C.; Grunwald, G., D. *SAR QSAR Environ. Res.* 1994, 2, 289.

- Basak, S., C.; Grunwald, G., D. *SAR QSAR Environ. Res.* 1995, 3, 265.
- Basak, S., C.; Gute, B., D. *SAR QSAR Environ. Res.* 1997, 7, 1.
- Basak, S., C.; Magnuson, V., R.; Niemi, G., J.; Regal, R., R. *Discrete Appl. Math.* 1988, 19, 17.
- Belsey, D., A.; Kuh, E.; Welsch, R., E. *Regression Diagnostics*; Wiley: New York, 1980.
- Benigni, R. 'Structure-activity relationships in mutagenesis and carcinogenesis', in M. Balls, A.M. van Zeller and M.E. Halder (eds) *Progress in the Reduction, Refinement and Replacement of Animal Experimentation*, Amsterdam: Elsevier Science, 2000, pp. 469-478.
- Benigni, R. and Richard, A.M. 'Quantitative structure-based modeling applied to characterization and prediction of chemical toxicity', *Methods.* 1998,14: 264-276
- Berwald, Y. & Sachs, L. In vitro transformation of normal cells to tumour cells by carcinogenic hydrocarbons. *Journal of the National Cancer Institute.* 1965, 35, 641-661.
- Boukamp, P., Dzarlieva-Petrusevska, R.T., Breitkreutz, D., Hornung, J., Markham, A. & Fusenig, N.E. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *Journal of Cell Biology.* 1988, 106, 761-771.
- Boukamp, P., Popp, S., Altmeyer, S., Hülsen, ., Fasching, C., Cremer, T. & Fusenig, N.E. Sustained non-tumorigenic phenotype correlates with a largely stable chromosome content during long-term culture of the human keratinocyte cell line HaCaT. *Genes Chromosomes and Cancer.* 1997, 19, 201-214.
- Breitkreutz, D., Schoop, V.M., Mirancea, N., Baur, M., Stark, H-J. & Fusenig, N.E. Epidermal differentiation and basement membrane formation by HaCaT cells in surface transplants. *European Journal of Cell Biology.* 1998, 75, 273-286
- Browder, A. *Mathematical Analysis. An Introduction.* Springer-Verlag, New York, Inc. 1996, pp 176-296
- Busacker, R., G.; Saaty, T. *Finite Graphs and Networks.* McGraw-Hill: New York, 1965.

C

- Cannon, R.E. The Tg.AC mouse model passes test by failing to respond. *Toxicological Sciences.* 2003, 74, 233-234.
- 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.

Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. Atom-based 3D-Chiral Quadratic Indices: 2. Prediction of the Corticosteroid-Binding Globulin Binding Affinity of the Thirty-one Benchmark Steroids data set. *Bioorg. Med. Chem.* 2006, *14*, 2398-2408.

Cativiela, C.; Garcia, J., I.; Fernandez-Alvarez, E.; Elorriaga, C. *Acta Cim. Hung.* 1992, *130*, 129.

Celeste-Vega, M.; Montero-Torres, A.; Marrero-Ponce, Y.; Rolón, M.; Gómez-Barrio, A.; Escario, J. A.; Arán, V. J.; Martínez-Fernández, A. R., Meneses-Marcel, A. New Ligand-Based Approach for the Discovery of Antitrypanosomal Compounds. *Bioorg. Med. Chem. Letter.* 2006, *16*, 1898–1904.

Codina, J. Ensayos microbianos para la detección de sustancias genotóxicas, 2004.

Columbano, Ledda-Columbano et. al. Chemically Induced Cell Proliferation and Carcinogenesis. In: Chemically induced cell proliferation: implications for risk assessment. B. E. Butterworth T. J. Slaga (eds.). Wiley-Liss Inc., New York. Páginas, 217-275. 1991.

Colville P y col: “Grow Factors in Angiogenesis: Current interest and therapeutic potential” *Molecular Medicine Today*. Jan/97. 14:23

Combes, R., Balls, M., Curren, R., Fischbach, M., Fusenig, N., Kirkland, D., Lasne, C., Landolph, J., LeBoeuf, R., Marquardt, H., McCormic, J., Muller, L., Rivedal, E., Sabbioni, E., Tanaka, N., Vasseur, P. & Yamasaki, H. 1999. Cell transformation assays as predictors of human carcinogenicity: The report and recommendations of ECVAM Workshop 39. 1999, *ATLA* 27, 745–767.

Combes, R., Schechtman, L., Stokes, W.S. & Blakey, D. The international symposium on regulatory testing and animal welfare: recommendations on best scientific practices for subchronic/chronic toxicity and carcinogenicity testing. *ILAR Journal* 43, 2002, Suppl. 1, 112–117.

Conolly, R.B., Reitz, R.H., Clewell, H.J. 3rd & Andersen, M.E. Pharmacokinetics, biochemical mechanism and mutation accumulation: a comprehensive model of chemical carcinogenesis. *Toxicology Letters* 43. 1988, 189–200.

Cronin, M., T., D.; Schultz, T., W. *J. Mol. Struct. (Theochem)*. 2003, 622, 39.

Cvetkovic, D., M.; Doob, M.; Sachs, H. *Spectra of Graphs. Theory and Applications*, 3rd edition, Johann Ambrosius Barth, Heidelberg, 1995.

D

Domínguez, L., Principios generales de carcinogénesis química y hormonal. 2004, 2-9

Duffus, J.H. (Ed.) "Environmental Toxicology and Ecotoxicology", Environmental Health Series No.10, World Health Organization, Regional Office For Europe, Copenhagen. 1986

E

Edwards, C., H.; Penney, D., E. *Elementary Linear Algebra*; Prentice-Hall, Englewood Cliffs: New Jersey, USA, 1988.

Egan, W., J.; Morgan, S., L. *Anal. Chem.* 1998, 70, 2372.

Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* 2000, 43, 3714-3717.

Estrada, E. *J. Chem. Inf. Comput. Sci.*, 35. 1995, 31.

Estrada, E., *J. Chem. Inf. Comput. Sci.*, 36. 1996, 844.

Estrada, E., Ramírez, A., *J. Chem. Inf. Comput. Sci.*, 36. 1996, 837.

Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L., De Clercq, E., *J. Med. Chem.* 43. 2000, 1975.

F

Fan, Y.; Shi, L., M.; Kohn, K., W.; Pommier, Y.; Weinstein, J., N. *J. Med. Chem* 44. 2001, 3254.

Fritzenschaf, H., Kohlpoth, M., Rusche, B. & Schiffmann, D. Testing of known carcinogens and noncarcinogens in the Syrian hamster embryo (SHE) micronucleus test *in vitro*: correlations with *in vivo* micronucleus formation and cell transformation. *Mutation Research*..1993, 319, 47-53.

Fusenig, N.E. & Boukamp, P. Carcinogenesis studies of human cells: reliable *in vitro* methods. In *Cell Culture in Pharmaceutical Research* (ed. N.E. Fusenig & H. Graf). 1994, pp. 79-102. Berlin, Germany: Springer.

Fusenig, N.E. & Boukamp, P. Multiple stages and genetic alterations in immortalization, malignant transformation, and tumor progression of human skin keratinocytes. *Molecular Carcinogenesis*.1998, 23, 144-158.

G

Gallo, Robert y col.: "Virus oncogénicos ARN". Seminarios de Oncología (1977) 126-148. Editorial Médica Panamericana. Buenos Aires.

García, P., Peña CC, Gutierrez C. Carcinogénesis Química. En "Toxicología de Postgrado" . M. Repetto. Ed. Área de Toxicología. Universidad de Sevilla. CD-ROM. Sevilla, 2003.

Gasteiger, J.; Marsilli, M., A New Model for Calculating Atomic Charge in Molecules. *Tetrahedron Lett.* 1978, 34, 3181-3184.

Gatehouse, D.G., Mutagenicity. In *International Pharmaceutical Product Registration — Aspects of Quality, Safety and Efficacy* (ed. A.C. Cartwright & B.R. Mathews). 1994, pp. 474-552. New York, USA: Ellis Horwood

Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* 1987, 27, 21-35.

Gironés, X.; Gallegos, A.; Ramon, C-D. *J. Chem. Inf. Comput. Sci.* 2000, 46, 1400.

Golbraikh, A.; Tropsha, A. *J. Comp. Aided Mol. Des.* 2002, 16, 357.

Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Modell.* 2002, 20, 269.

Gold, L. S.; Manley, N. B.; Slone, T. H.; Rohrbach, L. Supplement to the Carcinogenic Potency Database (CPDB): Results of Animal Bioassays Published in the General Literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ. Health Perspect. Suppl.* 1999, 107 (Suppl. 4), 527–602.

Gold, L. S.; Zeiger, E. *Handbook of Carcinogenic Potency and Genotoxicity Databases*; CRC: Boca Roca, FL, 1997.

Goodman, J.I. A perspective on current and future uses of alternative models for carcinogenicity testing. *Toxicologic Pathology* 29, Suppl. 2001, 1, 173–176.

Gorbátov, V., A. *Fundamentos de la Matematica discreta*. Ed., Mir. Moscú. URSS. Capitulo 3. 1988; p.102-180.

Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc, R.; Buelow, R. *Nature Biotech.* 1998, 16, 748.

Green M y col.: “Virus ADN: replicación, expresión del gene tumoral y papel en el cáncer humano”. *Seminarios de Oncología*. 1977, 103-125. Editorial Médica Panamericana. Buenos Aires.

Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986.

H

H. C. Pitot III and Y. P. Dragan. Chemical Carcinogenesis. In *Casaret and Doulls Toxicology* V edition. C D. Klaasen (eds). Mc Graw-Hill. New York. 1996, Páginas, 201-267..

Halachmi S y col: “Estrogen Receptor-Associated Proteins: Possible Mediators of Hormone-Induced Transcription”. *Science*; 1994. June3, Vol 264:1455-1458.

Harary, F. *Graph Theory*, Addison–Wesley, Reading, MA, 1971.

Heidelberger, C., Freeman, A.E., Pienta, R.J., Sivak, A., Bertram, J.S., Casto, B.C., Dunkel, V.C., Francis, M.W., Kakunaga, T., Little, J.B. & Schechtman, L.M. Cell transformation by chemical agents — a review and analysis of the literature. A report of the US Environmental Protection Agency Gene-Tox Program. *Mutation Research*, 1983, 114, 283-385.

Hunter, D. J., Hankinson, S. E., Laden, F., Colditz, G. A., Manson, J. E., Willett, W. C., Speizer, F. E., y Wolff, M. S. 1997. Plasma organochlorine levels and the risk of breast cancer. *N. Engl. J. Med.*, 337(18), 1253-1258.

I

I,A y col: “Human Papillomavirus DNA in Uterine Cervix Squamous Cell Carcinoma and Adenocarcinoma Deletec by Polymerase Chain Reaction” *Cancer*; 1996, June 1 Vol 77 (11) : 2275-2279.

IARC "IARC Monographs on the Evaluation of Carcinogenic Risk to Humans, Supplement 7", World Health Organization, International Agency for Research on Cancer, Lyon. 1987

IRIS "Integrated Risk Information System of the United States", EPA/600/6-86/032a, US. Environmental Protection Agency, Washington DC. 1986

K

Karelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons: New York. 2000.

Khan S y col.: “Estrogen Receptor Expression of Benign Breast Epithelium and Its Association with Breast Cancer” *Cancer Research*; 1994, February 15 (54): 993-997.

Kier L., B; Hall, L., H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, U. K, 1986.

Kier, L., B. *J. Pharm. Sci.* 1980, 69, 416.

Klein, D., J. Graph Theoretically Formulated Electronic–Structure Theory. *Internet Electron. J. Mol. Des.* 2003, 2, 814–834.

L

Laird P: “Oncogenic mechanisms mediated by DNA methylation. *Molecular Medicine Today*, May 1997. 223-229

Lajiness, M., S. In *Computational Chemical Graph Theory*; Rouvray, D. H. (Ed.); Nova Sci., New York. 299, 1990.

Landolph, J.R. Chemical transformation in C3H10T1/2 CC118 mouse embryo fibroblasts: historical background, assessment of the transformation assay, and evolution and optimization of the transformation assay protocol. In *Transformation Assay of Established Cell Lines: Mechanisms and Application*, IARC Scientific Publications No. 67 (ed. T. Kakunaga & H. Yamasaki). 1985, pp. 185-203. Lyon. France: IARC.

Landolph, J.R., Dews, P.M., Ozbun, L. & Evans, D.P. Metal-induced gene expression and neoplastic transformation. In *Toxicology of Metals* (ed. L. Magos & T. Suzuki). 1996, pp. 321-329. Boca Raton, FL, USA: CRC Press.

LeBoeuf, R.A., Kerckaert, G.A., Aardema, M.J., Gibson, D.P., Brauninger, R. & Isfort, R.J. The pH 6.7 Syrian hamster embryo cell transformation assay for assessing the carcinogenic potential of chemicals. *Mutation Research*. 1996, 356, 85-127.

LeBoeuf, R.A., Kerckaert, G.A., Aardema, M.J., Gibson, D.P., Brauninger, R. & Isfort, R.J. The pH 6.7 Syrian hamster embryo cell transformation assay for assessing the carcinogenic potential of chemicals. *Mutation Research*. 1996, 356, 85-127

LeBoeuf, R.A., Kerckaert, K.A., Aardema, M.J. & Isfort, R.J. Use of Syrian hamster embryo and Balb/c 3T3 cell transformation for assessing the carcinogenicity potential of chemicals. In *The Use of Short- and Medium-term Tests for Carcinogenic Hazard Evaluation*. IARC Scientific Publications No. 146 (ed. D.B. McGregor, J.M. Rice & S. Vemtt). 1999, pp. 409-425. Lyon, France: IARC.

Leon, J. y col. "Activación de los oncogenes por radiación y agentes químicos". Invest. y Ciencia, 143. Ago/88. 20-34.

Liotta, L. y col. "Factores de motilidad autocrina tumoral." Avances en Oncología. 1988. De Vita y Col. 33-49.

M

Maciag, Thomas: "Mecanismos moleculares y celulares de la angiogénesis." Avances en Oncología. 1990 De Vita y Col. 109-127.

Marković, S., Gutman, I., J. Mol. Struct. (Theochem), 235. 1991, 81.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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)

Marrero-Ponce, Y.; Castillo-Garit, J.A.; Nodarse, D. Linear Indices of the “Macromolecular Graph’s Nucleotides Adjacency Matrix” as a Promising Approach for Bioinformatics Studies. 1. Prediction of Paromomycin’s Affinity Constant with HIV-1 Ψ -RNA Packaging Region. *Bioorg. Med. Chem.* 2005, 13, 3397-3404.

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.

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.

Marrero-Ponce, Y.; Iyarreta-Veitia, 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.

Marrero-Ponce, Y.; Khan, M. T. H.; Casañola-Martin, G. M.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F. Atom-Based 2D Quadratic Indices in Drug Discovery of Novel Tyrosinase Inhibitors. Results of *in silico* Studies Supported by Experimental Results. *QSAR Comb. Sci.* Accepted for publication.

Marrero-Ponce, Y.; Khan, M. T. H.; Casañola-Martin, G. M.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F.; Rotondo, R. TOMOCOMD-CARDD Descriptors-Based Virtual Screening of Tyrosinase Inhibitors. 1. Evaluation of Different Classification Model Combinations Using Group, Atom-Type and Total Bilinear Indices. *J. Chem. Inf. Comput. Sci.* 2006. Submitted for publication.

Marrero-Ponce, Y.; Medina, R.; Castro, E. A.; de Armas, R.; González, H.; Romero, V.; Torrens, F. Protein Quadratic Indices of the “Macromolecular Pseudograph’s α -Carbon Atom Adjacency Matrix”. 1. Prediction of Arc Repressor Alanine-mutant’s Stability. *Molecules* 2004, 9, 1124–1147.

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.

Marrero-Ponce, Y.; Medina-Marrero, R.; Martínez, 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.

Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martínez, Y.; Romero-Zaldivar, V.; Castro, E. A. Atom, Atom-type, and Total Non-Stochastic and Stochastic Quadratic Fingerprints: A Promising Approach for Modeling of Antibacterial Activity. *Bioorg. Med. Chem.* 2005, 13, 2881-2899.

Marrero-Ponce, Y.; Meneses-Marcel, A.; Machado-Tugores, Y.; Escario, J. A.; Gómez-Barrio, A.; Montero Pereira, D.; Nogal-Ruiz, J. J.; Arán, V. J.; Martínez-Fernández, A. R.; Torrens, F.; Rotondo, R. Predicting Antitrichomonal Activity: A Computational Screening Using Atom-Based Bilinear Indices and Experimental Proofs. *Bioorg. Med. Chem.* Accepted for publication.

Marrero-Ponce, Y.; Meneses-Marcel, A.; Machado-Tugores, Y.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R., García Sánchez, R. N.; Montero-Torres, A.; and Francisco Torrens. A Computer-Based Approach to the Rational Discovery of New Antitrichomonas Drugs by Atom-Type Linear Indices. *Current Drug Discovery Technologies.* 2005, 2, 245-265.

Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitia, 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.

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.

Marrero-Ponce, Y.; Torrens, F. Bond, Bond-Type, and Total Linear Indices of the non-Stochastic and Stochastic Edge Adjacency Matrix. 1. Theory and QSPR Studies. *J. Phy. Chem. A.* Submitted for publication.

Marrero-Ponce, Y.; Torrens, F. Bond-Based Global and Local (Bond and Bond-Type) Quadratic Indices and Their Applications to Computer-Aided Molecular Design. 1. QSPR Studies of Octane Isomers. *J. Comput. Aided Mol. Des.* Accepted for publication.

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.

Marrero-Ponce, Y.; Torrens, F., Torres, L. Novel 2D TOMOCOMD-CARDD Descriptors: Atom-based Stochastic and non-Stochastic Bilinear Indices and their QSPR Applications. *Molecules*. Accepted for publication.

Masakazu T y col: "Quantitative Analysis of Vascular Endothelial Growth Factor in Primary Breast Cancer" *Cancer*; 1996, March 15, Vol 77 (6): 1101-1106.

Mauthe, R.J., Gibson, D.P., Bunch, R.T. & Custer, L. The Syrian hamster embryo (SHE) cell transformation assay: review of the methods and results. *Toxicologic Pathology*. 2001, 29, Suppl. 1, 138–146.

McClain, R.M., Keller, D., Casciano, D., Fu, P., MacDonald, J., Popp, J. & Sagartz, J. Neonatal mouse model: review of methods and results. *Toxicologic Pathology*. 2001, 29, Suppl. 1, 128–137.

McCormick, J.J. & Mäher, V.M. Malignant transformation of mammalian cells in culture, including human cells. *Environmental and Molecular Mutagenesis*. (1989), 14, 105-113.

Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R.; García Sánchez, R. N. A Linear Discrimination Analysis Based Virtual Screening of Trichomonacidal Lead-Like Compounds. Outcomes of *in silico* Studies Supported by Experimental Results. *Bioorg. Med. Chem Lett*. 2005, 17, 3838-3843.

Meyer, A.L. In vitro transformation assays for chemical carcinogens. *Mutation Research*. 1983, 115, 323-338.

Millar, K., J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* 1990, 112, 8533-8542.

Monro, A. (1996). Are lifespan rodent carcinogenicity studies defensible for pharmaceutical agents? *Experimental and Toxicologic Patholog.*, 1996, 48, 155–166

Montero-Torres, A.; Celeste Vega, A.; Marrero-Ponce, Y.; Rolón, M.; Gómez-Barrio, A.; Escario, J. A.; Arán, V. J.; Martínez-Fernández, A. R.; Meneses-Marcel, A. A Novel

Non-Stochastic Quadratic Fingerprints-based Approach for the “*in silico*” Discovery of New Antitrypanosomal Compounds. *Bioorg. Med. Chem.* 2005, 13, 6264-6275.

Montero-Torres, A.; García, R. N.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Nogal, J. J.; Martínez, A. R.; Arán, V. J.; Ochoa, C.; Meneses-Marcel, A.; Torrens, F. Non-stochastic Quadratic Fingerprints and LDA-Based QSAR Models in Hit and Lead Generation Through Virtual Screening: Theoretical and Experimental Assessment of a Promising Method for the Discovery of New Antimalarial Compounds. *Eur. J. Med. Chem.* 2006, 41, 483-493.

Morón, J., A.; Campillo, M.; Perez, V.; Unzeta, M.; Pardo, L. *J. Med. Chem.* 2000, 43, 1684.

Müller, L., Miltenburger, H.G., Marquardt, H. & Madle, S. The significance of *in vitro* cell transformation assays for routine testing. In *Current Issues in Genetic Toxicology*. 1993, pp. 102-113. Munich, Germany: Medizin-Verlag.

N

Norinder, U. *J. Chemometrics*. 1996, 10, 95.

Nurse, P., y Bissett, Y. Gene required in G1 for commitment to cell cycle and in G2 for control of mitosis in fission yeasts. *Nature*. 1981, 292, 558-560

O

Ogino, A.; Matsumura, S.; Fujita, T. *J. Med. Chem.* 1980, 23, 437.

P

Patel, H.; Cronin, M., T., D. *J. Chem. Inf. Comput. Sci.* 2001, 41, 1228.

Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: Ithaca (New York), 1939; 2-60.

Pienta, R.J., Poiley, J.A. & Lebherz, W.B. 3rd Morphological transformation of early passage golden Syrian hamster embryo cells derived from cryopreserved primary cultures as a reliable *in vitro* bioassay for identifying diverse carcinogens. *International Journal of Cancer*. 1977, 19, 642-655.

Potapov, V.M. Stereochemistry, Mir, Moscow, 1978.

R

Recanatini, M.; Cavalli, A.; Belluti, F.; Piazzzi, L.; Rampa, A.; Bisi, A.; Gobbi, S.; Valenti, P.; Andrisano, V.; Bartolini, M.; Cavrini, V. *J. Med. Chem.* 2000, 43, 2007.

REPETTO, M.; SANZ, P., Glosario de términos usados en toxicología. Notas para el usuario. 1995, 3- 162

Richard, A.M. ‘Application of SAR methods to non-congeneric data bases associated

with carcinogenicity and mutagenicity: issues and approaches', *Mutation Research*. 1994, 305: 73-97.

Robbins S y col: "Patología estructural y funcional" Capítulo VI "Trastornos de la Inmunidad" Capítulo VII "Neoplasia" 5ª Edición 1995.

Rodriguez.R. Las toxinas ambientales y el daño genetico. Volumen 3.1995

Romero, MP y col: "Cáncer y Metástasis, relación con radicales libres." Oncología. 1993, N° 3, Vol.3. 79-94.

Rose, K.; Hall, L., H.; Kier, L., B. *J. Chem. Inf. Comput. Sci.* 2002, 42, 651.

Rosemberg, S. y col.: "Linfocitos infiltrantes del tumor". Avances en Oncología. 1990 De Vita y Col. 37-63.

Rosemberg, Steven: "Inmunoterapia del cáncer" Invest. y Ciencia, 166. Jul/90. 26-42.

Rouvray, D., H. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976, pp 180-181.

S

S. C. Hasmall and R. A. Roberts. Hepatic ploidy, nuclearity and distribution of DNA synthesis: a comparison of nongenotoxic carcinogens with noncarcinogenic mitogens. *Toxicol Appl Pharmacol.* 144: 287-293. 1997.

Schneider BL y Kulezz-Martin M. (2004). Destructive cycles; the role of genomic instability and adaptation in carcinogenesis. *Carcinogenesis*.

Shao, J. *J. Amer. Stat. Assoc.* 1993. 88, 486.

Sidransky David: "Progresos en la detección del cáncer" - Investigación y Ciencia, 1996 noviembre, Pag: 54-59

STATISTICA ver. 5.5, Statsoft, Inc. 1999.

StatSoft, Inc., 2002. STATISTICA 6.0, version 6.0.

Stone, M. *J. Roy. Stat. Soc.* 1974, 36, 111.

Suzuki, T.; Ide, K.; Ishida, M.; Shapiro, S. *J. Chem. Inf. Comput. Sci.* 2001, 41, 718.

T

Teramoto B y col: "Epstein-Barr Virus Infección in the Neoplastic and Nonneoplastic Cells of Lymphoid Malignancies" *Cancer*, 1996; June 1, Vol 77 (11): 2339-2347

Thompson, K.L. & Sistare, F.D. (2003). Selection of drugs to test the specificity of the Tg.AC assay by screening for induction of the gadd153 promoter *in vitro*. *Toxicological Sciences*. 2003, 74, 260-270.

Todeschini, R.; Consonni, V. (Eds.) *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim (Germany), 2000.

Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; 2nd ed.; 1992; pp 32-33.

Tsutsui, T., Hayashi, N., Maizumi, H., Huff, J. & Barrett, J.C. Benzene-, catechol-hydroquinone- and phenol-induced cell transformation, gene mutations, chromosome aberrations, aneuploidy, sister-chromatid exchanges and unscheduled DNA synthesis in Syrian hamster embryo cells. *Mutation Research*. 1997, 373, 113-123.

V

van de Waterbeemd, H. (Ed.); *Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry, Vol. 2)*, John Wiley & Sons: New York. 1995.

van de Waterbeemd, H.; Carter, R., E.; Grassly, G.; Kubinyi, H.; Martin, Y., C.; Tute, M., S.; Willett, P. *Annu. Rep. Med. Chem.* 1998, 33, 397.

W

Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Perspect. Drug Discov. Des.* 2000, 19, 47-66.

WHO "Principles and Methods for Evaluating the Toxicity of Chemicals, Part I", Environmental Health Criteria 6, World Health Organization, Geneva.1978a

WHO "Agreed Terms on Health Effects Evaluation and Risk and Hazard Assessment of Environmental Agents", Internal Report of a Working Group, (EHE/EHC/79.19), World Health Organization, Geneva.1979.

WHO "Glossary of Terms on Chemical Safety for Use in IPCS Publications", World health Organization, Geneva.1989a

Wilson, R., J. *Introduction to Graph Theory*. Oliver and Boyd: Edinburgh, 1972.

Wold, S. *Technometrics*. 1978, 20, 397.

Wold, S; Erikson, L. Statistical Validation of QSAR Results. Validation Tools. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995; p. 309-318.

Y

Yuspa, S.H. & Poirier, M.C. Chemical carcinogenesis: from animal models to molecular models in one decade. *Advances in Cancer Research*, 1988, 50, 25-70.

Z

Zhang, H., Borman, H.D. & Myhr, B.C. Enhancement of the morphological transformation of Syrian hamster embryo (SHE) cells by reducing incubation time of the target cells. *Mutation Research*. 2004, 548, 1–7.

Anexo 1

Tabla 1: Categorías de peligro de toxicidad aguda y valores (aproximados) de DL50/CL50 que definen las categorías respectivas

Vía de exposición	Categoría 1	Categoría 2	Categoría 3	Categoría 4	Categoría 5
Oral (mg/kg de peso corporal)	5	50	300	2000	5000
Cutánea (mg/kg de peso corporal)	50	200	1000	2000	<i>Véanse criterios detallados en nota e)</i>
Gases (ppm V) véase nota a)	100	500	2500	5000	
Vapores (mg/l) véanse notas a), b), y c)	0,5	2,0	10,0	20,0	
Polvos y nieblas (mg/l) véanse notas a) y b)	0,05	0,5	1,0	5,0	

NOTA: La concentración de los gases se expresa en partes por millón de volumen (ppmV).

Notas a la tabla 1

a) Los valores de corte/límites de concentración para la toxicidad por inhalación que figuran en la tabla se basan en una exposición de 4 horas. Para convertir los datos de modo que respondan a una exposición de 1 hora, hay que dividirlos por 2 para gases y vapores y por 4 para polvos y nieblas

b) Algunas reglamentaciones utilizan la concentración de vapor saturado como un elemento adicional para protecciones específicas de salud y seguridad (por ejemplo, las Recomendaciones de las Naciones Unidas relativas al transporte de mercancías peligrosas, Reglamentación Modelo)

c) En algunos productos químicos, la atmósfera del ensayo no será sólo un vapor sino que consistirá en una mezcla de fases líquidas y gaseosas. En otros productos químicos, esa atmósfera podrá consistir en un vapor próximo al estado gaseoso. En estos últimos casos, la clasificación (en ppmV) será la siguiente: Categoría 1 (100 ppmV), Categoría 2 (500 ppmV), Categoría 3 (2500 ppmV), Categoría 4 (5000 ppmV). En el ámbito del Programa de la OCDE sobre las directrices de ensayo deberían definirse mejor los términos “polvos”, “nieblas” y “vapores” en relación con los ensayos de toxicidad por inhalación;

d) Los valores para polvos y nieblas deberían revisarse para adaptarse a futuros cambios de las directrices de la OCDE con respecto a las limitaciones técnicas en la generación, mantenimiento y medición de las concentraciones de polvos y nieblas en forma respirable

e) Los criterios de la Categoría 5 se proponen identificar las sustancias que presenten un peligro relativamente bajo de toxicidad aguda, pero que en determinadas circunstancias puedan suponer un peligro para poblaciones vulnerables. La DL50 de esas sustancias se sitúa en el rango de 2000-5000 mg/kg de peso corporal y en dosis equivalentes para la inhalación. Los criterios específicos de la Categoría 5 son:

1) La sustancia se clasifica en esta categoría si ya se dispone de información fidedigna que indique que la DL50 (CL50) corresponde al rango de valores de la Categoría 5 o cuando otros estudios con animales o sobre los efectos tóxicos agudos en seres humanos constituyen un motivo de preocupación para la salud humana

2) La sustancia se clasificará en esta categoría, mediante extrapolación, estimación o medición de datos, cuando no esté justificada su asignación a una categoría de mayor peligro; y:

- ✓ se disponga de información fidedigna sobre la existencia de efectos tóxicos significativos en los seres humanos
- ✓ se observe mortalidad en los ensayos sobre exposición por vía oral o cutánea o por inhalación, con valores hasta de la Categoría
- ✓ cuando la opinión de los expertos confirme la aparición de síntomas clínicos de toxicidad significativos (excepto diarrea, piloerección o aspecto descuidado) en ensayos realizados con valores hasta de la Categoría 4
- ✓ cuando tal opinión confirme la existencia de información fidedigna sobre efectos agudos potencialmente significativos procedentes de otros estudios con animales.

Habida cuenta de la necesidad de proteger a los animales, se desaconsejan los ensayos con éstos en la Categoría 5, que sólo deberían contemplarse cuando sea muy probable que los resultados de esos ensayos guarden relación directa con la protección de la salud humana.

Anexo 2

Tabla 2. Estructuras Moleculares de la Base de Datos utilizada.

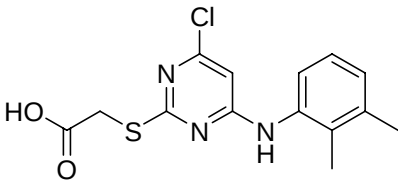
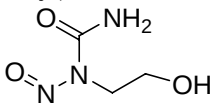
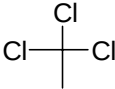
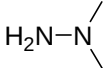
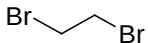
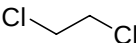
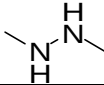
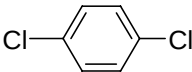
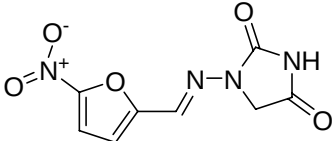
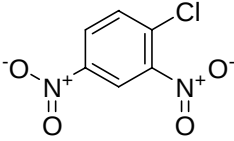
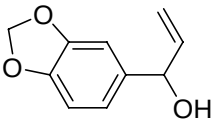
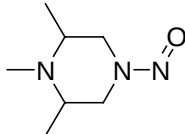
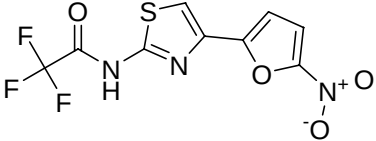
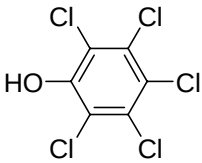
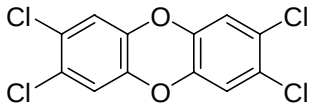
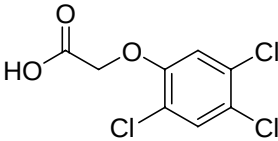
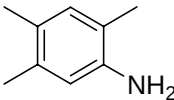
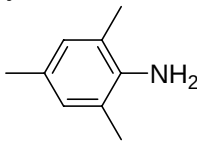
001 [4-Chloro-6-(2,3-xylidino)-2-pyrimidinylthio]acetic acid 	002 1-(2-Hydroxyethyl)-1-nitrosourea 
003 1,1,1-Trichloroethane(grado tec.) 	004 1,1-Dimethylhydrazine 
005 1,2-Dibromoethane 	006 1,2-Dichloroethane 
007 1,2-Dimethylhydrazine.2HCl 	008 1,4-Dichlorobenzene 
009 1-[(5-Nitrofurfurylidene)amino]hydantoin 	010 1-Chloro-2,4-dinitrobenzene 
011 1'-Hydroxysafrole 	012 1-Nitroso-3,4,5-trimethylpiperazine 
013 2,2,2-Trifluoro-N-[4-(5-nitro-2-furyl)-2-thiazolyl]acetamide 	014 2,3,4,5,6-Pentachlorophenol (Dowicide EC-7) 
015 2,3,7,8-Tetrachlorodibenzo-p-dioxin 	016 2,4,5-Trichlorophenoxyacetic acid 
017 2,4,5-Trimethylaniline.HCl 	018 2,4,6-Trimethylaniline.HCl 

Table 1. Cont.

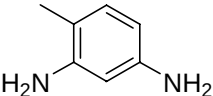
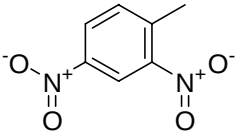
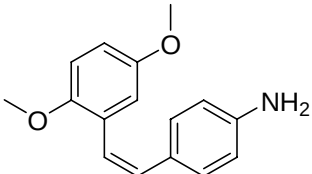
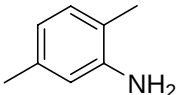
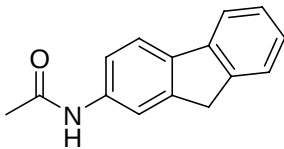
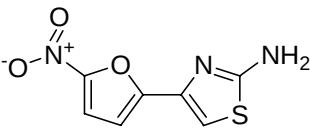
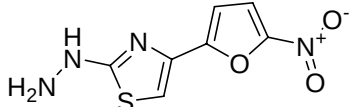
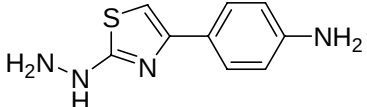
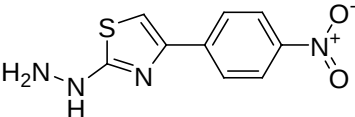
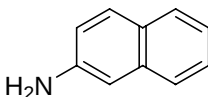
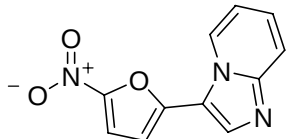
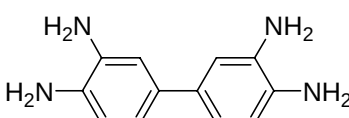
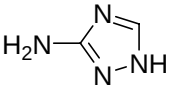
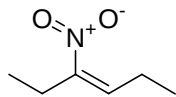
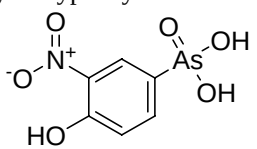
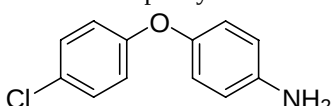
019 2,4-Diaminotoluene.2HCl 	020 2,4-Dinitrotoluene, practical grade 
021 2,5-Dimethoxy-4'-aminostilbene 	022 2,5-Xylidine.HCl 
023 2-Acetylaminofluorene 	024 2-Amino-4-(5-nitro-2-furyl)thiazole 
025 2-Hydrazino-4-(5-nitro-2-furyl)thiazole 	026 2-Hydrazino-4-(p-aminophenyl) thiazole 
027 2-Hydrazino-4-(p-nitrophenyl) thiazole 	028 2-Naphthylamine 
029 3-(5-Nitro-2-furyl)-imidazo(1,2-alpha) pyridine 	030 3,3',4,4'-Tetraaminobiphenyl.4HCl 
031 3-Aminotriazole 	032 3-Nitro-3-hexene 
033 3-nitro-4-hydroxyphenylarsonic acid 	034 4-Chloro-4'-aminodiphenylether 

Table 1. Cont.

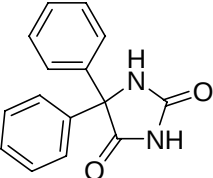
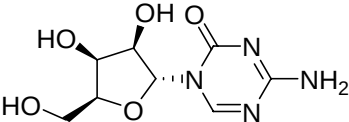
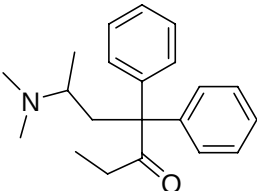
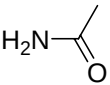
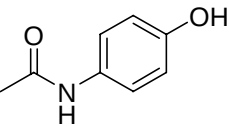
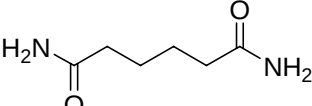
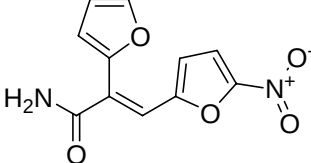
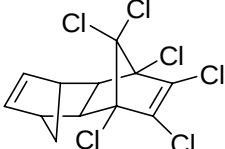
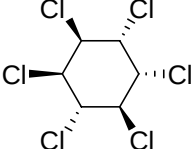
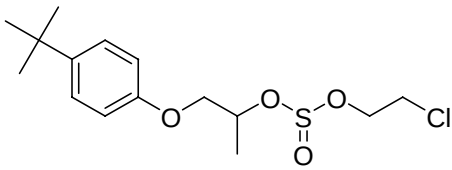
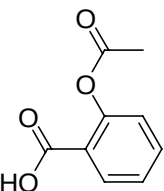
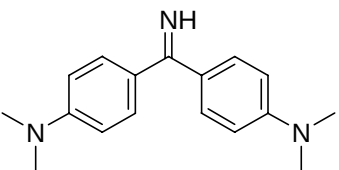
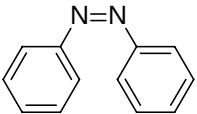
035 5,5-Diphenylhydantoin 	036 5-Azacytidine 
037 6-Dimethylamino-4,4-diphenyl-3-heptanone.HCl 	038 Acetaldehyde 
039 Acetamide 	040 Acetaminophen 
041 Adipamide 	042 AF-2 
043 Aldrin 	044 Alpha-1,2,3,4,5,6-Hexachlorocyclohexane 
045 Aramite 	046 Aspirin 
047 Auramine-O 	048 Azobenzene 

Table 1. Cont.

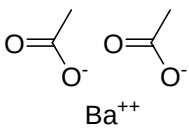
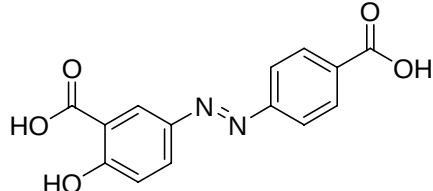
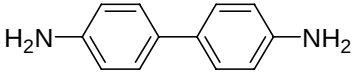
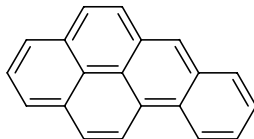
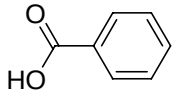
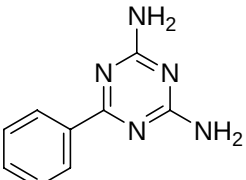
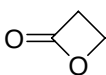
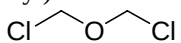
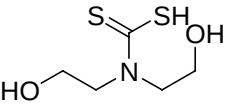
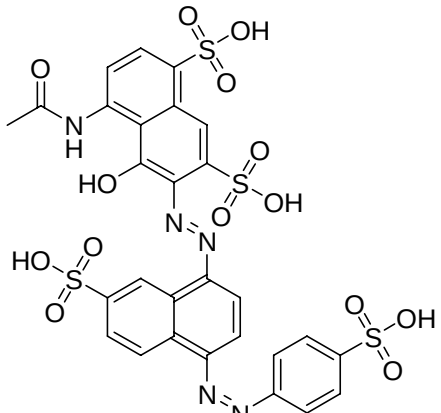
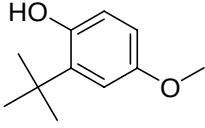
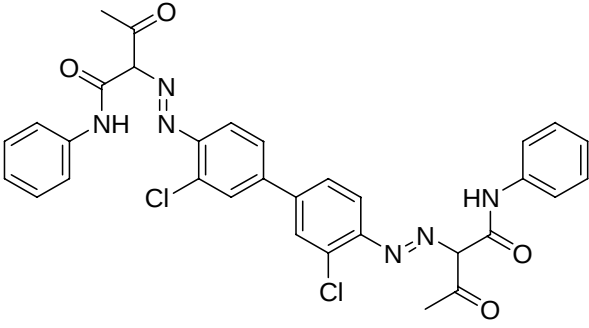
049 Barium acetate 	050 Benzalazine 
051 Benzidine 	052 Benzo(a)pyrene 
053 Benzoate, sodium 	054 Benzoguanamine 
055 Beta-Propiolactone 	056 Bis-(chloromethyl)ether 
057 Bis-2-hydroxyethylthiocarbamic acid,potassium 	058 Black PN 
059 Butylated hydroxyanisole 	060 C.I. pigment yellow 12 

Table 1. Cont.

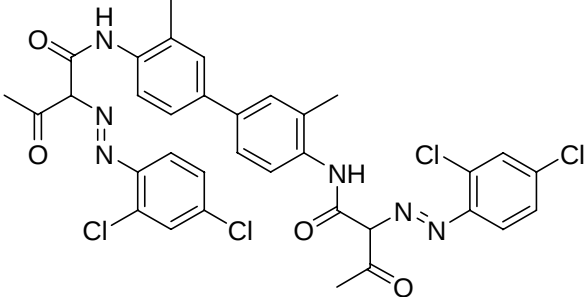
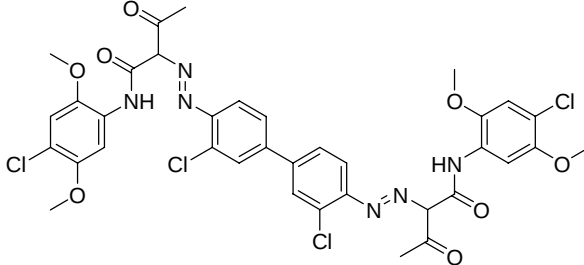
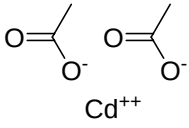
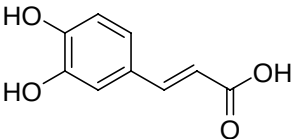
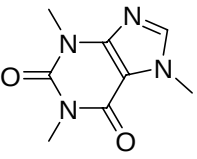
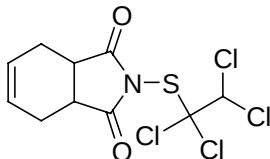
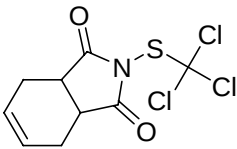
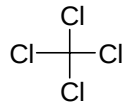
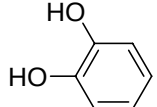
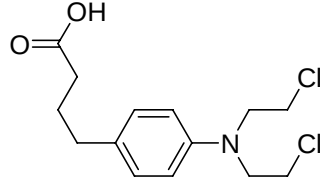
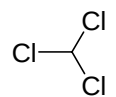
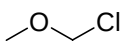
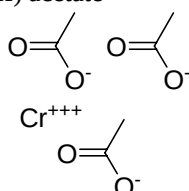
061 C.I. pigment yellow 16 	062 C.I. pigment yellow 83 
063 Cadmium acetate 	064 Caffeic acid 
065 Caffeine 	066 Captafol 
067 Captan 	068 Carbon tetrachloride 
069 Catechol 	070 Chlorambucil 
071 Chlorodifluoromethane 	072 Chloroform 
073 Chloromethyl methyl ether 	074 Chromium (III) acetate 

Table 1. Cont.

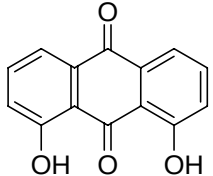
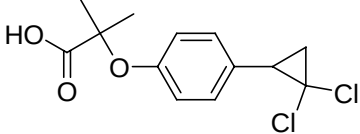
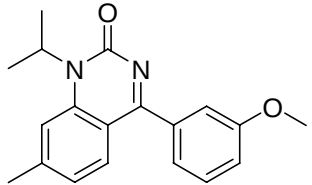
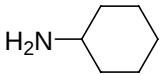
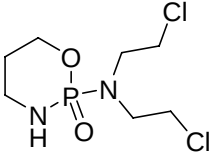
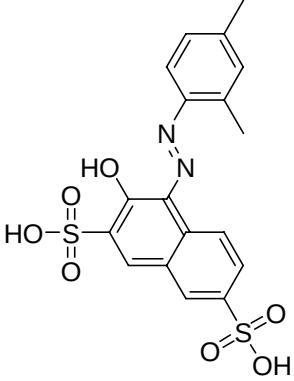
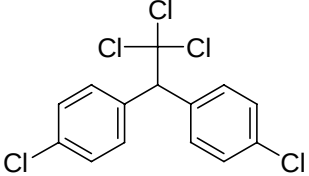
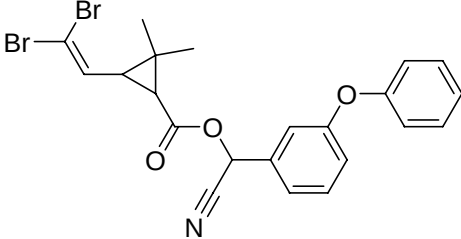
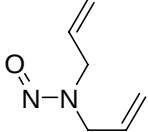
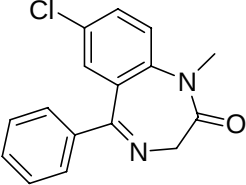
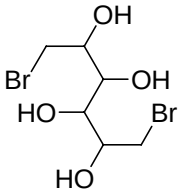
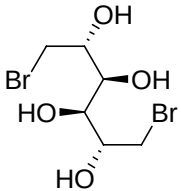
075 Chrysazin	076 Ciprofibrate
	
077 Compound 50-892	078 Cyclohexylamine.HCl
	
079 Cyclophosphamide	080 D & C red no. 5
	
081 DDT	082 Deltamethrin
	
083 Diallylnitrosamine	084 Diazepam
	
085 Dibromodulcitol	086 Dibromomannitol
	

Table 1. Cont.

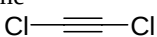
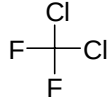
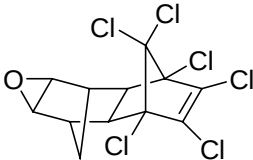
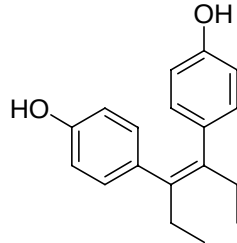
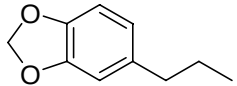
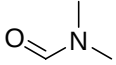
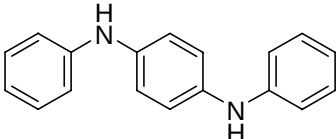
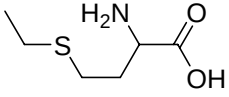
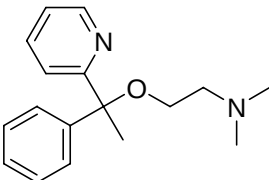
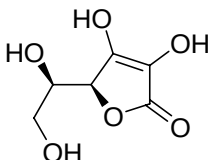
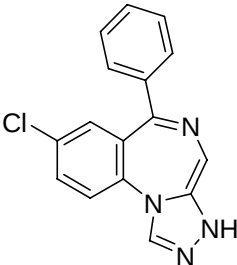
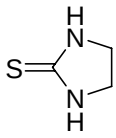
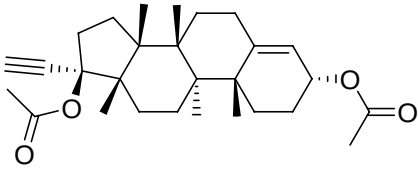
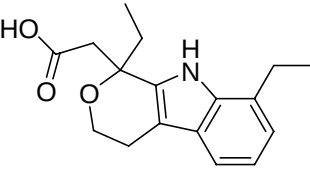
087 Dichloroacetylene 	088 Dichlorodifluoromethane 
089 Dieldrin 	090 Diethylstilbestrol 
091 Dihydrosafrole 	092 Dimethylformamide 
093 Diphenyl-p-phenylenediamine 	094 dl-Ethionine 
095 Doxylamine succinate 	096 Erythorbate, sodium 
097 Estazolam 	098 Ethylene thiourea 
099 Ethynodiol diacetate 	100 Etodolac 

Table 1. Cont.

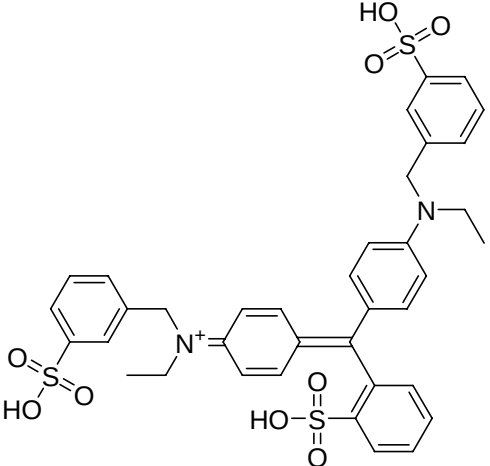
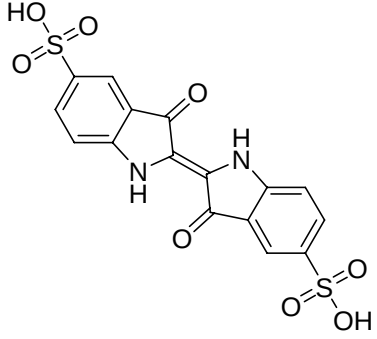
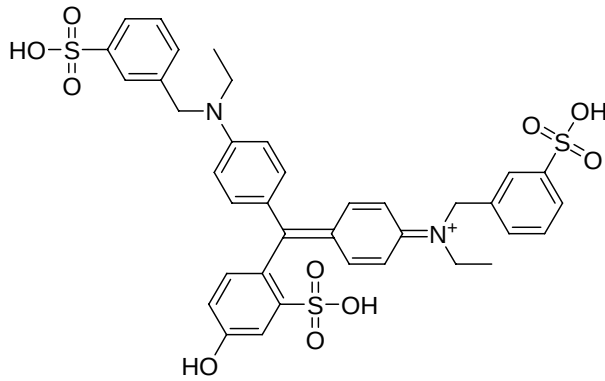
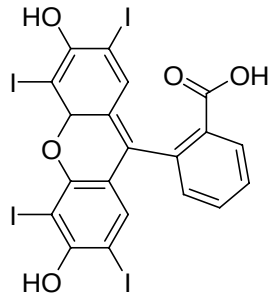
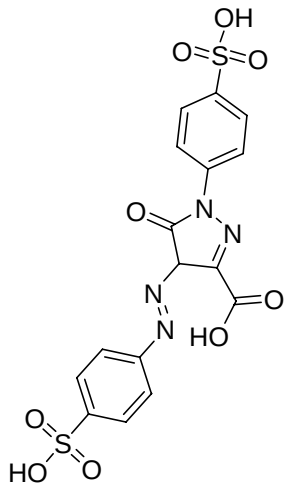
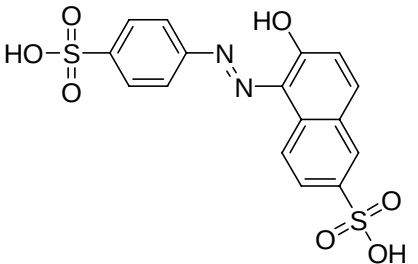
101 FD & C blue no. 1 	102 FD & C blue no. 2 
103 FD & C green no. 3 	104 FD & C red no. 3 
105 FD & C yellow no. 5 	106 FD & C yellow no. 6 

Table 1. Cont.

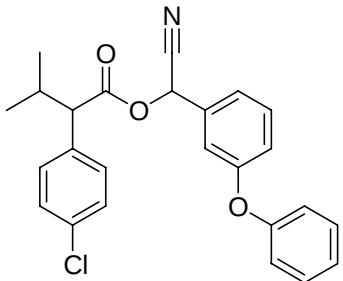
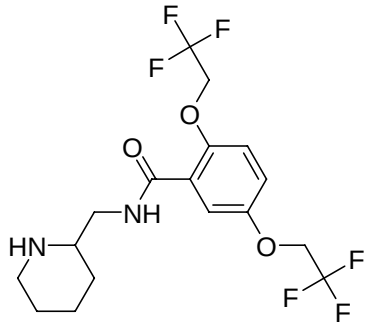
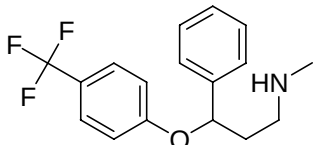
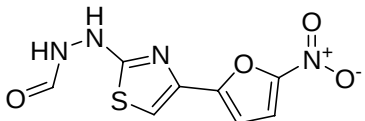
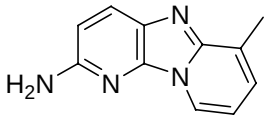
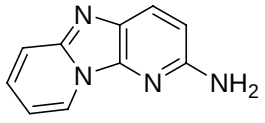
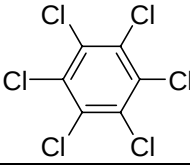
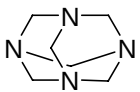
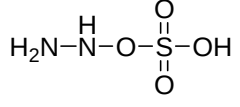
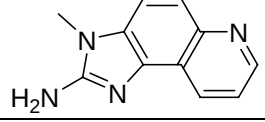
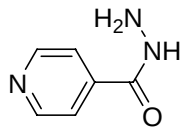
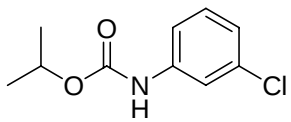
107 Fenvalerate	108 Flecainide acetate
	
109 Fluoxetine.HCl	110 Formaldehyde
	=O
111 Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	112 Glu-P-1
	
113 Glu-P-2	114 Hexachlorobenzene
	
115 Hexamethylenetetramine	116 Hydrazine
	$\text{H}_2\text{N}-\text{NH}_2$
117 Hydrazine sulfate	118 IQ
	
119 Isoniazid	120 Isopropyl-N-(3-chlorophenyl) carbamate
	

Table 1. Cont.

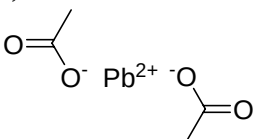
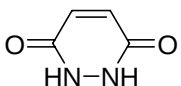
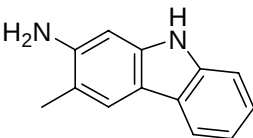
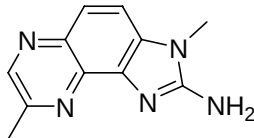
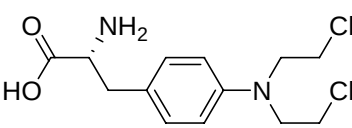
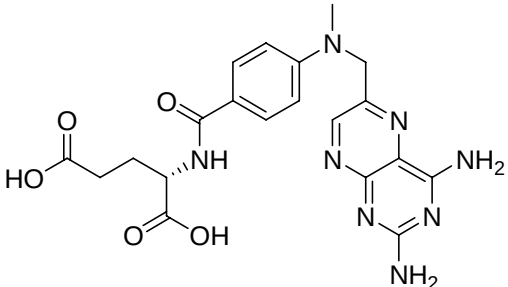
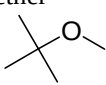
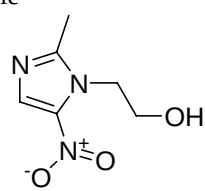
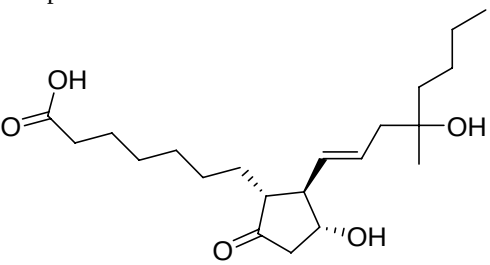
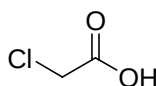
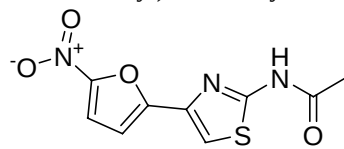
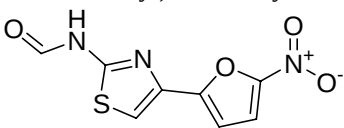
121 Lead acetate, basic 	122 Maleic hydrazide 
123 MeA-alpha-C acetate 	124 MeIQx 
125 Melphalan 	126 Methotrexate 
127 Methyl tert-butyl ether 	128 Metronidazole 
129 Misoprostol 	130 Monochloroacetic acid 
131 N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide 	132 N-[4-(5-Nitro-2-furyl)-2-thiazolyl]formamide 

Table 1. Cont.

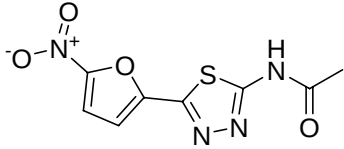
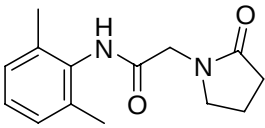
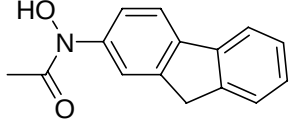
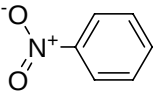
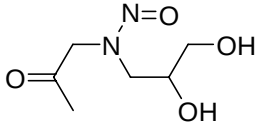
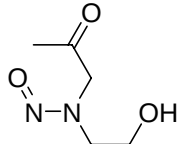
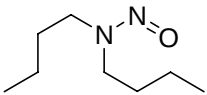
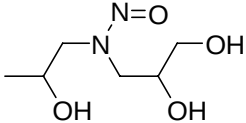
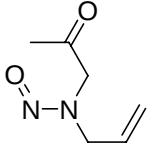
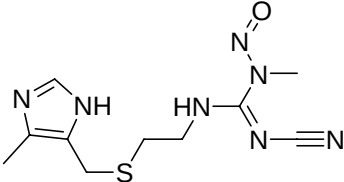
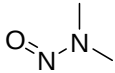
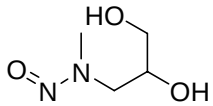
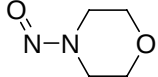
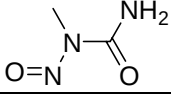
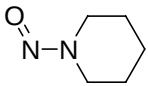
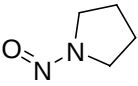
133 N-[5-(5-Nitro-2-furyl)-1,3,4-thiadiazol-2-yl]acetamide 	134 Nefiracetam 
135 N-Hydroxy-2-acetylaminofluorene 	136 Nitrobenzene 
137 Nitroso-2,3-dihydroxypropyl-2-oxopropylamine 	138 Nitroso-2-oxopropylethanolamine 
139 Nitrosodibutylamine 	140 N-Nitroso-2,3-dihydroxypropyl-2-hydroxypropylamine 
141 N-Nitrosoallyl-2-oxopropylamine 	142 N-Nitrosocimetidine 
143 N-Nitrosodimethylamine 	144 N-Nitrosomethyl-2,3-dihydroxypropylamine 
145 N-Nitrosomorpholine 	146 N-Nitroso-N-methylurea 
147 N-Nitrosopiperidine 	148 N-Nitrosopyrrolidine 

Table 1. Cont.

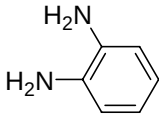
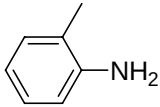
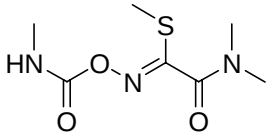
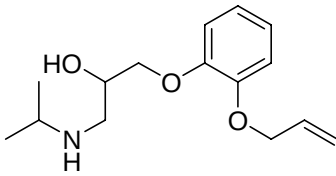
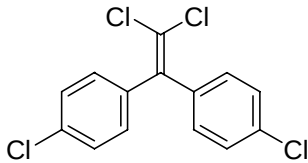
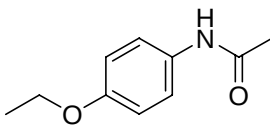
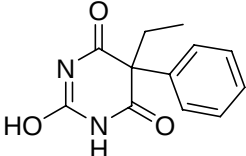
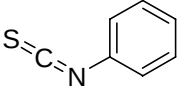
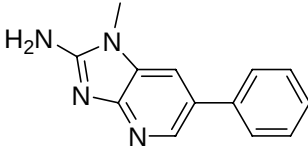
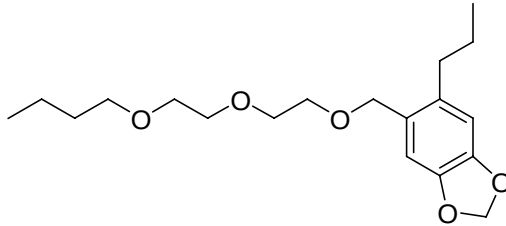
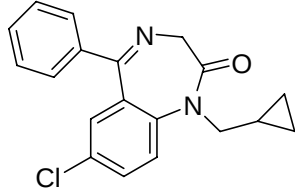
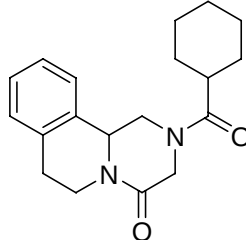
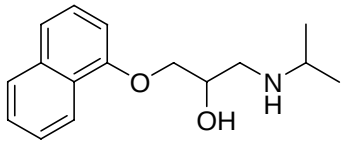
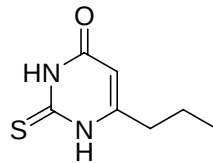
149 o-Phenylenediamine.2HCl 	150 o-Toluidine.HCl 
151 Oxamyl 	152 Oxprenolol.HCl 
153 p,p'-DDE 	154 Phenacetin 
155 Phenobarbital, sodium 	156 PHENYL ISOTHIOCYANATE 
157 PhIP.HCl 	158 Piperonyl butoxide 
159 Prazepam 	160 Praziquantel 
161 Propranolol.HCl 	162 Propylthiouracil 

Table 1. Cont

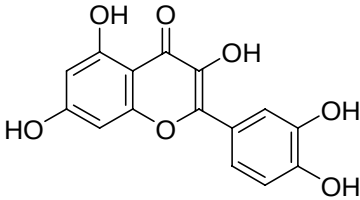
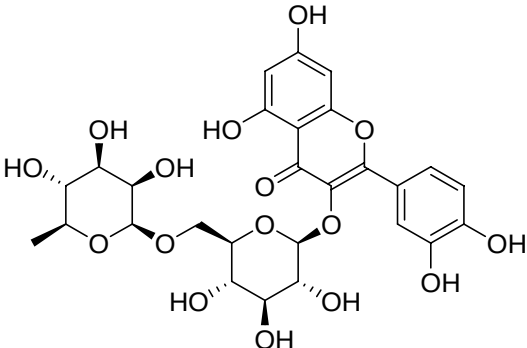
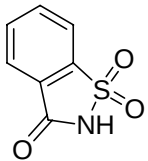
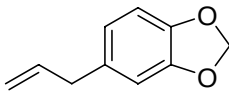
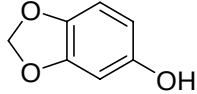
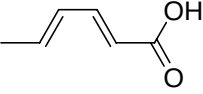
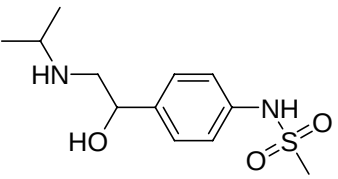
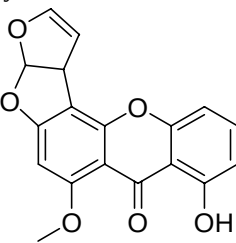
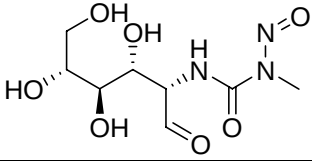
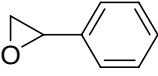
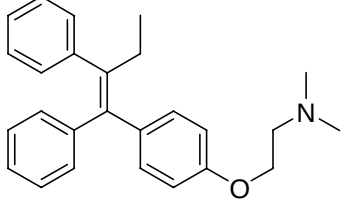
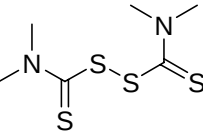
163 Quercetin dihydrate 	164 Rutin trihydrate 
165 Saccharin 	166 Safrole 
167 Sesamol 	168 Sorbic acid 
169 Sotalol.HCl 	170 Sterigmatocystin 
171 Streptozotocin 	172 Styrene oxide 
173 Tamoxifen citrate 	174 Tetramethylthiuram disulfide 

Table 1. Cont

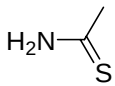
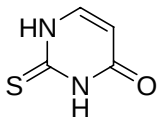
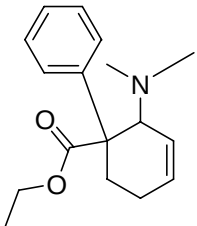
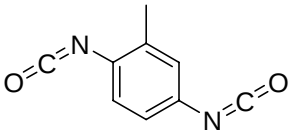
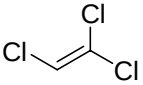
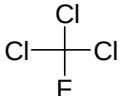
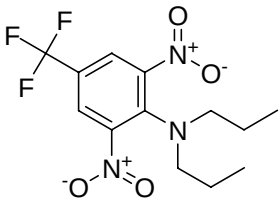
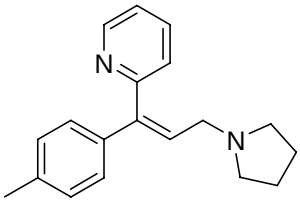
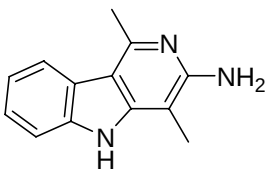
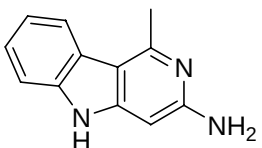
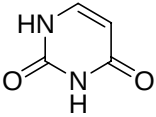
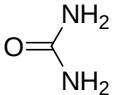
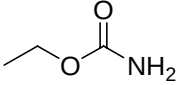
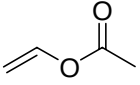
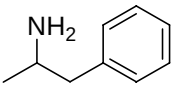
175 Thioacetamide 	176 Thiouracil 
177 Tilidine fumarate 	178 TOLUENE DIISOCYANATE 
179 Trichloroethylene 	180 Trichlorofluoromethane 
181 Trifluralin, technical grade 	182 Triprolidine.HCl monohydrate 
183 Trp-P-1 acetate 	184 Trp-P-2 acetate 
185 Uracil 	186 Urea 
187 Urethane 	188 Vinyl acetate 
189 Vinyl chloride 	190 dl-amphetamine sulfate 

Table 1. Cont

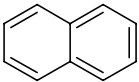
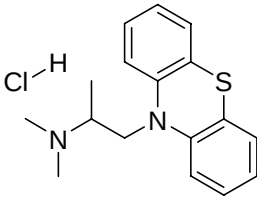
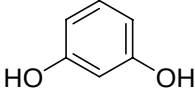
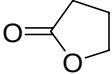
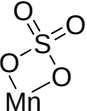
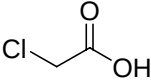
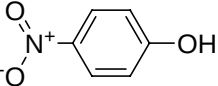
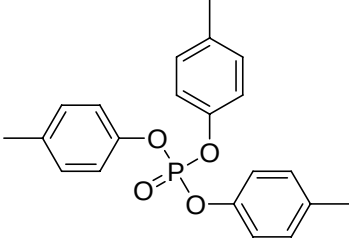
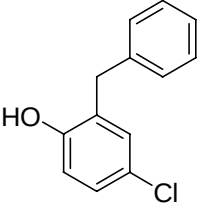
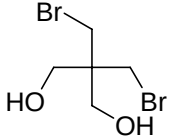
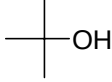
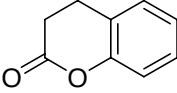
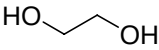
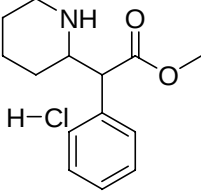
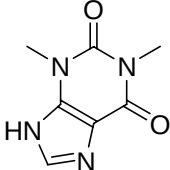
191 Naphthalene 	192 Promethazine hydrochloride 
193 Resorcinol 	194 Gamma-butyrolactone 
195 Manganese sulfate monohydrate 	196 Monochloroacetic acid 
197 p-nitrophenol 	198 Tricresyl phosphate 
199 o-benzyl-p-chlorophenol 	200 2,2-bis(bromomethyl)-1,3-propanediol 
201 t-butyl alcohol 	202 3,4-dihydrocoumarin 
203 Ethylene glycol 	204 Mercuric chloride Cl-Hg-Cl
205 Methylphenidate Hydrochloride 	206 Theophyllin 

Table 1. Cont

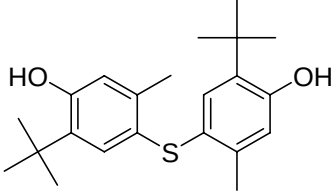
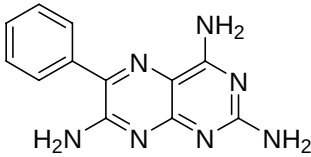
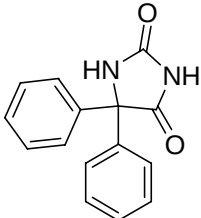
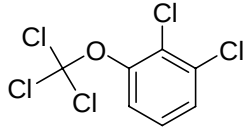
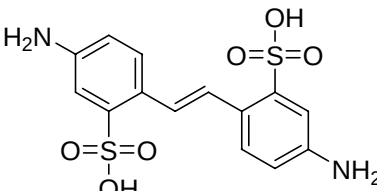
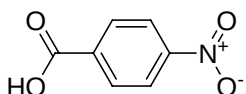
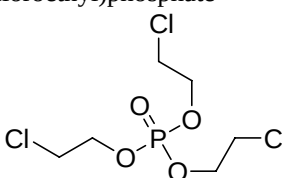
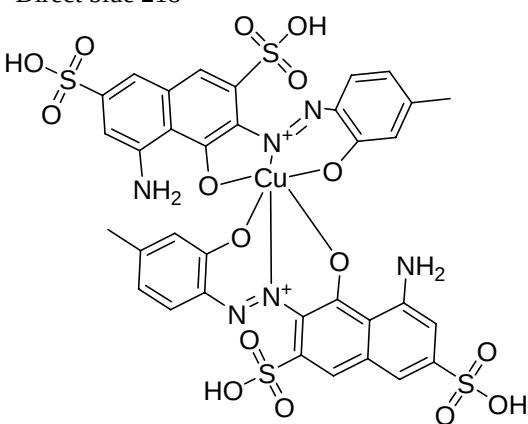
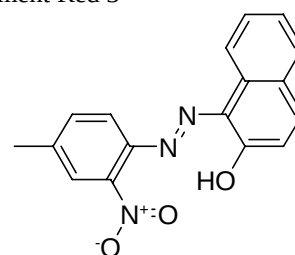
207 4,4'-thiobis(6-t-butyl-m-cresol) 	208 Triamterene 
209 Diphenylhydantoin 	210 Pentachloroanisole 
211 Chloramine $\text{H}_2\text{N}-\text{Cl}$	212 4,4'-diamino-2,2'-stilbene disulfonic acid 
213 Methyl bromide $-\text{Br}$	214 p-nitrobenzoic acid 
215 Sodium azide $\text{N}=\text{N}-\text{N}-\text{Na}$	216 Tris(2-chloroethyl)phosphate 
217 Direct blue 218 	218 C.I. Pigment Red 3 

Table 1. Cont

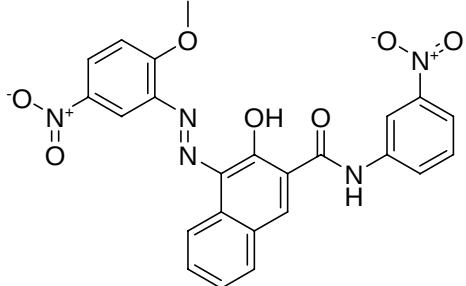
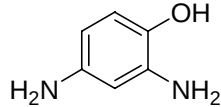
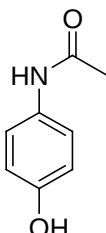
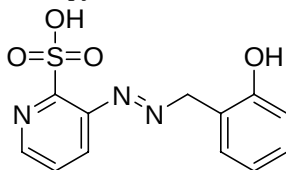
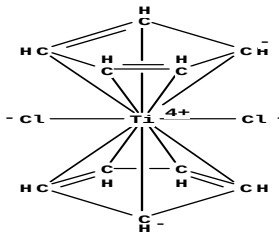
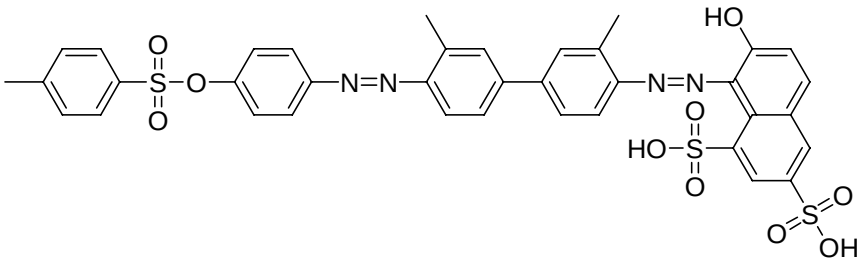
219 C.I. pigment red 23 	220 2,4-diaminophenol dihydrochloride 
221 4'-Hydroxyacetanilide 	222 Salicylazosulfopyridine 
223 Titanocene dichloride 	
224 C.I. Acid Red 114 	

Table 1. *Cont*

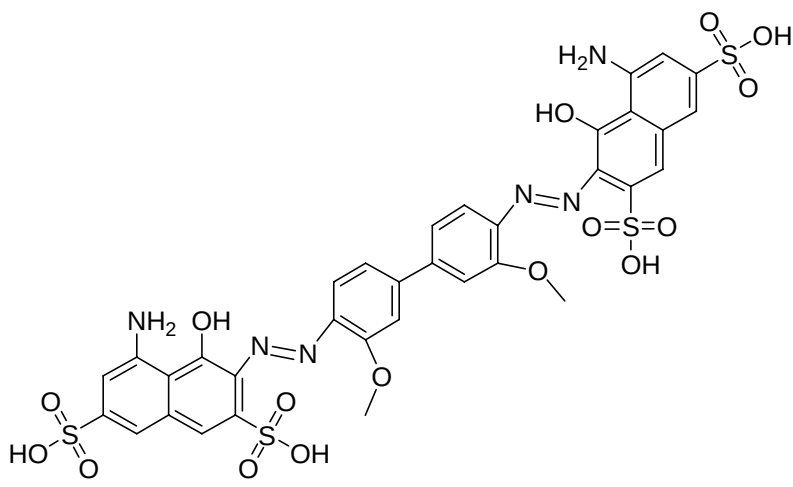
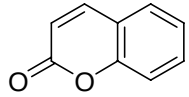
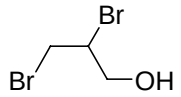
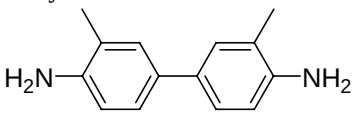
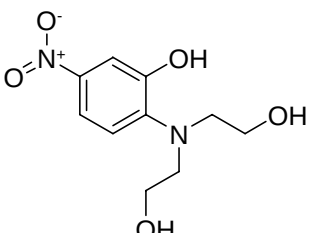
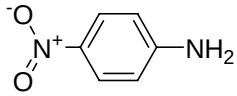
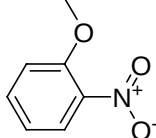
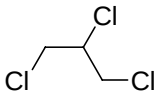
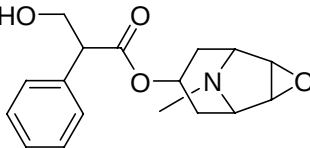
225 C.I. direct blue 15 	
226 Coumarin 	227 2,3-dibromo-1-propanol 
228 3,3'-dimethylbenzidine 	229 HC yellow 4 
230 p-nitroaniline 	231 o-nitroanisole 
232 1,2,3-trichloropropane 	233 Scopolaminehydrobromide trihydrate 

Table 1. Cont

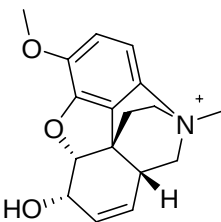
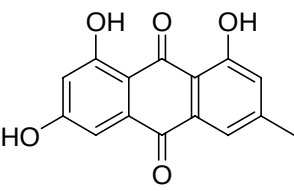
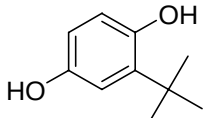
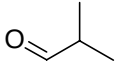
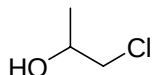
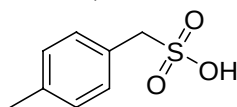
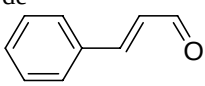
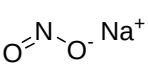
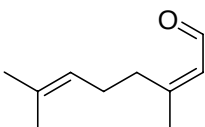
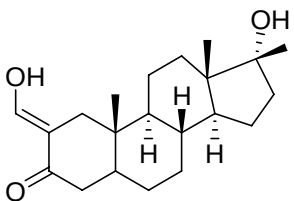
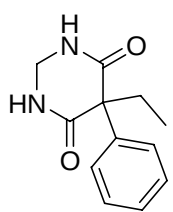
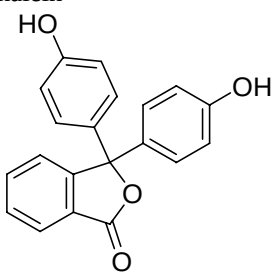
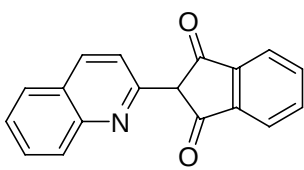
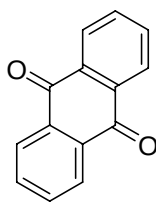
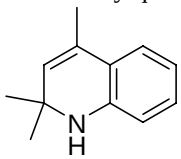
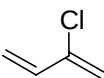
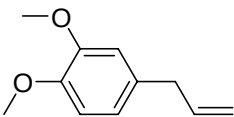
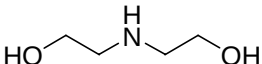
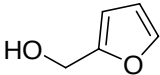
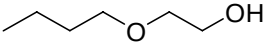
234 Codeine 	235 Emodin 
236 t-butylhydroquinone 	237 iso-butylaldehyde 
238 1-chloro-2-propanol 	239 Xylene sulfonic acid, sodium salt 
240 Cinnamaldehyde 	241 Sodium nitrite 
242 Citral 	243 Oxymethalone 
244 Primaclone 	245 Phenolphthalein 
246 D&C yellow no.11 	247 Anthraquinone 

Table 1. *Cont*

248 1,2-dihydro-2,2,4-trimethylquinoline 	249 Chloroprene 
250 isobutene 	251 Methyleugenol 
252 Nitromethane 	253 Pyridine 
254 Diethanolamine 	255 Furfuryl alcohol 
256 Tetrahydrofuran 	257 Ethylene glycol monobutyl ether 
258 Ethylbenzene 