



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE QUÍMICA

Trabajo de Tesis Doctoral

**APLICACIONES BIOLÓGICAS DE LA TEORÍA QSAR
EN EL CONTROL DEL MOSQUITO *Aedes aegypti* L.**

Tesista: Laura Marcela Saavedra Reyes

Directores: Pablo Román Duchowicz

y

Gustavo Pablo Romanelli

2020

El presente trabajo de Tesis se desarrolló en el Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), dependiente del Departamento de Química de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata (UNLP).

Se presenta en consideración de las autoridades de dicha Facultad para acceder al grado académico de Doctora de la Facultad de Ciencias Exactas Área Química.

Esta Tesis está dedicada con especial afecto a mi hija Majo,

A mis bisabuelos Jaime y Edelmira.

Y a la memoria de Nora, mi querida madre.

*“A veces es la gente de la que nadie puede imaginar
quién hace las cosas que nadie puede imaginar”*

Alan Turing

AGRADECIMIENTOS

Quiero agradecer a todas aquellas personas que contribuyeron directa o indirectamente en el desarrollo de la presente Tesis Doctoral, especialmente a:

Mis directores, el Dr. Pablo R. Duchowicz y el Dr. Gustavo P. Romanelli, por su inagotable paciencia, apoyo constante y por confiar en mis capacidades.

A los actuales integrantes del grupo “Estudio Teórico de Sistemas Químicos, Físicos y Biológicos” del INIFTA, el Dr. Eduardo A. Castro, el Dr. Francisco M. Fernández y la Dra. Ofelia B. Oña, por los apreciables momentos de enriquecimiento intelectual impartidos en cada seminario.

Al Consejo Nacional de Investigaciones Científicas y Técnicas- CONICET, por otorgarme la Beca Interna Doctoral que facilitó el inicio, desarrollo y finalización de la presente Tesis.

A la Facultad de Ciencias Exactas de UNLP, Departamento de Química por aprobar mi inscripción y admisión a la carrera de Doctorado.

ÍNDICE

	Pág.
LISTA DE ABREVIATURAS	I
MATERIAL ANEXO	V
PRÓLOGO	1
CAPÍTULO 1: FUNDAMENTOS DE LAS RELACIONES CUANTITATIVAS	
ESTRUCTURA–ACTIVIDAD (QSAR)	5
1.1 Introducción	5
1.2. Principio de la teoría QSAR	6
1.4. Principales objetivos de los modelos QSAR	10
1.5. Enfoques en QSAR	10
1.5.1. <i>El estudio SAR</i>	10
1.5.2. <i>Análisis MLR de Hansch</i>	11
1.5.3. <i>Análisis CoMFA</i>	12
1.6. Elementos y etapas del modelado QSAR	13
1.6.1. <i>Preparación</i>	13
1.6.2. <i>Procesamiento</i>	15
1.6.3. <i>Validación y predicción</i>	15
1.6.4. <i>Interpretación</i>	15
1.7. Referencias	17
CAPÍTULO 2: DESCRIPTORES MOLECULARES	23
2.1 Introducción	23
2.2. Conceptos básicos de los descriptores moleculares	25
2.3. Representación de la estructura molecular	26
2.4. Dimensionalidad de los descriptores moleculares	27
2.5. Descriptores adimensionales (0D)	28
2.6. Descriptores unidimensionales (1D) e indicadores	29
2.7. Descriptores bidimensionales (2D)	30
2.7.1. <i>Índice W de Wiener</i>	33
2.7.2. <i>Índice de conectividad de Randić</i>	34
2.7.3. <i>Índice de conectividad de Kier & Hall</i>	34
2.7.4. <i>Índice de conectividad J de Balaban</i>	35
2.7.5. <i>Índice de Zagreb</i>	36
2.7.6. <i>Índice de Rücker</i>	36
2.7.7. <i>Índice de estado electrotopológico</i>	36

2.8. Descriptores tridimensionales (3D)	37
2.8.1. <i>Descriptores WHIM</i>	39
2.8.2. <i>Descriptores GETAWAY</i>	39
2.8.3. <i>Descriptores 3D-MoRSE</i>	40
2.8.4. <i>Descriptores RDF</i>	40
2.8.5. <i>Descriptores EVA</i>	41
2.9. Descriptores basados en grillas (4D)	41
2.10. Ventajas y limitaciones según la dimensionalidad de los descriptores	43
2.11. Programas para calcular descriptores moleculares	43
2.12. Referencias	47
CAPITULO 3: MÉTODOS ESTADÍSTICOS EN EL MODELADO QSAR	55
3.1 Introducción.....	55
3.2. Pretratamiento de los datos	56
3.3 Estrategias convencionales de selección de variables	57
3.4. Estrategias basadas en clasificación	59
3.4.1. <i>Análisis de agrupamientos (CA)</i>	59
3.4.2. <i>Análisis Discriminante Lineal (LDA)</i>	59
3.5. Estrategias basadas en regresión	60
3.5.1. <i>Regresión Lineal Múltiple (MLR)</i>	60
3.5.2. <i>Regresión de Componentes Principales (PCR)</i>	62
3.5.3. <i>Cuadrados Mínimos Parciales (PLS)</i>	63
3.6. Técnicas de selección basadas en regresiones	63
3.6.1. <i>Búsqueda exacta (FS)</i>	63
3.6.2. <i>Técnica de regresión “de a pasos”</i>	64
3.6.3. <i>Método del Reemplazo (RM)</i>	65
3.7 Principales estrategias de validación en QSAR.....	66
3.7.1. <i>Validación Cruzada interna LOO y LMO</i>	67
3.7.2. <i>Prueba de Aleatorización-Y</i>	67
3.7.3. <i>Factor de Inflación de la Varianza (VIF)</i>	68
3.7.4. <i>Dominio de Aplicación (AD)</i>	68
3.8. Criterios de validación externa	69
3.8.1. <i>La selección de los subconjuntos</i>	70
3.8.2. <i>Método de subconjuntos balanceados (BSM)</i>	70
3.8.3. <i>Parámetro R^2_{pred}</i>	72
3.8.4. <i>Parámetro Q^2</i>	72
3.8.5. <i>Parámetro Q^2_{F1}</i>	73
3.8.6. <i>Parámetro Q^2_{F2}</i>	73
3.8.7. <i>Parámetro Q^2_{F3}</i>	74

3.8.8. Coeficiente de Correlación de Concordancia (CCC)	74
3.8.9. Parámetros de Golbraikh & Tropsha	74
3.8.10. Parámetro R_m^2	75
3.8.11. Error Absoluto Medio (MAE)	75
3.9. Lineamientos regulatorios del modelado QSAR	76
3.9.1. Punto final definido	76
3.9.2. Algoritmo inequívoco para la obtención del modelo	77
3.9.3. Dominio de aplicación definido	77
3.9.4. Medidas apropiadas de bondad de ajuste, robustez y predictibilidad	77
3.9.5. Interpretación del mecanismo involucrado	77
3.10. Diagnóstico de un modelo de regresión QSAR	78
3.11. Referencias	80
CAPITULO 4: ANTECEDENTES QSAR ORIENTADOS SOBRE EL CONTROL DEL MOSQUITO <i>Aedes aegypti</i> L.	87
4.1 Introducción	87
4.2 Las Enfermedades Transmitidas por Vectores (VBDs)	89
4.3 Principales arbovirus transmitidos por mosquitos del género <i>Aedes</i>	90
4.3.1. El virus del dengue (DENV)	91
4.3.2. El virus de chikungunya (CHIKV)	92
4.3.3. El virus de zika (ZIKV)	93
4.4. Principales vectores del género <i>Aedes</i>	95
4.5. El mosquito <i>Aedes aegypti</i> Linnaeus (Diptera: Culicidae)	96
4.5.1. Taxonomía	96
4.5.2. Morfología	97
4.5.3. Ciclo de vida	99
4.5.4. Comportamiento alimentario	100
4.5.5. Sitios de ovoposición	101
4.6 Interacciones vector-virus	102
4.7. Transmisión de los arbovirus a través del mosquito	103
4.8 Estrategias actuales en el control de vectores	104
4.8.1. Control biológico	106
4.8.2. Control genético	107
4.8.3. Control físico o del entorno	107
4.8.4. Control químico de las etapas inmaduras del mosquito	108
4.8.5. Control químico del mosquito adulto	109
4.9 El desarrollo de la resistencia a los insecticidas	110
4.10. Modos de acción y mecanismo de resistencia a los insecticidas	111
4.11. Estrategias de gestión y prevención de la resistencia	114
4.12. Estrategias <i>in silico</i> en el control del <i>Aedes aegypti</i> L.	115

4.13. Compuestos naturales	116
4.13.1. <i>Benzoquinonas</i>	116
4.13.2. <i>Monoterpenos</i>	116
4.14. Compuestos modificados químicamente	118
4.14.1. <i>Compuestos naturales y sus derivados</i>	118
4.15. Compuestos de síntesis en el laboratorio	122
4.15.1. <i>Compuestos organoestánicos</i>	122
4.16. Referencias	127

CAPITULO 5: APLICACIONES BIOLÓGICAS QSAR DE INTERÉS EN EL

CONTROL DEL <i>Aedes aegypti</i> L.	143
5.1. Introducción	143
5.2. Análisis QSAR en inhibidores de la enzima acetilcolinesterasa (AChE) derivados de Tacrina	145
5.2.1. <i>Datos de la actividad inhibitoria de la AChE</i>	147
5.2.2. <i>Descriptores moleculares</i>	148
5.2.3. <i>Selección de los modelos de regresión QSAR</i>	148
5.2.4. <i>Validación interna y externa de los modelos</i>	148
5.2.5. <i>Resultados y Discusión</i>	149
5.2.6. <i>Conclusiones</i>	156
5.3. Las Relaciones Cuantitativas Estructura–Actividad antifúngica (QSAR) de compuestos derivados de cinamato.	157
5.3.1. <i>Datos de la actividad antifúngica (GI)</i>	158
5.3.2. <i>Descriptores moleculares</i>	158
5.3.3. <i>Selección del modelo lineal QSAR</i>	159
5.3.4. <i>Validación interna y externa del modelo</i>	159
5.3.5. <i>Resultados y Discusión</i>	160
5.3.6. <i>Conclusiones</i>	168
5.4. Análisis QSAR de compuestos derivados de las plantas con actividad larvicida sobre el <i>Aedes aegypti</i> L., vector del dengue, chikungunya y el zika	168
5.4.1. <i>Datos de la actividad larvicida (LC₅₀)</i>	169
5.4.2. <i>Descriptores moleculares</i>	171
5.4.3. <i>Selección del modelo QSAR</i>	171
5.4.4. <i>Validación interna y externa del modelo</i>	171
5.4.5. <i>Resultados y Discusión</i>	172
5.4.6. <i>Conclusiones</i>	180
5.5. Análisis QSAR de agentes derivados de las plantas con actividad larvicida sobre el <i>Aedes aegypti</i> L. mediante descriptores de acceso libre	181
5.5.1. <i>Datos de la actividad larvicida (LC₅₀)</i>	181
5.5.2. <i>Descriptores moleculares</i>	181

5.5.3. Selección del modelo QSAR	182
5.5.4. Validación interna y externa del modelo	183
5.5.5. Resultados y Discusión	183
5.5.6. Conclusiones	188
5.6. Estudio QSAR no conformacional de 263 larvicidas derivados de las plantas sobre el vector del dengue y el zika, el <i>Aedes aegypti</i> L.	189
5.6.1. Datos de la actividad larvicida (LC_{50})	189
5.6.2. Descriptores moleculares	190
5.6.3. Selección del modelo QSAR	191
5.6.4. Validación interna y externa del modelo	191
5.6.5. Resultados y Discusión	192
5.6.6. Conclusiones	199
5.7 Referencias.....	200
CONCLUSIONES GENERALES Y PERSPECTIVAS FUTURAS	207
PUBLICACIONES Y TRABAJOS PRESENTADOS EN EVENTOS CIENTÍFICOS..	211

LISTA DE ABREVIATURAS

ACh	Neurotransmisor Acetilcolina
AChE	Enzima Acetilcolinesterasa
AD	Dominio de Aplicación
AD	Enfermedad de Alzheimer
Anti-HIV	Anticuerpos del Virus de Inmunodeficiencia Humana
ARN	Ácido Ribonucleico
BSM	Método de Subconjuntos Balanceados
CCE	Carboxil/Colinesterasas
CHIKV	Virus de Chikungunya
CNS	Sistema Nervioso Central
CoMASA	Análisis Comparativo del Sitio Activo Molecular
CoMFA	Análisis Comparativo de Campo Molecular
CoMSIA	Análisis Comparativo de Índice de Similitud Molecular
CV	Validación Cruzada
DCV	Validación Cruzada Doble
DDT	Dicloro-Difenil-Tricloroetano
DEET	<i>N, N</i> -dietil-3-metilbenzamida
DENV	Virus del Dengue
ECHA	Agencia de Químicos Europea
EC ₅₀	Concentración Efectiva en el 50% de la población
ECSA	Oeste de África, Este, Centro, y Sur de África
EIP	Periodo de Incubación Extrínseco
ELISA	Prueba de Inmunoadsorción Enzimática
EPI Suite	Interface del Programa de Estimación
ETA	Índice de Átomo Topoquímico Extendido
FET	Toxicidad Embrionaria en Peces
FN	Falsos Negativos
FP	Falsos Positivos
GA	Algoritmos Genéticos
GBS	Síndrome de Guillan-Barré
GST	Glutación Transferasas
HINT	CoMFA con hidrofobicidad
IC ₅₀	Concentración Inhibitoria en el 50% de la población
IG	Inhibición del Crecimiento
IgG	Presencia de Inmunoglobulina G
IgM	Presencia de Inmunoglobulina M
IGR	Reguladores del Crecimiento de Insectos

InChI	Identificador Internacional de Compuestos Químicos
<i>In vivo</i>	Respuesta que ocurre dentro de un organismo
<i>In vitro</i>	Respuesta de un medio de cultivo controlado ¹ fuera del organismo
<i>In silico</i>	Estudios de simulación por computadora
IsAChE	Inhibidores de la Enzima Acetilcolinesterasa
ITs	Índices Topológicos
IUPAC	Unión Internacional de Química Pura y Aplicada
KDR	Resistencia al desplome
<i>k</i> -MCA	Análisis de Agrupamientos <i>k</i> -medias
LC ₅₀	Concentración Letal en el 50% de la población
LDA	Análisis Discriminante Lineal
LMO	Técnica de validación cruzada Dejar-más-afuera
LOO	Técnica de validación cruzada Dejar-uno-afuera
MDL	Lenguaje de Definición de Modelo
MEB	Barrera de Escape del Intestino Medio
MIA	Análisis de Imagen Multivariante
MIB	Barrera de Infección del Intestino Medio
MIF	Campos de Interacción Molecular
MLR	Regresión Lineal Múltiple
Mold ²	Programa de Descriptores Moleculares bidimensionales
MOA	Modo de acción
MR	Refractividad Molar
NMR	Resonancia Magnética Nuclear
NTD	Enfermedades Tropicales Desatendidas
OMS	Organización Mundial de la Salud
OPS	Organización Panamericana de la Salud
OPs	Selección de Predictores Ordenados
PaDEL	Programa Laboratorio de Exploración de Datos Farmacéuticos
PCA	Análisis de Componentes Principales
PLISA	Plataforma de Información en Salud para las Américas
PLS	Cuadrados Mínimos Parciales
P450s	Citocromo Monooxigenasas P450
OCDE	Organización para la Cooperación y el Desarrollo Económico
QuBiLs	Programa Mapas Cuadráticos, Bilineales y N-Lineales
QSAR	Relaciones Cuantitativas Estructura-Actividad
QSPR	Relaciones Cuantitativas Estructura-Propiedad
QSTR	Relaciones Cuantitativas Estructura-Toxicidad
RM	Método del Reemplazo
RMN	Resonancia Magnética Nuclear

RMS	Error Cuadrático Medio
RT-PCR	Polimerasa con Transcriptasa inversa
SDF	Archivo de Datos de Estructural
SAR	Relaciones Estructura–Actividad
SGEB	Barrera de Escape de las Glándulas Salivales
SGIB	Barrera de Infección de las Glándulas Salivales
SMILES	Sistema de Entrada Molecular Lineal Simplificado
TN	Verdadero Negativo
TP	Verdadero Positivo
UR	Regresión Univariada
US EPA	Agencia de Protección Ambiental de EU
US FDA	Centro de Administración de Medicamentos y Alimentos de EU
VBDs	Enfermedades Transmitidas por Vectores
WHOPES	Esquema de Evaluación de Pesticidas
YFV	Virus de Fiebre Amarilla
ZIKV	Virus del Zika

MATERIAL ANEXO

En el CD que se adjunta al ejemplar de Tesis se encuentran disponibles todas las Tablas que detallan las bases de datos utilizadas en las diversas aplicaciones biológicas QSAR. Además, se incluye la versión electrónica de todos los artículos científicos publicados.

PRÓLOGO

Los continuos avances en biología, química y computación han contribuido al desarrollo acelerado de los estudios de simulación por computadora o también conocidos como estudios *in silico*; los cuales fueron propuestos inicialmente para orientar la búsqueda de nuevos fármacos o identificar los agentes contaminantes presentes en el ambiente, mediante el ajuste de un modelo.¹

Entre los diferentes métodos *in silico* existentes, la teoría de las Relaciones Cuantitativas Estructura-Actividad (QSAR) propuesta por Corwin Hansch y Toshio Fujita en el año 1964,^{2, 3} se ha destacado por ofrecer elementos útiles para abordar diferentes cuestiones químicas involucradas en los complejos procesos y mecanismos de los sistemas bióticos y abióticos.

El fundamento de teoría QSAR reside en elucidar la respuesta química de un compuesto en términos de sus características estructurales. Por ello, los modelos matemáticos QSAR proponen un paralelismo racional entre la estructura molecular de una sustancia, que es representada por diversos descriptores moleculares y el efecto final que dicha sustancia produce sobre un determinado organismo o entorno.⁴ Es así como el formalismo QSAR permite estimar diferentes parámetros de respuesta como la actividad, toxicidad o propiedad de nuevas moléculas, a partir del conocimiento de la estructura molecular involucrada.^{5,6}

Por más de 55 años, el crecimiento exponencial del número de publicaciones científicas basadas en la teoría QSAR ha sido muestra de su constante evolución, atribuida principalmente al aumento de las bases de datos químicas de acceso público y a la inclusión de múltiples técnicas estadísticas sencillas y complejas, capaces de establecer correlaciones entre numerosas características estructurales de las moléculas y diversas propiedades fisicoquímicas, organolépticas, biológicas o toxicológicas.⁷

A raíz de esto, los estudios QSAR han sido ampliamente utilizados en la academia, la industria y las agencias gubernamentales como una herramienta computacional que permite realizar predicciones de los posibles efectos que los productos químicos, los materiales y los nanomateriales provocan sobre la salud humana y los ecosistemas.⁸

Actualmente, una de las más novedosas aplicaciones de los modelos QSAR es con fines regulatorios. A medida que aumenta el número y complejidad de los compuestos comerciales, diferentes instituciones estatales como: la Organización para la Cooperación y el Desarrollo Económico (OCDE),⁹ la Agencia de Químicos Europea (ECHA),¹⁰ la Agencia de Protección Ambiental (US EPA) y el Centro de Administración de Medicamentos y Alimentos de los Estados Unidos (US FDA),^{11, 12} iniciaron labores para establecer sistemas de búsqueda y herramientas computacionales que

identifiquen y predigan rápidamente múltiples riesgos vinculados a la seguridad, eficacia, calidad y rendimiento de las sustancias nuevas o hipotéticas. Hasta ahora, existen numerosos modelos QSAR enfocados en clasificar e identificar los principales elementos estructurales asociados a las respuestas peligrosas o tóxicas producidas por millones de sustancias comerciales.¹³

Sin embargo, cabe destacar que la hipótesis principal de la teoría QSAR no consiste en explicar el mecanismo de acción, la toxico-cinética o toxico-dinámica de las sustancias químicas; por el contrario, se enfoca en calcular el efecto final que una estructura química produce sobre un individuo o ambiente definido. Aunque, si el modelo logra una cuantificación acertada de la actividad del compuesto, puede brindar información del fenómeno involucrado; tal como especificar si una molécula es activa o inactiva, tóxica o no tóxica; esto según los límites de detección propuestos a nivel experimental.^{14, 15}

Por otra parte, en aquellos casos que se cuenta con información de la conformación molecular del compuesto, dada por las técnicas de cristalografía de rayos X o Resonancia Magnética Nuclear (RMN), es posible establecer modelos de regresión QSAR basados en descriptores geométricos, que interpretan la actividad involucrada con base en la disposición tridimensional de la molécula; a este enfoque se le conoce como QSAR-3D.¹⁶

A partir de numerosas aplicaciones de los modelos QSAR, se logró evidenciar su eficacia para tratar grandes conjuntos moleculares y predecir cuantiosos efectos finales definidos por protocolos de ensayo de difícil medición.¹⁷ Estas condiciones permitieron a los investigadores discernir tendencias en los datos existentes más relevantes para la salud pública o el ambiente, y luego extrapolar tendencia a compuestos no sintetizados o carentes de información química, física o biológica.¹⁶

Es así como el enfoque QSAR ha contribuido en novedosas investigaciones orientadas en los compuestos químicos provenientes o inspirados en la naturaleza, reconocidos por presentar distintas propiedades biológicas plausibles como: antioxidante, antibacterial, anticonvulsiva, anticancerígena, inhibición enzimática, larvívica, repelente, entre otras propiedades de gran importancia para el diseño de nuevos productos químicos de interés comercial, medicinal, agrícola y sanitario.¹⁸⁻²⁵

No obstante, el paradigma de la teoría QSAR no solo proporciona modelos que cuantifican propiedades fisicoquímicas o biológicas. También se enfatiza en orientar la síntesis de nuevas moléculas conforme a los 12 principios de la química verde,²⁶ reducir el tiempo y los gastos que implican los ensayos *in vivo* e *in vitro*, y minimizar los experimentos en animales.²⁷

Por lo anterior, la presente Tesis Doctoral se fundamenta en el estudio, diseño y aplicación de modelos de la teoría QSAR que proporcionen una guía computacional en el diseño sustentable, racional y selectivo de nuevos compuestos con actividad

específica, inocuos en mamíferos y de alta eficiencia en el control de enfermedades neurológicas e infecciosas que afectan la salud pública mundial.

Actualmente, entre los compuestos químicos más utilizados en el tratamiento de diversos trastornos neurodegenerativos, como la enfermedad de Alzheimer (AD), se destacan los inhibidores de la enzima acetilcolinesterasa (IsAChE).

Este tipo de compuestos químicos actúan en el Sistema Nervioso Central (CNS), incrementando la actividad colinérgica cerebral. Este modo de acción también se encuentra en los pesticidas más novedosos y eficientes para controlar las plagas y los vectores de enfermedades.

Conforme a lo anterior, se examinó la capacidad inhibitoria *in vitro* de la enzima acetilcolinesterasa (AChE) de diferentes conjuntos moleculares basados en derivados de Tacrina, los cuales representan una alternativa en el diseño de nuevos fármacos de bajo costo para tratar las enfermedades neurológicas y reducir los efectos secundarios producidos por los medicamentos actuales. Asimismo, estas moléculas proveen información estructural relevante en el desarrollo de fitosanitarios selectivos y eficaces contra los vectores de enfermedades.

Ante las múltiples ventajas que tienen los productos de origen natural y sus derivados semi-sintéticos sobre la salud y el cuidado de los ecosistemas; se evaluó la actividad antifúngica medida a través de la Inhibición del Crecimiento (IG) de un conjunto molecular conformado por derivados de cinamato; los cuales representan una novedosa librería molecular bioactiva e inocua para la salud humana, obtenida a través de procedimientos ecocompatibles

Por último, se abordó a través del modelado QSAR el estudio de diferentes conjuntos moleculares constituidos por diversas moléculas naturales o semi-sintéticas con prominente actividad larvicida sobre el mosquito *Aedes aegypti* L., el principal vector de diversas enfermedades endémicas, como el dengue, la fiebre amarilla, la fiebre de chikungunya y el zika.

CAPÍTULO 1: FUNDAMENTOS DE LAS RELACIONES CUANTITATIVAS ESTRUCTURA–ACTIVIDAD (QSAR)

1.1 Introducción

Los estudios *in silico* han resultado determinantes para acceder, examinar, deducir, modelar y proyectar enormes conjuntos de datos químicos, sin requerir la síntesis, caracterización y ensayo de cada compuesto. A raíz de las numerosas ventajas que brindan las simulaciones por computadora, fueron postuladas como herramientas esenciales para identificar y diseñar estratégicamente nuevas entidades químicas con efectos finales específicos.^{2,3}

Conforme a esto, la teoría QSAR constituye una de las principales herramientas computacionales diseñadas para establecer modelos de correlación racionales, basados en los datos de respuesta final de los compuestos químicos y las características estructurales de los mismos.²⁸ Así, las aplicaciones QSAR intentan cuantificar el vínculo entre uno o varios aspectos de la estructura química y la actividad o propiedad impartida por dicha estructura, con el objetivo de revelar el potencial activo de las moléculas nuevas o carentes de esta información.

En este contexto, es posible inferir que los modelos de la teoría QSAR son el resultado final de un proceso computacional que inicia con la extensa caracterización de la estructura molecular mediante descriptores moleculares, que codifican la información de las propiedades fisicoquímicas y de las características estructurales simples (constitucionales y topológicas) y complejas (electrónicas y geométricas) de la molécula; para finalmente obtener un modelo que plantea algunas inferencias, hipótesis y predicciones del efecto final que genera un compuesto químico sobre un sistema ambiental o biológico determinado.^{4,6}

Las correlaciones resultantes de un análisis QSAR, por lo general, toman la forma de ecuaciones matemáticas desarrolladas a través de diferentes técnicas estadísticas lineales o no lineales, como Regresión Univariada (UR), Regresión Lineal Múltiple (MLR), Cuadrados Mínimos Parciales (PLS), Análisis Discriminante Lineal (LDA) o Análisis de Agrupamiento, que proponen paralelismos racionales para cuantificar o clasificar con mayor certeza la respuesta final bajo análisis.

Conforme a lo anterior, es posible inferir que la teoría QSAR tiene la capacidad de guiar el diseño racional, rápido y selectivo de nuevos precursores químicos con

propiedades inéditas, basándose únicamente en la información proveniente de la estructura molecular. Es así como el diseño y uso de un modelo QSAR puede orientar la búsqueda de entidades químicas alternativas, que puedan sustituir el uso indiscriminado de agrotóxicos comerciales o mejorar el poder farmacológico de los medicamentos actuales.

En los últimos años, el constante aumento de las investigaciones enfocadas en QSAR, revelan su importante agilidad para determinar mediante una relación matemática lógica, el efecto de las características estructurales sobre el potencial activo de las moléculas nuevas o hipotéticas.¹⁴ Este enfoque resulta particularmente útil cuando las medidas experimentales son extensas, laboriosas, de alto costo e implican el uso de animales de prueba.^{18, 27}

La teoría QSAR fue diseñada inicialmente para estimar el efecto final que tiene un conjunto de moléculas sobre un organismo o entorno definido. Los últimos avances en el modelado QSAR proponen incorporar descriptores basados en la información experimental disponible, como el potencial de unión al receptor, la capacidad de activación e inhibición enzimática, o la velocidad de formación de un complejo de proteína, con el fin de elucidar algunos procesos claves asociados a la respuesta final exhibida por las moléculas.

Acorde con lo anterior, el presente capítulo de Tesis fue elaborado para detallar los principales conceptos involucrados en el estudio, diseño y aplicación de los modelos QSAR. Asimismo, abordar diferentes paradigmas asociados con la utilidad, la eficacia, los alcances y las perspectivas futuras de las investigaciones basadas en la teoría QSAR.

1.2. Principio de la teoría QSAR

La teoría QSAR se fundamenta en el postulado básico de presentar la respuesta final generada por una sustancia química en términos de sus características estructurales y propiedades fisicoquímicas, definidas de manera teórica o experimental.

Ciertamente, los parámetros fisicoquímicos, como el punto de ebullición, el punto de fusión o la tensión superficial, son características específicas de un compuesto que brindan información relevante para explicar el efecto final generado por el mismo. Por ello, desde una perspectiva matemática, la propiedad o actividad de un compuesto puede ser representada como una función de su estructura molecular y sus propiedades fisicoquímicas, codificadas a través de los denominados descriptores moleculares, Ec. 1.1.^{1, 6, 7}

$$\text{Actividad / Propiedad} = f(\text{descriptores moleculares}) \quad (\text{Ec. 1.1})$$

Este concepto surgió en el año 1964, con la propuesta de los científicos C. Hansch and T. Fujita de diseñar modelos matemáticos de regresión, que correlacionaban pequeños conjuntos moleculares con algunas propiedades fisicoquímicas sencillas, con el único objetivo de estimar su potencial terapéutico para el diseño de nuevos fármacos.^{2,3}

El modelo QSAR expresa una relación matemática multivariable entre un conjunto de características fisicoquímicas o estructurales, codificadas por los descriptores moleculares (X_{ij}), y una función experimental de tipo biológico, fisicoquímico o toxicológico (Y_i). Dicho modelo tiene la capacidad de estimar el valor de una delimitada respuesta final. Para una molécula i , la ecuación lineal relaciona los descriptores moleculares (X_1, X_2, \dots, X_n) con la actividad de interés (Y), por medio de la siguiente ecuación:

$$Y_i = X_{i1}b_1 + X_{i2}b_2 + \dots + X_{in}b_n + e_i \quad (\text{Ec. 1.2})$$

Donde b_n corresponde a las pendientes lineales que expresan la correlación entre una característica molecular dada (X_{in}) y la actividad del compuesto i (Y_i); el parámetro e_i representa una constante. Las pendientes y constantes son calculadas por medio de un análisis de regresión. En la presente Tesis, se desarrollaron diferentes modelos QSAR basados en una única variable dependiente (Y).

Los estudios QSAR son estrategias rigurosas y precisas para anticipar respuestas fisicoquímicas, biológicas, organolépticas o toxicológicas, por medio del análisis exhaustivo de numerosas variables independientes (descriptores moleculares e indicadores de fragmentos), ejecutado por un algoritmo inequívoco de selección de variables, con capacidad de cuantificar la respuesta final exhibida por un delimitado grupo de moléculas conocidas o aún inexploradas.

Por esta razón, las herramientas computacionales basadas en QSAR han incrementado de forma escalonada su participación en diferentes campos de la investigación, la industria, y particularmente en la regulación de las sustancias peligrosas.^{15,29}

1.3. Influencia de la similaridad en los modelos QSAR

El paradigma central que subyace del modelado QSAR es el principio de similitud estructura-actividad. Este paradigma establece que los compuestos homólogos tienen actividades o propiedades similares. Por ello, se puede deducir que el análisis y estimación del potencial activo de un producto químico, conduce a la respuesta final de muchos otros sin evaluar.

En QSAR, el término de similaridad hace referencia a todos los compuestos químicos que presentan estructuras moleculares, modos de acción o funciones similares.

En este contexto, la Figura 1.1 exhibe dos estructuras moleculares, pertenecientes a los insecticidas deltametrina (A) y cipermetrina (B). Estos agentes insecticidas son análogos estructurales y funcionales, ya que comparten la subestructura del ácido ciclopropanocarboxílico común en los piretroides y presentan un poder insecticida similar.

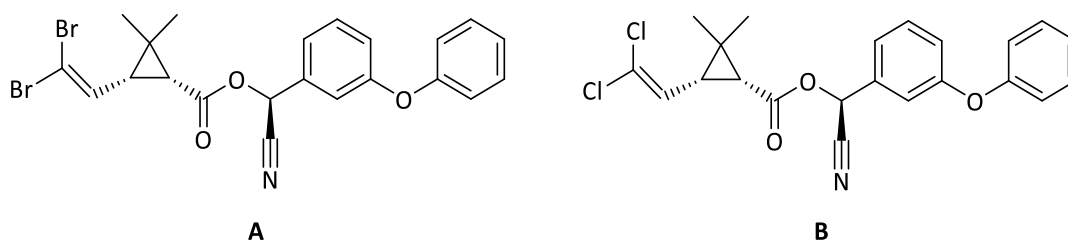


Figura 1.1. Compuestos análogos estructurales y funcionales. Fuente: Saavedra, L.M.

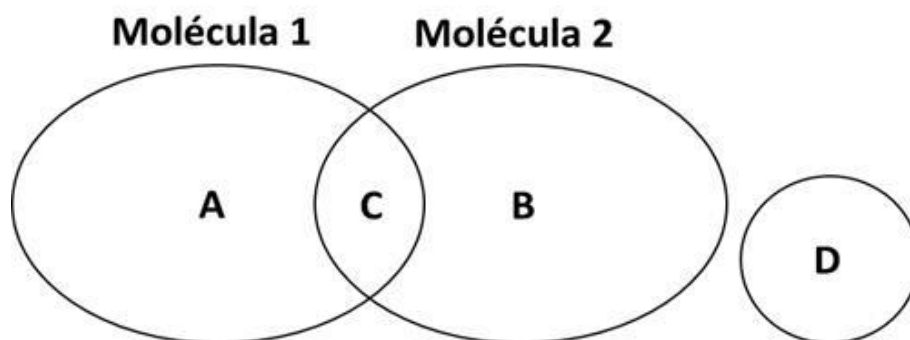
La posibilidad de definir análogos estructurales involucra un enfoque matemático común que emplea índices de similaridad para medir el grado de coincidencia entre los patrones detallados en la Figura 1.2; encargados de codificar las características estructurales de las moléculas.

Cabe resaltar que un alto grado de similaridad basado en los índices matemáticos de similitud no indica que exista similitud en el modo de acción (MOA) o que sean análogos funcionales.

Los índices de similitud utilizan la información bidimensional o tridimensional de la estructura química para establecer diferentes tipos de índices de correlación, tal como el índice de Tanimoto, el índice de Ricards Hodgkin, el índice de similitud de coseno vinculado a la medida de similitud y diferencia de Hamman, y el índice de distancia euclídea, basado en la distancia ordinaria entre las características estructurales consideradas.³⁰

Particularmente, los análogos funcionales se caracterizan por presentar respuestas fisicoquímicas o biológicas semejantes, sin necesidad de ser análogos estructurales.³¹

Un importante ejemplo de similaridad funcional son los productos químicos más utilizados en el control de insectos plaga y vectores de enfermedades, los cuales pertenecen a diferentes clases químicas con similaridad funcional, como se detallan en la Tabla 1.1.



Similitud entre el compuesto 1 y 2:

A: Número de características presentes en 1 y ausentes en 2

B: Número de características presentes en 2 y ausentes en 1

C: Número de características presentes en 1 y 2

D: Número de características ausentes en 1 y 2

Figura 1.2. Patrones considerados en los índices de similitud. Fuente: Monev, 2004.³⁰

A raíz del constante incremento de los estudios basados en QSAR, se consideró la similitud molecular como una estrategia válida para generar descriptores moleculares. La primera aplicación de similitud molecular acoplada a QSAR fue propuesta por Martín *et al.*, 1983.³² Desde entonces, las medidas de similitud molecular han sido incorporadas en los análisis QSAR, ya sea dentro de la formulación clásica de los índices de Carbó, los índices de similitud topológica, entre otros.^{33, 34}

Tabla 1.1. Modos de acción asociados a las clases químicas de los pesticidas comerciales.³⁵

Modo de Acción	Grupo químico
Inhibidor de acetilcolinesterasa (AChE)	Carbamatos Organofosforados
Antagonista de los canales de cloruro activados por GABA	Derivados clorados de ciclodienos Fenilpirazoles
Modulador del canal de sodio	Organoclorados Piretroides
Agonista del receptor nicotínico de acetilcolina	Neonicotinoides
Imitador de hormona juvenil	Análogos de hormona juvenil Carbamatos Piridinas reguladoras del crecimiento en insectos

1.4. Principales objetivos de los modelos QSAR

El principal objetivo de un modelo QSAR es estimar racionalmente el potencial activo contenido en moléculas nuevas o hipotéticas, sin requerir la síntesis, caracterización y prueba de grandes conjuntos moleculares inactivos o peligrosos.^{36, 37} Adicionalmente, el diseño de un modelo QSAR tiene como finalidad cumplir otros objetivos, como:

1. Orientar el diseño rápido y selectivo de nuevas entidades químicas con propiedades plausibles, minimizando las prácticas experimentales, el tiempo y los costos.
2. Revelar las variables teóricas, semiempíricas o experimentales asociadas al comportamiento activo de diferentes conjuntos moleculares.
3. Reducir o reemplazar las pruebas *in vivo* que involucran el sacrificio o maltrato animal.
4. Proveer el rango de concentración o dosis necesario para que una sustancia química genere un determinado efecto final.
5. Elucidar una aproximación del mecanismo de acción de los compuestos químicos en los seres vivos o los ecosistemas.
6. Establecer la búsqueda selectiva basada en ligandos o receptores de amplios conjuntos moleculares con moléculas candidatas a fármaco.
7. Realizar un análisis de clasificación con base en las predicciones, considerando las categorías de actividad establecidas experimentalmente para la respuesta final evaluada.
8. Optimizar las estructuras moleculares con prominente actividad o propiedad hacia un perfil de respuesta particular y efectivo.

1.5. Enfoques en QSAR

Entre las estrategias más utilizadas para estimar de forma cualitativa o cuantitativa las propiedades almacenadas en las moléculas, se destacan el análisis de las Relaciones Estructura–Actividad (SAR), el Análisis de Regresión Lineal Múltiple (MLR) de Hansch y el Análisis Comparativo de Campo Molecular (CoMFA), descritos brevemente a continuación:

1.5.1. *El estudio SAR*

Un estudio SAR se basa en el conocimiento básico de estimar la respuesta fisicoquímica o biológica producida por las moléculas (generalmente orgánica) en función de los cambios en las estructuras químicas de las mismas. De este modo,

establece que las moléculas análogas tienen respuestas similares, si se encuentran bajo las mismas condiciones de experimentación.

Esencialmente, el análisis SAR compara los aspectos experimentales cualitativos o cuantitativos por medio de un análisis de varianza matemática, tal como el análisis de regresión, el análisis discriminante, el análisis factorial o el método de reconocimiento de patrones.

Seguidamente, con base en los datos experimentales de las moléculas probadas, se utilizan las correlaciones con mayor ajuste para diseñar una expresión matemática, con capacidad de anticipar la actividad o propiedad de las moléculas análogas inexploradas o no probadas.

Sin embargo, aunque la estructura molecular contiene toda la información química, sólo una parte de esa información puede ser cuantificada por técnicas experimentales. Por ello, es necesario utilizar variables numéricas, conocidas como descriptores moleculares, con capacidad de codificar numerosas características estructurales de los sistemas moleculares.

Se puede inferir que un modelo QSAR busca cuantificar las relaciones SAR, mediante un paralelismo matemático racional, diseñado a partir de distintas técnicas de análisis de varianza matemática acopladas a las herramientas de la química computacional, caracterizadas por examinar diferentes representaciones moleculares, tal como la geometría euclídea, la teoría de grafos y la representación de vectores.^{38, 39}

1.5.2. Análisis MLR de Hansch

El análisis de Hansch y Fujita,⁴⁰ también conocido como método extra-termodinámico, es una aproximación de energía libre lineal para el diseño de fármacos inspirados en series de compuestos contemplados en la formulación de ecuaciones matemáticas de predicción, basadas en múltiples combinaciones lineales de parámetros fisicoquímicos sencillos.

Esta metodología, emplea el análisis MLR para determinar los mejores mínimos cuadrados que ajusten la variable dependiente (actividad o propiedad) a una combinación lineal de variables independientes (descriptores moleculares). La mejor correlación alcanzada, expresada en la Ec. 1.3, se utiliza para cuantificar la actividad de los compuestos químicos sin datos experimentales.⁵

$$\log \frac{1}{C} = -a\pi^2 + b\pi + \rho\sigma + cE_s + dS + e \quad (\text{Ec. 1.3})$$

Donde C es la actividad o propiedad evaluada; $a, b, \rho, c, y d$ son los coeficientes de regresión derivados de la curva lineal que los ajusta; π , σ y E_s son constantes de hidrofobicidad, electrónica y estérica del sustituyente, conceptualizadas por Hansch, Hammett y Taft respectivamente,^{3, 41, 42} S indica un factor topológico que vincula las

variaciones de forma que ocurren en las moléculas debido a la interacción con receptores; e es la constante de la ecuación.

En este enfoque, la respuesta final evaluada se expresa como $\log 1/C$. Por ello, cuanto menor sea el valor de $\log 1/C$, el potencial biológico, farmacológico o toxicológico de la molécula es mayor.

Para el modelo de Hansch y Fujita, el uso de descriptores π , σ , E_s o Refractividad Molar (MR), permitió recolectar información farmacológica contenida en compuestos orgánicos simples, y realizar predicciones en sistemas moleculares más complejos de manera rápida, confiable y a un bajo precio. Asimismo, logró obtener información cuantitativa de la actividad de nuevas moléculas dentro del límite de confiabilidad de la técnica MLR.

La principal ventaja de los métodos de regresión lineal es que presentan menos problemas de sobreajuste en la selección de los descriptores moleculares más representativos para la actividad, ya que no requieren la optimización de varios parámetros durante el diseño del modelo (solo un coeficiente de regresión por descriptor); por tanto, se obtienen ecuaciones matemáticas más simples de interpretar.⁴³

1.5.3. Análisis CoMFA

Varias investigaciones QSAR sugieren que la selectividad biológica resulta de cada objetivo que forma interacciones altamente específicas, tal como enlaces de hidrógeno con un ligando.⁷ No obstante, otras observaciones han indicado que las preferencias de unión al ligando, se deben principalmente a los efectos de campo no covalente ejercidos en el área espacial de los ligandos.

En la década de los 80s, esta visión motivo a proponer los estudios QSAR-3D, enfocados sobre el espacio de los Campos de Interacción Molecular (MIF), donde el tipo de representación de las estructuras moleculares, la cantidad de descriptores moleculares requeridos para caracterizar un campo de ligando y la disponibilidad de los datos de respuesta de las actividades, fueron determinantes al momento de establecer modelos con capacidad de predecir distintos efectos biológicos finales.⁴⁴⁻⁴⁶

En particular, el análisis CoMFA determina cuando las interacciones droga-receptor no son covalente, y si las variaciones en la actividad dependen de los cambios en los campos electrostáticos o estéricos de los fármacos.¹ Este método utiliza la estructura molecular optimizada de una sustancia química para definir el enrejado tridimensional alrededor de la misma, con el fin de medir las interacciones de los átomos que actúan como sonda en cada punto del enrejado, y así definir el contorno que reúne los puntos con valores de interacción similar.

En CoMFA se emplean variables de interacciones estéricas y electrónicas asociadas a miles de puntos del enrejado para describir el fenómeno que conduce a la respuesta final evaluada. Los resultados obtenidos generan expresiones matemáticas de predicción y mapas de contornos utilizados para interpretar las regiones más importantes para la actividad.

Actualmente, se encuentra disponible una serie de programas para desarrollar modelos QSAR-3D basados en métodos como CoMFA, HINT (CoMFA con hidrofobicidad), Análisis Comparativo de Índices de Similaridad Molecular (CoMSIA) y Análisis Comparativo del Sitio Activo Molecular (CoMASA).^{1, 47-49}

1.6. Elementos y etapas del modelado QSAR

El formalismo QSAR emplea diferentes técnicas estadísticas para diseñar los modelos de correlación, encargados de relacionar la información cuantitativa de la estructura molecular con un delimitado comportamiento activo de las mismas. La información requerida para diseñar un modelo QSAR proviene de dos fuentes:

- Cientos de descriptores moleculares derivados de las características estructurales y fisicoquímicas de las moléculas.

- Los valores experimentales de la actividad o propiedad de interés.

En general, el desarrollo, validación y aplicación de un modelo QSAR consta de 7 pasos generales de diseño, ilustrados en la Figura 1.3. Los cuales son considerados durante las etapas de preparación, procesamiento, predicción y validación, e interpretación del modelo; definidas brevemente en las siguientes subsecciones:

1.6.1. Preparación

Una adecuada preparación de los datos inicia con la selección de un conjunto molecular con datos experimentales confiables, generados bajo estrictas medidas de laboratorio que conducen a obtener información sobre una delimitada respuesta final.

Al modelar una propiedad biológica o de toxicidad, se evalúa principalmente la concentración requerida para evidenciar una respuesta puntual. En este caso, los parámetros de medida más utilizados son la IC₅₀ (concentración que produce el 50% de inhibición o de antagonismo al receptor), la EC₅₀ (concentración responsable del 50% de acción efectiva) o la LC₅₀ (concentración que causa la muerte al 50% de la población evaluada).

Cada medida mencionada anteriormente se obtiene a través de distintos protocolos de ensayos, que difieren tanto en las condiciones experimentales recomendadas como en los modelos animales o cultivos celulares evaluados.⁵⁰

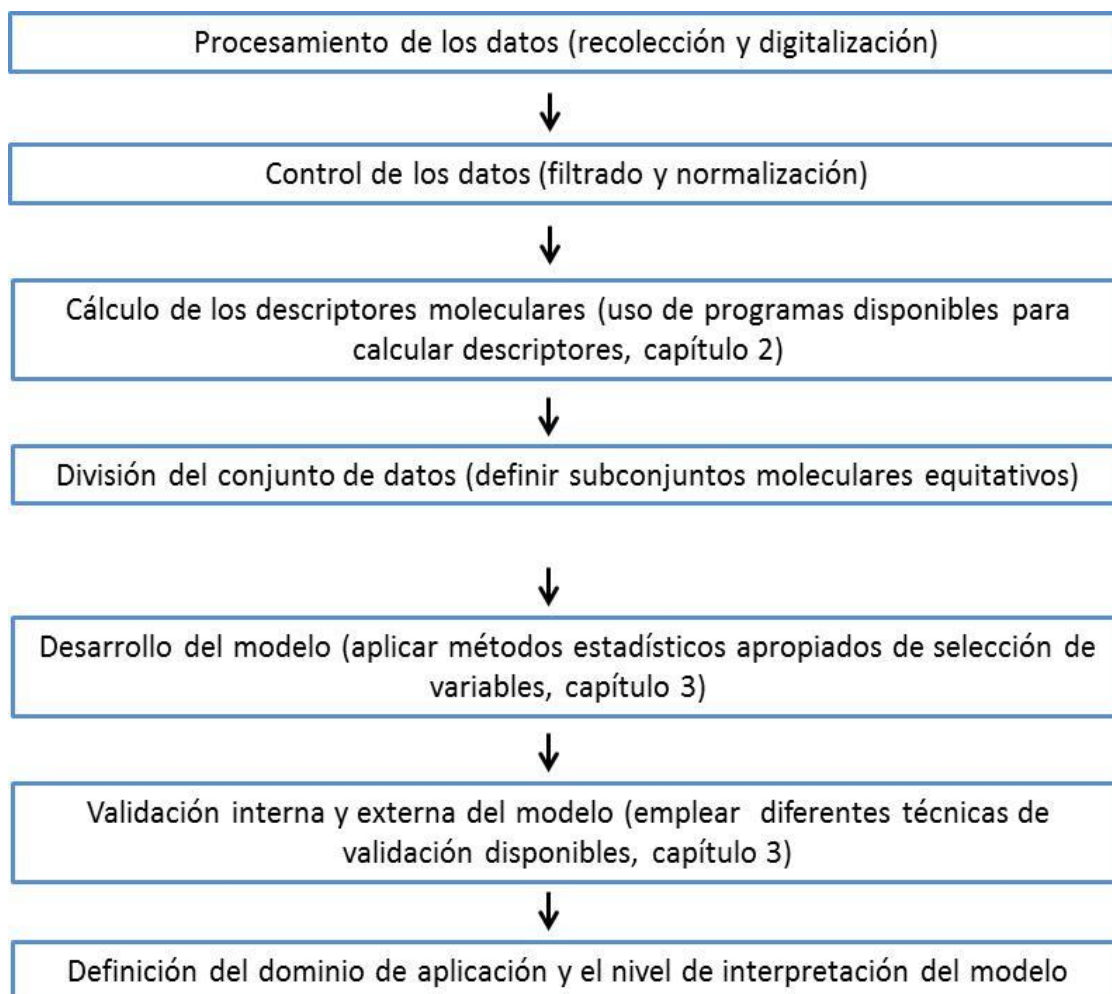


Figura 1.3. Esquema general del modelado QSAR. Fuente: Saavedra, L.M.

Un concepto fundamental sobre las medidas biológicas *in vivo* o *in vitro*, es que tiende a presentar mayor incertidumbre en relación con las medidas fisicoquímicas, ya que involucran más variables ajustables, que incorporan incerteza experimental al evaluar la respuesta producida por uno o varios sistemas vivos.⁵¹

Para analizar la mayor cantidad de datos disponibles de una propiedad (p) biológica o toxicológica, es necesario aplicar una transformación logarítmica, expresada como $\log 1/p$ o $-\log p$, para mantener los valores de respuesta en un rango numérico más cercano.

Otro aspecto clave en la preparación de los datos, está centrado en la representación adecuada de las estructuras moleculares utilizadas para calcular los descriptores moleculares necesarios para modelar la actividad. Por último, en esta etapa de diseño se crea una matriz de datos QSAR para iniciar el procesamiento y análisis de la información suministrada al modelo.

1.6.2. *Procesamiento*

La teoría QSAR es supervisada por un enfoque estadístico encargado de brindar predicciones confiables y consistentes de la propiedad o actividad modelada. Por esto, es importante verificar la contribución e idoneidad de los datos incorporados al modelo a través del criterio de pretratamiento de los datos, donde se eliminan los descriptores moleculares indeterminados, constantes o casi constantes (con un único valor diferente) y se excluyen las moléculas duplicadas.

En el desarrollo del modelo QSAR, es necesario dividir el conjunto molecular total en tres subconjuntos moleculares conocidos como calibración, validación y predicción, los cuales se emplean para diseñar, supervisar y probar el modelo, respectivamente.⁵² Además, cuando se cuenta con información adicional de moléculas inexploradas o sin datos experimentales de la propiedad evaluada, resulta oportuno establecer un nuevo conjunto molecular denominado aplicación, con el cual se mide la capacidad de aplicación del modelo.

El conjunto molecular inicial se puede dividir equitativamente considerando las características estructurales o los valores de la propiedad. También, suelen emplearse diferentes técnicas quimiométricas (método de análisis de agrupamientos *k*-medias (*k*-MCA), exclusión de esfera), que proveen subconjuntos equilibrados en la relación estructura-actividad.

1.6.3. *Validación y predicción*

Para considerar un modelo QSAR válido, es indispensable superar múltiples pruebas de validación interna, como la Validación Cruzada (CV) dejar-uno-afuera (LOO) o dejar-más-afuera (LMO). También, es necesario realizar una adecuada validación externa, que brinda información acerca de la capacidad del modelo para predecir la actividad del conjunto molecular de predicción.

Puntualmente, un modelo QSAR no está habilitado para predecir todo el universo de moléculas existentes o aún inexploradas. Por ello, es importante que cada modelo defina su dominio de aplicación (AD); es decir, que revele el alcance químico de sus predicciones. Hasta el momento, el enfoque QSAR cuenta con numerosos métodos de validación propuestos. Además, es un área en constante evolución, que propone pruebas cada vez más rigurosas y difíciles de cumplir.⁵³⁻⁵⁵

1.6.4. *Interpretación*

Actualmente, los modelos QSAR más novedosos están enfocados en predecir la actividad modelada, brindar una posible interpretación del fenómeno responsable de la respuesta y deducir algunos aspectos del mecanismo de acción involucrado, basándose en los datos experimentales o predichos de otras respuestas finales asociadas con la actividad evaluada. Como en el caso de los inhibidores irreversibles de

la AChE, que no solo promueven la actividad cerebral en los pacientes con AD; también están asociados con efectos insecticidas sobre distintos artrópodos.

En este contexto, es importante resaltar que el objetivo de un análisis QSAR es establecer predicciones acertadas de un delimitado efecto final generado por un conjunto de moléculas nuevas o aún inexploradas, a partir de la información contenida en la estructura molecular de las mismas.

Aunque, si el modelo presenta una base estadística sólida, cumple con diferentes parámetros de validación y anticipa acertadamente la actividad del conjunto externo de predicción, podría asistir el diseño racional y selectivo de nuevos agentes químicos con propiedades específicas.

1.7. Referencias

1. Roy, K., Kar, S., Nas, R.N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier: London. **2015**.
2. Hansch, C and Leo, A., *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*. Vol 1. American Chemical Society (ACS) Professional Reference Book: Washington DC. **1995**, p. 139–205.
3. Hansch, C & Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal American Chemical Society*. **1964**, 86 (8), 1616–1626.
4. Ramsden, C.A. *Quantitative drug design. Comprehensive medicinal chemistry*. In Hansch, C., Sammes, P.G., Taylor, J.B., Eds. The rational design, mechanistic study; therapeutic application of chemical compounds. Pergamon Press: Oxford. **1990**.
5. Hansch C.L.A. *Substituent constants for correlation analysis in chemistry and biology*. John Wiley and Sons: New York. **1979**.
6. Muhammad, U., Uzairu, A., Arthur, D.E., Review on: Quantitative Structure Activity Relationship (QSAR) modelling. *Journal Analytical & Pharmaceutical Research*. **2018**, 7(2), p.240–242.
7. Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To?. *Journal of Medicinal Chemistry*. **2014**, 57, p.4977–5010.
8. Burello, E. Review of (Q)SAR models for regulatory assessment of nanomaterials risks, *NanoImpact*. **2017**, 8, p.48–58.
9. Organization for Economic Co-operation and Development (OECD) [En línea]. OECD Quantitative Structure-Activity Relationships Project [(Q)SARs]., The OECD QSAR Project, 2019. [Citado el 17 de Abril de **2019**]. Disponible en: www.oecd.org/chemicalsafety/risk-assessment/oecdquantitativestructure-activityrelationshipsprojectqsars.htm
10. European Chemical Agency (ECHA). Guidance on information requirements and chemical safety assessment chapter 6: *QSARs and grouping of chemicals*. ECHA: Sofia and Bucharest, **2008**. Disponible en: <https://echa.europa.eu/es/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment>
11. United State Environmental Protection Agency (US EPA). *(Q)uantitative Structure Activity Relationship [(Q)SAR] guidance document*. Manibusan, M., Peterson, J., Kent, R., Chen, J., *et al.*, Eds. North American Free Trade Agreement (NAFTA) Technical Working Group on Pesticides (TWG). US EPA: Atlanta, **2012**. Disponible

en: <https://www.epa.gov/pesticide-registration/quantitative-structure-activity-relationship-qsar-guidance-document>

12. U.S. Food and Drug Administration (U. S. FDA) [En línea]. Applied Regulatory Science: Regulatory Science at CDER. Department of Health and Human Services, 2017 [Citado el 18 de Abril de 2019]. Disponible en: <https://www.fda.gov/drugs/scienceresearch/ucm294603.htm#Computational>
13. Demchuk, E., Ruiz, P., Chou, S., Fowler, B.A. SAR/QSAR methods in public health practice. *Toxicology and Applied Pharmacology*. **2011**, 254, p. 192–197.
14. Dearden, J.C. The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relationships*. **2016**, 1(1), p.1–44.
15. Benfenati, E., Diana, R.G., Gini, G., Cardamone, L., Gocieva, M., Mancusi, M., Padovani, R., Tamellini, L. *Introduction on in vivo, in vitro and in silico methods*. En Benfenati, E., Ed. The e-book on QSAR and REACH: theory, guidance and applications. Orchestra: Milan, **2012**, p. 5-37. Disponible en: http://ebook.insilico.eu/insilico-ebook-orchestrabenfenati-ed1_rev-June2013.pdf.
16. Kubinyi, H. QSAR and 3D QSAR in drug design. Part 1: methodology. *Drug Discovery Today*. **1997**, 2(11), p. 457–467.
17. Chen, B., Zhang, T., Bond, T., Gan, Y. Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: A review of methods and resources. *Journal of Hazardous Materials*. **2015**, 299, p.260–279.
18. Scotti, L., Ishiki, H.M., Cavalcante Duarte, M., Branquinho Oliveira, T., Scotti, M.T. *Computational Approaches in Multitarget Drug Discovery*. En Nicolotti, H., Ed. Computational Toxicology: Methods and Protocols. Humana Press: Bari, **2018**, p. 327-345.
19. Erzincan, P. Saçan, M.T., Yüce-Dursun, B., Daniş, Ö., Demir, S., Erdem, S.S., Ogan, A. QSAR models for antioxidant activity of new coumarin derivatives. *SAR and QSAR Environmental Research*. **2015**, 26(7-9), p. 721–737.
20. Khalafi-Nezhad, A., Soltani Rad, M.N., Mohabatkar, H., Asrari, Z., Hemmateenejad, B. Design, synthesis, antibacterial and QSAR studies of benzimidazole and imidazole chloroaryloxyalkyl derivatives. *Bioorganic Medical Chemistry*. **2005**, 13(6), p.1931–1938.
21. Garro Martinez, J.C., Vega-Hissi, E.G., Andrada, M.F., Estrada, M.R. QSAR and 3D-QSAR studies applied to compounds with anticonvulsant activity. *Expert Opinion on Drug Discovery*. **2015**, 10(1), p. 37–51.
22. Shoombuatong, W., Schaduangrat, N., Nantasenamat, C. Towards understanding aromatase inhibitory activity via QSAR modelling. *EXCLI Journal*. **2018**, 17, p. 688–708.
23. Gharaghani, S., Khayamian, T., Ebrahimi, M. Molecular dynamics simulation study and molecular docking descriptors in structure-based QSAR on

- acetylcholinesterase (AChE) inhibitors. *SAR and QSAR Environmental Research*. **2013**, 23 (9), p. 773–794.
24. Devillers, J. 2D and 3D structure–activity modelling of mosquito repellents: a review. *SAR and QSAR Environmental Research*. **2018**, 29(9), p. 693–723.
 25. Devillers, J., Lagneau, C., Lattes, A., Garrigues, J.C., Clémenté, M.M., Yébakima, A. In silico models for predicting vector control chemicals targeting *Aedes aegypti*. *SAR and QSAR Environmental Research*. **2014**, 25(10), p. 805–835.
 26. Anastas, P.T., Warner, J.C. *Green Chemistry: Theory and Practice*. Oxford University Press: New York. **1998**, p.30.
 27. Illing, P. General *Overview of the Safety Evaluation of Chemicals*. In Hester, R.E., Harrison, M.R., Eds. *Alternatives to Animal Testing*. Environmental Science and Technology. Vol 23. The Royal Society of Chemistry: Cambridge, **2006**, p.3–24.
 28. Cronin, M. Chapter 10: *Quantitative structure-activity relationships (QSARs) — applications and methodology*. In Puzyn, T., Leszczynski, J., Cronin, M.T.D., Eds. *Recent Advances in QSAR Studies: Methods and Applications*. Springer Press: Heidelberg. **2010**, p. 3–11.
 29. Dearden, J.C., Cronin, M.T.D., Kaiser, K.L.E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*. **2009**, 20, p. 241–266.
 30. Monev, V. Introduction to similarity searching in chemistry. *MATCH Communications in Mathematical and in Computer Chemistry*. **2004**, 51, p. 7–38.
 31. Saliner, G., Patlewicz, G., Worth, A.P. A similarity based approach for chemical category classification. European Commission Joint Research Centre Report number: EUR 21867 EN. Institute for Health and Consumer Protection, Toxicology and Chemical Substances Unit, European Chemicals Bureau, I-21020 Ispra. **2005**.
 32. Martín, M., Sanz, F., Campillo, M., Pardo, L., Péres, J., Turmo, J., Aulló, J.M. Quantum chemical structure-activity relationship on b-carbolines as natural monoamine oxidase inhibitors. *International Journal of Quantum Chemistry*. **1983**, 83, p. 1643–1652.
 33. Richard, A.M., Quantitative comparison of molecular electrostatic potential for structure activity studies. *Journal of Computational Chemistry*. **1991**, 12, p. 959–969.
 34. Rum, G., Herndon, W.C., Molecular similarity concepts. Analysis of steroid-protein binding constants. *Journal of American Chemical Society*. **1991**, 113, p. 9055–9060.
 35. OECD environment health and safety publications series on testing and assessment No. 69 Guidance document on the validation of (Quantitative) Structure - Activity Relationship [(Q)SAR] models., 2007. [Citado el 21 de Abril de **2019**]. Disponible en:[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote5env/jm/mono\(2007\)2&doclanguage5en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote5env/jm/mono(2007)2&doclanguage5en)

36. Walker, J.D., Jaworska, J., Comber, M.H.I., Schultz, W., Dearden, J.C Guidelines for developing and using quantitative structure–activity relationships. *Environmental Toxicology and Chemistry*. **2003**, 22 (8), p. 1653–1665.
37. Lessigiarska, I. Development of Structure-Activity Relationship for Pharmacotoxicological Endpoints Relevant to European Union Legislation. Thesis of the Liverpool John Moores University for the PhD Degree. School of Pharmacy and Chemistry & Institute of Health and Consumer Protection. Liverpool & Ispra. **2006**.
38. Johnson, A., M., Maggiora, G, M., Concepts and Applications of Molecular Similarity. *Journal of Computational Chemistry*. **1992**, 13, p. 539–540.
39. Carbó-Dorca, R., Robert, D., Amat, L. I., Jirones, X., Besalú, E. *Molecular Quantum Similarity in QSAR and Drug Research*. Springer-Verlag: Berlin. **2000**.
40. Hansch, C., Maloney P.P., Fujita, T., Muir, R.M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*. **1962**, 194, p. 178–180.
41. Hammett, L.P. Some relations between reaction rates and equilibrium constants. *Chemical Review*. **1935**, 17(1), p.125–136.
42. Taft Jr, R.W. Linear free-energy relationships from rates of esterification and hydrolysis of aliphatic and ortho-substituted benzoate esters. *Journal of American Chemical Society*. **1952**, 74, p. 2729–2732.
43. Karcher, W., Devillers, J. SAR and QSAR in environmental chemistry and toxicology: Scientific tool or whimsical thinking?. En *SAR and QSAR in environmental chemistry and toxicology*, Karcher, W., Devillers, J. Kluwer Academic Publishers: Dordrecht. **1990**, p. 1–12.
44. Verloop, A., Hoogenstraaten, W., Tipker, J. Development and Application of New Steric Substituent Parameters. En *Ariens, E.J., Ed. Academic Press: New York*. **1976**, p 165–207.
45. Cramer, R.D., III., Patterson, D.E., Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). Effect of Shape on Binding of Steroids to Carrier Proteins. *Journal of American Chemical Society*. **1988**, 110, p. 5959–5967.
46. Kubinyi, H. QSAR and 3D QSAR in drug design. Part2: Applications and problems. *DDT*. **1997**, 2, p. 538–546.
47. Kellogg, G.E., Semus, S.F., Abraham, D.J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *Journal of Computer-aided Molecular Design*. 1991, 5(6), 545–552.
48. Klebe, G. *Comparative Molecular Similarity Indices Analysis: CoMSIA*. En *3D QSAR in Drug Desing. Three-Dimensional Quantitative Structure Activity Relationships*, vol 3. Kubinyi, H., Folkers, G., Martin, Y.C., Eds. Springer Press: Dordrecht. **1998**, p. 87-104.
49. Kotani, T., Higashiura, K. Comparative Molecular Active Site Analysis (CoMASA). 1. An Approach to Rapid Evaluation of 3D QSAR. *Journal of Medicinal Chemistry*. **2004**, 47(11), p. 2732–2742.

50. Van Leeuwen, K., Schultz, T. W., Henry, T., Diderich, B., Veith, G.D. Using Chemical Categories To Fill Data Gaps in Hazard Assessment. SAR & QSAR in Environmental Research. **2009**, 20, p. 207–220.
51. Peter, S.C., Dhanja, J.K., Malik, V., Radhakrishnan, N., Jayakanthan, M., Sundar, D. *Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications*. En: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds. Encyclopedia of Bioinformatics and Computational Biology. Vol 2, Elsevier: Amsterdam. **2019**, p. 661-676.
52. Benfenati, E., Clook, M., Fryday, S. Hart, A. *QSARs for regulatory purposes: the case for pesticide authorization*. En: Benfenati, E., Ed. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes, Elsevier Science Ltd: Amsterdam. **2007**, p. 1–57.
53. Gramatica, P. Principles of QSAR models validation: internal and external. Molecular Informatics. **2007**, 26, p. 694–701.
54. Chirico, N., Gramatica, P. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. Journal of Chemical Information and Modeling. **2011**, 51, p. 2320–2335.
55. Roy, K., Mitra, I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Combinatorial Chemistry High Throughput Screening. **2011**, 14, p. 450–474.

CAPÍTULO 2: DESCRIPTORES MOLECULARES

2.1 Introducción

Los estudios cuantitativos de la química se enfocan en identificar los componentes básicos de los sistemas moleculares y buscar una relación con la lógica matemática para resolver los problemas de la naturaleza. En este contexto, los descriptores moleculares han promovido y facilitado la construcción de bases teóricas funcionales que asocien de manera racional las características estructurales de los compuestos químicos con los efectos benignos o adversos exhibidos por los mismos.

Cada descriptor molecular es un patrón de información codificada asociado a una pequeña parte de la estructura molecular de una delimitada molécula.¹⁻³ En QSAR, los descriptores moleculares inicialmente codificaban algunas propiedades fisicoquímicas comunes, como el coeficiente de partición octanol/agua o Log *P*, el punto de ebullición, fusión y condensación y la densidad; medidas a través de procedimientos estandarizados.

Con el sobresaliente progreso de la química y la computación surgieron decenas de descriptores de carácter teórico, basados en algoritmos matemáticos bien definidos y ajustados para identificar de forma automática un determinado aspecto estructural o medida experimental, y luego codificarlo en número útil e interpretable.

En general, los descriptores moleculares se dividen en dos principales grupos, los descriptores basados en los aspectos conformacionales de la molécula o los descriptores independientes de la conformación molecular; los cuales se pueden generar mediante diversos programas para calcular descriptores moleculares de acceso libre o comercial.

Particularmente los descriptores no conformacionales especifican las características invariantes de la estructura molecular; es decir, los aspectos atómicos, constitucionales y topológicos de una molécula.

Estos descriptores se dividen en 3 diferentes categorías: los descriptores adimensionales (0D) codifican los aspectos más simples de la estructura molecular, tal como la fórmula molecular. Seguidamente los descriptores unidimensionales (1D) o indicadores de fragmentos interpretan numéricamente las subestructuras presentes, con base en el conteo de los grupos funcionales. Luego, se encuentran los descriptores bidimensionales (2D) basados en la teoría de grafos, logran cuantificar la topología y la conectividad entre los átomos que conforman la molécula.⁴

Por otra parte, los descriptores conformacionales se caracterizan por definir la estructura molecular de forma tridimensional (3D) para codificar la mayor cantidad de características geométricas, electrónicas y espaciales presentes en la estructura molecular dada. Otros estudios más avanzados de la conformación molecular han incluido otros niveles de representación molecular (4D, 5D, y 6D) para resolver tanto la estructura molecular como el entorno interactuante.

Para calcular los descriptores geométricos de manera más precisa, es indispensable conocer la conformación activa de las moléculas, identificada por cristalografía de rayos X o espectrometría de Resonancia Magnética Nuclear (RMN).⁵ De lo contrario, es necesario utilizar las estructuras moleculares optimizadas bajo un método de estructura electrónica apropiado, que resuelva una aproximación de la conformación activa en el espacio (coordenadas x , y , z), para iniciar el cálculo de los descriptores geométricos.^{6,7}

Si bien los descriptores dependientes de la conformación brindan una representación más específica de la molécula, que por lo general, se asocia con la disposición real de los compuestos químicos en la naturaleza, también involucran más parámetros ajustables, que incrementan la incertidumbre en las predicciones de los modelos QSAR.⁸

A medida que aumentaba la cantidad, facilidad y rapidez de cálculo de los descriptores moleculares, incrementó la posibilidad de caracterizar más aspectos estructurales o experimentales de las moléculas, esenciales para desarrollar modelos prospectivos con capacidad de estimar los efectos fisicoquímicos, tóxicos, organolépticos o biológicos contenidos en diversos agentes químicos nuevos o hipotéticos.^{9,10}

Por ello, es posible destacar que los descriptores moleculares han contribuido de forma directa a identificar y revolver varios problemas en la salud humana o los ecosistemas, asociados a los efectos nocivos de los productos químicos comerciales.¹¹ Adicionalmente, en los campos de la toxicología y farmacología computacional, los descriptores moleculares han permitido identificar decenas de candidatos a fármaco, así como estimar anticipadamente la toxicidad *in vivo* aguda y crónica, sin necesidad de realizar un extenuante número de ensayos en diferentes modelos animales.¹²⁻¹⁷

Actualmente, existe una amplia disponibilidad de descriptores moleculares capacitados para representar las características estructurales más complejas, particularmente localizadas en un complejo proteico o complejo proteína-ligando; así como, evaluar la toxicidad y el destino ambiental de un agente químico antropogénico,^{18, 19} e interpretar las propiedades sensoriales más relevantes para la elaboración de nuevos productos cosméticos o de alimentos.^{20,21}

Por lo anterior, en este capítulo de Tesis Doctoral se detallaron los aspectos más relevantes de los descriptores moleculares, los principales esquemas de

representación molecular utilizados en el modelado QSAR, las características generales y específicas de las clases de descriptores, así como las principales ventajas y limitaciones de cada clase. Además, se definieron las principales herramientas computacionales o plataformas en línea para calcular miles de los descriptores moleculares de forma libre o con pago previo.

2.2. Conceptos básicos de los descriptores moleculares

Una de las tareas fundamentales para comprender los complejos procesos de la naturaleza, radica en poder estimar e interpretar los efectos finales de las sustancias químicas, a través de representaciones simbólicas de las estructuras moleculares, codificadas por los descriptores moleculares.

La primera definición de los descriptores moleculares, señalaba que: *un descriptor molecular es el resultado de un procedimiento lógico matemático encargado de transformar la información química codificada dentro de una representación simbólica de la molécula en un número útil o el resultado de algún experimento estandarizado.*^{22, 23}

En particular, cada descriptor molecular codifica una pequeña parte de toda la información química contenida en la estructura molecular. Este enfoque, ha promovido el constante interés de la comunidad científica por incrementar el número de descriptores moleculares disponibles, con el fin de generar mayor conocimiento acerca del funcionamiento de los sistemas químicos y biológicos de difícil acceso experimental o aún inexplorados.²⁴

En consecuencia, la extensa variedad de descriptores moleculares existentes hace referencia a las diversas formas o perspectivas de caracterizar un sistema molecular, considerando diferentes elementos y disposiciones de la estructura química. Es así como el conjunto global de descriptores moleculares postulados hasta el momento ha contribuido en el desarrollo de múltiples estudios computacionales, operados por medio de herramientas estadísticas, quimiométricas y quimioinformáticas.

Sin embargo, no cualquier variable matemática puede ser considerada un descriptor molecular. Los descriptores deben aportar, en mayor o menor medida, información química útil, cumplir algunas propiedades básicas de invariancia, y derivarse de métodos bien definidos, reproducibles y de fácil repetitividad en otros conjuntos moleculares.

Ante la complejidad que involucra predecir cualquier respuesta biológica, toxicológica, fisicoquímica o sensorial generada por una sustancia química determinada, resulta importante postular descriptores moleculares que conserven un balance entre el grado de interpretación del descriptor y la capacidad de predecir una propiedad o actividad de manera acertada.^{25, 26} De este modo, aumenta la

probabilidad de obtener un modelo QSAR conformado por una combinación óptima de descriptores moleculares, fácilmente interpretables y con capacidad de predecir la actividad de forma coherente.²⁷

Por otra parte, es evidente que un descriptor molecular o un pequeño conjunto de estos, no está capacitado para interpretar completamente la complejidad molecular que interviene en los procesos fisicoquímicos o las interacciones biológicas responsables de las propiedades.

Por ello, al considerar un modelo basado en descriptores moleculares, las predicciones generadas son una aproximación al resultado final del mecanismo de acción involucrado. Aunque en muchos casos, es más factible desarrollar un modelo de predicción QSAR para evaluar de manera preliminar las respuestas finales generadas por un gran conjunto de moléculas, que ejecutar los costosos e imprácticos procedimientos experimentales basados en la prueba y el error en cada una de estas.²⁸

En términos generales, los descriptores moleculares seguirán desempeñando un rol importante en los avances científicos. De hecho, la creciente disponibilidad de un gran número de descriptores moleculares con alto contenido de información química extraída de la estructura molecular, acompañada de la inclusión de nuevos algoritmos matemáticos que los resuelven, y la automatización de los programas disponibles para calcularlos, han generado un extenso conocimiento acerca de las relaciones existentes entre los sistemas moleculares y la evidencia experimental.

2.3. Representación de la estructura molecular

Antes de iniciar el cálculo de los descriptores moleculares es indispensable digitalizar las estructuras químicas en un formato de representación molecular acorde con el propósito del estudio. Actualmente, existen diferentes formatos gráficos o de texto para digitalizar una molécula, que difieren según el contenido de información estructural.⁴ Las formas más comunes para codificar pequeñas moléculas son:

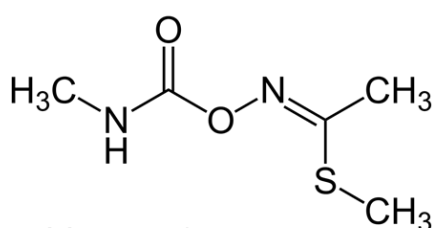
- Identificador Internacional de Compuestos Químicos (InChI)
- Sistema de Entrada Molecular Lineal Simplificado (SMILES)
- Sistema de información MDL de datos moleculares (MOL)
- Sistema de información MDL de datos estructurales (SDF)

Los formatos moleculares SMILES e InChI son notaciones lineales de cadena que codifican la conectividad entre los átomos individuales de una molécula;^{29, 30} pero no logran capturar toda la información conformacional de la misma.

Si bien el formato SMILES está capacitado para codificar decenas de estructuras moleculares de forma rápida y sencilla, utiliza diferentes formalismos de escritura (canónico o isomérico) que pueden generar ambigüedad. Por el contrario, la notación InChI expresa de forma inequívoca algunos aspectos de la estereoquímica de una

molécula. Un caso particular de esta discrepancia se expresa en la Figura 2.1, para la estructura molecular del pesticida Metomil.³¹

Por otra parte, para interpretar la estructura tridimensional de una sustancia química se emplean los esquemas gráficos MDL de tipo sdf o mol, encargados de identificar y codificar la posición de los átomos de una molécula en el sistema de coordenadas cartesianas.



Nombre Comercial= Metomil
Fórmula molecular= C₅H₁₀N₂O₂S
Número CAS= 16752-77-5
Nombre IUPAC= S-metil -N-(metilcarbamoiloxi)tioacetimidato
SMILE canónico= CC(=NOC(=O)NC)SC
SMILES isomérico= C/C(=N\OC(=O)NC)/SC
InChI=1S/C5H10N2O2S/c1-4(10-3)7-9-5(8)6-2/h1-3H3,(H,6,8)/b7-4+

Figura 2.1. Notaciones lineales SMILES e InChI del pesticida Metomil. Fuente: PPDB.³¹

La representación molecular MDL mol logra capturar toda la información acerca de la conectividad, posición y tipo de átomos y enlaces que constituyen una molécula. En general, el formato MDL mol cuenta con dos versiones de codificación denominadas V2000 y V3000. La versión V2000 captura la información contenida en hasta 999 átomos y enlaces; mientras que la versión V3000 es compatible para más de 999 átomos y enlaces; esta versión suele emplearse para representar macromoléculas en los estudios de modelado molecular y acoplamiento molecular (Docking).¹

El esquema gráfico MDL sdf proporciona de manera gráfica la localización espacial (x, y, z) de los átomos individuales que conforman el compuesto químico. En algunos casos, el archivo sdf contiene información adicional sobre el estado de hibridación y de conectividad molecular. En particular, los formatos gráficos de representación molecular mol y sdf son compatibles con la mayoría de los programas para calcular descriptores moleculares.^{1,7}

2.4. Dimensionalidad de los descriptores moleculares

Para que una sustancia química genere un determinado efecto final en un organismo o entorno expuesto, deben ocurrir una serie de procesos que dependen de las características estructurales de dicha sustancia. La codificación correcta de esas características representa la base del modelado QSAR.

Existen diferentes formas para describir una estructura molecular de acuerdo con el punto final evaluado y las características del conjunto molecular elegido para diseñar el modelo.³² La información capturada por los descriptores moleculares varía desde las propiedades químicas más simples hasta las características conformacionales más complejas. Puntualmente, el contenido de información de un descriptor molecular depende directamente del nivel de representación (dimensionalidad) utilizado para definir la estructura molecular.

Por esta razón, la dimensionalidad es el principal criterio de clasificación de los descriptores moleculares, dividiéndolos entre descriptores no conformacionales (0D, 1D y 2D) y descriptores basados en la conformación (3D). Cada categoría comprende un delimitado nivel de incertidumbre, discriminación, complejidad de cálculo, contenido de información e interpretación del sistema molecular. Por este motivo, se definen de forma detallada a continuación:

2.5. Descriptores adimensionales (0D)

La representación más simple de un descriptor adimensional (0D) es la fórmula química, que permite identificar el tipo y la cantidad de elementos químicos que conforman una molécula. Un ejemplo particular se muestra para el compuesto químico 2,3,7,8-Tetraclorodibenzo-p-dioxina, reconocido por sus efectos tóxicos para la salud humana y el ambiente; los descriptores 0D utilizan su fórmula química $C_{12}H_4Cl_4O_2$,³³ para identificar la presencia de 12 átomos de carbono, 4 de hidrógeno, 4 de cloro y 2 de oxígeno.

Esta forma de representar una sustancia química carece de cualquier conocimiento sobre conectividad atómica o tipos de enlace. Por lo tanto, los descriptores moleculares 0D solo pueden capturar las características estructurales más generales de la molécula.³⁴ Los descriptores 0D más conocidos son: el número de átomos de cualquier elemento químico, el peso molecular (MW), la suma o el promedio de las propiedades atómicas, tal como los volúmenes atómicos de van der Waals o la electronegatividad.

En general, para generar los descriptores 0D no es necesario disponer de algún programa de cálculo de descriptores, ya que codifican los aspectos más básicos de la estructura molecular, fáciles de cuantificar e interpretar. No obstante, esta clase de descriptores se caracteriza por su bajo contenido de información química y alto grado de degeneración, lo que indica que pueden tener valores iguales para diferentes moléculas.

2.6. Descriptores unidimensionales (1D) e indicadores

A medida que aumente la dimensionalidad de los descriptores, incrementa la complejidad de cálculo, el contenido de información y la capacidad discriminante entre estructuras similares, tal como lo exhibe la Figura 2.2.

Los descriptores unidimensionales (1D) o indicadores de fragmentos interpretan las moléculas a través de las subestructuras, los grupos funcionales o los fragmentos de las mismas. Para generar estos descriptores moleculares no es necesario disponer de la conformación de las estructuras moleculares.

La representación 1D de la molécula se refleja mediante los descriptores indicadores de fragmentos o de frecuencia de ocurrencia, encargados de codificar la presencia o ausencia de una subestructura dada a través del código binario.

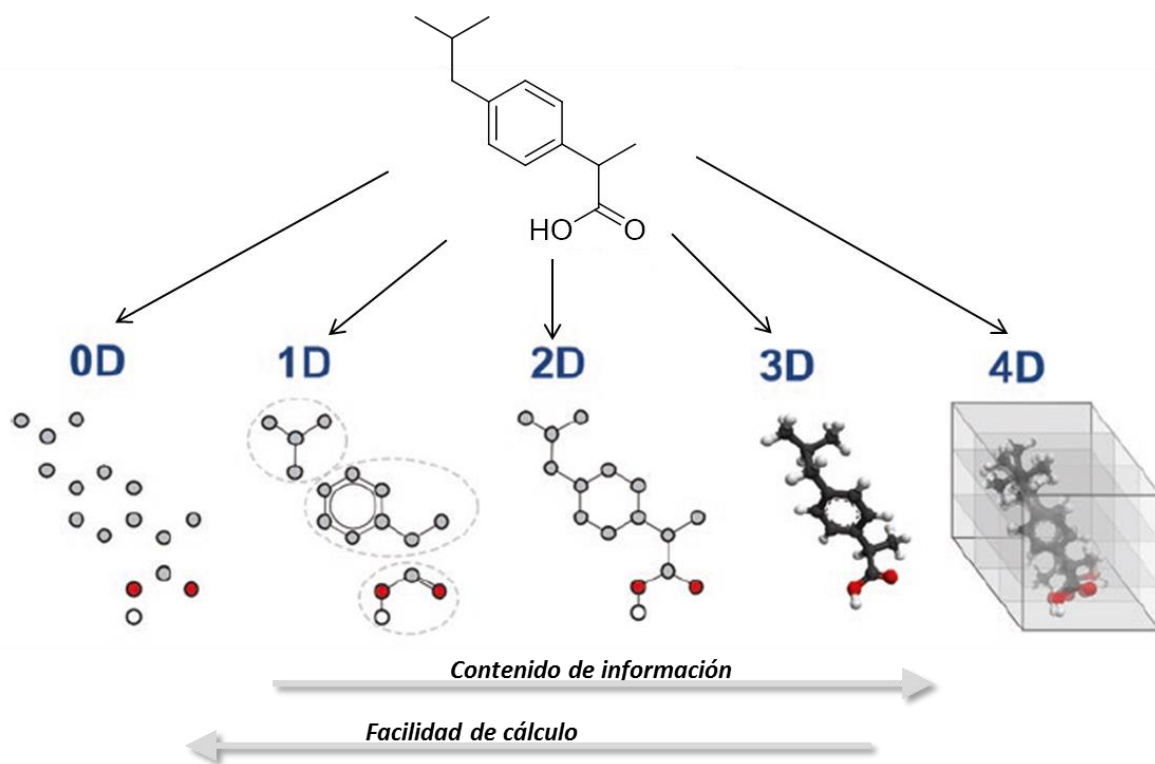


Figura 2.2. Niveles de representación del ácido 2-(4-isobutilfenil) propiónico, según la dimensión, facilidad de cálculo y contenido de información. Fuente: Grisoni *et al.*, 2018.³⁴

Los descriptores 1D permiten considerar las características estructurales basadas en los grupos funcionales o conjuntos de átomos que influyen en las respuestas biológicas de las moléculas. Estos descriptores también se conocen como variables ficticias o descriptores booleanos, debido a su capacidad de establecer nexos lógicos matemáticos, que explican la relación de presencia o ausencia de uno o varios elementos químicos en la estructura molecular, para codificar la posición o el número de sustituyentes en la misma.

De acuerdo con un principio simple de aditividad, donde las modificaciones químicas orientadas en la adición o eliminación de grupos sustituyentes alteran el efecto biológico exhibido inicialmente por la molécula; los descriptores 1D permiten cuantificar el grado de contribución que tiene un grupo sustituyente en la respuesta biológica evaluada.

Esta estrategia fue propuesta por Free & Wilson en 1964, y se conoce actualmente como el método de Free-Wilson.³⁵ Consiste en generar una matriz de datos conformada por valores de 1 y 0; cada columna de la matriz corresponde a un sustituyente particular en una posición específica de la molécula, y se considera una variable independiente. En este caso, la variable dependiente ubicada en una de las columnas de la matriz de datos, incluye los valores de la actividad de cada molécula.

Seguidamente, se aplica la técnica de Regresión Lineal Múltiple (MLR) para seleccionar las variables independientes más representativas para la variable dependiente, conforme a los parámetros estadísticos clásicos para medir la bondad de ajuste. Los coeficientes de regresión del modelo representan la contribución que tienen los sustituyentes de la molécula sobre la actividad.

La estrategia Free-Wilson ha sido utilizada para desarrollar diversos modelos de la teoría QSAR, con resultados de predicción acertados y con una interpretación mecanicista bastante sencilla.³⁶⁻³⁸ No obstante, a diferencia de los modelos QSAR clásicos derivados de algunas propiedades fisicoquímicas de las moléculas; los modelos QSAR basados en el método de Free-Wilson no pueden estimar el efecto biológico generado por aquellos sustituyentes no contemplados durante el diseño del modelo.

Por esta razón, otros análisis QSAR han considerado los descriptores basados en la frecuencia de aparición de los átomos o grupos funcionales específicos. En este caso, la frecuencia debe ser lo suficientemente alta para que los valores de los coeficientes de regresión sean confiables y representativos para estimar la propiedad.^{39, 40}

En este contexto, algunos científicos han recurrido al uso de diversos descriptores indicadores de fragmentos como alertas estructurales, ya que son independientes y específicos para algunas propiedades, tal como la actividad carcinogénica con al menos 35 alertas estructurales bien definidas, basadas en subestructuras de riesgo propuestas por Benigni *et al.*, 2008.⁴¹

Conforme a lo anterior, es posible destacar que la información contenida en los descriptores 1D o indicadores de fragmentos es útil cuando el conjunto molecular analizado es mayormente homogéneo y la respuesta final evaluada se asociada a los aspectos más simples de la estructura molecular.

2.7. Descriptores bidimensionales (2D)

Básicamente, los descriptores 2D basados en la teoría de grafos, capturan la información relacionada a la conectividad de los átomos en términos de presencia y

naturaleza de enlace químico para codificar las conexiones en la estructura molecular, cuyos vértices y aristas son los átomos y enlaces, respectivamente.

En general, un grafo molecular (G) se refiere a un tipo de representación molecular basado en la topología de un compuesto químico, que se calcula de la siguiente forma:

$$G = (V, E) \quad (2.1)$$

En la Ec. 2.1 el parámetro V corresponde al conjunto de vértices, y E es el conjunto de elementos que representan la relación binaria entre pares de vértices. En un grafo molecular, los pares de vértices (pares de átomos) son considerados una arista, es decir un enlace entre dichos átomos.

Por lo general, los grafos moleculares consideran todos los elementos y enlaces químicos que comprenden una molécula; solo en algunos casos, el grafo molecular procurará excluir los átomos de hidrógeno para simplificar la información topológica cuantificable; este tipo de grafo se conoce como grafo molecular carente de hidrógenos.

Para cuantificar la información contenida en un grafo molecular, es común utilizar matrices grafo-teóricas, que codifican la información estructural proporcionada por el grafo molecular. Las matrices teóricas de un grafo molecular pueden ser de varios tipos:

- Matriz de vértices: conformada por filas y columnas que representan los vértices (átomos) del grafo, y buscan representar diferentes propiedades de los pares de vértices.

- Matriz de aristas: conformada por filas y columnas basadas en las aristas del grafo (los enlaces intramoleculares), para codificar algunas propiedades de pares de aristas.

- Matriz de adyacencia: codifica diferente información sobre los pares de vértices, tal como su conectividad.

- Matriz de distancia: identifica las distancias topológicas entre aristas.

- Matriz ponderada: suma los pesos de los vértices presentes en un camino topológico conectado

- Matriz aumentada: las entradas diagonales de la matriz pueden ser iguales a 0, o codificar información química sobre los vértices.

A diferencia de las propiedades atómicas más comunes, las invariantes locales de vértice son cantidades numéricas derivadas de la topología molecular, que se utilizan para caracterizar otras propiedades de los átomos, tal como el grado de vértice, suma

de distancias del vértice o excentricidad del átomo; estas propiedades son comúnmente dadas como ponderaciones atómicas.

Por otra parte, las invariantes de un grafo molecular son cantidades matemáticas derivadas de una representación gráfica de la molécula, comúnmente empleadas para identificar propiedades teóricas de los grafos, asentadas por isomorfismo; es decir, propiedades con valores idénticos para grafos moleculares isomorfos.

En los últimos años, el aumento en el número de fórmulas y algoritmos disponibles para tratar la información contenida en los grafos moleculares, por medio de diferentes matrices moleculares y esquemas de ponderación, ha dado lugar a nuevas clases de invariantes de grafo útiles.⁴²

Todos los índices derivados del grafo molecular se conocen como Índices Topológicos (ITs); estos codifican de forma directa y sin ambigüedades diferentes propiedades topológicas específicas como la adyacencia o la conectividad de una molécula, y más generales como el tamaño y peso molecular, la forma, la simetría, el grado de ramificación y la ciclicidad.

Adicionalmente, tienen la capacidad de considerar propiedades químicas específicas de los átomos, como la presencia de átomos electroatrayentes o electrodadores de electrones, la polarizabilidad y la masa atómica.⁴³⁻⁴⁵ En particular, los ITs se dividen en dos principales grupos:

- Índices topoestructurales: describen la información acerca de la adyacencia y las distancias de enlace directo entre dos átomos.
- Índices topoquímicos: describen información topológica de la molécula, pero también detallan aspectos químicos específicos de los átomos, como el estado de hibridación y la identidad química.

En general, los ITs se enfocan en cuantificar principalmente las distancias entre átomos calculados por el número de enlaces intermedios, por lo tanto se consideran índices de enlace directo.

Otro tipo de invariante del grafo molecular son los índices de autocorrelación, propuestos Moreau & Broto en 1980.⁴⁶ Estos índices se encargan de definir una relación entre los vértices en función de su separación espacial.

Los índices de autocorrelación más comunes se obtienen a partir de un conjunto de puntos discretos en el espacio (los átomos de la molécula) y una función evaluada en esos puntos (una propiedad atómica definida).

Cada descriptor de autocorrelación corresponde a la integración de los productos de la función calculada en el átomo x y el átomo $x+k$; donde k es la distancia topológica. Este tipo de descriptores expresan como se correlacionan los valores de la función a intervalos iguales de distancia topológica.

Por otra parte, Randić en 1991,⁴⁷ propuso que los ITs deben presentar ciertas características específicas para interpretar de forma coherente la topología molecular; éstas consisten en:

1. Interpretación estructural
2. Alta correlación con una o varias propiedades
3. Capacidad de discriminar isómeros
4. Destinados a la estructura local
5. Ser independientes
6. Detallar información simple
7. No interpretar propiedades experimentales
8. Mínima intercorrelaciones con otros ITs
9. Provenientes de algoritmos de fácil reproducibilidad y repetitividad
10. Considerar aspectos estructurales simples
11. Exhibir una dependencia apropiada
12. Presentar cambios graduales acorde con los cambios estructurales.

La gran mayoría de los descriptores topológicos propuestos hasta el momento presentan las características anteriores. Debido a esto, han logrado intervenir exitosamente en decenas de estudios QSAR orientados en medir el efecto que tienen las similitudes o diferencias estructurales de las moléculas sobre la actividad o propiedad evaluada.

Hasta el momento, los ITs más utilizados en QSAR son: el índice W de Wiener, índice de conectividad de Randić, índice de conectividad de Kier & Hall, índice de conectividad J de Balaban, el índice Zagreb, índice de Rucker e índices de estado electrotopológico,⁴⁸⁻⁵⁰ los cuales se detallan individualmente a continuación.

2.7.1. Índice W de Wiener

En la década de los 40s, el científico Wiener postuló uno de los primeros descriptores moleculares basados en la topología molecular de un conjunto de alcanos. De este modo, Wiener logró establecer el descriptor de conteo de caminos, basado en la suma del número de enlaces unidos a todos los pares de átomos.⁵¹

Años más tarde, el nombre de dichos descriptores fue modificado debido a que los caminos pueden presentar longitudes variables en el grafo molecular. Por ello, se planteó el denominado índice W de Wiener, definido como la mitad de la suma de todas las entradas en la matriz de distancia acorde con la Ec. 2.2.

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij} = \frac{1}{2} \sum_{i=1}^n S_i \quad (2.2)$$

A partir del índice W , se derivaron una gran cantidad de fórmulas para analizar diversas clases de grafos moleculares. Luego, Wiener postuló el descriptor de polaridad p de un grafo, bajo el concepto de número de pares de vértices separados por tres aristas; es decir, el número de distancias de longitud tres.⁵²

2.7.2. Índice de conectividad de Randić

Los índices de conectividad son los descriptores más populares en términos de análisis del grafo molecular. El primer índice de conectividad fue propuesto por Randić en 1975.⁵³ A este descriptor se le conoce como índice de conectividad de Randić (χ) o índice de ramificaciones, y se define por medio de la Ec. 2.3.

$$\chi = \sum_{\text{todas aristas}} (\delta_i \delta_j)^{-\frac{1}{2}} \quad (2.3)$$

Donde δ_i y δ_j corresponden a los grados de vértices de los átomos incidentes en el enlace considerado, a partir de los enlaces i y j de la molécula. El índice χ no sólo posee buena degeneración, también alta correlación con diversas propiedades fisicoquímicas o bioquímicas. Acorde con su definición, cuanto más compacto sea el grafo molecular mayor será el valor de χ .

2.7.3. Índice de conectividad de Kier & Hall

Años más tarde, los científicos Kier & Hall plantearon un esquema general basado en el índice de conectividad de Randić para calcular descriptores de orden cero y de orden superior, con el fin de obtener índices de conectividad de i -ésimo orden. Comúnmente, estos descriptores son conocidos como índices de conectividad de Kier & Hall, y se pueden obtener a través de las ecuaciones 2.4–2.7.⁵⁴⁻⁵⁶

$${}^0\chi = \sum_{\text{todas aristas}} \delta^{-\frac{1}{2}} \quad (2.4)$$

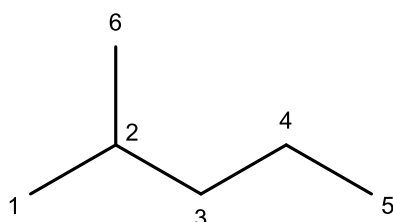
$${}^1\chi = \sum_{\text{todas aristas}} (\delta_i \delta_j)^{-\frac{1}{2}} \quad (2.5)$$

$${}^2\chi = \sum_{k=1}^n (\delta_i \delta_l \delta_j)_k^{-\frac{1}{2}} \quad (2.6)$$

$${}^m\chi_q = \sum_{k=1}^K \left(\prod_{a=1}^n \delta_a \right)_k \quad (2.7)$$

En este caso, el parámetro k recorre todos los subgrafos de i -ésimo orden conformadas por n átomos ($m+1$ para subgrafos acíclicos). K es el número total de subgrafos de i -ésimo orden presentes en el grafo molecular, y en el caso de los subgrafos de camino es igual al recuento de aristas de i -ésimo orden.^{22, 23}

El producto está sobre los vértices simples de grado δ , que involucran a todos los vértices de cada subgrafo. La variable q , representa el tipo de subgrafo molecular considerado. En general, se usa ch para identificar anillos o cadenas lineales; pc para camino de grupo; c para grupo y p para camino. Un ejemplo de aplicación de los índices de conectividad de Kier & Hall para la molécula 2-metilpentano se muestra en la Figura 2.3.



$${}^0\chi = (\delta_1\delta_1)^{-\frac{1}{2}} + (\delta_2\delta_2)^{-\frac{1}{2}} + (\delta_3\delta_3)^{-\frac{1}{2}} + (\delta_4\delta_4)^{-\frac{1}{2}} + (\delta_5\delta_5)^{-\frac{1}{2}} + (\delta_6\delta_6)^{-\frac{1}{2}} =$$

$$(1 \cdot 1)^{-\frac{1}{2}} + (3 \cdot 3)^{-\frac{1}{2}} + (2 \cdot 2)^{-\frac{1}{2}} + (2 \cdot 2)^{-\frac{1}{2}} + (1 \cdot 1)^{-\frac{1}{2}} + (1 \cdot 1)^{-\frac{1}{2}} = 4.333$$

$${}^1\chi = (\delta_1\delta_2)^{-\frac{1}{2}} + (\delta_2\delta_3)^{-\frac{1}{2}} + (\delta_3\delta_4)^{-\frac{1}{2}} + (\delta_4\delta_5)^{-\frac{1}{2}} + (\delta_2\delta_6)^{-\frac{1}{2}} =$$

$$(1 \cdot 3)^{-\frac{1}{2}} + (3 \cdot 2)^{-\frac{1}{2}} + (2 \cdot 2)^{-\frac{1}{2}} + (2 \cdot 1)^{-\frac{1}{2}} + (3 \cdot 1)^{-\frac{1}{2}} = 2.770$$

$${}^2\chi = (\delta_1\delta_2\delta_3)^{-\frac{1}{2}} + (\delta_2\delta_3\delta_4)^{-\frac{1}{2}} + (\delta_3\delta_4\delta_5)^{-\frac{1}{2}} + (\delta_1\delta_2\delta_6)^{-\frac{1}{2}} + (\delta_3\delta_2\delta_6)^{-\frac{1}{2}} =$$

$$(1 \cdot 3 \cdot 2)^{-\frac{1}{2}} + (3 \cdot 2 \cdot 2)^{-\frac{1}{2}} + (2 \cdot 2 \cdot 1)^{-\frac{1}{2}} + (1 \cdot 3 \cdot 1)^{-\frac{1}{2}} + (2 \cdot 3 \cdot 1)^{-\frac{1}{2}} = 2.183$$

Figura 2.3. Índices de conectividad de Kier & Hall para la molécula 2-metilpentano. Fuente: Todeschini *et al.*, 2000.²²

2.7.4. Índice de conectividad J de Balaban

La complejidad del cálculo de los ITs se encuentra directamente asociada al tamaño del grafo molecular analizado. A raíz de esto, Balaban entre 1979 y 1985 realizó una serie de modificaciones de la fórmula del índice χ propuesto por Randić,⁵⁷⁻⁵⁹ enfocadas en reemplazar los grados de vértices por suma de distancias, e introdujo un factor de corrección que involucra el número ciclomático para obtener el índice de conectividad J , expresado en la siguiente ecuación:

$$J = \frac{e}{\mu + 1} \sum_{\text{todas aristas}} (\delta_i\delta_j)^{-\frac{1}{2}} \quad (2.8)$$

Donde el parámetro e representa el número de aristas del grafo y la sumatoria es sobre todos los pares de aristas adyacentes i y j . La degeneración de este índice es

mucho menor que los ITs mencionados anteriormente. Además, se estableció que el grafo molecular más compacto es el que posee mayor índice J .

2.7.5. Índice de Zagreb

En la década de los 70s, el científico croata Zagreb propuso dos ITs derivados de la matriz de adyacencia, con el objetivo de establecer otras aplicaciones químicas de la teoría de grafos, a través de las ecuaciones 2.9 y 2.10.⁶⁰

$$M_1 = \sum_{i=2}^n v_i^2 \quad (2.9)$$

$$M_2 = \sum_{\text{todas aristas}} v_i v_j \quad (2.10)$$

2.7.6. Índice de Rucker

Los científicos Rucker & Rucker en 1993,⁶¹ postularon una serie de descriptores topológicos basados en el recuento de todos los caminos de mayor longitud, enfocándose en todos los caminos, excepto en los que regresan a ellos mismos. Si bien estos descriptores se obtienen bajo un razonamiento sencillo de la estructura molecular, como en un procedimiento de suma tipo Morgan; tienen un alto poder discriminante entre los compuestos isómeros o la disposición de los átomos en la estructura molecular.

El IT basado en el recuento total de caminos se conoce como *twc*, y consiste en un código para grafos moleculares fundamentado en todos los caminos, que otorga un alto poder discriminante y menor degeneración que el índice de conectividad de Balaban.

2.7.7. Índice de estado electrotopológico

El estado electrotopológico S_i del i -ésimo átomo de la molécula, denominado índice de estado electrotopológico o estado-E, proporciona información relacionada con el estado electrónico y topológico del átomo en la molécula,^{57, 62} expresado como:

$$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^A \frac{I_i - I_j}{(d_{ij} + 1)^k} \quad (2.11)$$

Donde I_i indica el estado intrínseco del i -ésimo átomo y ΔI_i es el efecto de campo en el i -ésimo átomo calculado como perturbación del estado intrínseco del i -ésimo átomo por todos los demás átomos de la molécula; el parámetro d_{ij} es la distancia topológica entre el i -ésimo y el j -ésimo átomo; A es el número de átomos. El

exponente k modifica la influencia de los átomos distantes o cercanos en estudios particulares; regularmente $k=2$.²²

El valor del estado-E codifica el estado electrónico intrínseco del átomo influenciado por el ambiente electrónico de todos los otros átomos dentro de la estructura molecular. Para calcular el estado intrínseco del i -ésimo átomo se aplica la siguiente ecuación:

$$I_i = \frac{\left(\frac{2}{L_i}\right)^2 \cdot \delta_i^v + 1}{\delta_i} \quad (2.12)$$

En este caso, el parámetro L_i es el número cuántico principal (2 para los átomos de C, N, O y F, y 3 para los átomos de Si, S y Cl); δ_i^v es el número de electrones de valencia (grado de vértice de valencia), y δ_i es el número de electrones sigma (grados de vértice) del i -ésimo átomo en un grafo molecular carente de hidrógenos.

2.8. Descriptores tridimensionales (3D)

Un nivel adicional de complejidad se ve reflejado al percibir la molécula no solo en términos de tipo de átomo, conectividad y adyacencia, sino también observándola como un objeto geométrico en el espacio, caracterizado por la configuración espacial de los átomos; es decir, cuando la molécula se define en términos de tipos de átomos y su posición en el sistema de coordenadas cartesianas x, y, z .

Los descriptores moleculares basados en la representación tridimensional (3D) de la molécula tienen un alto contenido de información que logra explicar adecuadamente el efecto de la estructura molecular sobre diversas respuestas biológicas, toxicológicas, organolépticas y fisicoquímicas.⁶³

Sin embargo, para calcular de manera precisa los descriptores geométricos, es necesario conocer los datos de la conformación activa de cada molécula, determinados por técnicas de cristalografía de rayos X o espectrometría de RMN.⁵ Aunque, si esta información no está disponible, se deben utilizar las estructuras moleculares optimizadas bajo un nivel de teoría de estructura electrónica apropiado para realizar el cálculo de los descriptores 3D.⁶⁴⁻⁶⁶

Con base en lo anterior, durante el proceso de optimización de las estructuras moleculares, es importante reconocer y superar algunos problemas emergentes, como:

1. La influencia del método, el funcional o la base seleccionada para obtener la conformación del mínimo de energía global en el sistema de coordenadas x, y, z .

2. La posibilidad de encontrar en la Superficie de Energía Potencial (PES) más de un conformero con energía mínima similar, particularmente en las moléculas con alta capacidad de interconversión.

3. La diferencia existente entre la conformación activa y la conformación optimizada; donde el nivel de deformación depende directamente de los grados de libertad conformacionales de la molécula.

En general, los descriptores geométricos se dividen en 4 principales grupos, puntualizados en la Figura 2.4.

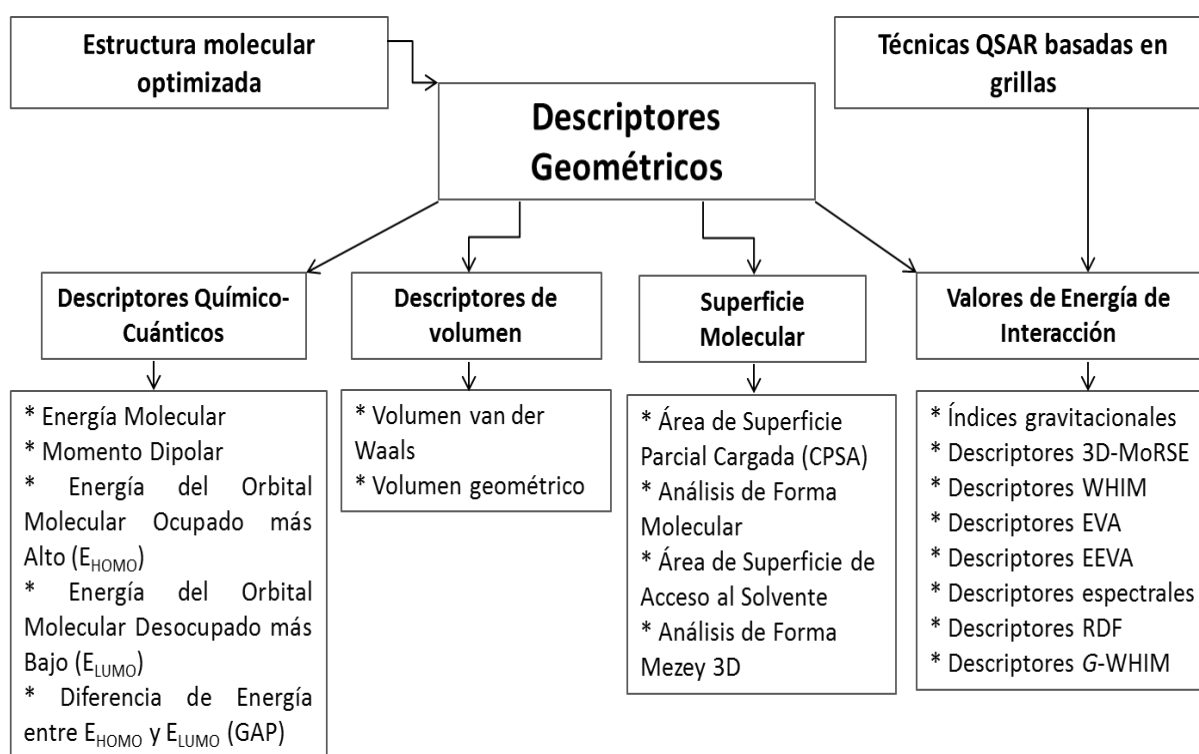


Figura 2.4. Clasificación general de los descriptores 3D. Fuente: Todeschini *et al.*, 2000.²²

Entre estos grupos, los descriptores basados en los valores de energía de interacción se han destacado en el modelado QSAR por su capacidad de correlación con diversas actividades y su habilidad discriminante; donde se destacan las siguientes clases de descriptores:

- Descriptores Invariantes Holísticos Ponderados (WHIM)⁶⁷
- Descriptores de Ensamblado de Geometría, Topología y Pesos Atómicos (GETAWAY)⁶⁸
- Descriptores de Representación Molecular 3D de la estructura basada en difracción de electrones (3D-Morse)^{69, 70}
- Descriptores de la Función de Distribución Radial (RDF)⁷¹
- Descriptores de Autovalores (EVA)⁷²

Para comprender el enfoque conformacional abordado por cada una de estas clases de descriptores, y su importante relación con diversas respuestas finales modeladas; se dispone a continuación una definición breve de las mismas:

2.8.1. Descriptores WHIM

Los descriptores *WHIM* están basados en índices estadísticos calculados sobre las proyecciones de los átomos a lo largo de los ejes principales de la molécula. Estos descriptores se construyen de tal forma que capturen información molecular tridimensional relevante enfocada en el tamaño, forma, distribución y simetría de los átomos con respecto a los marcos de referencia invariantes de la molécula.

En principio, el cálculo de los descriptores *WHIM* implica realizar un Análisis de Componentes Principales (PCA) en las coordenadas cartesianas centradas de una molécula, empleando una matriz de covarianza ponderada, constituida por diferentes esquemas de ponderación para los átomos. Para cada esquema de ponderación, se calcula un conjunto de índices estadísticos sobre los átomos proyectados en cada componente principal.

2.8.2. Descriptores GETAWAY

Los descriptores GETAWAY son derivados de la matriz de influencia molecular (**H**), que representa la estructura molecular a través de la Ec. 2.13.

$$\mathbf{H} = \mathbf{M}(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \quad (2.13)$$

Donde el parámetro **M** es la matriz molecular que contiene las coordenadas cartesianas centradas (x, y, z) de los átomos de la molécula, en una conformación optimizada elegida.

Es importante destacar que las coordenadas atómicas se calculan con respecto al centro de geometría de la molécula para obtener la invarianza traslacional. La matriz de influencia molecular es una matriz simétrica $n \times n$, donde n representa el número de átomos, y muestra la invarianza rotacional con respecto a las coordenadas de la molécula, lo que resulta independiente de las reglas de alineación.

Los elementos diagonales de la matriz (h_{ii}) presentan valores en el rango de 0 a 1 para codificar la información atómica relacionada con la “influencia” de cada átomo en la molécula, y así determinar la conformación molecular. Específicamente, los átomos más superficiales de la estructura molecular siempre tienen valores h_{ii} más altos que los átomos más próximos al centro de la molécula.⁶⁸

Los descriptores de tipo GETAWAY se obtienen a través de funciones de autocorrelación con doble ponderación, donde un esquema de ponderación es el

parámetro de influencia y el otro una propiedad atómica, tal como la polarizabilidad o la masa atómica.

2.8.3. Descriptores 3D-MoRSE

El conjunto de descriptores 3D-MoRSE fue propuesto por Selzer *et al.*, 1996,⁶⁹ con el objetivo de acceder a la información de las coordenadas atómicas tridimensionales, a través de la transformación utilizada en los estudios de difracción de electrones para preparar curvas de dispersión teóricas, considerada en la Ec. 2.14.

$$I(s) = \sum_{i=1}^{A-1} \sum_{j=i+1}^A \varpi_i \varpi_j \frac{\text{sen}(sr_{ij})}{sr_{ij}} \quad (2.14)$$

Donde $I(s)$ es la intensidad de electrones dispersos; ϖ una propiedad atómica; el parámetro r_{ij} es la distancia interatómica entre los átomos i -ésimo y j -ésimo; y A es la cantidad de átomos.

2.8.4. Descriptores RDF

Los descriptores de Función de Distribución Radial (RDF) se basan en la distribución de distancias en la representación geométrica de una molécula para generar un código RDF, que exhibe ciertas características en común con el código 3D-MoRSE.

Formalmente, el código RDF de un conjunto de átomos A puede interpretarse como la probabilidad de encontrar un átomo en un conjunto esférico de radio R . La forma general de un código RDF está representada por:

$$g(R) = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A \varpi_i \cdot \varpi_j \cdot e^{-\beta(R-r_{ij})^2} \quad (2.15)$$

Donde f es un factor de escalado; el parámetro ϖ representa las propiedades atómicas de los átomos i y j ; r_{ij} es la distancia interatómica entre el i -ésimo átomo y el j -ésimo átomo; A es el número de átomos. El término exponencial contiene la distancia r_{ij} entre los átomos i y j y el parámetro normalizado β , que define la probabilidad de la distancia interatómica individual. En general, el término $g(R)$ se calcula en varios puntos discretos con intervalos definidos.

2.8.5. Descriptores EVA

Los descriptores de Autovalores (EVA) fueron propuestos por Ferguson *et al.*, 1997,⁷² para extraer la información química estructural de los espectros de infrarrojo (IR) y Raman. Este enfoque consiste en utilizar como descriptor multivariante las frecuencias vibratorias de una molécula y una propiedad molecular fundamental, caracterizada de forma fácil y confiable a partir de la función de energía potencial.

Cada descriptor EVA es una función de los valores propios obtenidos de la matriz de coordenadas cartesianas, las cuales corresponden a las frecuencias vibratorias fundamentales de la molécula, que pueden ser calculadas por medio de métodos mecanocuánticos o métodos moleculares estándares de química computacional.

2.9. Descriptores basados en grillas (4D)

Un mayor nivel de complejidad y contenido de información se puede lograr al analizar la molécula a través de un enfoque 4D, con el objetivo de identificar y caracterizar cuantitativamente las interacciones entre las moléculas y uno o varios sitios activos de un receptor biológico.

En este contexto, se puede obtener una representación molecular basada en la disposición de la molécula dentro de un enrejado tridimensional compuesto por varios miles de puntos espaciados uniformemente, utilizando una sonda (estérica, electrostática, hidrofílica, hidrofóbica, entre otras) para mapear la superficie de la molécula.

A partir del trabajo publicado por Cramer *et al.*, 1988,⁷³ se planteó el uso de las propiedades de campo de las moléculas en el espacio tridimensional para desarrollar y aplicar modelos de la teoría QSAR. Este método fue denominado Análisis Comparativo de Campo Molecular (CoMFA); conocido como el primer enfoque QSAR-3D que continúa vigente hasta ahora.^{74, 75}

El fundamento básico del método CoMFA consiste en muestrear los campos estéricos (van der Waals) y electrostáticos (Coulombicos) alrededor de un conjunto de moléculas alineadas, con el fin de capturar toda la información necesaria para explicar la respuesta final exhibida por las mismas. El muestreo consiste en calcular las energías de interacción de las moléculas, mediante sondas apropiadas localizadas en el enrejado tridimensional dispuesto alrededor de las estructuras moleculares.

Un paso crítico del método CoMFA es la apropiada alineación de las moléculas, que puede tomar tiempo y requiere el conocimiento previo de la conformación molecular precisa. Por esta razón, es necesario analizar las moléculas más activas del conjunto de datos y usarlas como plantilla.

Adicionalmente, cuando el conjunto de datos presenta moléculas con varios grados de libertad conformacional, la selección de la conformación activa suele ser un

obstáculo significativo en el modelado QSAR-3D. Por ello, se debe ejecutar una búsqueda conformacional sistemática previa, para definir el conformero más estable según la Superficie de Energía Potencial (PES) generada.

Actualmente, existen diferentes métodos computacionales para mejorar el rendimiento en la alineación de la técnica CoMFA.⁷⁶⁻⁷⁹ Un método análogo a CoMFA es el Análisis Comparativo de Índices de Similaridad Molecular (CoMSIA).⁸⁰

El método CoMSIA utiliza la similitud entre los átomos de la molécula junto con una sonda localizada en el enrejado tridimensional alrededor del sistema molecular, para evaluar cinco propiedades relacionadas con los campos estéricos, electrostáticos, hidrofóbicos, donadores y receptores de enlaces de hidrógeno.

La posibilidad de emplear más campos ha permitido analizar de forma más precisa las regiones espaciales donde las contribuciones de los diferentes campos tienen un importante efecto sobre la actividad de las moléculas. Otros métodos derivados de CoMFA, son el Análisis Comparativo del Sitio Activo Molecular (CoMASA),⁸¹ y el Patrón de Reactividad Común (COREPAN).⁸²

El método CoMASA se caracteriza por ejecutar un análisis QSAR-3D rápido, eficiente y sencillo, basado en 3 principales ventajas:

(i) Genera mapas de fácil interpretación, para transformarlos en farmacóforos o atributos estructurales requeridos en los patrones de búsqueda de las bases de datos.

(ii) No es obligatorio considerar la orientación de la molécula dentro de un enrejado tridimensional o una grilla en particular.

(iii) Se puede ejecutar de manera rápida y simple en procesadores comunes y accesibles, ya que tiene funciones de puntuación y reducción de puntos de interacción eficaces.

Con respecto al método COREPAN, su principal ventaja radica en evadir los problemas asociados a la alineación de las moléculas y la selección de un átomo o fragmento de farmacóforo específico; ya que realiza un análisis de la distribución conformacional de la molécula, basado en los parámetros de reactividad global y local potencialmente asociados con el punto final estudiado.

Durante los últimos 15 años, se han incorporado nuevos descriptores multidimensionales en el modelado QSAR, denominados descriptores 5D y 6D; estos se basan en los parámetros estructurales asociados a la flexibilidad del sitio de unión al receptor junto con la topología del ligando.

Puntualmente, los descriptores 5D son calculados a partir de múltiples conformaciones, orientaciones, estados de protonación y estereoisómeros del ligando analizado. En el caso de los descriptores 6D, es necesario tener en cuenta los escenarios de solvatación del complejo, los ligandos y el entorno interactuante.⁸³⁻⁸⁵

2.10. Ventajas y limitaciones según la dimensionalidad de los descriptores

Los descriptores moleculares, ya sean conformacionales (3D) o independientes de la conformación (0D–2D) codifican una cantidad delimitada de características estructurales de la molécula, que por lo general, se pueden asociar racionalmente con la actividad o propiedad exhibida por la misma.

Sin embargo, cada enfoque presenta una serie de ventajas y limitaciones detallada en la Tabla 2.1, que deben ser consideradas al momento de abordar un conjunto de datos a través del modelado QSAR.^{86, 87}

Tabla 2.1. Principales ventajas y limitaciones de los tipos de descriptores moleculares. Fuente: Duchowicz, 2019.⁸⁶

Detalle	Descriptores 0D–2D	Descriptores 3D
Complejidad	Involucra un cálculo directo de los descriptores, a partir de una notación lineal en cadena o un esquema gráfico de la estructura molecular	Requiere las estructuras moleculares optimizadas bajo un método de estructura electrónica apropiado para calcular los descriptores
Tiempo	Poco tiempo de cálculo	Mayor tiempo de cálculo
Requisitos	No requiere el conocimiento de la conformación molecular	Requiere información sobre la conformación de la molécula, dada por técnicas apropiadas
Contenido de información	Información estructural basada en la composición atómica y la conectividad	Información estructural basada en la composición atómica, la conectividad y varios parámetros conformacionales
Isomería	Los isómeros se consideran mediante indicadores de presencia o ausencia de ciertos elementos geométricos	Se realiza un cálculo directo de los descriptores de isómeros
Capacidad predictiva	Evita ambigüedades 3D e incrementa el poder predictivo del modelo QSAR, con valores estables de la propiedad predicha	Si se conoce la conformación activa de cada molécula, el modelo QSAR presentará alto poder predictivo y discriminante

2.11. Programas para calcular descriptores moleculares

Los continuos avances en las investigaciones de química computacional, en particular de la teoría QSAR, han favorecido la creciente producción del material intelectual basado en nuevos programas o herramientas virtuales para calcular miles

de descriptores moleculares e indicadores de fragmentos de forma fácil y rápida, a través de entornos de programación amigables de uso libre o comercial.

Ante la extensa diversidad de programas disponibles para calcular descriptores moleculares, en la Tabla 2.2 se recopilaron las principales herramientas o plataformas disponibles hasta 2019, empleadas en los estudios QSAR para generar miles de descriptores teóricos, experimentales, semiempíricos o indicadores de fragmentos de forma automática.

Tabla 2.2. Lista de los principales programas y plataformas disponibles para calcular descriptores moleculares e indicadores de fragmentos. Fuente: Saavedra, L.M.

Programa o Plataforma	Descripción	Acceso
ACD/LogP freeware	Calcula Log <i>P</i> a partir de los fragmentos estructurales	Libre
ACD/Tox Suite	Predice puntos finales de toxicidad	Comercial
ADAPT	Calcula descriptores topológicos, geométricos y electrónicos	Comercial
ADMET predictor	Calcula propiedades ADMET	Comercial
AFGen	Calcula descriptores basados en fragmentos topológicos	Libre
BioPPsy	Calcula propiedades farmacocinéticas para estimar compuestos candidatos a fármacos	Libre
BioTriangle	Calcula descriptores moleculares de dimensionalidad 0D-3D	Libre
BlueDesc	Calcula descriptores a partir de los aspectos estructurales más generales de la molécula	Libre
CACTVS	Calcula descriptores 0D-2D y geométricos (3D)	Libre
CDK	Calcula descriptores topológicos, geométricos, constitucionales y de carga	Libre
ChemDes	Calcula 3679 descriptores moleculares y 59 tipos de indicadores de fragmentos	Plataforma web
CODESSA PRO	Estima descriptores químico-cuánticos, topológicos y electrónicos	Comercial
CoFFer	Estima moléculas y genera alertas de los fragmentos asociados a la propiedad evaluada	Libre
CORAL	Permite el diseño de modelos QSAR basados en descriptores flexibles	Libre
Danish	Genera descriptores semiempíricos de propiedades toxicológicas en más de 166.000 moléculas	Libre

DRAGON	La versión DRAGON 7 calcula 5270 descriptores OD-3D e indicadores fragmentos	Comercial
ECOSAR	Determina propiedades vinculadas a toxicidad acuosa.	Libre
E-DRAGON	Calcula 1600 descriptores moleculares incluidos en el programa DRAGON	Libre
EPI Suite™	Programa de la EPA para estimar propiedades fisicoquímicas con destino ambiental	Libre
eTOXlab	Realiza modelado para generar información útil para el monitoreo de nuevos productos químicos	Libre
GRID	Calcula el campo de fuerza empírica GRID en el punto de la cuadrícula	Comercial
GUSAR	Calcula propiedades de toxicidad aguda y crónica en ratas	Libre
HYBOT-PLUS	Estima enlaces de hidrógeno y factores de energía libre	Comercial
ISIDA-Fragmentor	Calcula fragmentos subestructurales y fragmentos marcados con la propiedad ISIDA	Libre
MD-LOVIs	Genera 352.000 descriptores moleculares conforme al vector de pesos atómicos e invariantes de vértices locales	Libre
MolCode Toolbox	Estima propiedades biológicas y toxicológicas de monitoreo ambiental y sanitario	Comercial
MOLCONN-Z	Calcula los descriptores topológicos comunes, como electrotopológicos	Comercial
Mold²	Calcula 777 descriptores basados en la topología molecular	Libre
MOLE db	Brinda información de 1124 descriptores moleculares calculados en 234.773 moléculas	Plataforma web
MOLGEN	Genera 708 descriptores aritméticos, topológicos y geométricos	Libre
MolInfo	Estima propiedades estructurales y fisicoquímicas de la molécula	Plataforma web
MolSig	Calcula descriptores según la estereoquímica de la molécula	Libre
OCHEM	Proporciona información teórica o experimental de varias propiedades fisicoquímicas o biológicas	Libre

Open3dqsar	Ejecuta un análisis quimiométrico de alto rendimiento de los campos de interacción molecular	Libre
PaDEL	Calcula 1.875 descriptores 1D, 2D y 3D, así como 12 tipos de descriptores indicadores de fragmentos (16.092 bits)	Libre
Pattern Match Counter	Realiza el conteo de los grupos funcionales o los fragmentos estructurales de la molécula	Plataforma web
PowerMV	Permite el análisis estadístico, visualización de la molécula, cálculo de descriptores y búsqueda de similitud estructural	Libre
QSARpro	Calcula más de 1000 descriptores moleculares topológicos y geométricos	Comercial
QuBILS-MAS	Calcula descriptores moleculares algebraicos de las formas bilineales (2D), cuadráticas y lineales	Libre
SiRMS	Genera descriptores 2D para moléculas aisladas, mezclas, semi-mezclas y reacciones	Libre
SYBYL-X	Calcula descriptores geométricos EVA relacionados al análisis de campo CoMFA y CoMSIA	Comercial
T.E.S.T.	Estima diferentes puntos finales de toxicidad a partir de modelos QSAR	Libre
TMACC	Genera descriptores 2D de correlación cruzada máxima topológica	Libre
VEGA HUB	Herramienta <i>in silico</i> de predicción de propiedades con fines regulatorios o académicos	Libre
VIDEAN	Análisis de descriptores visuales e interactivos	Libre

2.12. Referencias

1. Roy, K., Kar, S., Nas, R.N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier: London. **2015**.
2. Testa, B., Kier, L.B. The concept of molecular structure in structure-activity relationship studies and drug design. *Medicinal Research Review*. **1991**, 11(1), p. 35–48.
3. Kier, L.B. & Hall, L.H. *Molecular connectivity in structure-activity analysis*. Research Studies. **1986**.
4. Benfenati, E., Diana, R.G., Gini, G., Cardamone, L., Gocieva, M., Mancusi, M., Padovaní, R., Tamellini, L. Introduction on *in vivo*, *in vitro* and *in silico* methods. En Benfenati, E., Ed. The e-book on QSAR and REACH: theory, guidance and applications. Orchestra: Milan, **2012**, p. 5–37. Disponible en: http://ebook.insilico.eu/insilico-ebook-orchestrabenfenati-ed1_rev-June2013.pdf.
5. Kubinyi, H. QSAR and 3D QSAR in drug design. Part 1: methodology. *Drug Discovery Today*. **1997**, 2(11), p. 457–467.
6. Randic, M. Molecular profiles novel geometry-dependent molecular descriptors. *New Journal Chemistry*. **1995**. 19, p. 781–791.
7. Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To?. *Journal of Medicinal Chemistry*. **2014**, 57, p.4977–5010.
8. Thomsen, M. QSAR in environmental risk assessment – Interpretation and validation of SAR/QSAR based on multivariate data analysis. Department of Environmental Chemistry in National Environmental Research Institute & Department of Life Science and Chemistry, Roskilde. **2001**.
9. European Chemical Agency (ECHA). Guidance on information requirements and chemical safety assessment chapter 6: *QSARs and grouping of chemicals*. ECHA: Sofía and Bucharest, **2008**. Disponible en: <https://echa.europa.eu/es/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment>
10. United State Environmental Protection Agency (US EPA). *(Q)uantitative Structure Activity Relationship [(Q)SAR] guidance document*. Manibusan, M., Peterson, J., Kent, R., Chen, J., Eds. North American Free Trade Agreement (NAFTA) Technical Working Group on Pesticides (TWG). US EPA: Atlanta, **2012**. Disponible en: <https://www.epa.gov/pesticide-registration/quantitative-structure-activity-relationship-qsar-guidance-document>
11. Burello, E. Review of (Q)SAR models for regulatory assessment of nanomaterials risks, *NanoImpact*. **2017**, 8 p.48–58.

12. Höfer, T., Gerner, I., Gundert-Remy, U., Liebsch, M., Schulte, A., Spielmann, H., Vogel, R., Wettig, K. Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Archives of Toxicology*. **2004**, 78, p. 549–564.
13. Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., Zakharov, A., Worth, A., Richard, A.M., Grulke, C.M., Trisciuzzi, D., Fourches, D., Horvath, D., Benfenati, E., Muratov, E., Wedebye, E.B., Grisoni, F., Mangiatordi, G.F., Incisivo, G.M., Hong, H., Ng, H.W., Tetko, I.V., Balabin, I., Kancherla, J., Shen, J., Burton, J., Nicklaus, M., Cassotti, M., Nikolov, N.G., Nicolotti, O., Andersson, P.L., Zang, Q., Politi, R., Beger, R.D., Todeschini, R., Huang, R., Farag, S., Rosenberg, S.A., Slavov, S., Hu, X., Judson, R.S. CERAPP: collaborative estrogen receptor activity prediction project. *Environmental Health Perspectives*. **2016**, 124 (7), p. 1023–1033.
14. Tong, W., Fang, H., Hong, H., Xie, Q., Perkins, R., Sheehan, D.M. Receptor-mediated toxicity: QSARs for estrogen receptor binding and priority setting of potential estrogenic endocrine disruptors. CRC Press, Boca Raton. **2004**.
15. Grisoni, F., Reker, D., Schneider, P., Friedrich, L., Consonni, V., Todeschini, R., Koeberle, A., Werz, O., Schneider, G. Matrix-based molecular descriptors for prospective virtual compound screening. *Molecular Informatics*. **2017**, 36, p. 1–7.
16. Ekins, S., Mestres, J., Testa, B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*. **2007**, 152, p. 9–20.
17. Jacob, L., Vert, J.P. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*. **2008**, 24, p. 2149–2156.
18. Carlsen, L., Walker, J.D. QSARs for prioritizing PBT substances to promote pollution prevention. *QSAR & Combinatorial Science*. **2003**, 22, p. 49–57.
19. Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R., Consonni, V. Quantitative structure–activity relationship models for ready biodegradability of chemicals. *Journal of Chemical Information and Modeling*. **2013**, 53, p. 867–878.
20. Martinez-Mayorga, K., Medina-Franco, J.L. Chapter 2: Chemoinformatics—applications in food chemistry. *Advances in Food and Nutrition Research*. **2009**, 58, p. 33–56
21. Rojas, C., Todeschini, R., Ballabio, D., Mauri, A., Consonni, V., Tripaldi, P., Grisoni, F. A QSTR-based expert system to predict sweetness of molecules. *Frontiers in Chemistry*. **2017**, 5, p. 53. doi: 10.3389/fchem.2017.00053
22. Todeschini, R., Consonni, V. *Handbook of molecular descriptors: Methods and Principles in Medicinal Chemistry*. Vol 11. Wiley-VCH: Weinheim. **2000**.
23. Todeschini, R., Consonni, V. *Molecular Descriptors for Chemoinformatics*. Vol 1. WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim. **2009**.
24. Kubinyi, H. QSAR in Drug Design. In *Handbook of Chemoinformatics*. Gasteiger, J. Ed. Wiley-VCH: Weinheim. **2003**, p. 1532–1554.

25. Karcher, W., Devillers, J. Eds. Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology. Kluwer Academic: London. **1992**.
26. Steinberg, C.E.W., Haitzer, M., Brüggemann, R., Perminova, I.V., Yashchenko, N.Y., Petrosyan, V.S. Towards a quantitative structure-activity relationship (QSAR) of dissolved humic substances as detoxifying agents in freshwater. *International Review of Hidrobiology*. **2000**, 85, p. 253–266.
27. Schultz, T.W., Cronin, M.T.D., Walker, J.D., Aptula, A.O. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure THEOCHEM*. **2003**, 622 (1-2), p. 1–22-
28. Todeschini, R., Consonni, V., Gramatica, P. Chemometrics in QSAR. In: *Comprehensive Chemometrics*. Elsevier: Oxford. **2009**, p 129–172.
29. The IUPAC International Chemical Identifier (InChI). Disponible en: <http://www.iupac.org/home/publications/e-resources/inchi.html>
30. Daylight SMILES (Simplified Molecular Input Line Entry System) Disponible en: <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>
31. Pesticide Properties DataBase (PPDB). Agriculture & Environment Research Unit (AERU)., University of Hertfordshire, UK. Disponible en: <https://sitem.herts.ac.uk/>
32. Deviller, J. Chapter 1: *Methods for building QSARs*. En Reisfeld, B., & Mayeno, A.N. Eds. *Computational Toxicology*., Vol 2. Humana Press: New York. **2013**, p. 13–37.
33. Mukerjee, D., Derosa, C., Stara, J. *2,3,7,8 Tetraclorodibenzo-p-dioxina : evaluación de efectos sobre la salud*. Centro Panamericano de Ecología Humana y Salud., Organización Panamericana de la Salud (OPS) & Organización Mundial de la Salud (OMS): Metepec. **1987**.
34. Grisoni, F., Ballabio, D., Todeschini, R., Consonni, V. *Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach*. En Nicolotti, O., Ed. *Computational Toxicology: Methods and Protocols*. Methods in Molecular Biology., Vol. 1800, Humana Press: New York. **2018**, p. 15–66.
35. Free, S.M., Wilson, J.W. A mathematical contribution to structure–activity studies. *Journal of Medicinal Chemistry*. **1964**, 7, p. 395–399.
36. Globisch, C., Pajeva, I.K., Wiese, M. Structure–activity relationships of a series of tariquidar analogs as multidrug resistance modulators. *Bioorganic & Medicinal Chemistry*. **2006**, 14, p. 1588–1598.
37. Alkorta, I., Blanco, F., Elguero, J. Application of Free-Wilson matrices to the analysis of the tautomerism and aromaticity of azapentalenes: a DFT study. *Tetrahedron*. **2008**, 64, p. 3826–3836.
38. Baggiani, C., Baravalle, P., Giovannoli, C., Anfonssi, L., Giraudi, G. Molecularly imprinted polymers for corticosteroids: analysis of binding selectivity. *Biosensors & Bioelectronics*. **2010**, 26, p. 590–595.

39. Hall, L.H., Kier, L.B., Phipps, G. Structure–activity relationship studies on the toxicities of benzene derivatives: I. An additivity model. *Environmental Toxicology and Chemistry*. **1984**, 3, p. 355–365.
40. Duewer, D.L. The Free-Wilson paradigm redux: significance of the Free-Wilson coefficients, insignificance of coefficient “uncertainties” and statistical sins. *Journal of Chemometrics*. **1990**, 4, p. 299–321.
41. Benigni, R., Bossa, C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutation Research*. **2008**, 659 (3), p. 248–261.
42. Ivanciuc, O. QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs. *Journal of Chemical Information and Computer Sciences*. **2000**, 40 (6), p. 1412–1422.
43. Consonni, V., Todeschini, R. *Multivariate analysis of molecular descriptors*. En: Dehmer, M., Varmuza, K., Bonchev, D., Eds. *Statistical modelling of molecular descriptors in QSAR/ QSPR*. Wiley-VCH Verlag GmbH & Co, KGaA. **2012**, p. 111–147.
44. Reutlinger, M., Koch, C.P., Reker, D., Todoroff, N., Schneider, P., Rodrigues, T., Schneider, G. Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for “orphan” molecules. *Molecular Informatics*. **2013**, 32, p. 133–138.
45. Fechner, U., Franke, L., Renner, S., Schneider, P., Schneider, G. Comparison of correlation vector methods for ligand-based similarity searching. *Journal of Computer-Aided Molecular Design*. **2003**, 17:687–698.
46. Moreau, G., Broto, P. Autocorrelation of Molecular Structures. Application to SAR Studies. *Nouveau Journal Chimie*. **1980**, 4, p. 757–764.
47. Randić, M. Generalized Molecular Descriptors. *Journal of Mathematical Chemistry*. **1991**, 7 (1), 155–168.
48. Boethling, R.S., Howard, P.H., Meylan, W.M. Finding and estimating chemical property data for environmental assessment. *Environmental Toxicology and Chemistry*. **2004**, 23 (10), p. 2290–2308.
49. Cronin, M.T.D., Livingstone, D.J. *Calculation of physicochemical properties*. En: Cronin, M.T.D., Livingstone, D.J., Eds. *Predicting chemical toxicity and fate*. CRC: Boca Raton **2004**.
50. Devillers, J., Balaban, A.T. *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers: Amsterdam. **1990**.
51. Wiener, H. Structural determination of paraffin boiling points. *Journal of American Chemical Society*. **1947**, 69, p. 17–20.
52. Wiener, H. Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. *Journal of American Chemical Society*. **1947**, 69, p. 2636–2638.

53. Randić, M. Characterization of molecular branching. *Journal of American Chemical Society*. **1945**, 97, p. 6609–6615.
54. Kier, L.B., Hall, L.H., Murray, W.J., Randić, M. Molecular connectivity I: Relationship to non-specific local anesthesia. *Journal of Pharmaceutical Sciences*. **1975**, 64, p. 1971–1974.
55. Kier, L.B. & Hall, L.H. An electrotopological-state index for atoms in molecules. *Pharmaceutical Research*. **1990**, 7, p. 801–807.
56. Hall, L.H. & Kier, L.B. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*. **1995**, 35 (6), p. 1039–1045.
57. Balaban, A.T. Chemical graphs. *Theoretical Chemical Acta*. **1979**, 53, p. 355–375.
58. Balaban, A.T. Highly discriminating distance-based topological index. *Chemical Physics Letters*. **1982**, 89, p. 399–404.
59. Balaban, A.T., Ionescu-Pallas, N., Balaban, T.S. Asymptotic Values of Topological Indices J and J' (Average Distance Sum Connectivities) for Infinite Acyclic and Cyclic Graphs. *MATCH Communication in Mathematical and in Computer Chemistry*. **1985**, 17, p. 121–146.
60. Gutman, I., Ruščić, B., Trinajstić, N., Wilcox, C.F. Graph theory and molecular orbitals. XII. Acyclic polyenes. *The Journal of Chemical Physics*. **1975**, 62 (9), p. 3399–3405.
61. Rücker, G. & Rücker, C. Counts of all walks as atomic and molecular descriptors. *J. Chemical Information and Computer Sciences*. **1993**, 33, p. 683–695.
62. Hall, L.H. & Story, C.T. Boiling point and critical temperature of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks. *Chemical Information and Computer Sciences*. **1996**, 36, p. 1004–1014.
63. Kubinyi, H. *3D QSAR in drug design*. En: *Theory methods and applications*, vol 1. Springer Science & Business Media: Berlin. **1994**.
64. Schuur, J.H., Selzer, P., Gasteiger, J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*. **1996**, 36, p. 334–344.
65. Rybinska, A., Sosnowska, A., Barycki, M., Puzyn, T. Geometry optimization method versus predictive ability in QSPR modelling for ionic liquids. *Journal of Computer-Aided Molecular Design*. **2016**, 30, p. 165–176.
66. Nettles, J.H., Jenkins, J.L., Bender, A., Deng, Z., Davies, J.W., Glick, M. Bridging chemical and biological space: “target fishing” using 2D and 3D molecular descriptors. *Journal of Medicinal Chemistry*. **2006**, 49 (23), p. 6802–6810.
67. Todeschini, R., Lasagni, M., Marengo, E. New Molecular Descriptors for 2D- and 3D-Structures, Theory. *Journal of Chemometrics*. **1994**, 8, p. 263–273.

68. Consonni, V., Todeschini, R., Pavan, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. Part 1. Theory of the Novel 3D Molecular Descriptors. *Journal of Chemical Information and Computer Sciences*. **2002**, 42, p. 682–692.
69. Selzer, P., Schuur, J., Gasteiger, J. *Simulation of IR Spectra with Neural Networks Using the 3D-MORSE Code*. En: *Software Development in Chemistry*, vol. 10. Gasteiger, J., ed. Fachgruppe Chemie-Information-Computer (CIC): Frankfurt. **1996**, p. 293-302.
70. Schuur, J., Selzer, P., Gasteiger, J. The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. *Journal of Chemical Information and Computer Sciences*. **1996**, 36, p. 334–344.
71. Raevsky, O.A., Trepalin, S.V., Razdol'skii, A.N. New QSAR Descriptors Calculated from Interatomic Interaction Spectra. *Pharmaceutical Chemistry Journal*. **2000**, 34, 646–649.
72. Ferguson, A.M., Heritage, T.W., Jonathon, P., Pack, S.E., Phillips, L., Rogan, J., Snaith, P.J. EVA: A New Theoretically Based Molecular Descriptor for Use in QSAR/QSPR Analysis. *Journal of Computer-Aided Molecular Design*. **1997**, 11, p. 143–152.
73. Cramer, R.D., Patterson, D.E., Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of American Chemical Society*. **1988**, 110, p. 5959–5967.
74. Geladi, P., Tosato, M.L. *Multivariate latent variable projection methods: SIMCA and PLS*. En: Karcher, W., Devillers, J. Eds. *Practical applications of quantitative structure– activity relationships (QSAR) in environmental chemistry and toxicology*. Springer Science+Business Media B.V: Dordrecht. **1990**.
75. Doucet, J.P., Panaye, A. *Three dimensional QSAR: applications in pharmacology and toxicology*. CRC: Boca Raton. **2010**.
76. Kearsley, S.K., Smith, G.M. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron Computer Methodology*. **1990**, 3: 615–633.
77. Feher, M., Schmidt, J.M. Multiple flexible alignment with SEAL: a study of molecules acting on the colchicine binding site. *Journal of Chemical Information and Computer Sciences*. **2000**, 40, p. 495–502.
78. Korhonen, S.P., Tuppurainen, K., Laatikainen, R., Peräkylä, M. FLUFF-BALL, a template-based grid-independent superposition and QSAR technique: validation using a benchmark steroid data set. *Journal of Chemical Information and Computer Sciences*. **2003**, 43, p. 1780–1793.
79. Korhonen, S.P., Tuppurainen, K., Laatikainen, R., Peräkylä, M. Comparing the performance of FLUFF-BALL to SEAL-CoMFA with a large diverse estrogen data set: from relevant superpositions to solid predictions. *Journal of Chemical Information and Modeling*. **2005**, 45 (6), p. 1874–1883.

80. Klebe, G., Abraham, U., Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry*. **1994**, 37, p. 4130–4146.
81. Kotani, T., Higashiura, K. Comparative Molecular Active Site Analysis (CoMASA). 1. An Approach to Rapid Evaluation of 3D QSAR. *Journal of Medicinal Chemistry*. **2004**, 47(11), p. 2732-2742.
82. Serafimova, R., Walker, J., Mekenyan, O. Androgen receptor binding affinity of pesticide “active” formulation ingredients. QSAR evaluation by COREPA method. *SAR & QSAR in Environmental Research*. **2002**, 13, p. 127–134.
83. Vedani, A., Dobler, M. 5D-QSAR: the key for simulating induced fit? *Journal of Medicinal Chemistry*. **2002**, 45, p. 2139–2149.
84. Vedani, A., Dobler, M., Lill, M.A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of Medicinal Chemistry*. **2005**, 48, p. 3700–3703
85. Polanski, J. Receptor Dependent Multidimensional QSAR for Modeling Drug–Receptor Interactions. *Current Medicinal Chemistry*. **2009**, 16 (25), p. 3243–3257.
86. Duchowicz, P.R. QSPR studies on water solubility, octanol-water partition coefficient and vapour pressure of pesticides. *SAR & QSAR in Environmental Research*. **2019**, 31 (2), p. 135-148.
87. Nikolov, N., Pavlov, T., Niemelä, J.R., Mekenyan, O., *Accessing and Using Chemical Databases*. En: Reisfeld, B., Mayeno, A.N., Eds. *Computational Toxicology*, vol 2. Humana Press: New York. **2013**.

CAPITULO 3: MÉTODOS ESTADÍSTICOS EN EL MODELADO QSAR

3.1 Introducción

Diversas investigaciones de la teoría QSAR se han enfocado en evaluar la importancia que tienen las subestructuras y los grupos funcionales de una molécula interactuante en un entorno u organismo determinado. De este modo, siempre que las predicciones o hipótesis planteadas por el modelo sean acertadas, resulta viable deducir algunos aspectos del mecanismo de acción involucrado.¹

Con el progresivo desarrollo y la constante innovación de los modelos QSAR se han incrementado las posibilidades de reconocer las relaciones existentes entre la respuesta biológica final generada por un agente químico y la estructura molecular del mismo. Esto a su vez, suministró información relevante acerca del papel que cumplen las moléculas orgánicas específicas en la naturaleza o un punto final dado (actividad o propiedad).²

Un paso trascendental en la construcción de los modelos QSAR radica en hallar un conjunto de descriptores moleculares que exhiban los mejores parámetros estadísticos y representen de manera fehaciente las variaciones estructurales asociadas al poder activo de las moléculas. Para esto, es imprescindible emplear métodos estadísticos con capacidad de definir las relaciones matemáticas existente entre miles de características estructurales, codificadas por los descriptores moleculares, y un delimitado efecto biológico final.

Debido a la progresiva incorporación de distintas estrategias estadísticas en los estudios QSAR, fue necesario establecer tres principales categorías de análisis, conocidas como técnicas de regresión, técnicas de clasificación y técnicas de aprendizaje automático.³⁻⁵ La finalidad del modelado QSAR es poder predecir cualquier tipo de actividad que cuente con datos experimentales válidos, ya sean lineales o no lineales.

Los métodos estadísticos lineales, también conocidos como supervisados, son utilizados para modelar respuestas cuantitativas a través de técnicas como la Regresión Univariada (UR), Regresión Lineal Múltiple (MLR) o regresión parcial. Por el contrario, los métodos de clasificación o no supervisados se utilizan en el tratamiento y reconocimiento de las categorías a las que pertenecen los conjuntos de datos químicos con respuestas cualitativas, tal como ocurre con la técnica de Análisis Discriminante Lineal (LDA).

Las herramientas estadísticas basadas en aprendizaje automático, conocidas como semi-supervisadas,⁶ son utilizadas para resolver cuestiones basadas en regresión o clasificación, ya que proveen modelos QSAR basados en inteligencia artificial, como la estrategia de Redes Neuronales Artificiales (ANN), con capacidad de establecer relaciones no lineales a partir de conjuntos de datos con información química compleja, extensa y variable.^{7,8}

Otra etapa del modelado QSAR que requiere la asistencia de los métodos estadísticos es la validación del modelo. Una vez que se logra identificar la mejor solución matemática, con un número apropiado de descriptores moleculares para predecir la respuesta final evaluada, es indispensable ejecutar diversas pruebas de validación multipropósito, que verifiquen la veracidad, predictibilidad y aplicabilidad del modelo obtenido.^{9,10}

Por lo general, para validar un modelo QSAR es indispensable contar con un conjunto externo de moléculas con datos experimentales de la actividad bajo estudio; este conjunto se conoce como predicción (N_{pred}). Además, se deben aplicar diversas técnicas de validación interna como la Validación Cruzada (CV) dejar-uno-afuera (LOO) y dejar-más-afuera (LMO), la técnica de Aleatorización-Y, y el Factor de Inflación de la Varianza (VIF); así como, definir el dominio de aplicación (AD) del modelo. A través de estas estrategias se puede caracterizar de forma acertada la bondad predictiva del modelo, requerida para determinar el grado de aceptación del mismo.¹¹⁻¹⁵

No obstante, proponer nuevos métodos de validación es un área de investigación activa y en permanente desarrollo. Por ello, resulta útil investigar sobre nuevas herramientas de validación basadas en otras medidas de confiabilidad, que puedan revelar el grado de confianza y precisión de la información recopilada por el modelo para extrapolarla sobre nuevas moléculas no incluidas en el ajuste.^{16,17}

Por lo anterior, en el presente capítulo de Tesis Doctoral se cuenta con información detallada acerca de los diferentes métodos estadísticos involucrados durante el tratamiento de los datos, la selección de las variables óptimas, el desarrollo del modelo, la validación y aplicación del mismo. Además, es importante reconocer los criterios básicos que justifican el uso de un método estadístico determinado; ya que este influirá en la complejidad, aplicabilidad, reproducibilidad y predictibilidad de cada modelo QSAR.

3.2. Pretratamiento de los datos

El pretratamiento de la matriz de datos QSAR puede ejecutarse en uno o varios pasos, dependiendo de la naturaleza, cantidad y calidad del conjunto de datos. Con este paso, se pretende analizar la homogeneidad de la varianza en los datos y examinar la hipótesis que concierne a los puntos finales.

Aunque en la mayoría de los casos, obtener datos biológicos con variaciones homogéneas es casi improbable. Por lo tanto, se busca que la varianza en los datos de respuesta sea baja y lo más homogénea posible.¹⁸

Cuando se extrae un conjunto de datos de literatura especializada o una base de datos química, es indispensable postular como requisito básico que los valores de ese conjunto de datos presenten una varianza estadística asociada a la función de distribución gaussiana.¹⁹

Asimismo, un adecuado pretratamiento de los datos implica tener conocimiento del tipo y el nivel de incertidumbre en los puntos finales específicos. La influencia del conjunto de datos inicial, es decir, las variaciones en los puntos finales específicos son controladas a través de un análisis de varianza simple y de reconocimiento de patrones, donde se aplican técnicas como Mínimos Cuadrados Parciales (PLS) acoplada a Regresión de Componentes Principales (PCR).^{2, 20}

Para garantizar que los programas de cálculo de descriptores interpreten correctamente las moléculas, es fundamental representarlas simbólicamente por medio de procedimientos formales y reglas convencionales de representación estructural. Tanto el método como la representación simbólica influyen directamente en la cantidad y el tipo de información química preservada, y esto a su vez, decide el tipo de descriptores moleculares computables.

Cuando se trata de una percepción bidimensional o 2D, las representaciones más utilizadas son: (1) El grafo molecular: con el cual se pueden obtener descriptores moleculares de forma manual. (2) Las anotaciones de cadena lineal, como la notación del Sistema de Entrada Molecular Lineal Simplificado (SMILES),²¹ que constituyen un modo de entrada basado en un formato de texto para calcular automáticamente descriptores moleculares.²²

Por el contrario, para calcular descriptores tridimensionales o 3D, se debe realizar una adecuada optimización de la estructura molecular bajo un método de estructura electrónica apropiado.

Después de calcular miles de descriptores moleculares, es importante verificar la contribución e idoneidad de la información incorporada por los mismos. Por ello, se debe ejecutar una estrategia de filtrado apropiada para eliminar los descriptores indeterminados, constantes o casi constantes (con un único valor diferente); así como excluir las moléculas duplicadas o importadas erróneamente.

3.3 Estrategias convencionales de selección de variables

Inicialmente, los modelos QSAR fueron planteados por C. Hansch y T. Fujita,²³⁻²⁵ como ecuaciones lineales conformadas por unos pocos parámetros estructurales relacionados a la respuesta activa. Sin embargo, la producción y comercialización acelerada de los productos químicos, promovió las investigaciones orientadas en

delimitar más elementos estructurales y fisicoquímicos asociados al potencial activo o tóxico de las moléculas.

En consecuencia, la posibilidad de calcular miles de descriptores moleculares e indicadores de fragmentos definidos bajo diferentes teorías, impulsó el uso de diversas estrategias estadísticas para seleccionar las variables óptimas de manera fácil, rápida y sistemática.

Hasta la fecha, existen varias técnicas estadísticas diseñadas para encontrar la mejor relación entre la información estructural de las moléculas y la actividad o propiedad exhibida por las mismas. En particular, se destacan las estrategias basadas en regresión y en clasificación, puntualizadas en la Figura 3.1.

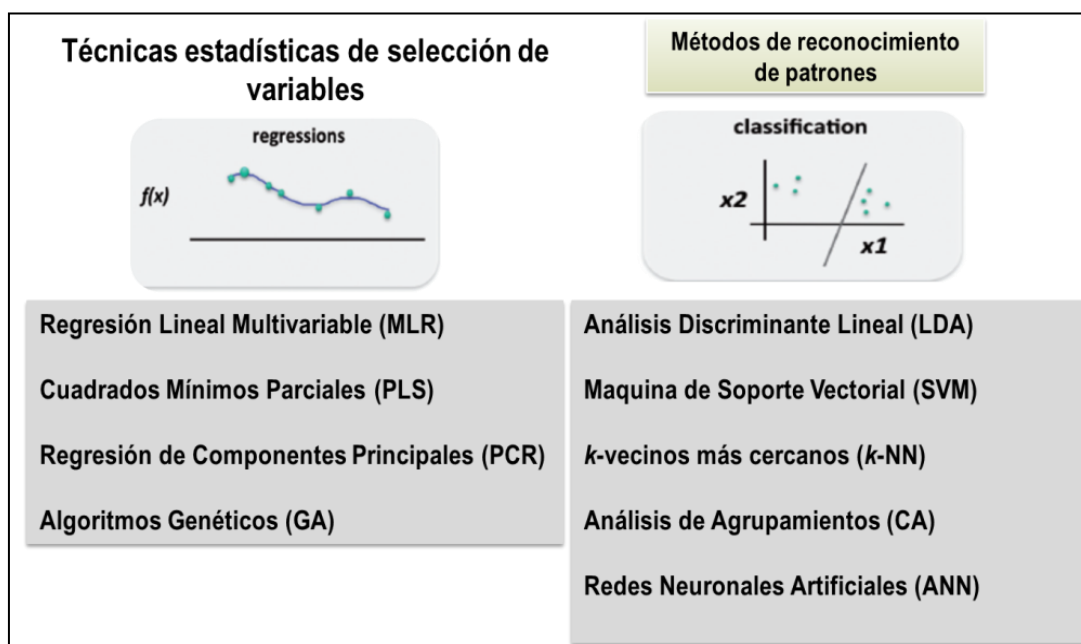


Figura 3.1. Principales técnicas estadísticas añadidas en QSAR. Fuente: Benfenati *et al.*, 2012.²⁶

Entre las técnicas basadas en regresión, sobresale la Regresión Lineal Múltiple (MLR), por correlacionar de forma simple y comprensible la respuesta biológica final de una molécula con su estructura molecular. El enfoque MLR ha sido empleado en el diseño de todos los modelos QSAR propuestos en la presente Tesis Doctoral. Por ello, se brindará más detalles de dicha técnica en la siguiente sección.

Adicionalmente, otras herramientas basadas en regresión, tal como PLS, PCR, y Algoritmos Genéticos (GA), han permitido obtener paralelismos racionales entre los datos experimentales de respuesta final y la información química codificada por los descriptores moleculares, comúnmente usados para predecir cuantitativamente la actividad de las sustancias químicas en una escala continua.^{26, 27}

Por el contrario, los enfoques de clasificación involucran otras técnicas que buscan reconocer patrones solo en los datos de los descriptores moleculares. Las técnicas de clasificación más utilizadas en QSAR son el Análisis Discriminante Lineal

(LDA), Redes Neuronales Artificiales (ANN), Máquina de Soporte Vectorial (SVM), k -vecinos más cercanos (k -NN) y el Análisis de Agrupamientos (CA). Las cuales han resultado útiles en los estudios de Relaciones Estructura-Actividad (SAR), orientados en evaluar las características estructurales de las moléculas para definir la categoría de respuesta a la que pertenece.

No obstante, es importante destacar que las categorías de clasificación hacen referencia a diferentes umbrales; es decir, cada categoría presenta un rango de valores específico para cada propiedad.²⁸

3.4. Estrategias basadas en clasificación

3.4.1. Análisis de agrupamientos (CA)

En principio, la técnica CA no requiere el conocimiento de previo sobre la localización de los elementos en cada grupo, ya que realiza el análisis completo de los datos incluidos. Esta herramienta de clasificación involucra diferentes tipos de agrupamientos, como el agrupamiento jerárquico o el agrupamiento de k -medias. El agrupamiento jerárquico fija los grupos en función de las diferencias calculadas por medio de las distancias entre los objetos, por lo general, distancias euclídeas.

Mientras que el Método de Análisis de Agrupamiento k -medias (k -MCA), es un enfoque no-jerárquico comúnmente empleado cuando se conoce el número de grupos presentes en los objetos, esta técnica es una estrategia no supervisada basada en k -centroides por agrupamiento, localizados lo más distante posible el uno del otro.

Seguidamente, se considera cada compuesto de un conjunto de datos dado, y se asocia al centroide más cercano. Este procedimiento se repite n iteraciones hasta que cada centroide no varíe. Además, en los estudios QSAR suelen utilizarse otras herramientas de clasificación basadas en agrupamientos, tal como regresión logística o LDA.^{6, 29, 30}

3.4.2. Análisis Discriminante Lineal (LDA)

El objetivo principal de la herramienta LDA es encontrar las variables óptimas que permitan una discriminación coherente entre dos o más clases de objetos, como lo expresa la Ec. 3.1.^{10, 15}

$$FD = a_0 + \sum a_i X_i \quad (3.1)$$

Donde FD es el valor de la función discriminante relacionada con una actividad en particular; X_i son los descriptores moleculares; a_0 es la constante de ecuación; y a_i son los coeficientes discriminantes o los niveles de contribución de cada variable independiente.

El enfoque LDA intenta modelar la diferencia entre las clases de datos. De modo que considera dos agrupamientos, que se dividen entre activos e inactivos, según el valor umbral de la actividad. Luego, la capacidad discriminante de la función seleccionada se mide a través de:

- La matriz de clasificación, en la que cada componente se localiza en un agrupamiento acorde con la función de clasificación.
- La matriz clasificada por una prueba de corte, en la que cada componente se clasifica en un agrupamiento, según con las funciones de clasificación calculadas a partir del conjunto de datos, a excepción del mismo.
- Una prueba de conjunto externo de validación.

3.5. Estrategias basadas en regresión

3.5.1. Regresión Lineal Múltiple (MLR)

Por el grado de simplicidad, reproducibilidad, claridad y aplicación, la estrategia de MLR es ampliamente utilizada para seleccionar las mejores correlaciones entre la variable de respuesta (actividad/propiedad) y los descriptores moleculares calculados para un determinado sistema molecular. Un modelo de regresión lineal busca predecir la variable dependiente (actividad), únicamente con la variable independiente óptima (descriptor) y su respectivo coeficiente de regresión, según la Ec. 3.2.^{31,32}

$$Y = a_0 + a_1 X_1 \quad (3.2)$$

Donde Y vincula la variable de respuesta bajo estudio; a es la constante; X_1 es el descriptor óptimo; y a_1 es el coeficiente de regresión.

Por el contrario, las soluciones basadas en MLR, indicadas en la Ec. 3.3, utilizan más de una variable independiente para estimar de manera lógica el efecto biológico final evaluado. El método MLR busca establecer una relación lineal entre la respuesta final (Y) y 2 o más descriptores moleculares (X_i).

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n \quad (3.3)$$

En este caso, cada descriptor es indispensable para explicar el fenómeno global responsable de la respuesta final. Aunque ninguno de los X_i puede presentar una intercorrelación muy alta.

En MLR, cada variable X_i tiene una delimitada contribución sobre la actividad, que varía de acuerdo con el valor del coeficiente de correlación y su signo algebraico. Cada coeficientes de regresión deberá ser significativo en $p < 0.05$, de modo que pueda verificarse mediante una prueba 't'. La calidad de un modelo MLR es determinada por

los parámetros estadísticos más comunes, como el coeficiente de determinación (R^2), la desviación estándar (S), el error cuadrático medio (RMS) y la relación de varianza (F); los cuales se describen a continuación:

1- Coeficiente de determinación (R^2)

El parámetro R^2 de un modelo determina la calidad de ajuste del mismo; es decir, la similitud entre los valores observados de la actividad y los predichos por el modelo. El parámetro R^2 es definido en la Ec. 3.4.

$$R^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \quad (3.4)$$

Donde Y_{obs} e Y_{pred} son los valores de respuesta observados y predichos, respectivamente; \bar{Y}_{obs} es el promedio de los valores de respuesta observados. Para un modelo QSAR es importante que el valor de R^2 sea lo más cercano a 1. De lo contrario, la calidad de ajuste tiende a deteriorar.⁶

2- Desviación estándar (S)

Para un modelo QSAR la estimación del error estándar de la variable Y debe ser lo más bajo posible. El parámetro S se define por medio de la Ec. 3.5. Cabe destacar que los grados de libertad son delimitados por la expresión $N - d - 1$.

$$S = \sqrt{\frac{\sum (Y_{obs} - Y_{pred})^2}{N - d - 1}} \quad (3.5)$$

3- Error Cuadrático Medio (RMS)

El valor de RMS es la representación del error entre la media de los valores de la actividad observados y predichos. Para admitir un modelo QSAR, este parámetro debe ser $RMS_{cal} < 0.3$, calculado con la Ec. 3.6.

$$RMS = \sqrt{\frac{\sum (Y_{obs} - Y_{pred})^2}{N}} \quad (3.6)$$

4- Relación de varianza (F)

Para verificar la importancia general de los coeficientes de regresión es necesario estimar el parámetro F a través de la Ec. 3.7. La relación de varianza F mide los grados de libertad en d y $N-d-1$.

$$F = \frac{\frac{\sum (Y_{pred} - \bar{Y})^2}{d}}{\frac{\sum (Y_{obs} - Y_{pred})^2}{N - d - 1}} \quad (3.7)$$

En este caso, el parámetro d es el número de descriptores moleculares incluidos en el modelo establecido. Para que F tenga significación general de los coeficientes de regresión su valor debe ser alto.

Para verificar la veracidad, estabilidad y predictibilidad de un modelo de regresión QSAR, se deben superar distintas pruebas de validación interna, como la CV dejar-uno-afuera (LOO) o dejar más-afuera (LMO), la prueba de Aleatorización-Y, entre otras.³³⁻³⁵

3.5.2. Regresión de Componentes Principales (PCR)

El análisis PCR intenta explicar la conformación de las variaciones de los descriptores moleculares (X_i), mediante la proyección de las variables originales en el espacio dimensional m , en un hiperespacio dimensional interceptado por un pequeño número de variables latentes, llamadas componentes principales.

En general, en este tipo de estudios la matriz de datos QSAR se transforma para obtener distribuciones altamente homogéneas, antes de aplicar un PCR. El procedimiento de PCR está orientado en descomponer la matriz de descriptores X_i , bajo la expresión bilineal de la Ec. 3.8.^{19, 36}

$$X = \sum_{a=1}^A (t_a * p_a^T) + E_a \quad (3.8)$$

Donde t_a son los valores de puntaje de los N compuestos en los hiperplanos abarcados por un número significativo de componentes principales p_a^T . Para definir el número óptimo de componentes principales se utilizan varios criterios de validación, tal como la técnica CV.

Mientras que, la matriz de datos de respuesta Y se debe descomponer en la forma bilineal planteada en la Ec. 3.9. Una regresión de u_a sobre t_a representa la regresión del componente principal.

$$Y = \sum_{a=1}^A (u_a * q_a^T) + F_A \quad (3.9)$$

3.5.3. Cuadrados Mínimos Parciales (PLS)

La estrategia PLS suele aplicarse para diseñar modelos QSAR que cuentan con un conjunto de datos inicial significativamente menor al número de descriptores moleculares, altamente intercorrelacionados.³⁷ En general, un análisis PLS se realiza después de aplicar la técnica PCR, a menos que la descomposición de la matriz de descriptores X_i y la matriz de respuesta Y se realice de forma simultánea e iterativa, para extraer la variación más relevante en X_i con respecto a Y .

En este caso, las Variables Latentes (LVs), es decir u_a y t_a , dependen de los espacios X e Y intercorrelacionados bajo la Ec. 3.10.

$$u_a = b_a t_a + e_a \quad (3.10)$$

El parámetro b_a indica el coeficiente de regresión de mínimos cuadrados y e_a es el residuo. Usualmente, la estrategia PLS se utiliza para complementar los estudios QSAR-3D, al igual que las técnicas CoMFA o CoMSIA.^{6, 38} También, se utiliza como criterio de validación, particularmente de los modelos de regresión QSAR.³⁹

3.6. Técnicas de selección basadas en regresiones

En la actualidad, existen varios enfoques matemáticos basados en regresión que buscan resolver el problema de encontrar d variables independientes óptimas $\mathbf{d} = (X_1, X_2, X_3, \dots, X_d)$, con los mejores parámetros estadísticos mencionados anteriormente, que dan origen al modelo a partir de un conjunto de descriptores moleculares D de magnitud $D > d$.

A continuación, se describen las técnicas de selección basadas en regresiones más reconocidas e implementadas en el área; así como los algoritmos usados para seleccionar los modelos QSAR propuestos en la presente Tesis Doctoral.

3.6.1. Búsqueda exacta (FS)

Si bien la búsqueda exacta (FS) permite obtener de manera simple las mejores d variables, también implica un costo computacional alto, ya que al explorar la mejor expresión con d variables derivadas de un conjunto más extenso denominado D , realiza $D! / [d!(D-d)!]$ regresiones lineales; es decir, considera todos los casos posibles.

En particular, si se elige el parámetro S como criterio de búsqueda, la técnica FS brindará un conjunto con d variables de menor S , que será la mejor solución posible dentro del conjunto D analizado.

En términos matemáticos, FS debe encontrar el mínimo global de $S(\mathbf{d})$ en un conjunto de $D!/[d!(D-d)!]$ "puntos" $\mathbf{d}_i = \{X_{i1}, X_{i2}, \dots, X_{id}\}$, donde cada descriptor X_i es un vector de N componentes, de modo que cada punto \mathbf{d}_i es una matriz de $d \times N$.

Generalmente, para tratar un amplio conjunto inicial con D variables, la técnica FS resulta impráctica, costosa e insuficiente; a menos que sea ejecutada por una supercomputadora; aunque cualquier resultado obtenido será el producto de un procedimiento sumamente costoso.

Ante esta falencia, hoy en día se cuenta con otros métodos aproximados que brindan soluciones similares a las obtenidas mediante FS.

3.6.2. Técnica de regresión "de a pasos"

En QSAR, la técnica de regresión de a pasos ha permitido desarrollar numerosos modelos de manera rápida, sencilla, acertada y accesible en diferentes programas computacionales. Esta técnica presenta tres variantes principales, fundamentadas sobre el método de inclusión de a pasos (SI), exclusión de a pasos (SE), y de a pasos (SW).

La estrategia SI, busca calcular en una primera etapa el mejor modelo de una variable, luego en cada etapa subsiguiente adiciona una nueva variable óptima, que busca mejorar la calidad del modelo; así, el modelo óptimo se localiza cuando no es posible encontrar una mejor relación.

Contrariamente, la estrategia SE inicia con un modelo conformado por varias variables y en cada etapa elimina la variable que no contribuye a mejorar la correlación. Este proceso termina cuando se alcanza un conjunto óptimo, donde la exclusión de cualquiera de las variables empeora la calidad del modelo.

Por otra parte, el método de regresión de a pasos, basado en las estrategias SI y SE, busca en cada paso considerar cuatro alternativas viables: adicionar una variable, excluir una variable, intercambiar dos variables, o finalizar la búsqueda. A partir de estas opciones, se pueden establecer múltiples variantes de SW, condicionadas al modo de ejecución.⁴⁰

Sin embargo, el método SW no garantiza que la solución óptima obtenida coincida con la solución exacta; ya que cada vez que introduce o remueve una variable en el modelo, este ya cuenta con otras variables que limitan y determinan la calidad de la solución. De esta forma, aparece el denominado "efecto de mezcla de variables", originado cuando las variables se combinan entre sí para generar una determinada respuesta final sobre los parámetros estadísticos.

Debido a que SW no considera el efecto de mezcla de variables, es razonable apuntar que este método es una aproximación a la solución óptima. Una forma de reducir el efecto de mezcla de variables, es utilizar un conjunto inicial D con la menor

cantidad de variables intercorrelacionadas entre sí, que resulten en cierto grado ortogonales.

3.6.3. Método del Reemplazo (RM)

El método del Reemplazo (RM) fue diseñado y programado por Duchowicz *et al.*, 2006.⁴¹ Este método surgió ante la necesidad de establecer un algoritmo inequívoco para encontrar los mejores subconjuntos con d variables óptimas extraídas del conjunto inicial D , con un menor grado de incertidumbre con respecto a las estrategias mencionadas anteriormente.

La principal ventaja de utilizar el RM, es que ejecuta un menor número de regresiones lineales con respecto a la estrategia FS, y produce resultados finales cercanos o similares a los exactos. La técnica RM es una aproximación formulada de manera fácil, clara y aplicable, que considera el efecto de mezcla de variables mencionado antes.

El principal objetivo de RM es obtener un valor mínimo en S , considerando el error relativo (*der*) de los coeficientes de regresión del ajuste lineal con d descriptores, $\mathbf{d}=\{X_1, X_2, \dots, X_d\}$. El fundamento del camino utilizado por RM, consiste en que los modelos obtenidos mediante una FS producen *der* bajos en los coeficientes. Esta observación se tomó como criterio acertado para ir evaluando las d variables en el modelo.

Puntualmente, el RM se utilizó para diseñar y seleccionar cada uno de los modelos QSAR propuestos en la presente Tesis Doctoral. De modo que se detallará el procedimiento utilizado a continuación:

1. Selección aleatoria de un conjunto con d variables $\mathbf{d}=\{X_1, X_2, \dots, X_d\}$, originadas del conjunto D más amplio, y se realiza una regresión lineal.
2. Elegir uno de los descriptores del conjunto denominado X_i , y se lo reemplaza con cada uno de los D descriptores del conjunto inicial, excepto por sí mismo. Allí, se conserva la mejor solución con S mínimo.
3. Ya que en el primer paso se reemplaza cualquiera de las d variables del modelo inicial, se dispondrá de d rutas distintas para alcanzar la solución final.
4. Escoger una variable del modelo resultante que exhiba el mayor *der* en su coeficiente (a excepción de la variable reemplazada en el paso previo), y se retiene el mejor modelo resultante. Conforme a lo anterior, se busca reemplazar cada una de las d variables del modelo, omitiendo las variables reemplazadas previamente.
5. Una vez finalizado el proceso de reemplazo, se reinicia con la variable que presenta mayor *der* en su coeficientes, y de nuevo serán reemplazadas las variables remanentes. Este procedimiento se repite tantas veces como sea necesario hasta que

el subconjunto con d variables sea invariante. En este punto se cuenta con el mejor modelo para la ruta i .

6. En seguida se procede de la misma forma para todas las rutas posibles $i = (1, 2, \dots, d)$, y se elige la mejor relación de d variables con menor S .

7. Por último, se comparan los modelos obtenidos con este método, y se elige el mejor subconjunto de d variables como modelo QSAR; este debe superar rigurosas pruebas de validación, detalladas en la siguiente sección.

3.7 Principales estrategias de validación en QSAR

Desde que la validación del modelo QSAR se convirtió en una etapa determinante para su aceptación, diferentes enfoques de validación han permitido controlar la correlación aleatoria de los descriptores moleculares, evitar el ajuste excesivo de los datos, y evaluar la precisión, veracidad, aplicabilidad y capacidad predictiva de los modelos QSAR.¹²

En la actualidad, se dispone de múltiples estrategias para validar un modelo QSAR, que difieren de acuerdo al enfoque evaluado; es decir, si es un modelo de regresión o clasificación, donde cada enfoque cuenta con sus propios parámetros de validación.⁴²

Para los modelos de regresión es común utilizar las pruebas de validación interna, tal como la Validación Cruzada (CV) dejar-uno-afuera (LOO) y dejar-más-afuera (LMO), encargadas de evaluar las variaciones de las predicciones del modelo, en ausencia de una o varias moléculas del conjunto de calibración, respectivamente. Asimismo, miden la precisión de las predicciones en función de los parámetros estadísticos evaluados.¹¹⁻¹⁵

Aplicar un modelo QSAR para predecir un conjunto externo de moléculas no consideradas durante su diseño, permite determinar el poder predictivo del mismo.⁴³ También, considerar la técnica de Aleatorización-Y, y definir el dominio de aplicación (AD) del modelo, permite verificar la estabilidad y aplicabilidad del modelo obtenido.^{26-28, 33} Las principales estrategias de validación interna de los modelos de regresión QSAR son detalladas en la siguiente sección.

En cuanto a las pruebas de validación de los modelos de clasificación, se destaca la estadística de Wilks λ (λ), utilizada para probar la importancia de la función del modelo discriminante, y se calcula como la relación entre la suma de cuadrados dentro de la categoría y la dispersión total, con un valor que oscila entre $0 < \lambda < 1$. En esta técnica cuanto más se acerque λ a 0, mayor será el nivel de discriminación.

Además, se usa el Índice Canónico (R_c) para estimar la fuerza de relación entre varias variables dependientes e independientes, y Chi-cuadrado (χ^2) para determinar la calidad de los modelos de clasificación. También, se utiliza la distancia cuadrada de

Mahalanobis como medida de estabilidad, calculada mediante la aleatorización de los puntos de datos.^{6,11}

3.7.1. Validación Cruzada interna LOO y LMO

En la prueba de validación CV-LOO, una molécula del conjunto de calibración es extraída al azar; con las moléculas restantes se recalculan los coeficientes de regresión del modelo elegido, y se mantienen los descriptores seleccionados para predecir la molécula extraída.

Este proceso se repite tantas veces como número de moléculas en el conjunto de calibración (N_{cal}).²⁸ La bondad del ajuste alcanzado por la CV-LOO se mide con el coeficiente de determinación (R_{LOO}^2), calculado con la Ec. 3.11.

$$R_{LOO}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \quad (3.11)$$

Donde las variables Y_{obs} e Y_{pred} representan los valores de actividad observados y predichos de las moléculas del conjunto de calibración, respectivamente. La variable \bar{Y}_{obs} es la actividad media observada de las moléculas de calibración.

Si bien la prueba de CV-LMO se basa en el mismo principio que CV-LOO; la primera resulta más rigurosa, debido a que extrae al azar una porción mayor a una molécula del conjunto de calibración en cada ciclo; de tal forma que, cada modelo generado prediga la actividad de las moléculas ausentes. Una vez completado todos los ciclos, la estabilidad del modelo se evalúa con el parámetro R_{LMO}^2 , calculado con las predicciones obtenidas por la prueba.

Cabe resaltar que para las dos pruebas de validación, el valor de los parámetros R_{LOO}^2 y R_{LMO}^2 debe ser mayor a 0.5, según lo indica la literatura especializada.^{4, 5, 44}

3.7.2. Prueba de Aleatorización-Y

La prueba de Aleatorización-Y se emplea para verificar la estabilidad del modelo QSAR generado, mediante la permutación de los valores de la variable de respuesta Y, con respecto a la matriz de descriptores (X_i), que permanece fija. El resultado de esta prueba se verifica a través de la desviación estándar (S^{aleat}) o el Error Cuadrático Medio del modelo aleatorizado (RMS^{aleat}) del modelo aleatorizado.^{36, 45}

Para determinar que el modelo QSAR no es producto de una correlación fortuita del método, sino por el contrario, representa una autentica relación estructura-actividad, los parámetros mencionados previamente deben cumplir con los siguientes criterios, según corresponda:

$$S^{aleat} > S_{cal}$$

$$RMS^{aleat} > RMS_{cal}$$

3.7.3. Factor de Inflación de la Varianza (VIF)

Dada la necesidad de verificar el efecto de intercorrelación de los descriptores moleculares del modelo, es necesario realizar el análisis del Factor de Inflación de la Varianza (VIF), calculado por medio de la Ec. 3.12. Los valores obtenidos determinan cuanto se aumenta la varianza del coeficiente del descriptor, en comparación con el caso donde los descriptores moleculares son completamente ortogonales entre sí.⁴⁶

$$VIF = \frac{1}{1 - R_{ij}^2} \quad (3.12)$$

Donde, R_{ij}^2 es el coeficiente de determinación entre un descriptor dado y los restantes del modelo. Según la literatura, el valor ideal del parámetro VIF en cada descriptor del modelo debe ser menor a 10.⁴⁷

3.7.4. Dominio de Aplicación (AD)

Para delimitar la región teórica de aplicación de un modelo QSAR con respecto al espacio químico generado por los descriptores moleculares y la actividad modelada, se debe definir el Dominio de Aplicación (AD) del mismo. El análisis AD permite obtener información sobre el nivel de incertidumbre que tienen las predicciones del modelo en nuevos conjuntos moleculares, en función del grado de similitud que tienen con respecto al conjunto de calibración.¹²

Si bien un modelo QSAR puede predecir la actividad de cualquier entidad química existente o hipotética, solo las moléculas localizadas dentro del AD pueden ser predichas de manera confiable; por el contrario, se consideran extrapolaciones del modelo.

Para definir el AD de un modelo se pueden utilizar diferentes métodos de análisis, que varían de acuerdo a la tipología del modelo; los más comunes son:^{48, 49}

1. Métodos geométricos
2. Métodos basados en distancia
3. Rangos en el espacio de los descriptores moleculares
4. Rango de la variable de respuesta

En general, para establecer el AD sobre múltiples variables de predicción con una distribución normal del conjunto de datos, suelen emplearse los métodos basados en

distancia, tal como la técnica de influencia, estudiada y aplicada en los modelos QSAR desarrollados en la presente Tesis.

La técnica de influencia (h), permite verificar la región teórica de aplicación del modelo con respecto a un nuevo conjunto molecular, mediante la Ec. 3.13. Si cada molécula del nuevo conjunto se localiza dentro del AD del modelo, las predicciones obtenidas se consideran confiables, al ser el resultado de una interpolación con incertidumbre reducida.

$$h_i = x_i (\mathbf{X}^T \mathbf{X})^{-1} x_i^T \quad (3.13)$$

Donde h_i es la influencia para el compuesto i ; el parámetro x_i es el vector del descriptor para i ; \mathbf{X} representa la matriz del modelo para el conjunto de calibración. Por el contrario, si alguna molécula se ubica afuera del AD, la actividad calculada se tomará como el producto de una extrapolación del modelo; por lo tanto, este dato contiene alta incertidumbre y es poco confiable.

Por otra parte, si una molécula tiene un valor de influencia alto y pertenece al conjunto de calibración, esta reforzaría la confiabilidad del modelo. Aunque, si la misma molécula pertenece al conjunto externo de validación, el valor calculado se deriva de una sustancial extrapolación del modelo. Por este motivo, es necesario establecer el valor de influencia crítico (h^*), expresado en la Ec. 3.14:

$$h^* = 3(d+1)/N_{train} \quad (3.14)$$

La variable d es la cantidad de descriptores moleculares; N_{train} es el número de compuestos en el conjunto de calibración.

Cuando una sustancia química tiene un valor $h < h^*$, el valor calculado tiene un grado de certeza semejante a las predicciones del conjunto de calibración. En cambio, si una molécula presenta un valor $h > h^*$, y además pertenece al conjunto externo de validación, se puede deducir que es estructuralmente distante con respecto al conjunto de calibración.

3.8. Criterios de validación externa

Por medio de las pruebas de validación externa es posible determinar tanto la capacidad predictiva del modelo como su aplicabilidad, veracidad y estabilidad para estimar una delimitada respuesta biológica final contenida en un conjunto de moléculas nuevas, hipotéticas o carentes de información experimental.

Sin embargo, para determinar el poder predictivo de un modelo QSAR es necesario aplicarlo en un nuevo conjunto de moléculas no contempladas durante la construcción del mismo. Por ello, es imprescindible dividir el conjunto de datos inicial

en dos subconjuntos, denominados calibración (*cal*) y predicción (*pred*). El subconjunto *cal* se emplea para construir y verificar la calidad del modelo QSAR, mientras que el subconjunto *pred* se usa para evaluar el nivel predictivo y la exactitud del modelo en nuevas moléculas.

Cabe resaltar que, si el conjunto de datos inicial es extenso, resulta oportuno dividirlo en tres subconjuntos equitativos, adicionando un conjunto de validación (*val*) para corroborar la estabilidad y calidad interna del modelo.⁵⁰

3.8.1. La selección de los subconjuntos

Por lo general, una correcta división del conjunto de datos implica que los subconjuntos de calibración y predicción ocupan todo el espacio de los descriptores del modelo. Además, cada punto del subconjunto *pred* debe estar cerca de al menos un punto del subconjunto *cal*; asimismo, todos los componentes contemplados en *pred* deben estar dentro del rango de valores designado por *cal*.

Utilizar técnicas adecuadas para dividir de manera racional el conjunto de datos total, brinda una mejor calidad de las predicciones del modelo. Entre las distintas herramientas disponibles para seleccionar los subconjuntos moleculares, se destaca la selección aleatoria, la selección basada en la respuesta Y (actividad), y la selección basada en la respuesta X (descriptores). Esta última ha sido ampliamente utilizada para validar los modelos QSAR, bajo las siguientes técnicas:

- Agrupamiento *k-medias*
- Diseño molecular estadístico
- Exclusión de esferas
- Selección de Kennard-Stone
- Selección de mapas autoorganizados de Kohonen
- Selección por extrapolación enfocada al conjunto *pred*.

3.8.2. Método de subconjuntos balanceados (BSM)

Una apropiada validación externa del modelo inicia con una acertada selección de los subconjuntos moleculares. El Método de Subconjuntos Balanceados (BSM),⁵¹ propuesto y programado por nuestro grupo a través del programa Matlab,⁵² realiza la división del conjunto de datos total, de modo que, los subconjuntos de calibración, validación y predicción sean equitativos en cuanto a su relación estructura-actividad.

El método BSM considera los datos de la actividad y de los descriptores moleculares no conformacionales. En este caso, se omiten los descriptores indicadores binarios, por contener información que puede tergiversar la selección racional de los subconjuntos.

A través del BSM se ejecuta el análisis de agrupamientos k -medias (k -MCA) de compuestos en términos de mínima distancia euclídea entre ellos; de tal manera que, los compuestos similares se localizarán en el mismo grupo. El procedimiento involucrado en BSM se encuentra ilustrado en la Figura 3.2, y se describe a continuación:

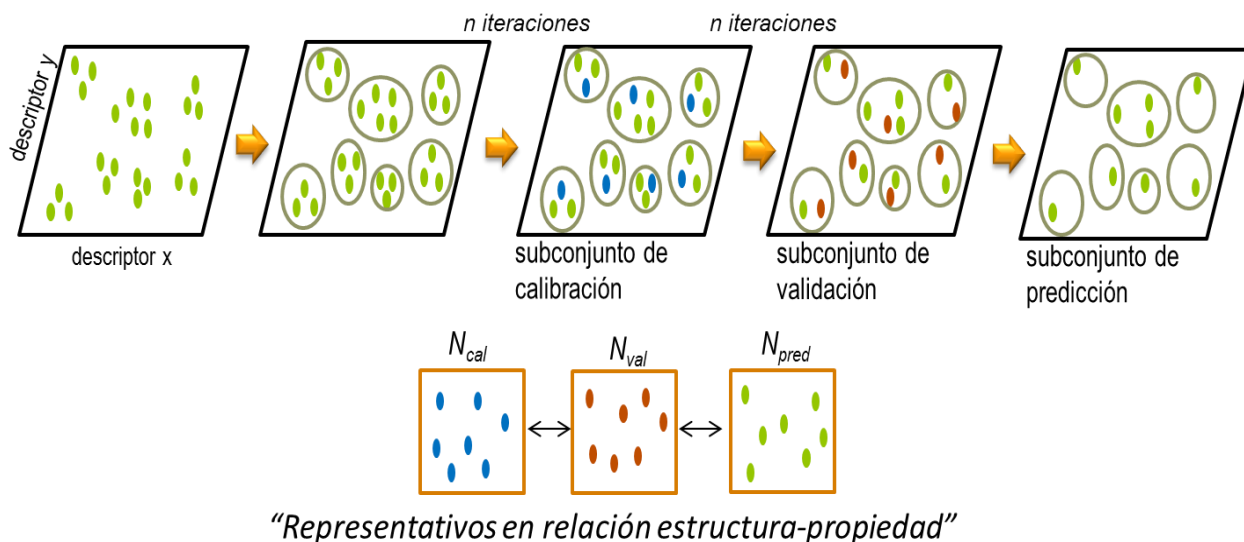


Figura 3.2. Selección de los subconjuntos moleculares mediante BSM. Fuente: Saavedra, L.M.

1. Crear una matriz (**C**) con los datos experimentales de la actividad (N) y el número de descriptores (d) no conformacionales calculados para el conjunto de datos total. Ahora la matriz **C** tiene una dimensión $N \times d$.

2. Excluir los descriptores moleculares con valores ausentes, constantes, casi constantes (con un solo valor diferente) y linealmente dependientes. Ahora se obtiene una matriz **C**₁ de dimensión $N \times d'$.

3. Estandariza la matriz **C**₁ para centrar y escalar sus elementos de matriz.

4. Crear N_{cal}^0 agrupamientos mediante k -MCA, utilizando la matriz **C**₁ en términos de distancia euclídea, junto con n iteraciones (>10.000 casos) para optimizar el algoritmo y lograr la mejor solución. Este proceso calcula N_{cal}^0 localizaciones de los centroides, cada uno de dimensión $1 \times d'$.

La variable N_{cal}^0 se encuentra definida como $N_{cal} - N_{\min \max}$, donde N_{cal} es el número de moléculas en el subconjunto de calibración obtenido, y $N_{\min \max}$ son las moléculas con el valor mínimo y máximo de la actividad evaluada.

5. El subconjunto N_{cal} está conformado por moléculas de todos los agrupamientos. En particular, por las moléculas más cercanas al centroide de cada agrupamiento. En N_{cal} también se incluyen las $N_{\min \max}$ moléculas.

6. Crear N_{val} agrupamientos con las restantes de $N - N_{cal}$ moléculas por medio el análisis k -MCA con n iteraciones (>10.000 casos), descrito previamente. Aquí se calculan N_{val} ubicaciones de los centroides.

7. Conformar el subconjunto de validación (N_{val}) a través de la inclusión de una molécula por agrupamiento; es decir, se extraen las moléculas más cercanas al centroide de cada agrupamiento obtenido.

8. Por último, el conjunto de predicción (N_{pred}) se conforma con las restante de $N - N_{cal} - N_{val}$ moléculas.

3.8.3. Parámetro R_{pred}^2

Comúnmente, la validación externa de un modelo de regresión QSAR calcula el parámetro del coeficiente de determinación del conjunto externo (R_{pred}^2) para evaluar la capacidad que tiene el modelo para reproducir los datos utilizados en su diseño, a través de la Ec. 3.15.^{44, 48}

$$R_{pred}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{obs(cal)})^2} \quad (3.15)$$

Por otra parte, diferentes situaciones revelaron que el parámetro R_{pred}^2 no es suficiente para determinar el poder predictivo del modelo, ya que su valor depende de la expresión $\sum (Y_{obs} - \bar{Y}_{obs(cal)})^2$; es decir, el valor medio de la actividad del conjunto de calibración puede conducir a una medida ficticia del poder predictivo del modelo con respecto a un nuevo conjunto de datos.⁵³

En la pretensión de corroborar la habilidad predictiva de los modelos QSAR sobre nuevas entidades químicas, se han propuesto diversas estrategias estadísticas que buscan verificar la bondad de ajuste del conjunto de calibración para generar nuevas predicciones. Entre los criterios de validación externa más utilizados se destacan los siguientes:

3.8.4. Parámetro Q^2

El parámetro Q^2 , propuesto por Shi *et al.*, 2001,⁵⁴ se utiliza para definir el rendimiento predictivo del ajuste definido por R_{pred}^2 . Con el parámetro Q^2 se contemplan los valores previsto por el modelo cuando los objetos no pertenecen al conjunto de calibración, tal como lo expresa la Ec. 3.16.⁵⁵

$$Q^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} = 1 - \frac{PRESS}{TSS} \quad (3.16)$$

En este caso, *PRESS* es la suma de los residuos al cuadrado; *TSS* es la suma de las desviaciones al cuadrado del valor medio del conjunto de datos, e \bar{Y}_{obs} indica el valor medio de la actividad observada para el conjunto de predicción.

3.8.5. Parámetro Q_{F1}^2

El parámetro Q_{F1}^2 fue propuesto por Tropsha *et al.* 2003,⁵⁶ con base en el parámetro Q^2 , y se calcula a través de la Ec. 3.17.

$$Q_{F1}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{TR})^2} = 1 - \frac{PRESS}{TSS_{EXT(\bar{Y}_{TR})}} \quad (3.17)$$

Aquí, $TSS_{EXT(\bar{Y}_{TR})}$ corresponde al cuadrado de la suma total de las desviaciones del conjunto externo, obtenidas con base a la media del conjunto de calibración. El parámetro de validación Q_{F1}^2 es considerado un criterio de aprobación de los modelos, ya que hace parte de los principios de la OCDE para diseñar y validar modelos de la teoría QSAR, detallados en la siguiente sección.⁵⁷⁻⁶⁰

3.8.6. Parámetro Q_{F2}^2

A partir de un estudio dirigido por Schüürmann *et al.*, 2008,⁶¹ se demostró que el parámetro Q_{F1}^2 arrojaba estimaciones optimistas del rendimiento predictivo de un modelo, manifestando que aumentaba a medida que incrementaba la diferencia entre \bar{Y}_{TR} e \bar{Y}_{EXT} . En su reemplazo, los autores propusieron el parámetro denominado Q_{F2}^2 , calculado a través de la Ec. 3.18.

$$Q_{F2}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \bar{Y}_{EXT})^2} = 1 - \frac{PRESS}{TSS_{EXT(\bar{Y}_{EXT})}} \quad (3.18)$$

Donde $TSS_{EXT(\bar{Y}_{EXT})}$ es el cuadrado de la suma total de las desviaciones del conjunto externo, calculadas con base en el valor medio del mismo conjunto. Así, todos los valores de esta ecuación dependen exclusivamente del conjunto externo de predicción.

3.8.7. Parámetro Q_{F3}^2

El parámetro Q_{F3}^2 , detallado en la Ec. 3.19, fue propuesto por Consonni *et al.* 2010,⁶² en consecuencia de ciertas limitaciones incorporadas en los parámetros Q_{F1}^2 y Q_{F2}^2 . Los autores determinaron que si el conjunto de predicción (*pred*) no cuenta con valores distribuidos de manera uniforme dentro del rango de datos contemplado por el conjunto de calibración (*cal*), los parámetros Q_{F1}^2 y Q_{F2}^2 presentan diversas inconsistencias para determinar el verdadero rendimiento predictivo del modelo.

$$Q_{F3}^2 = 1 - \frac{\left[\sum (Y_{obs} - Y_{pred})^2 \right] / N_{EXT}}{\left[\sum (Y_{obs} - \bar{Y}_{TR})^2 \right] / N_{TR}} = 1 - \frac{PRESS / N_{EXT}}{TSS / N_{TR}} \quad (3.19)$$

El parámetro de validación Q_{F3}^2 propone una expresión matemática independiente de la distribución o el tamaño del conjunto externo.

3.8.8. Coeficiente de Correlación de Concordancia (CCC)

El parámetro CCC propuesto por Lin *et al.*, 1989,⁶³ y expresado en la Ec. 3.20; se aplica para determinar la precisión y veracidad del modelo, basándose en el distanciamiento de las observaciones de la línea de regresión, y que tan lejos se desvía esta línea de la ordenada del origen.⁶⁴

$$CCC = \frac{2 \sum (Y_{obs} - \bar{Y})(Y_{pred} - \hat{Y})}{\sum (Y_{obs} - \bar{Y})^2 + \sum (Y_{pred} - \hat{Y})^2 + N_{EXT}(\bar{Y} - \hat{Y})^2} \quad (3.20)$$

Acorde con la Ec. 3.20, el parámetro CCC logra cuantificar la similitud entre los valores observados y los estimados como un criterio único.

3.8.9. Parámetros de Golbraikh & Tropsha

En 2002, los científicos Golbraikh y Tropsha,⁶⁵ postularon una serie de parámetros para determinar el poder predictivo real de un modelo QSAR a través del conjunto molecular externo. Acorde con los autores, solo se puede considerar un modelo QSAR aceptable si logra cumplir con los siguientes criterios:

- a). $R_{pred}^2 > 0.6$
- b). $\frac{R_{pred}^2 - R_0^2}{R_{pred}^2} < 0.1 \quad 0.85 \leq k \leq 1.15$
- c). $\frac{R_{pred}^2 - R_0'^2}{R_{pred}^2} < 0.1 \quad 0.85 \leq k' \leq 1.15$

$$d). |R_{pred}^2 - R_0'^2| < 0.3$$

Donde R_0^2 y $R_0'^2$ corresponden a los coeficientes de determinación entre los valores observados y predichos de los compuestos con y sin intercepto en el origen, respectivamente. El parámetro $R_0'^2$ contempla la misma información que R_0^2 , pero tiene los ejes invertidos. Los parámetros k y k' representan la pendiente de la línea de regresión.

3.8.10. Parámetro R_m^2

En general, para verificar la proximidad entre los valores observados y predichos, se emplea el parámetro R_m^2 propuesto por Roy *et al.*^{50, 66, 67} Este parámetro se calcula por medio de las siguientes ecuaciones:

$$R_m^2 = R_{pred}^2 \left(1 - \sqrt{(R_{pred}^2 - R_0^2)} \right) \quad (3.21)$$

$$R_m'^2 = R_{pred}^2 \left(1 - \sqrt{(R_{pred}^2 - R_0'^2)} \right) \quad (3.22)$$

$$\overline{R_m^2} = \frac{(R_m^2 + R_m'^2)}{2} \quad (3.23)$$

$$\Delta R_m^2 = |R_m^2 - R_m'^2| \quad (3.24)$$

Los significados de R_0^2 y R_{pred}^2 fueron definidos previamente. El parámetro R_m^2 usa los valores experimentales de la actividad en el eje de ordenadas, y $R_m'^2$ los usa en el eje de abscisas. El parámetro $\overline{R_m^2}$ es el promedio de R_m^2 y $R_m'^2$, y ΔR_m^2 representa la diferencia absoluta de los mismos. Solo cuando $\Delta R_m^2 < 0.2$ y $R_m^2 > 0.5$ se puede deducir que los valores predichos por el modelo tienen una proximidad aceptable con respecto a los valores observados.

3.8.11. Error Absoluto Medio (MAE)

Otro importante parámetro para validar el poder predictivo de los modelos QSAR es el Error Absoluto Medio (MAE), propuesto por Roy *et al.*, 2016;⁶⁸ expresado en la Ec. 3.25.

$$MAE = \frac{\sum |y_{obs} - y_{pred}|}{N} \quad (3.25)$$

El parámetro MAE se calcula a partir del rango de datos y la dispersión del error. En este caso, cuanto más bajo sea el valor de MAE, más equitativas serán las predicciones del modelo. Diferentes autores han determinado que el parámetro MAE

debe ser calculado en compañía de otras estrategias de análisis de dispersión del error, tal como *RMS* y *S*, mencionados previamente.⁶⁹

3.9. Lineamientos regulatorios del modelado QSAR

La teoría QSAR es básicamente una forma de extraer conocimiento a partir de las observaciones. En ese sentido, el grado de aceptación de los resultados obtenidos por los modelos QSAR depende de la calidad, veracidad, aplicabilidad e interpretabilidad de los mismos. En el estudio, diseño y aplicación de un modelo QSAR es importante reconocer y controlar las principales limitaciones de los elementos ajenos al modelado QSAR, tal como:

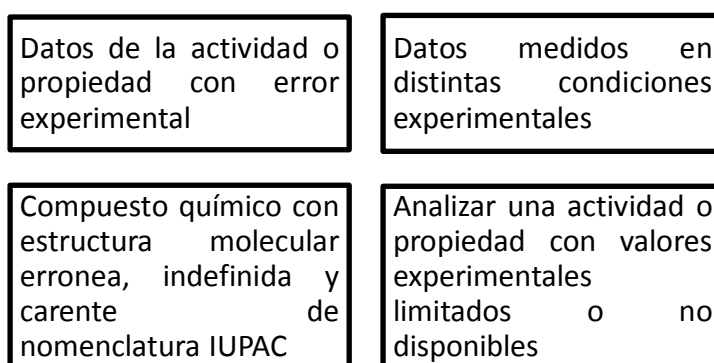


Figura 3.3. Principales limitaciones externas del modelado QSAR. Fuente: Saavedra, L.M.

A partir de 2007, la Organización para la Cooperación y el Desarrollo Económico (OCDE) postuló y adoptó 5 lineamientos claves para abordar apropiadamente el desarrollo de los modelos QSAR empleados con fines regulatorios o puramente científicos.^{70, 71} Cada lineamiento se especifica a continuación:

3.9.1. Punto final definido

El primer lineamiento enfatiza claramente que el punto final evaluado por un modelo QSAR debe estar bien definido. Asimismo, indica que para predecir acertadamente una respuesta final, ya sea biológica, fisicoquímica, sensorial o toxicológica, la misma debe presentar un procedimiento de medida preciso, que provea datos con el mínimo error experimental.

También, es necesario que todos los datos experimentales incorporados al modelo correspondan a medidas realizadas bajo las mismas condiciones de laboratorio o de campo, recomendadas por un protocolo de ensayo específico, preferiblemente estandarizado.

3.9.2. *Algoritmo inequívoco para la obtención del modelo*

Debido a la extensa diversidad de programas y métodos empleados para diseñar los modelos QSAR, el segundo lineamiento de la OCDE demanda utilizar un algoritmo inequívoco que garantice la reproducibilidad y aplicabilidad tanto del modelo QSAR como de las predicciones obtenidas por el mismo.

Asimismo, solicita que el desarrollo de los modelos QSAR se ejecute a través de programas o herramientas de acceso libre, que permitan replicar y comprobar los datos de predicción obtenidos para uno o varios conjuntos moleculares, independientemente del entorno de cálculo utilizado.

3.9.3. *Dominio de aplicación definido*

Dado que los modelos QSAR presentan limitaciones en términos estructurales, de propiedad y del mecanismo químico involucrado; la definición del dominio de aplicación es un elemento crucial para delimitar la capacidad predictiva del modelo QSAR en nuevos compuestos químicos aun inexplorados.⁴⁹

Este lineamiento indica que el límite teórico del modelo, definido por el dominio de aplicación puede variar según el método utilizado para definirlo y el rango de amplitud adquirido entre las variables independientes del modelo y la confiabilidad general de los datos de predicción.

3.9.4. *Medidas apropiadas de bondad de ajuste, robustez y predictibilidad*

Este lineamiento se refiere a la necesidad de validar los modelos QSAR respecto a su ajuste, estabilidad y capacidad predictiva. Existen diversas técnicas de validación implementadas en QSAR, que proveen dos tipos de información: a) el comportamiento interno del modelo, expresado a través del conjunto de calibración; y b) la predictibilidad del modelo, determinada por medio del conjunto de predicción.

El nivel de evaluación, control y respuesta de cada prueba es variable. Aunque, las pruebas de validación requeridas para aceptar un modelo QSAR se pueden basar en métodos comunes y simples, o métodos más novedosos, rigurosos y específicos. En particular, cuanto mayor sea el número de criterios de validación superados por el modelo, mayor será su credibilidad y aceptación.⁷²

3.9.5. *Interpretación del mecanismo involucrado*

La capacidad de interpretar el contenido de un modelo QSAR es un elemento útil para evidenciar el grado de reconocimiento y certeza que tiene el modelo frente al mecanismo químico, físico o biológico involucrado. Por ello, este lineamiento postula que los descriptores moleculares seleccionados deben presentar valores asociados al proceso responsable de la respuesta biológica observada.

Aunque, por distintas razones la interpretación de cada modelo QSAR es limitada. En primer lugar, existen miles de descriptores moleculares correlacionados tanto a las características estructurales como a las propiedades físicas o estéricas de las moléculas; esta correlación se puede considerar como una interpretación aproximada del mecanismo asociado. Además, los descriptores moleculares descartados del modelo pueden contener información más representativa para comprender el complejo proceso que conduce a la propiedad o actividad exhibida por las moléculas.

Incluso, existen diferentes modelos QSAR que utilizan descriptores moleculares completamente diferentes y apartados de las características biológicas contempladas por la respuesta final. Por estos motivos, se estima que la interpretación de cada modelo QSAR, es tan solo una hipótesis del mecanismo involucrado, y no representa una caracterización fehaciente del fenómeno natural envuelto.⁷³

3.10. Diagnóstico de un modelo de regresión QSAR

En un modelo QSAR diseñado con datos biológicos continuos, es necesario verificar la incertidumbre del valor predicho con respecto al valor observado de respuesta para cada molécula. Esta incertidumbre se conoce como valor residual. Un residuo demasiado grande, es un valor atípico del modelo.

Estadísticamente, el valor atípico representa un dato muy apartado del resto, en términos de valor absoluto. En general, se considera un valor atípico al valor de residuo con una diferencia mayor o igual a tres veces la desviación estándar del modelo ($\geq 3S$), y supera ampliamente la media de los valores de residuo.⁷⁴

En la práctica, los valores atípicos deben ser tratados con precaución, ya que son elementos anómalos del modelo que deben ser justificados. La causa de este comportamiento atípico puede atribuirse a uno o varios aspectos vinculados al diseño del modelo, tales como:

- Recolección y discriminación incorrecta del conjunto de compuestos químicos utilizados para desarrollar el modelo QSAR.
- Presencia de datos biológicos experimentales incorrectos, debido a discrepancias del protocolo de ensayo, criterios variables de experimentación, condiciones de laboratorio inestables, entre otros factores.
- La necesidad de involucrar nuevos descriptores moleculares que codifiquen otras características estructurales con nueva información para predecir la respuesta biológica estudiada.
- La sustancia química detectada como atípica interactúa por un mecanismo molecular diferente en su sitio bioquímico de acción, que difiere de las otras sustancias químicas estudiadas.

- La molécula atípica presenta uno o más productos de transformación metabólica o química que actúan por un mecanismo de acción bioquímico diferente a las otras moléculas del conjunto.
- El método estadístico utilizado para construir el modelo presenta importantes limitaciones para encontrar la relación funcional entre la actividad biológica y los descriptores moleculares seleccionados.

Después de una explicación lógica de un valor atípico, se puede perfeccionar el modelo QSAR y mejorar su poder predictivo. Aunque, es importante resaltar que la eliminación de un compuesto químico atípico solo se debe realizar cuando se identifica claramente un problema ajeno al modelado. De lo contrario, las estrategias más convenientes son para superar dicho valor atípico son:

- Incluir nuevas moléculas similares al elemento atípico
- Involucrar más descriptores moleculares
- Cambiar de herramientas estadísticas de modelado

Cabe destacar que el proceso para perfeccionar un modelo QSAR puede ser lento, especialmente cuando se deriva de métodos no lineales como las Redes Neuronales Artificiales (ANN).⁷⁵

3.11. Referencias

1. Ghafourian, T., Cronin, M.T. The impact of variable selection on the modelling of oestrogenicity. *SAR & QSAR Environmental Research*. **2005**, 16(1–2), p.171–190.
2. Miller, J.C., Miller, J.N. Eds. *Statistic for analytical chemistry*. 2nd Ed. Ellis Horwood: Chischester. **1988**.
3. Goodarzi, M., Dejaegher, B., Heyden, Y.V. Feature selection methods in QSAR studies. *Journal of AOAC International*. **2012**, 95, p. 636–651.
4. Roy, K., Kar, S. *How to judge predictive quality of classification and regression based QSAR models*. In: Haq, Z.U., Madura, J., Eds. *Frontiers of Computational Chemistry*. Sharjah, UAE: Bentham Science. **2015**. p.71–120.
5. Roy, K., Kar, S., Das, R.N. *Statistical methods in QSAR/QSPR. A Primer on QSAR/QSPR Modeling*. **2015**, p. 37–59.
6. Peter, S.C., Dhanja, J.K., Malik, V., Radhakrishnan, N., Jayakanthan, M., Sundar, D. *Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications*. En: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds. *Encyclopedia of Bioinformatics and Computational Biology*. Vol 2, Elsevier: Amsterdam. **2019**, p. 661–676.
7. Settles, B. *Active learning: Synthesis lectures on artificial intelligence and machine learning*. Morgan & Clay Pool: Long Island, NY. **2012**.
8. Levatic, J., Dzeroski, S., Supek, F., Smuc, T. Semi-supervised learning for quantitative structure – activity modeling. *Informatica*, **2013**, 37 (2), p.173.
9. Devillers, J. *Methods for building QSARs*. En Reisfeld, B., Ed. *Computational Toxicology*. Vol 2., Part I: Toxicological/Pharmacological Endpoint Prediction. Humana Press: London. **2013**, p. 19–35.
10. Devillers, J. *QSAR in Environmental and Health Sciences: Computational Design of Chemicals for the Control of Mosquitoes and Their Diseases*. CRC Press – Taylor & Francis group: Boca Raton. **2018**.
11. Roy, K., Mitra, I. Comb. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Chem. High Throughput Screen*. **2011**, 14, p. 450–474.
12. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Combinatorial Science*. **2007**, 26, p. 694–701.
13. Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.D., Lee, K.H., Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*. **2003**, 17, p. 241–253.

14. Konovalov, D.A., Llewellyn, L.E., Vander Heyden, Y., Coomans, D. Robust cross-validation of linear regression QSAR models. *Journal of Chemical Information and Modeling*. **2008**, 48, p. 2081–2094.
15. Wold, S., Eriksson, L., Clementi, S. *Statistical Validation of QSAR Results. Chemometrics. Methods in Molecular Design*. En Van de Waterbeemd. H. Eds.; Wiley VCH Verlag GmbH: Weinheim. **1995**.
16. Organization for Economic Co-operation and Development (OECD) [En línea]. OECD Quantitative Structure-Activity Relationships Project [(Q)SARs]., The OECD QSAR Project, 2019. [Citado el 17 de Abril de **2019**]. Disponible en: www.oecd.org/chemicalsafety/risk-assessment/oecdquantitativestructure-activityrelationshipsprojectqsars.htm
17. Schüürmann, G., Ebert, R.U., Chen, J., Wang, B., Kuhne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean. *Journal of Chemical Information and Modeling*. **2008**, 48 (11), p. 2140–2145.
18. Höskuldsson, A. *Prediction Methods in Science and Technology*. Thor publishing: Denmark. **1996**, p. 405.
19. Thomsen, M. QSAR in environmental risk assessment: Interpretation and Validations of SAR/QSAR based on multivariate data analysis. Thesis of Department of Life Science and Chemistry, Roskilde University for the Ph.D Degree. Department of Environmental Chemistry National Environmental Research Institute: Roskilde. **2001**.
20. Dearden, J.C., Cronin, M.T.D., Kaiser, K.L.E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*. **2009**, 20, p. 241–266.
21. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences*. **1988**, 28 (1), p. 31–36.
22. West, D.B. *Introduction to graph theory*. Pearson, Prentice hall Upper: Saddle River. **2001**.
23. Hansch, C and Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal American Chemical Society*. **1964**, 86 (8), 1616–1626.
24. Hansch, C.L.A. *Substituent constants for correlation analysis in chemistry and biology*. John Wiley and Sons: New York. **1979**.

25. Hansch, C., Leo, A. Exploring QSAR, Fundamentals and Applications in Chemistry and Biology. Ed by American Chemical Society, Washington DC, **1995**, p. 139–205.
26. Benfenati, E., Diana, R.G., Gini, G., Cardamone, L., Gocieva, M., Mancusi, M., Padovani, R., Tamellini, L. *Introduction on in vivo, in vitro and in silico methods*. En Benfenati, E., Ed. The e-book on QSAR and REACH: theory, guidance and applications. Orchestra: Milan, **2012**, p. 5–37. Disponible en: http://ebook.insilico.eu/insilico-ebook-orchestrabenfenati-ed1_rev-June2013.pdf.
27. Ekins, S. *Computational Toxicology: Risk Assessment for Pharmaceutical and Environmental Chemicals*. Wiley-Interscience: New Jersey. **2007**.
28. Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T., McDowell, R.M., Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARS. *Environmental Health Perspectives*. **2003**, 111, p. 1361–1375.
29. Agresti, A. *An Introduction to Categorical Data Analysis*. John Wiley: New York. **2007**.
30. Everitt, B., Landau, S., Leese, M. *Cluster analysis*, 4th ed. Arnold: London, **2001**.
31. Harrell, F. *Regression modelling strategies*. Springer Cross Ref Google Scholar: Nashville. **2001**.
32. Roy, P.P., Leonard, J.T., Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*. **2008**, 90, p. 31–42.
33. Veerasamy, R., Rajak, H., Jain, A., . Validation of QSAR models-strategies and importance. *International Journal of Drug Design & Discovery*. **2011**, 3, p. 511–519.
34. Hawkins, D.M., Basak, S.C., Mills, D. Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*. **2003**, 43, p. 579–586.
35. Stanton, D.T., Murray, W.J., Jurs, P.C. Comparison of QSAR and molecular similarity approaches for a structure-activity relationship study of DHFR inhibitors. *Quantitative Structure-Activity Relationship*. **1993**, 12, p. 239–245.
36. Rücker, C., Rücker, G., Meringer, M. Y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information Modeling*. **2007**, 47, p. 2345–2357
37. Wold, S., Sjöström, M. *Chemometrics and its Roots in Physical Organic Chemistry*. *Acta Chem. Scan*. **1998**, 52, p. 517–523.

38. Lorber, A., Wangen, L.E., Kowalski, B.R. A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*. **1987**, 1, p. 19–31.
39. Cramer, R.D. Partial least squares (PLS): Its strengths and limitations. *Perspectives in Drug Discovery and Design*. **1993**, 1, p. 269–278.
40. Dixon, W.J. *Biomedical computer programs P-series*. University of California Press. **1979**.
41. Duchowicz, P.R., Castro, E.A. & Fernández, F.M. Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Communications in Mathematical and in Computer Chemistry*. **2006**, 55 (1), p. 179–192.
42. Mota, S.G., Barros, T.F., Castilho, M.S. 2D QSAR studies on a series of bifonazole derivatives with antifungal activity. *Journal of the Brazilian Chemical Society*. **2009**, 20, p. 451–459.
43. Roy, K., Kar, S. How to judge predictive quality of classification and regression based QSAR models. En Haq, Z.U., Madura, J. Eds., *Frontiers of Computational Chemistry*. Sharjah, UAE: Bentham Science. **2015**, p.71–120.
44. Chirico, N., Gramatica, P. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*. **2011**, 51, p. 2320–2335.
45. Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H. Comparative Studies on some metrics for external validation of QSAR model. *Journal of Chemical Information Modeling*. **2012**, 52 (2), p.396–408.
46. Mullen, L.M.A., Duchowicz, P.R., Castro, E.A. QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. *Chemometrics Intelligent Laboratory Systems*. **2011**, 17, p. 269–275.
47. Rafiei, H., Khanzadeh, M., Mozaffari, S., Bostanifar, M.H., Avval, Z.M., Aalizadeh, R., Pourbasheer, E. QSAR study of HCV NS5B polymerase inhibitors using the Genetic Algorithm-Multiple Linear Regression (GA-MLR). *EXCLI Journal*. **2016**, 15, p. 38–53.
48. Gramatica, P., Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: a Clarification Concerning Metrics and Terminology. *Journal of Chemical Information and Modeling*. **2016**, 56 (6). p. 1127–1131.
49. Roy, K., Kar, S., Nas, R.N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier: London. **2015**.

50. Roy, K. On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opinion in Drug Design*. **2007**, 2, p. 1567–1577.
51. Duchowicz, P.R. Linear Regression QSAR Models for Polo-Like Kinase-1 Inhibitors. *Cells*. **2018**, 7 (2), p. 1–11.
52. Matlab 7.0. Massachusetts, USA: The MathWorks, Inc., <http://www.mathworks.com>
53. Muhammad, U., Uzairu, A., Arthur, D.E., Review on: Quantitative Structure Activity Relationship (QSAR) modelling. *Journal Analytical & Pharmaceutical Research*. **2018**, 7(2), p. 240–242.
54. Shi, L.M., Fang, H., Tong, W.D., Wu, J., Perkins, R., Blair, R.M., Branham, W.S., Dial, S.L., Moland, C.I., Sheehan, D.M. QSAR Models Using a Large Diverse Set of Estrogens. *Journal of Chemical Information and Computational Science*. **2001**, 41, p. 186–195.
55. Jaworska, J.S., Comber, M., Auer, C., Van Leeuwen, C.J. Summary of a workshop on regulatory acceptance of (Q) SARs for human health and environmental endpoints. *Environmental Health Perspectives*. **2003**, 111(10), p. 1358–1360.
56. Tropsha, A., Gramatica, P., Gombar, V.K. The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Combinatorial Science*. **2003**, 22, p. 69–77.
57. Todeschini, R., Consonni, V., Pavan, M. *MOBY DIGS-Models By Descriptors In Genetic Selection*; Talete srl, Milan, Italy, **2002**.
58. Consonni, V., Ballabio, D., Todeschini, R. Comments on the Definition of the Q(2) Parameter for QSAR Validation. *Journal of Chemical Information Modeling*. **2009**, 49, 1669–1678.
59. Ensuncho, A.E., Robles, J.R., Figueredo, S.F. Modeling Quantitative Structure-Activity Relationship (QSAR) of Derivatives 5-(5-Nitroheteroaryl)-1,3,4-thiadiazols with Leishmanicidal Activity. *Información Tecnológica*. 2017, 28 (2), doi:10.4067/S0718-07642017000200020
60. Duchowicz, P.R., Szewczuk, N.A., Pomilio, A.B. QSAR Studies of the Antioxidant Activity of Anthocyanins. *Journal of Food Science and Technology*. **2019**, 56(12), p. 5518–5530.
61. Schürmann, G., Ebert, R.U., Chen, J., Wang, B., Kuehne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient–Test Set Activity Mean vs Training Set Activity Mean. *Journal of Chemical Information and Modeling*. **2008**, 48, p. 2140–2145.
62. Consonni, V., Ballabio, D., Todeschini, R. Evaluation of Model Predictive Ability by External Validation Techniques. *Journal of Chemometrics*. **2010**, 24, p. 194–201.
63. Lin, L.I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, 45, p. 255–268.

64. Lin, L.I.-K. Assay Validation Using the Concordance Correlation Coefficient. *Biometrics* **1992**, 48, p. 599–604.
65. Golbraikh, A., Tropsha, A. Beware of q^2 !. *Journal of Molecular Graphics and Modelling*. **2002**, 22, p. 269–276.
66. Roy, P.P., Paul, S., Mitra, I., Roy, K. On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules (Basel Switzerland)*. **2009**, 14 (5), p. 1660–1701.
67. Roy, K.; Mitra, I.; Kar, S.; Ojha, P. K.; Das, R. N.; Kabir, H. Comparative Studies on Some Metrics for External Validation of QSPR Models. *Journal of Chemical Information and Modeling*. **2012**, 52, p. 396–408.
68. Roy, K., Das, R.N., Ambure, P., Aher, R.B. Be Aware of Error Measures. Further Studies on Validation of Predictive QSAR Models. *Chemometrics and Intelligent Laboratory Systems*. **2016**, 152, p. 18–33.
69. Chai, T., Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments against Avoiding RMSE in the Literature. *Geoscientific Model Development*. **2014**, 7 (3), p. 1247–1250.
70. OECD environment health and safety publications series on testing and assessment No. 69 Guidance document on the validation of (Quantitative) Structure - Activity Relationship [(Q)SAR] models., 2007. [Citado el 21 de Abril de **2019**]. Disponible en: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote5env/jm/mono\(2007\)2&doclanguage5en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote5env/jm/mono(2007)2&doclanguage5en).
71. OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models. Organization for economic co-operation and development, **2004**. [Citado el 26 de Abril de 2019]. Disponible en: <http://www.oecd.org/env/ehs/riskassessment/37849783.pdf>
72. Gini, G. Chapter 3 QSAR: *What else?*. En Nicolotti, O., Ed. *Computational Toxicology: Methods and Protocols*, Molecular Biology, vol. 1800, Humana Press: New York. **2018**, p. 79–105.
73. Draper, N., Smith, H. *Applied regression analysis*, 2nd edn. Wiley: New York. **1981**.
74. Devillers, J., Boule, P., Vasseur, P., Prevot, P., Steiman, R., Seigle-Murandi, F., Benoit-Guyod, J.L., Nendza, M., Grioni, C., Dive, D., Chambon, P. Environmental and health risks of hydroquinone. *Ecotoxicology and Environmental Safety*. **1990**, 19 (3), p. 327–354.
75. Devillers, J. *Neural networks in QSAR and drug design*. Academic: London. **1996**.

CAPITULO 4: ANTECEDENTES QSAR ORIENTADOS SOBRE EL CONTROL DEL MOSQUITO *Aedes aegypti* L.

4.1 Introducción

Los vectores de enfermedades son aquellos insectos ectotérmos con capacidad de transmitir por medio de la picadura diferentes bacterias, parásitos o virus que causan importantes afecciones a los humanos. Las denominadas Enfermedades Transmitidas por Vectores (VBDs) están catalogadas como Enfermedades Tropicales Desatendidas (NTD), propagadas principalmente por mosquitos, moscas, flebótomos, chinches, garrapatas, ácaros y piojos.¹

Anualmente, alrededor de 700.000 muertes son atribuidas a las principales VBDs, como el paludismo, el dengue, la filariasis linfática (elefantiasis), la malaria, la fiebre de chikungunya, la fiebre amarilla, el zika, la enfermedad de Chagas y la esquistosomiasis. Entre estos patógenos, el dengue constituye la principal amenaza para la salud pública global, con 390 millones de contagios reportados por año, de los cuales el 20% son mortales.² En este contexto, se estima que el 80% de la población mundial se encuentra en riesgo de contraer alguna VBD y el 40% el virus del dengue (DENV).

Durante los últimos 50 años, el DENV aumento 30 veces su virulencia, debido a diversos factores ambientales, demográficos y sociales que favorecieron la propagación de su principal vector, el mosquito *Aedes aegypti* Linnaeus (Diptera: Culicidae); localizado en las regiones tropicales y subtropicales del planeta, donde se adapta fácilmente a las zonas urbanas y tiene la capacidad de transmitir múltiples *flavivirus*, como el DENV, la fiebre amarilla, el virus de chikungunya (CHIKV) y el virus del zika (ZIKV).³

A causa de una proliferación sin precedentes de los patógenos transmitidos por el mosquito *A. aegypti* L., particularmente los brotes del ZIKV reportados en la mayor parte del hemisferio Occidental en el período 2014-2016; la Organización Mundial de la Salud (OMS) propuso diversas estrategias públicas para el control de vectores, recopiladas en la denominada *Respuesta mundial para el control de vectores 2017-2030*.⁴ Allí, se dispuso que el *A. aegypti* L. es una grave amenaza para la salud pública mundial.

Entre los múltiples esfuerzos por monitorear a los mosquitos de orden *Diptera*, se han vinculado técnicas integrales de control de tipo físico, biológico y químico, tal

como las trampas con cebos aromáticos, el uso de cepas de *Wolbachia* y las fumigaciones con diversos insecticidas de síntesis.

La implementación de insecticidas sintéticos para el control de vectores ha generado un eficiente control poblacional de estos insectos. Sin embargo, dicha estrategia repercutió en el desarrollo de múltiples casos de resistencia en mosquitos y otros artrópodos, así como en la bioacumulación y biomagnificación en los organismos no objetivos y sus ecosistemas.

Por estas razones, tanto en Norteamérica como en el continente Europeo decretaron a los fabricantes de productos químicos realizar una extensa evaluación toxicológica, ecotoxicológica y biológica de cada sustancia comercial que se destine al control de insectos, particularmente de aquellas pertenecientes al grupo organofosforado, organoclorado, piretroide o carbamato, extensamente aplicadas en los planes de defensa, control y erradicación de los vectores de enfermedades.⁵

A partir de estos requerimientos fue posible evidenciar la inminente demanda de la sociedad por prevenir los peligros provenientes de los pesticidas sintéticos y buscar alternativas químicas que favorezcan la protección de la salud humana y eludan el actual fenómeno de resistencia de los artrópodos. La exploración de nuevas estrategias de control vectorial, condujo la atención de los investigadores sobre los compuestos de origen natural y sus análogos, caracterizados por contener propiedades toxicológicas, organolépticas y biológicas reconocidas y otras aún inexploradas.

En los procesos de evaluación y determinación de la respuesta activa de una molécula, se emplean diferentes ensayos *in vivo* e *in vitro* que brindan información acerca de los efectos benignos o nocivos que estas entidades químicas generan sobre un organismo animal o un cultivo celular.

Sin embargo, este tipo de pruebas son consideradas de alta complejidad, ya que al momento de evaluar grandes conjuntos de compuestos, requieren mayor tiempo e insumos de alto costo. Además, la mayor parte de las pruebas *in vivo* son éticamente inaceptables ante la comunidad científica internacional por involucran el sacrificio o maltrato animal.

Otra de las técnicas utilizadas para estimar las propiedades activas de las moléculas apunta a las investigaciones de simulación por computadora, las cuales han sido altamente determinantes para acceder, examinar, deducir, modelar y proyectar enormes conjuntos moleculares, sin requerir la síntesis, caracterización y ensayo de cada molécula. Estos atributos postulan a los estudios *in silico* como instrumentos esenciales para identificar y diseñar estratégicamente nuevas entidades químicas con actividades plausibles.^{6,7}

En este contexto, la teoría QSAR constituye una de las principales herramientas computacionales diseñadas para establecer paralelismos racionales entre la estructura molecular de una sustancia y el efecto biológico final producido por la misma.⁸⁻¹⁰

Un modelo de la teoría QSAR es considerado el resultado final de un proceso computacional que inicia con la extensa descripción de la estructura molecular, donde las características simples (constitucionales y topológicos) y complejas (electrónicas y geométricos) de las moléculas son detalladas mediante los descriptores moleculares; y finaliza con los valores de predicción de la propiedad evaluada, utilizados para establecer algunas inferencias o hipótesis sobre el comportamiento de las moléculas en un sistema ambiental o biológico definido.^{11, 12}

El diseño y aplicación de un modelo QSAR resulta particularmente útil cuando las pruebas y medidas experimentales son extensas, laboriosas y de alto costo; ya que provee información relevante para guiar el diseño racional, rápido y selectivo de nuevos precursores químicos con actividad o propiedad deseada, basándose en la información química, biológica y física de los compuestos existentes y ensayados. Por esta razón, la teoría QSAR constituye un primer paso computacional necesario en la búsqueda de nuevas estructuras químicas con valores favorables de toxicidad aguda sobre el principal transmisor de diversas enfermedades endémicas, el mosquito *A. aegypti* L.

Por tanto, este capítulo de Tesis Doctoral busca profundizar sobre el impacto de las VBDs sobre la salud pública mundial. Asimismo, resulta trascendental destacar el rol del mosquito *Aedes aegypti* L., en la transmisión de los virus DENV, CHIKV y ZIKV en el hemisferio Occidental. Por ello, fue necesario puntualizar sobre las estrategias de control o erradicación vectorial que han repercutido en efectos secundarios nocivos para la salud humana y el entorno.

Por último, para indagar sobre el estado actual y la trascendencia de las herramientas computacionales y no computacionales para estimar la actividad larvicida sobre el mosquito *A. aegypti* L.; se recopilaron de la literatura especializada diferentes investigaciones generadas en la última década (2009–2019), enfocadas en predecir la toxicidad aguda LC₅₀ o IC₅₀ de diferentes conjuntos moleculares altamente heterogéneos sobre el mosquito *A. aegypti* L en estadio larval.¹³⁻¹⁷

4.2 Las Enfermedades Trasmitidas por Vectores (VBDs)

Una primicia crítica de la epidemiología es que las enfermedades no ocurren aleatoriamente en una población. Por el contrario, se deben a factores ambientales, demográficos y sociales minuciosamente organizados para contagiar a los individuos más vulnerables. Un importante uso de los estudios epidemiológicos es identificar el comportamiento de los factores de riesgo que conducen a las poblaciones a contraer alguna enfermedad. La mayoría de las enfermedades requieren diferentes balances e interacciones de los factores de riesgo.

En el caso particular de las Enfermedades Trasmitidas por Vectores (VBDs), como el paludismo, el dengue, la malaria, la fiebre de chikungunya, la fiebre amarilla, el zika,

la filariasis linfática (elefantiasis), la enfermedad de Chagas y la esquistosomiasis, constituyen el 17% de todas las enfermedades infecciosas, caracterizadas por una compleja dinámica y naturaleza, que dificulta la posibilidad de estimar las repercusiones existentes, reemergentes o eventuales sobre la salud humana. Esta imprevisibilidad, trae como consecuencia alrededor de 700.000 muertes por año.^{2, 18}

Para el desarrollo práctico y efectivo de las estrategias de control y prevención de las VBDs, usualmente se requiere la evaluación de los componentes y las interacciones para elucidar la compleja tríada ecológica entre el patógeno, el vector, el huésped y el ecosistema Figura 4.1.

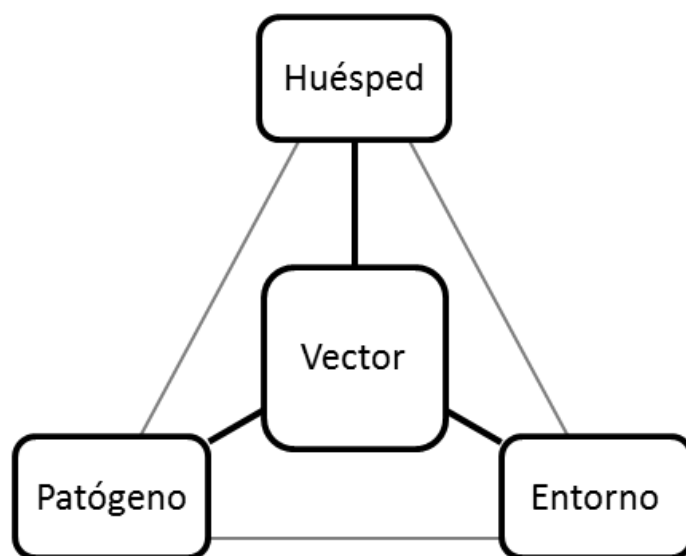


Figura 4.1. Tríada epidemiológica de las VBDs. Fuente: Saavedra, L.M.

Los arbovirus, son los virus presentes en la naturaleza a causa de la transmisión biológica de los artrópodos hematófagos a sus huéspedes vertebrados. En la actualidad, se conocen más de 130 tipos de arbovirus que causan infecciones a los vertebrados más susceptibles y han promovido grandes epidemias de enfermedades reemergentes en las últimas décadas.

4.3 Principales arbovirus transmitidos por mosquitos del género *Aedes*

La expansión global de los arbovirus es precedida por la propagación descontrolada de los insectos vectores como las moscas, flebótomos, chinches, garrapatas, ácaros, piojos y los mosquitos, en especial del géneros *Aedes*, *Anopheles*, *Culex* y *Sabethes*.^{1, 19} Entre los diversos arbovirus transmitidos por los mosquitos del género *Aedes*, los de mayor impacto para la salud humana son los virus de DENV, CHIKV y ZIKV, al no contar con una vacuna o cura efectiva.

4.3.1. El virus del dengue (DENV)

El virus del dengue (DENV), fue aislado y estudiado por primera vez durante la epidemia de Nagasaki en el año 1943. Allí, se identificó que el DENV estaba conformado por cuatro serotipos inmunológicos estrechamente asociados pero antigénicamente distintos, denominados DENV-1, DENV-2, DENV-3 y DENV-4; pertenecientes al género *flavivirus* y la familia *flaviviridae*. Estos cuatro serotipos comparten aproximadamente el 65% de sus genomas; es decir, cada infección producida por un determinado serotipo del DENV, presenta la misma sintomatología clínica y la inmunidad permanente contra el serotipo involucrado.²⁰

Los cuatro serotipos del DENV presentan ciclos de transmisión tanto silvestre como urbana. En el ciclo de transmisión urbana, el principal vector del virus es el mosquito *Aedes aegypti* L., seguido por los mosquitos *Aedes polynesiensis* y *Aedes albopictus*, localizados en las zonas tropicales y subtropicales del planeta. Este tipo de transmisión ha promovido que el DENV este asociado con los problemas económicos y sociales de las regiones endémicas, donde anualmente se registran casos infecciosos cada vez más severos.

Por más de 50 años, la enfermedad del dengue alcanzó progresivamente el estado de pandemia, con 390 millones de contagios anuales, que conducen a 70 millones de muertes a nivel mundial. Asimismo, se identificó que aproximadamente 5 millones de contagios del DENV involucran niños, niñas y adolescentes, y que al menos 70 mil mueren al contraer dengue hemorrágico.^{1, 2, 21-23}

En la última década, la OMS ha generado boletines periódicos de monitoreo y control del DENV. Para el período interanual 2012-2013, se determinó que este virus había incrementado su área de influencia a 111 países con aproximadamente 3,9 billones de habitantes bajo amenaza de contagio. En el año 2017, la OMS publicó el primer planisferio con la disposición geográfica de los casos clínicos confirmados del DENV en 2016, exhibido en la Figura 4.2.

La mayor parte del hemisferio occidental registró infecciones por dengue, con cifras alarmantes en países de Suramérica como Brasil, Colombia, Perú, Venezuela, Ecuador, Paraguay, Bolivia, Guayana Francesa y Surinam. También las regiones de Centro América y el Caribe reportaron infecciones por el DENV, especialmente México y Puerto Rico con más de 100 mil casos confirmados.²⁴

Acorde con la Figura 4.2, el continente insular Oceanía y parte del continente Asiático presentan una proliferación del DENV. A partir de este planisferio, la OMS reveló que 40% de la población mundial se encuentra en áreas endémicas del DENV y vive constantemente bajo alerta de contagio por dengue y dengue hemorrágico.

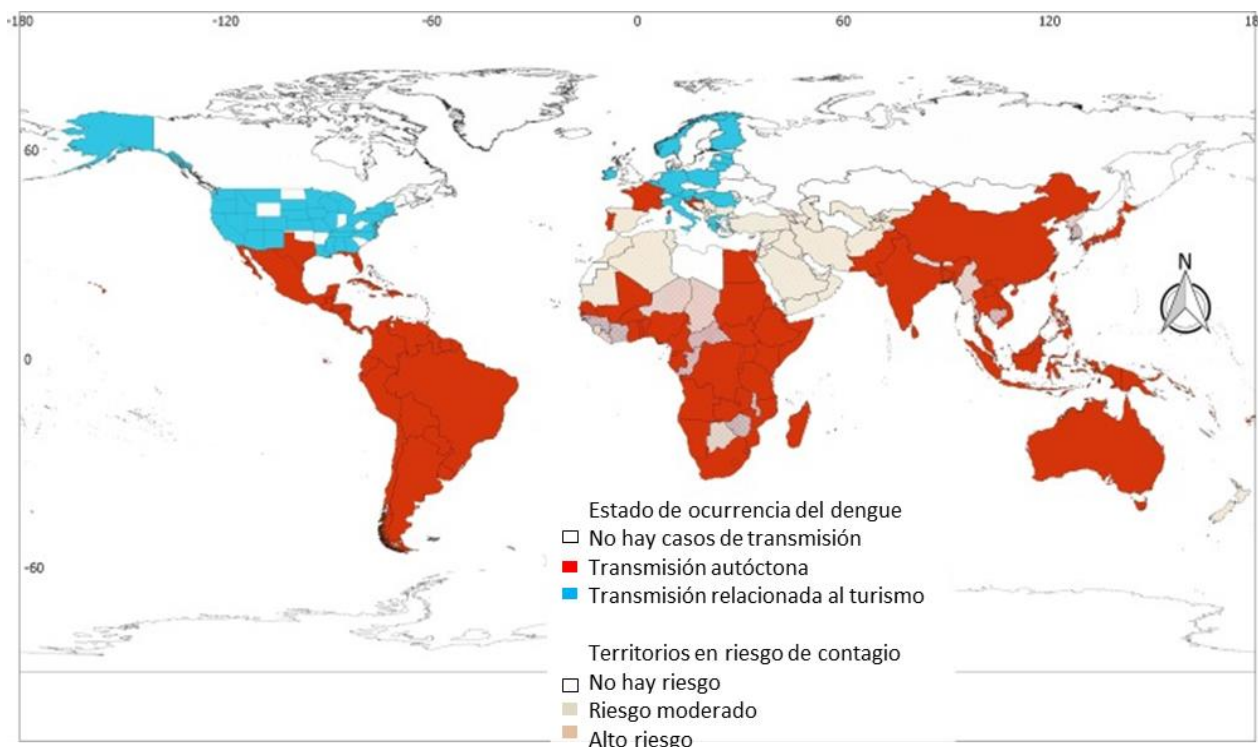


Figura 4.2. Distribución del dengue a nivel Mundial en 2016. Fuente: WHO, 2017.²⁴

4.3.2. *El virus de chikungunya (CHIKV)*

El virus de chikungunya (CHIKV) fue descubierto, aislado y analizado en el año 1953, a partir de muestras sanguíneas de personas infectadas durante el brote de enfermedad febril ocurrido en Tanzania. El CHIKV es un Alfavirus que hace parte de la familia *Togaviridae*. Este virus posee cuatro genotipos distintos clasificados como: Oeste de África, Este, Centro, y Sur de África (ECSA), Asia y Océano Indico.²⁵

La fiebre de chikungunya es transmitida por mosquitos del género *Aedes*, en particular *Aedes aegypti* y *Aedes albopictus*. Los primeros síntomas de contagio del CHIKV ocurren 4 u 8 días después de la picadura del vector; en este caso el huésped infectado presenta síntomas clínicos similares a los producidos por el DENV.

La sintomatología común de la fiebre de chikungunya involucra un fuerte cuadro febril acompañado por dolor muscular y articular en las extremidades, dolor de cabeza y erupciones cutáneas de color rojizo; estos síntomas persisten por 1 u 8 semanas. Aunque, en los casos más irregulares pueden persistir por varios meses o años. Otras complicaciones menos frecuentes derivadas de esta enfermedad vírica, implican problemas oculares, neurológicos y cardíacos.²⁶⁻²⁸

Si bien no existen fármacos o vacunas comerciales que limiten la replicación del CHIKV, los tratamientos clínicos actuales buscan contrarrestar y aliviar los síntomas

mediante terapias antipiréticas y antálgicas basadas en los fármacos favipiravir, ribavirina e interferón, con acción *in vitro* sobre los virus con genoma ARN.

El virus del chikungunya exhibió sus primeros brotes endémicos en la década del 60. El intercambio comercial favoreció la propagación del CHIKV en diversos países del Sudeste Asiático. Los buques de carga provenientes de África transportaban mercadería, tripulantes y contenedores de agua que hospedaban al vector del virus.²⁵

Históricamente, las epidemias atribuidas al CHIKV han mostrado un comportamiento cíclico, con períodos interepidémicos que oscilan en 4 y 30 años. Desde 2004, el CHIKV incremento su distribución geográfica, alcanzando una magnitud sin precedentes de casos infecciosos en Asia y África. Durante el período 2005–2010, el CHIKV tuvo una amplia proliferación en el Océano Índico y el sur de Europa.^{29, 30}

Posterior a esto, surgió el rebrote del CHIKV en América, lo cual generó la inquietud acerca del impacto de este virus sobre la salud pública mundial. El primer caso importado se dio en la isla de Saint Martin en el Caribe. Pocos meses después, la Organización Panamericana de la Salud (OPS) en alianza con la OMS registraron más de 4 mil casos confirmados en los países del Caribe.^{26, 31}

La propagación sin precedentes del CHIKV llegó a los 150 mil contagios autóctonos en más de 43 países del hemisferio Occidental. Aunque, el mayor impacto del virus recayó sobre Brasil, con el 97% de los casos infecciosos registrados, seguido por países como Ecuador, Bolivia y México. En 2017, la alianza OPS/OMS registró cerca de 122 mil casos de morbilidad atribuidos al CHIKV. Desde entonces, la región Andina y parte del cono sur fueron declaradas zonas endémicas de alto riesgo por contagio de la fiebre de chikungunya.³¹⁻³³

Durante el último año, la carga de morbilidad atribuida al CHIKV presentó una disminución escalonada como resultado de las múltiples estrategias de vigilancia entomológica, control vectorial y prevención del virus impuestas en diferentes países tropicales y subtropicales. No obstante, diversos estudios proyectan que el cambio climático y la propagación indiscriminada del mosquito *A. aegypti*, promueven el riesgo de enfrentar rebrotes del virus.^{34, 35}

4.3.3. El virus de zika (ZIKV)

El virus de zika (ZIKV) es un arbovirus transmitido por vectores, principalmente por el mosquito *A. aegypti*; este patógeno pertenece al género *flavivirus* y a la familia *flaviviridae*, al igual que el virus de la fiebre amarilla (YFV) y el virus del dengue (DENV).¹⁹

El ZIKV fue detectado por primera vez en Uganda en 1947. Desde entonces, los contagios por ZIKV ha sido esporádicos en las poblaciones Africanas, y solo hasta 1960 irrumpió en el continente Asiático. Posteriormente, en 2007 ocurrió el primer brote

endémico de zika en la Isla de Yap de los Estados Federados de Micronesia, donde el 73% de la población contrajo el ZIKV. En el período interanual 2013–2014 surgieron múltiples rebrotes en varias islas del océano pacífico y la Polinesia Francesa, con más de 28 mil afectados.

Eventualmente, en 2015 se registraron los primeros brotes del ZIKV en el Caribe, Centro y Sur América. Acorde con la Plataforma de Información en Salud para las Américas (PLISA), tan solo un año después, se confirmaron 199 mil contagios de ZIKV en más de 50 países de la región, donde el 65% de estos casos se concentraron en la región del cono sur.³⁶⁻³⁹ Por ello, la OMS catalogó a la enfermedad de zika de alto riesgo para la salud pública mundial.

A través de los múltiples esfuerzos de vigilancia y control vectorial, se logró disminuir las tasas de morbilidad atribuida al ZIKV. Según la plataforma PLISA (<http://www.paho.org/data/index.php/es/temas/indicadores-zika.html>) los casos de ZIKV han disminuido gradualmente; en 2017 se reportaron cerca de 20 mil contagios confirmados, en 2018 la cifra fue menor a 4 mil casos. Aunque, la tendencia no se mantuvo, el impacto del virus aumento a 6 mil contagios confirmados y 4 muertes durante el último año.

Por otra parte, el ZIKV posee cepas clasificadas en tres principales linajes: Este de África, Oeste de África y Asia.⁴⁰ Al igual que el DENV, este virus posee dos ciclos de transmisión ilustrados en la Figura 4.3., denominados silvestre y urbana. El ciclo de transmisión silvestre se basa en la interacción entre primates y mosquitos del género *Aedes* que habitan naturalmente en la selva, tal como el *A. africanus*, *A. bromeliae*, *A. dalzieli*, *A. furcifer*, *A. vittatus*, *A. luteocaphalus*, *A. opok*, *A. taylori* y *A. unilineatus*.

En el ciclo de transmisión urbana, el ser humano es el principal reservorio y huésped del virus. Este ciclo presenta dos formas de transmisión, ya sea por medio de los vectores *A. aegypti* y *A. albopictus* o mediante fluidos corporales como sangre, saliva, orina, leche materna, líquido amniótico, el líquido que envuelve tanto el encéfalo como la médula espinal y por transmisión sexual.³⁷⁻⁴²

La infección del ZIKV presenta sintomatología leve, que se manifiesta pocos días después del contagio; regularmente los casos infecciosos manifiestan conjuntivitis, dolores musculares, articulares y cansancio, que persisten por 2 o 7 días bajo tratamiento médico con antipiréticos y antiespasmódicos de uso común.⁴²

Recientemente, diversos estudios científicos postularon la posible relación entre los brotes del ZIKV desatados en Brasil y la Polinesia Francesa, con el aumento repentino de diversas enfermedades neurológicas; en particular, el Síndrome de Guillan-Barré (GBS) en adultos y diferentes malformaciones congénitas *in utero* como microcefalia fetal, disfunción del tallo encefálico, contractura de los miembros y anomalías cerebrales, auditivas y visuales.⁴³⁻⁴⁵

Ante esta situación, la OMS catalogo estos trastornos como una emergencia de la salud pública internacional, denominada *síndrome congénito asociado a la infección por el virus de zika*.⁴⁶

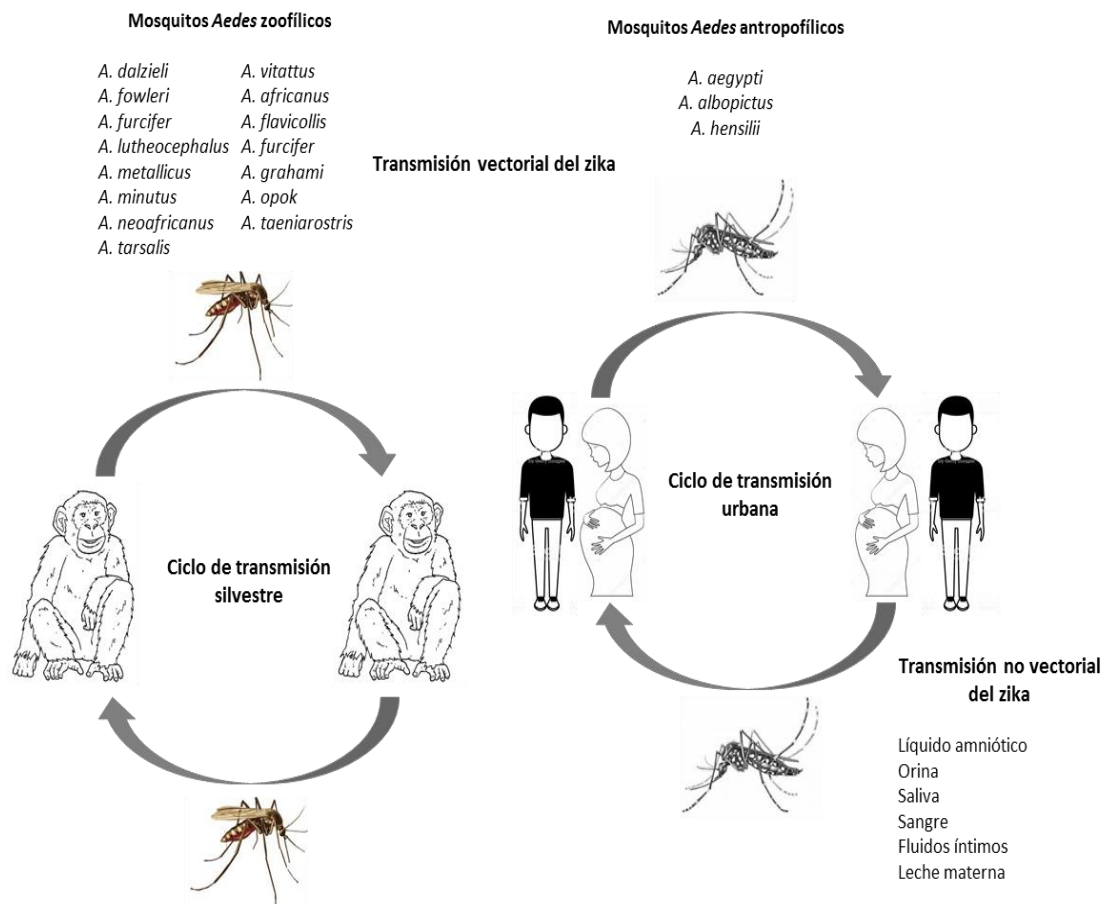


Figura 4.3. Ciclos de transmisión del virus de zika (ZIKV). Fuente: Saavedra, L.M.

En consecuencia, existe una amplia necesidad de proponer nuevas estrategias de vigilancia entomológica, control de vectores, evaluación del impacto y virulencia del ZIKV, como base fundamental de las acciones en materia de salud pública global, destinadas a prevenir y limitar el impacto de este virus y sus repercusiones congénitas en las poblaciones más vulnerables del planeta.

4.4. Principales vectores del género *Aedes*

Entre los artrópodos más abundantes se encuentran los insectos de orden Diptera, familia Culicidae y subfamilia Culicinae. Esta subfamilia contiene 3.050 especies pertenecientes a 109 géneros. Entre estos géneros, se destaca el *Aedes*, *Culex*, *Mansonia*, *Haemagogus*, *Sabethes* y *Psorophora*, por transmitir enfermedades a los humanos.⁴⁷ En particular, los insectos del género *Aedes* se encuentran divididos en cuatro subgéneros: *Aegypti*, *Stegomyia*, *Aedimorphus* y *Fredwardsius*, que abarcan alrededor de 154 especies de vectores.

Los mosquitos *A. aegypti* y *A. albopictus*, comúnmente conocidos como “mosquito de la fiebre amarilla” y “mosquito tigre Asiático”, constituyen una amenaza para la salud pública mundial, debido a su capacidad de transmitir múltiples virus a los humanos. Particularmente, estos vectores se caracterizan por su extensa distribución en las regiones tropicales y subtropicales del planeta. Aunque, solo el *A. albopictus* se localiza en Asia y parte de Europa; en cambio, el *A. aegypti* habita las regiones selváticas de África y América.⁴⁸⁻⁵⁰

Otro elemento característico de los vectores *A. aegypti* y *A. albopictus* radica en que ambas especies pican en exteriores, pero solo el *A. aegypti* pica en interiores, comúnmente donde habitan los humanos. El mosquito *A. aegypti* se asemeja en tamaño, color y morfología al *A. albopictus*; pero difieren en su capacidad de distribución, reproducción, alimentación y transmisión de los virus.

La descripción anterior fue necesaria para delimitar los estudios QSAR abordados en la presente Tesis Doctoral, orientados en establecer modelos de predicción de propiedades biológicas para el control del mosquito *A. aegypti* L.; responsable de transmitir múltiples patógenos asociados al aumento de los casos de morbilidad y mortalidad notificados en el hemisferio occidental durante los últimos años.

4.5. El mosquito *Aedes aegypti* Linnaeus (Diptera: Culicidae)

El mosquito *Aedes aegypti* Linnaeus (Diptera: Culicidae), es el primer transmisor de múltiples arbovirus que afectan la salud humana. Para el *A. aegypti*, el intercambio comercial proveniente de África, facilitó su expansión territorial, la interacción con los humanos y el surgimiento de grandes epidemias.

4.5.1. Taxonomía

El constante incremento de la población humana, su migración a zonas silvestres y la continua evolución del comportamiento de los mosquitos, representan algunos de los fenómenos claves en la formación de subpoblaciones domésticas de mosquitos.⁵¹ La especie *Aedes aegypti* fue clasificada por Carl Nilsson Linnæus en 1758 dentro de la clase Insecto y el orden Diptera. Para 1818, el entomólogo Alemán Johann Meigen vinculó al *A. aegypti* dentro del género *Aedes*; y en 1829, Gabrius Stephens catalogó a esta especie en la familia Culicidae. Para el *Aedes aegypti* existen dos subespecies:

- La primera subespecie es el *Aedes aegypti formosus*, conocido como el ancestro del *Aedes aegypti* doméstico. Esta subespecie todavía vive en los bosques y las pequeñas vegetaciones del sub-Sahara Africano. Suele alojarse en ambientes húmedos naturales y presenta actividad hematófaga sobre los animales salvajes propios de su habitat.
- La segunda subespecie es el *Aedes aegypti aegypti*, localizado principalmente en las regiones tropicales y subtropicales asociadas con la actividad humana.

Este mosquito habita los contenedores artificiales de agua situados en interiores domésticos, y así favorece su actividad hematófaga basada en la sangre humana como principal recurso de proteína.

4.5.2. Morfología

Las características del *A. aegypti* describen los diferentes estadios que presenta durante todo su ciclo de vida. El ciclo de vida de este vector inicia con el huevo, que al eclosionar se convierte en una larva acuática (de I-IV estadios), luego se transforma en pupa (dimorfismo sexual), para finalmente emerger como mosquito adulto (macho o hembra).⁵² Cada etapa del ciclo de vida del mosquito *A. aegypti* se detalla a continuación:

A. *Huevos*: presentan forma ovalada, son lisos, largos y con un tamaño de aproximadamente 1 milímetro. Se tornan de color claro en el momento de la ovoposición, y 2 horas después producto de la melanización se tornan oscuros.

Los huevos son depositados individualmente por la hembra adulta en un contenedor de agua, ya sea natural o artificial; cada huevo es ubicado aleatoriamente sobre una superficie en contacto con el agua, para omitir variables ambientales que puedan interferir en el proceso de eclosión. En la especie *A. aegypti*, los huevos sobreviven en estado de desecación y conservan sus propiedades por varios meses; de modo que, pueden eclosionar normalmente al entrar en contacto con agua.

B. *Larva*: el *A. aegypti* en etapa larval tiene una morfología similar respecto a otras especies larvas. En general, presentan cabeza ovalada, un tórax y un abdomen compuesto por nueve segmentos, como lo exhibe la Figura 4.4.

El segmento perineal de la larva contiene branquias de cuatro lóbulos que le proporcionan la regulación osmótica, y un sifón corto en forma de barril con un solo par de crestas subventrales que le permiten respirar en la superficie del agua.^{52, 53}



Figura 4.4. Vista dorsal del *A. aegypti* en etapa larval. Fuente: García, *et al.*, 2009.⁵⁴

Otros elementos morfológicos particulares en la larva *A. aegypti*, son al menos tres pares de setas en el cepillo ventral, antenas curvadas y ausencia de grandes setas en el tórax. Estos atributos facilitan la identificación de la especie, ya que son propios del género *Aedes*.

La posición de reposo de las larvas en el agua es una conducta propia de cada especie. En el caso de la especie *Anopheles*, se ubican paralelas a la superficie del agua; mientras que

en la especie *Culex* se orientan en diagonal a la superficie. Las larvas *Aedes* se orientan perpendiculares a la superficie del agua, como se exhibe en la Figura 4.5.

Otras características morfológicas que facilitan la identificación de las larvas *A. aegypti* radican en dos espinas laterales en cada lado del tórax, una fila recta de 7 a 12 escamas en el octavo segmento abdominal y una espina media rodeada de varias espinas robustas, ausentes en otras especies de larvas *Aedes*.

Las larvas *A. aegypti* atraviesan de I-IV estadios, conocidos como *instar*; cada *instar* involucra cambios de tamaño, forma, color y movilidad, que culminan en el cambio de larva a pupa.

C. Pupa: la pupa es la etapa del mosquito que antecede el último estadio larvario y precede la etapa adulta. Una pupa, presenta forma de coma y se compone de dos secciones principales, el cefalotórax, es decir cabeza y tórax fusionados, y el abdomen.⁵² En la base del cefalotórax se encuentran ubicados los tubos respiratorios con forma de trompeta, los cuales perforan la superficie del agua para captar oxígeno.



Figura 4.5. Larvas *Aedes* (A), *Anopheles* (B) y *Culex* (C). Fuente: García *et al.*, 2009.⁵⁴

En la etapa de pupa, el mosquito define su género sexual. Las pupas hembras suelen presentar mayor tamaño que el macho. Aunque, esta característica no es constante, y puede variar acorde a los diferentes factores bióticos y abióticos interactuantes.

Por ello, la forma más común para determinar el género sexual es mediante la observación del extremo final del abdomen; allí, la pupa cuenta con un par de aletas para nadar, que para las hembras son anchas y se superponen, y para los machos son estrechas y separadas.^{55, 56}

D. Adulto: El cuerpo del mosquito *Aedes aegypti* en la etapa adulta, está compuesto por cabeza, tórax y abdomen. Su comportamiento alimentario se basa en los recursos naturales de carbohidratos, tal como el néctar de las flores o los frutos

Las hembras y los machos de esta especie tienen una apariencia similar. Aunque, difieren en algunas características morfológicas. El macho presenta menor tamaño que la hembra, tiene antenas plumosas y un par de palpos maxilares con bandas blancas más desarrollados que la hembra. Aunque, la hembra presenta una probóscide más corta que el macho, adaptada principalmente para traspasar la piel de un mamífero y obtener la sangre.⁵⁵⁻⁵⁷

4.5.3. Ciclo de vida

Todas las especies de mosquitos tienen un ciclo de vida complejo, basado en drásticos cambios de forma, función y hábitat. Los mosquitos de orden Diptera, presentan cuatro etapas generales de metamorfosis en su ciclo de vida, exhibidas en la Figura 4.6.

Las hembras *A. aegypti* son hematófagas, lo que indica que están adaptadas para abastecerse con sangre proveniente de animales vertebrados. Este comportamiento se evidencia únicamente en la etapa de reproducción, donde requiere la ingesta de sangre para madurar los huevos e iniciar el proceso de ovoposición. Por lo general, una hembra de *Aedes aegypti* L. tiene la capacidad de producir entre 100 y 200 huevos.

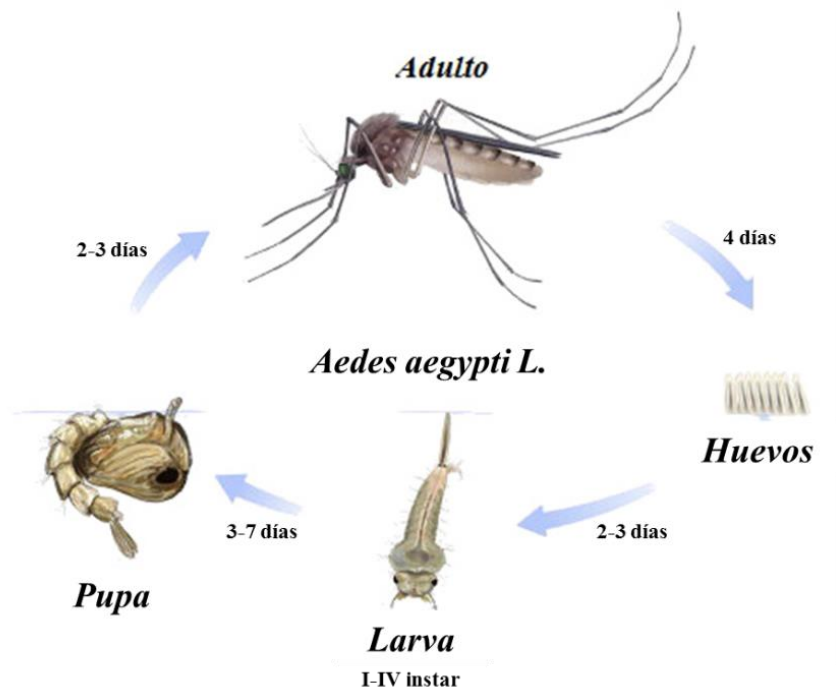


Figura 4.6. Ciclo de vida del mosquito *Aedes aegypti* L. Fuente: Saavedra, L.M.

El ciclo de vida del mosquito *A. aegypti* inicia con la ovoposición de una hembra adulta, encargada de instalar los huevos sobre una superficie húmeda en contacto intermitente con el agua; pero solo 2 o 3 días después de la ovoposición se realiza la

embriogénesis.⁵⁸ La velocidad de eclosión de cada lote de huevos depende de la temperatura ambiente, la humedad relativa y el tiempo de desecación.⁵⁹

No obstante, los huevos instalados y secos pueden sobrevivir hasta un año sin eclosionar. Este particular proceso de conservación ha facilitado la amplia proliferación del vector *A. aegypti* L. y las enfermedades transmitidas por el mismo.

La etapa larval, constituye el principal período de alimentación y crecimiento de las etapas inmaduras. Las larvas se alimentan del material orgánico presente en el medio acuático que habitan.⁶⁰ El tiempo de desarrollo de una larva, puede variar según la disponibilidad de alimento, la densidad poblacional y la temperatura del agua. En condiciones óptimas, el desarrollo completo tarda 1 semana.⁶¹ Después de esto, las larvas adquieren suficiente energía y tamaño para convertirse en pupas.

Una pupa representa la última etapa inmadura del ciclo de vida del mosquito. Esta etapa tiene una duración de 2 a 3 días, aunque puede alargarse o acortarse según las condiciones del medio; por lo general, los machos suelen desarrollarse más rápido que las hembras. A diferencia de una larva, altamente susceptible al entorno, la mortalidad de las pupas es casi nula. De manera que, la cantidad de pupas es directamente proporcional a la cantidad de potenciales vectores de enfermedades.⁵³

Las primeras 24 horas del mosquito adulto o imago son vitales para su supervivencia. Durante este período, se consolida su estado fisiológico, endurece el exoesqueleto y madura sexualmente. Comúnmente, el imago tiene una actividad crepuscular y su expectativa de vida varía según el género. Las hembras suelen vivir entre 10 y 35 días, por el contrario los machos solo sobreviven 4 o 6 días.

La etapa adulta de un mosquito es trascendental para la supervivencia de la especie; ya que debe garantizar la distribución, alimentación, reproducción e iniciación del nuevo ciclo de vida.⁶²⁻⁶⁴

4.5.4. *Comportamiento alimentario*

Si bien el comportamiento alimentario de los machos y las hembras *A. aegypti* L. es igual, solo la hembra es hematófaga.^{65, 66} A pesar de su habilidad para sobrevivir con soluciones azucaradas provenientes de las flores o frutas, la hembra requiere la ingesta de sangre para iniciar el proceso de maduración de los huevos en cada ciclo de desarrollo ovárico, denominados ciclos gonotróficos.⁶⁷

Cada ciclo gonotrófico requiere el consumo de sangre, es decir, cada hembra *A. aegypti* L. realiza múltiples picaduras, con una ingesta aproximada de 2,5 µl de sangre extraída de distinto individuos, preferiblemente humanos. Este comportamiento resulta trascendental para el diagnóstico epidemiológico, ya que intensifica la transmisión de virus a los humanos.⁶⁸ Las hembras *A. aegypti* L. suelen picar durante

todo el período diurno, intensificando la actividad hematófaga durante el amanecer y el atardecer.

Los mosquitos tienen la habilidad de localizar al huésped mediante factores visuales, olfativos y térmicos; aunque, el olfato predomina en este propósito. El vector es guiado a través de las emisiones de dióxido de carbono (CO₂) o ácido láctico generadas por el huésped a distancias variables. Por lo general, el mosquito realiza la picadura en cualquier parte del cuerpo; pero las manos y las piernas al concentrar mayor sudoración son más vulnerables a los ataques por picaduras. Incluso, la frecuencia de picadura se puede intensificar o minimizar acorde con los componentes del sudor corporal del individuo objetivo.⁶⁹

Desde la picadura del huésped, la sangre adquirida permanece entre 38 y 48 horas en el intestino medio de la hembra. Posterior a esto, la proteína proveniente de la sangre, provee los aminoácidos necesarios para la síntesis de vitelogenina, la proteína precursora de los huevos.⁷⁰⁻⁷²

También, se ha comprobado que la especie *A. aegypti* L. puede realizar una transmisión transovárica de los virus para garantizar la subsistencia del patógeno en la naturaleza. Este comportamiento causa la permanencia del virus en las siguientes siete generaciones del mosquito, o en los huevos remanentes del período de eclosión.⁷³

4.5.5. *Sitios de ovoposición*

Para los mosquitos, la ubicación de un lugar adecuado para depositar los huevos requiere un conjunto de señales visuales, táctiles y olfativas que gobiernan a la hembra antes iniciar la ovoposición. La habilidad de la hembra para encontrar un sitio apropiado resulta transcendental en el crecimiento, desarrollo y supervivencia de sus descendientes.^{73, 74}

Las hembras *A. aegypti* fijan los sitios de ovoposición en contenedores artificiales de agua limpia, oscuros y húmedos. También, debido a su comportamiento endófilo y antropófilo, esta especie prefiere realizar la ovoposición en ambientes urbanos o suburbanos cerca de la actividad humana, los cuales proveen elementos vitales para el desarrollo de la siguiente generación.

Entre las condiciones de reproducción del mosquito *A. aegypti* L., involucra un recipiente de almacenamiento de agua limpio, estático y sellado. Por lo general, los mosquito *A. aegypti* L. fijan el inicio del ciclo de vida en los recipientes de uso doméstico, como jarras, macetas, piscinas plásticas, gomas de las rudas, bebederos de animales, canaletas obstruidas o trampas de desagüe. Aunque, todos los contenedores de agua son potenciales criaderos para las etapas inmaduras de esta especie.⁷⁵⁻⁷⁷

4.6 Interacciones vector-virus

En el ciclo de transmisión de los arbovirus los artrópodos son infectados cuando ingieren sangre de un huésped virémico. El ciclo de los arbovirus requiere la replicación celular del vector. El Período de Incubación Extrínseco (EIP), comprende el tiempo que transcurre entre la ingesta de sangre infectada hasta la presencia de partículas virales en la secreción salival del vector portador.⁷⁸

Después del EIP, el insecto vector tiene la capacidad de transmitir el virus a huéspedes vertebrados sanos. El EIP es un factor epidemiológico importante, ya que la vida útil del vector se relaciona con este factor; asimismo, la propagación del arbovirus dependerá que la longevidad del vector supere el EIP.

La transmisión de los arbovirus al vector susceptible se realiza a través de las partículas virales; éstas deben adherirse a los receptores celulares de las células diana en el epitelio del intestino medio del insecto para fijar la infección, tal como lo exhibe la Figura 4.7.

Los viriones deben ingresar a las células epiteliales por medio de las microvilas antes de que el alimento sanguíneo sea rodeado por la matriz peritrófica. De lo contrario, el virus no podrá infectar el intestino medio del insecto, debido a que el tamaño del poro de la matriz peritrófica es más pequeño (20-30 nm) que todos los arbovirus.⁷⁹⁻⁸¹

En el mosquito *A. aegypti* L. la matriz peritrófica surge entre 4-8 horas después de la ingesta de sangre, y logra una textura madura a las 12 horas. Los patrones de infección de las células epiteliales del intestino medio varían según la relación entre el virus y la especie de mosquito. Hacia una infección productiva del mosquito se requiere la ingesta de suficiente virus para infectar el intestino medio del insecto y garantizar el ciclo de transmisión a otros huéspedes vertebrados.

A medida que el virus se introduce en las células epiteliales del intestino medio, inicia el proceso de replicación. Los viriones deben atravesar la lámina basal del epitelio del intestino medio para ingresar al hemocoel, la cavidad del cuerpo de los artrópodos que contiene los órganos, los músculos y el sistema circulatorio abierto de fluido hemolinfa. Una vez que los viriones del intestino medio atraviesan el hemocoel, suelen diseminarse dentro del tejido nervioso y los órganos secundarios del insecto.⁸²⁻⁸⁵

Para un vector susceptible, la infección culmina cuando las partículas víricas son instaladas en las glándulas salivales, ubicadas en la parte lateral del tórax del insecto. Cada glándula salival se divide en tres lóbulos, dos laterales y uno medio, conectados simultáneamente al ducto salival principal.

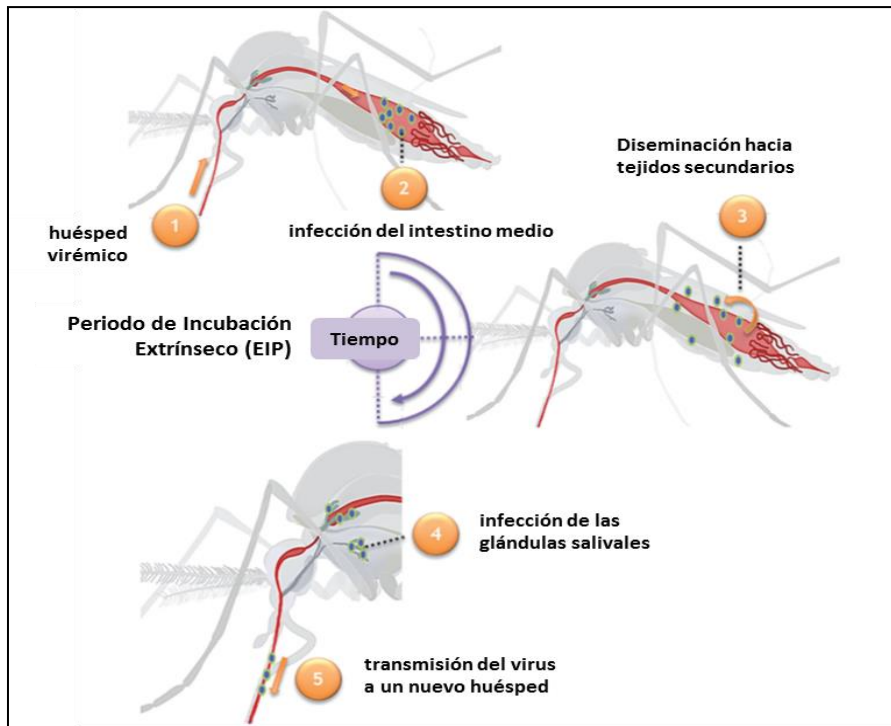


Figura 4.7. Interacción infecciosa virus–mosquito susceptible. Fuente: Saavedra, L.M.

La infección del virus generalmente acontece en los lóbulos laterales de las glándulas salivales. Aunque, para los virus DENV y CHIKV el proceso de infección consecutivo termina en los tres lóbulos de las glándulas salivales del *A. aegypti* L.⁸⁵⁻⁸⁷

Al ocurrir la replicación del virus en el organismo vector, se liberan viriones en los conductos salivales para iniciar la transmisión horizontal a un nuevo huésped sano. La saliva de los artrópodos facilita la transmisión y regula la respuesta del huésped a la replicación del virus, inyectando múltiples mezclas complejas de proteínas peptídicas, tales como glicosidasas, antimicrobianos, antihemostáticos, proteínas con potencial antiinflamatorio y moduladores inmunológicos.

4.7. Transmisión de los arbovirus a través del mosquito

El Período de Incubación Extrínseco (EIP) es el principal indicador del potencial de transmisión y propagación de un virus. Este período depende de los atributos genéticos del virus, como la patogenicidad del virus y la cantidad de virus ingerido por el vector. Asimismo, el EIP obedece tanto a los factores ambientales (temperatura y humedad) como a los factores intrínsecos de competencia del vector y el genotipo viral involucrado.^{76, 82-86}

La transmisión biológica de un arbovirus a un insecto susceptible, implica superar una serie de barreras físicas y fisiológicas para lograr la infección del organismo en una sola toma de sangre. Existen barreras en el insecto capaces de evitar que el virus se replique y propague a las glándulas salivales.

Estas barreras son definidas como inmunidad innata, Barrera de Infección del Intestino Medio (MIB), Barrera de Escape del Intestino Medio (MEB), Barrera de Infección de las Glándulas Salivales (SGIB) y Barrera de Escape de las Glándulas Salivales (SGEB).^{88, 89}

Las variaciones genéticas intraespecíficas en las poblaciones de mosquitos influyen sobre el funcionamiento de varias de las barreras mencionadas anteriormente, previniendo o permitiendo la replicación y propagación del virus. Dichas variaciones genéticas, propias entre las poblaciones de mosquitos, contribuyen significativamente sobre el potencial de transmisión de los arbovirus, el EIP y el poder epidemiológico contra los humanos.

Entre los diversos arbovirus transmitidos por el mosquito *A. aegypti* L., los cuatro serotipos del DENV, presentan diferente intensidad de réplica y poder de propagación. Además, se ha reportado que las glándulas salivales de una misma población de vectores, exhiben amplias variaciones en la competencia del mosquito por transmitir el dengue. El *A. aegypti* L., después realizar la ingesta de sangre de un individuo virémico, presenta un EIP de 7 a 14 días antes de poder transmitir el virus a un nuevo huésped susceptible.^{78, 82, 88}

Por otra parte, la transmisión de los virus también involucra numerosos factores entomológicos del mosquito, fuertemente asociados a la competencia entre vectores. Entre estos factores, sobresalen la capacidad y regularidad de picadura, la densidad poblacional, la habilidad de supervivencia y desplazamiento durante la vida adulta.^{85, 89}

4.8 Estrategias actuales en el control de vectores

Los mosquitos son reconocidos como una amenaza para la salud pública mundial por ser los principales transmisores de diversas enfermedades a los humanos. A raíz de esto, los objetivos del control vectorial se basan en prevenir la picadura del mosquito, mantener la densidad poblacional en niveles normales, evitar la interacción mosquito-huésped y reducir la longevidad de las hembras hematófagas.⁹⁰

El control de enfermedades endémicas se refiere a reducir o eliminar los casos de incidencia, prevalencia, morbilidad y mortalidad generados por las mismas. Bajo este contexto, un efectivo control vectorial proporciona estrategias con capacidad de limitar o erradicar el insecto transmisor de patógenos, y evadir la resistencia a los insecticidas de uso comercial.⁹¹

Los controles actuales sobre el mosquito *A. aegypti* L., se basan en planes integrales de vigilancia entomológica y erradicación de vectores, consensuados por cada país bajo amenaza de contagio.⁹² Por ello, es importante comprender las características del entorno donde ocurren los brotes epidémicos para emplear con éxito las estrategias de control existentes, detalladas en la Tabla 4.1.

Tabla 4.1. Estrategias de control vectorial sobre *A. aegypti* L. Fuente: WHO, 2007.⁹²

Método	Descripción	Ejemplos
Control Químico	Estados inmaduros	Tratamientos en los sitios de reproducción con larvicidas como temefos Reguladores del Crecimiento (IGR) como metopreno, piriproxifeno y spinosad
	Adultos	Fumigaciones aéreas y tratamientos superficiales de las zonas afectadas
	Protección Personal	Repelentes naturales y sintéticos, pinturas y materiales tratados con insecticidas
Control Biológico	Estados inmaduros y adultos (el total de la población)	Insectos: peces, libélulas, copépodos y <i>Toxorhynchites</i> Bacterias: <i>Wolbachia</i>
Control Genético (autocontrol)	Estados inmaduros y adultos (el total de la población)	Limitación propia del insecto y técnicas de esterilización
Control Genético (reemplazo de la población)	Modificaciones genéticas en la población del vector	Sistema de control genético basados en <i>Homing Endonuclease Genes (HEG)</i> y CRISPR, <i>Wolbachia</i>
Administración y control de ambientes	Modificación: permanente transformación en algunas características propias del hábitad de reproducción Manipulación: Cambios temporales que afectan el comportamientos de los sitios de reproducción Cambios estructurales tanto en el hábitad como en el comportamiento de los humanos	Airear, limpiar, reciclar, reducir, desechar los sitios de reproducción vectorial y potencial hábitad de larvas. Realizar una apropiada disposición de los residuos sólidos orgánicos y reciclables Limpieza permanente de las red de alcantarillado Campañas públicas para eliminar los potenciales criaderos urbanos del vector Instalación de mosquiteros tratados con insecticida en interiores y accesos

4.8.1. Control biológico

El control biológico se encarga de introducir organismos que compiten, eliminan o contaminan la población objetivo, con el fin de reducir su población. En este contexto, el *Bacillus thuringiensis* var. *israelensis* (*Bti*), es una bacteria entomopatógena disponible comercialmente para atacar los sitios de reproducción y ovoposición del *A. aegypti* L.^{93,94} La cepa de *B. thuringiensis* AM65-52, está avalada por el Esquema de Evaluación de Pesticidas (WHOPES) de la OMS, para tratar los contenedores de agua domésticos o naturales que contengan larvas *A. aegypti* L.

Entre los organismos acuáticos más utilizados en el control de larvas *A. aegypti* L., se destacan los peces *Poecilia reticulata*, conocidos como “Peces Guppy”, y los crustáceos copépodos de orden Cyclopoidea.⁹⁵⁻⁹⁷ Ambas especies reducen o eliminan las larvas presentes en pequeños contenedores de agua. Sin embargo, el pez Guppy también funciona en medios acuáticos de gran volumen.

Por otra parte, el uso de la bacteria *Wolbachia* surgió como una estrategia biológica para reducir o modificar la población del mosquito *A. aegypti* L. La estrategia consiste en liberar machos *A. aegypti* L. infectados con *Wolbachia*, para iniciar el proceso de reproducción con hembras silvestres y producir huevos infértiles; lo que conduce a la inminente interrupción del ciclo de vida del vector.

No obstante, dicha estrategia involucra algunos riesgos, tal como una infección cruzada con *Wolbachia* en hembra *A. aegypti* L. silvestres, con una capacidad de transmisión maternal del 100%, favoreciendo la proliferación de *Wolbachia* en el ambiente.⁹⁸

Si bien la infección con la bacteria *Wolbachia* ocurre de manera natural en diversas especies de mosquitos, como el *Culex pipiens*, *Culex quinquefasciatus*, *Aedes fluviatilis* y *Aedes albopictus*. En el caso del *A. aegypti* L., se requiere un procedimiento de infección artificial de la bacteria, ya sea mediante micro inyección embrionaria del citoplasma infectado con *Wolbachia* o la extracción y purificación de *Wolbachia* a partir de hospedadores naturales de la bacteria, como el insecto *Drosophila melanogaster*.⁹⁹⁻¹⁰¹

Emplear *Wolbachia* como estrategia de reducción poblacional del *A. aegypti* L., involucra el uso de distintas cepas de la bacteria con diversos modos de acción. La cepa denominada wMe1Pop, fue introducida en las poblaciones silvestres del mosquito para acortar el tiempo de vida útil del adulto, minimizando su longevidad hasta el EIP de los virus.¹⁰²

Por otro lado, las cepas wMe1 y wMe1Pop-CLA (línea celular adaptada), han demostrado conferir propiedades antivirales al vector *A. aegypti* L.; ya que reduce la capacidad de diseminación del virus desde el intestino medio hasta las glándulas

salivales del mosquito, reduciendo el potencial de trasmisión de los flavivirus, particularmente del DENV.

Con base en lo anterior, se puede resaltar la habilidad de las cepas de *Wolbachia* para minimizar la vida útil del *A. aegypti* L, invadir la población del mismo a través de una incompatibilidad citoplasmática e interferir con la replicación y diseminación de los patógenos que afectan la salud humana.

4.8.2. Control genético

El control genético de vectores realiza estudios basados en técnicas clásicas de esterilización de insectos o de autolimitación de la especie. La tecnología que busca esterilizar insectos es utilizada por países como México, Sri Lanka, Malasia, Indonesia y Tailandia. Esta técnica se enfoca en la producción masiva de mosquitos machos de la especie *A. aegypti* L., esterilizados por rayos X o irradiación gamma.

Particularmente, las medidas de control genético han beneficiado al sector agrícola y productor de alimentos, con problemas de plagas y pestes cada vez más complejos.¹⁰³ Para el control del vector *A. aegypti* L. no existen registros de su eficacia.

Una novedosa estrategia de control genético de insectos es la autolimitación de la especie, mediante mosquitos alterados genéticamente para transferir a sus descendientes un gen que ocasiona la muerte, antes de llegar a la etapa adulta. Esta técnica fue desarrollada para el control específico del *A. aegypti* L., basándose en la liberación de mosquitos diseñados para iniciar el proceso de reproducción con mosquitos silvestres localizados en áreas endémicas y procrear generaciones con baja longevidad y mínima persistencia en los ecosistemas.

Una de las técnicas de autolimitación más conocida, es la línea transgénica de mosquitos OX513A y sus descendientes, caracterizados por un marcador fluorescente DsRed2, que permite identificar las larvas y pupas OX513A en condiciones de laboratorio.¹⁰⁴

El despliegue de vectores OX513A autolimitados, condujo a obtener resultados favorables en la reducción poblacional del *A. aegypti* L., particularmente en países bajo amenaza de brotes endémicos, tal como Brasil, las Islas Caimán y Panamá. Esta técnica es ampliamente aceptada por los organismos reguladores de salud pública de la Unión Europea y Estados Unidos.¹⁰⁵⁻¹⁰⁷

4.8.3. Control físico o del entorno

Todos los enfoques propuestos para limitar la interacción huésped – vector, minimizar la densidad poblacional del mosquito o prevenir su proliferación en nuevos entornos, son conocidos como métodos de control físico y ambiental. Estos métodos están destinados a reducir, reutilizar, modificar o eliminar los contenedores artificiales

de agua presentes en zonas urbanas o suburbanas, que actúan como criaderos de las etapas inmaduras del *A. aegypti* L.

Todas las estrategias de vigilancia física y ambiental contra el *A. aegypti* L., forman parte de los enfoques estratégicos orientados al control de vectores.¹⁰⁸⁻¹¹⁰ En este contexto, se determinaron tres controles ambientales sobre mosquitos, detallados a continuación:

- *Modificación ambiental*: Consiste en modificar los espacios físicos de larga duración para reducir la densidad poblacional del vector. Algunas de estas modificaciones se enfocan en eliminar los contenedores artesanales de agua, instalar redes hídricas apropiadas y aptas para el consumo humano, y diseñar sistemas de alcantarillado aptos para captar escorrentías superficiales intensas.
- *Manipulación ambiental*: puntualiza sobre los cambios temporales del hábitat que aloja al mosquito en cada etapa de su ciclo de vida. En este caso, se alteran los contenedores artificiales de agua, ya sea por limpieza o vaciado de los recipientes de uso doméstico.

Además, se realiza un proceso de remoción de residuos sólidos o plantas ornamentales que puedan almacenar agua. Con esta medida se logró disminuir los casos de dengue (DENV) en las áreas endémicas.^{22, 27}

- *Cambios en el comportamiento humano*: evade el contacto huésped-vector, mediante la instalación de trampas para mosquitos en los espacios domésticos.

4.8.4. Control químico de las etapas inmaduras del mosquito

El control químico de las etapas inmaduras busca tratar los contenedores naturales o artificiales de agua dulce habitados por larvas y pupas con productos químicos de acción larvicida y pupicida. Las sustancias larvicidas o pupicidas difieren en su modo de acción y efecto residual.

Los compuestos de síntesis como diflubenzuron, novaluron, piriproxifeno, fention, pirimifos metilo, spinosad y temefos, son comúnmente empleados en las fuentes hídricas de uso doméstico para erradicar prematuramente el mosquito *A. aegypti* L.¹³⁻¹⁵

El uso de larvicidas implica realizar fumigaciones terrestres y aéreas. Sin embargo, las larvas de *A. aegypti* L. suelen habitar en contenedores reducidos (menores a 200 mL); de modo que, tratar de manera razonable estos criaderos altamente diseminados es un paradigma dentro de las estrategias globales de control de vectores.

Una novedosa alternativa antivectorial involucra la liberación de mosquitos machos estériles, provenientes de criaderos controlados e impregnados con piriproxifeno, para contaminar a las hembras silvestres por transferencia venérea.¹¹¹

Esta estrategia supero las pruebas de campo a pequeña escala sobre el mosquito *A. albopictus*, promoviendo la apertura de un nuevo eje de investigación orientado al control del *A. aegypti* L.

4.8.5. *Control químico del mosquito adulto*

En el marco del control vectorial, se busca abordar la vigilancia y eliminación de los mosquitos adultos por acción de los insecticidas, ya sea por el tratamiento de superficies o el tratamiento espacial (aerosoles de bajo volumen o nebulizadores térmicos). En ambos casos, los resultados esperados implican la disminución de la población, longevidad y capacidad de dispersión.¹¹²

Entre los insecticidas más utilizados contra vectores, se destacan por su efectividad los piretroides, los organofosforados y los carbamatos. Estos compuestos de síntesis están recomendados por WHOPES, para el tratamiento espacial de interiores domésticos y exteriores urbanos.⁹⁵ En el caso de las áreas con brotes epidémicos, un adecuado tratamiento vectorial involucra fumigaciones masivas eficientes, específicas y de efecto residual prolongado.

Uno de los tratamientos más implementados en áreas con brote víricos, consiste en fumigar los interiores residenciales con sustancias de acción residual, focalizada en lugares oscuros y sombríos, usualmente habitados por el vector.¹¹³ Por lo general, esta técnica emplea pequeñas dosis de agentes químicos de potencial toxico prolongado, con una aplicación de una o dos veces por año. Otro tipo de tratamiento vectorial es la fumigación inmediata de interiores a través de una niebla insecticida con efecto residual a corto plazo, reduciendo su utilidad.¹¹⁴

Otra forma de tratar los exteriores urbanos es la fumigación focalizada. En este caso, el insecticida se distribuye como un aerosol emitido por un vehículo que circula por los accesos terrestres de las regiones infestadas de mosquitos. De este modo, los países con proliferación indiscriminada de vectores han minimizado la densidad poblacional del mismo, logrando tasas sostenibles.

Por otra parte, el uso de compuestos químicos de aplicación espacial o tópica, con capacidad de evitar la interacción entre el huésped y el vector, representa la estrategia más utilizada a nivel global. Entre estas sustancias, prevalecen los insecticidas domésticos, repelentes naturales o sintéticos, materiales impregnados con insecticida o pinturas elaboradas con sustancias de acción repelente, como el piretroide denominado metoflutrin.¹¹⁵

En general, los piretroides son los principales ingredientes activos de los repelentes corporales e insecticidas domésticos de venta libre, por su eficacia sobre los mosquitos y baja toxicidad en mamíferos.

Actualmente, existen numerosos repelentes comerciales con una amplia variedad de formulaciones y modos de aplicación, como aerosoles o ungüentos que se

aplican en la ropa o en la piel. Estos compuestos contienen ingredientes activos basados en productos naturales o sintéticos. Entre los repelentes de venta libre más utilizados, sobresalen aquellos que tienen como ingrediente activo DEET (*N, N*-dietil-3-metilbenzamida), IR3535 (3- [N-acetil-N-butil] – éster etílico del ácido aminopropiónico) o Icaridina (ácido 1-piperidincarboxílico 2-(2- hidroxietil)-1-metilpropilester).²³

Otra de las estrategias antivectoriales consiste en instalar trampas tratadas con insecticidas de acción tóxica selectiva, tal como los compuestos piretroides. Esta estrategia se caracteriza por brindar una acción repelente de 1 a 3 años.

El despliegue de las trampas para mosquitos impregnadas con agentes tóxicos logró reducir considerablemente las enfermedades transmitidas por el *A. aegypti* L.; asimismo, minimizó su longevidad y capacidad de propagación. Aunque, estos artefactos de control presentan diversas limitaciones, ya que protegen únicamente áreas específicas de las viviendas.

4.9 El desarrollo de la resistencia a los insecticidas

Acorde a la OMS, la resistencia a los insecticidas se define como la capacidad de un insecto para resistir los efectos tóxicos producidos por un agente químico, ya sea por mutación o selección natural.^{116, 117} El estudio de la resistencia a los insecticidas requiere diferentes herramientas biológicas, bioquímicas y moleculares para identificar los complejos mecanismos que envuelven el desarrollo de la resistencia en el insecto, tanto a nivel individual como poblacional.

A través de diversos estudios científicos enfocados sobre la resistencia, fue posible establecer diferentes modos de resistencia en más de 500 especies de insectos alrededor del mundo. De este modo, se comprobó que al menos 20 especies de mosquitos del género *Aedes* (Diptera: Culicidae) presentaban algún tipo de resistencia a los insecticidas, atribuida principalmente a la aplicación frecuente y excesiva de un mismo conjunto de sustancias químicas sobre una determinada población de vectores.¹¹⁸

Frecuentemente, los casos de resistencia en insectos se producen sobre una clase química específica, que por lo general, es el ingrediente activo del producto. Es decir, los efectos de resistencia suelen manifestarse sobre todos los productos químicos comerciales basados en una misma clase química; este fenómeno se conoce como resistencia cruzada.^{119, 120}

Si bien la resistencia es un fenómeno que depende de la naturaleza del insecto y su entorno, diversos factores antropogénicos generados por los agroquímicos empleados en el tratamiento de los cultivos o en los programas de control de enfermedades, han afectado de manera indirecta la susceptibilidad de los mismos, provocando otro modo de resistencia, denominada resistencia múltiple. Este modo de

resistencia radica en la habilidad de los insectos para soportar los efectos de tóxicos producidos por diferentes clases químicas de compuestos, que poseen mecanismos de acción similares, como es el caso de los organofosforados y los carbamatos.

Hasta ahora, la cantidad de productos químicos seguros y eficaces para el tratamiento, control o eliminación de los vectores es limitada y con bajas expectativas de aumento en un futuro cercano.¹⁰⁹ Es importante reconocer que el desarrollo de la resistencia a los insecticidas de uso común constituye una amenaza para combatir la proliferación de virus transmitidos por insectos.¹²¹

En consecuencia, existe una profunda necesidad por modificar los insecticidas actuales o diseñar nuevos recursos químicos racional y selectivamente; de tal forma que, sea posible eludir los problemas de resistencia actuales. Para esto, resulta vital promover las investigaciones enfocadas en identificar los mecanismos y modos de acción de los diferentes casos de resistencia en insectos reportados hasta ahora; en particular, acerca de las diversas formas de resistencia desarrollada por el principal vector de los virus DENV, CHIKV y ZIKV, el mosquito *A. aegypti* L.¹⁻⁵

4.10. Modos de acción y mecanismo de resistencia a los insecticidas

A nivel general, los programas de vigilancia y control de mosquitos se enfocan en intervenir todas las etapas del ciclo de vida del vector. Las acciones de monitoreo y control de las etapas inmaduras del *A. aegypti* L., se concentran en aplicar de manera conjunta o individual, diferentes herramientas biológicas, químicas o ambientales diseñadas para disminuir o eliminar la densidad poblacional de larvas o pupas; ya sea por fumigación o alteración de los criaderos.²⁻⁴

Por otra parte, en las gestiones enfocadas en el control de la etapa adulta se realizan fumigaciones sobre las áreas endémicas por brotes de enfermedades transmitidas por vectores (VBDs).^{2, 122} Entre los insecticidas más eficientes para el control de mosquitos, sobresalen los compuestos organoclorados, organofosforados, piretroides y carbamatos, por su eficiencia al atacar el Sistema Nervioso Central (CNS) del insecto. Estos agentes químicos presentan dos principales mecanismos de acción en el CNS, que provocan la sobreexcitación y muerte del organismo atacado.

En el caso particular de los compuestos organofosforados y carbamatos, ambos presentan un mecanismo de acción similar, basado en la inhibición de la acetilcolinesterasa (AChE), enzima reguladora encargada de transmitir los impulsos nerviosos entre las neuronas, mediante la hidrolización del neurotransmisor acetilcolina (ACh).¹²³ Por otra parte, los compuestos piretroides y organoclorados fabricados en el laboratorio, modulan los canales de sodio controlados por voltaje, desencadenando un rápido desequilibrio en el organismo atacado.¹²⁴

Ciertamente, se puede verificar la insuficiente diversidad de mecanismos de acción para eliminar mosquitos y otros artrópodos transmisores de patógenos. Si bien

existe una gran cantidad de insecticidas comerciales, estos se caracterizan por tener una eficacia y selectividad para controlar los vectores, que depende exclusivamente de dos modos de acción dentro del CNS del insecto.

Por el contrario, la variedad de sustancias químicas usadas en la protección de los cultivos, contemplan numerosos mecanismos de defensa y preservación, que benefician al sector agrícola y alimentario.¹²⁵

Esta particularidad, promovió la existencia de múltiples mecanismos alternativos, que le permiten al insecto tolerar la acción de los insecticidas actuales. Estos se pueden agrupar en cuatro categorías: (I) resistencia metabólica, (II) resistencia del sitio molecular objetivo, (III) resistencia del comportamiento por penetración reducida y (IV) perturbación del comportamiento. Cada categoría se detalla a continuación:

(I). *Resistencia metabólica*: La resistencia metabólica es el principal mecanismo de resistencia en insectos. Esta resistencia se enfoca sobre los sistemas enzimáticos que todos los insectos poseen como herramienta de desintoxicación frente a los agentes nocivos de origen natural o sintético.

Comúnmente, los sistemas de desintoxicación del insecto se relacionan con la evolución de los mecanismos de defensa de las plantas o con diferentes tipos de enzimas, que actúan como medios de protección contra diversas toxinas vegetales e insecticidas comerciales.¹²⁶ En este contexto, la sobreexpresión enzimática es un modo de defensa a los insecticidas, que realiza la sustitución de los aminoácidos presente en la enzima, con el fin de alterar su afinidad con uno o varios compuestos insecticidas, de manera que se genere la resistencia a los mismos.^{127, 128}

En el caso de los mosquitos, la resistencia metabólica más común es el incremento de la expresión de los genes que codifican las principales enzimas metabolizadoras de múltiples xenobióticos. El funcionamiento de este mecanismo se basa en la sobreexpresión de enzimas normalizadoras de toxinas, ya sea mediante la ampliación del gen, duplicación del gen o cambios en los elementos reguladores de la región promotora del gen.

Estas alteraciones genéticas traen como resultado un incremento significativo de la producción de enzimas con capacidad de metabolizar o degradar las sustancias insecticidas, antes de que estas puedan ejercer un efecto tóxico en el insecto.^{119, 129}

En consecuencia, existen tres familias de enzimas vinculadas en los procesos metabólicos de los insecticidas, las citocromo monooxigenasas P450 (P450s), las glutatión transferasas (GST) y las carboxil/colinesterasas (CCE). Cada familia se encarga de catalizar una amplia variedad de reacciones de desintoxicación; por tanto, son consideradas la primera defensa enzimática contra los agentes tóxicos. Además, estas familias de enzimas son responsables de eliminar numerosas sustancias metabólicas, ejercer funciones esenciales en diferentes rutas biosintéticas y participar en la comunicación química entre las células.¹³⁰

Recientemente, pese a los numerosos compuestos xenobióticos producidos por el hombre, principalmente los insecticidas; la acelerada expansión y diversificación de los genes de desintoxicación en los insectos, favorecieron su rápida adaptación en los nichos ecológicos y elevaron su longevidad.¹²⁹⁻¹³¹

(II). *Resistencia del sitio molecular objetivo*: La resistencia del sitio objetivo molecular ha sido registrada como el segundo mecanismo de resistencia a los insecticidas de uso común. Los insecticidas de tipo organofosforado, carbamato, piretroide y organoclorado, presentan como modo de acción la interacción con un sitio molecular específico del CNS del insecto. En las cepas de insectos resistentes, el sitio molecular objetivo presenta alteraciones genéticas que impiden su interacción con el insecticida y restringen los efectos tóxicos en el organismo tratado.¹³²

Principalmente, este tipo de resistencia se enfoca en presentar una sensibilidad reducida del sitio activo, atribuida a diferentes mutaciones del gen que codifica el sitio activo de la proteína, objetivo de los insecticidas. En particular, los insecticidas de tipo organofosforado y carbamato, despliegan como mecanismo de toxicidad la inhibición de la AChE, presente en la sinapsis colinérgica de las células nerviosas.

La AChE es una esterasa tipo B que cataliza la hidrólisis del neurotransmisor ACh, liberado en el espacio sináptico de las neuronas. La AChE tiene un sitio activo conformado por dos partes, la parte aniónica y la parte esteárica. Esta última parte, es el objetivo de los insecticidas organofosforados y carbamatos, encargados de inhibir la AChE mediante un enlace covalente.¹³³ En el caso de los insectos resistentes, se han encontrado diversas mutaciones que minimizan la interacción covalente entre el insecticida y la parte esteárica de la AChE.

Otro tipo de resistencia del sitio molecular objetivo, orientada principalmente sobre los compuestos organoclorados y piretroides, se denomina Resistencia *Knockdown* (KDR) o resistencia al desplome. Esta resistencia radica en la habilidad del insecto con genes dominantes para resistir prolongadamente a los insecticidas sin desvanecerse.

El desarrollo de la resistencia KDR se atribuye a las mutaciones no similares del gen del canal de sodio dependiente del voltaje, que buscan reducir la unión del insecticida con este canal localizado en la envoltura del nervio, y así eludir la pérdida de coordinación y eventual parálisis del insecto.^{134, 135}

Alrededor del mundo, existen numerosas mutaciones del gen del canal de sodio dependiente del voltaje, las cuales son responsables de la resistencia KDR. Entre las más conocidas, se encuentran las mutaciones S989P; I1,011M; I1,011V; V1,016I; V1,016G; F1,534C y D1,794Y. Todas estas mutaciones asociadas a la resistencia KDR fueron identificadas en el mosquito *A. aegypti* L, localizado en las regiones de Centroamérica, Suramérica y el Caribe.¹³⁴⁻¹⁴¹

(III). *Resistencia del comportamiento por absorción reducida*: La resistencia del comportamiento por absorción reducida ocurre cuando el insecto desarrolla un mecanismo hereditario que reduce o impide el ingreso de una toxina en el organismo atacado. Los insectos que presentan esta resistencia, por lo general, presentan modificaciones en el revestimiento de la cutícula o el tracto digestivo, que previene o retarda el ingreso o absorción de los insecticidas.^{142, 143}

Particularmente, este tipo de resistencia tiene un mecanismo no específico, que disminuye la efectividad de una amplia gama de insecticidas comerciales. Es decir, la baja absorción de un compuesto insecticida, frecuentemente se asocia a una resistencia cuticular, catalogada como un efecto de resistencia menor, que suele presentarse en los tratamientos aplicados por vía oral.

Usualmente, las estrategias químicas para el control de vectores administran los insecticidas de síntesis por vía espacial o superficial, donde la absorción del agente tóxico se realiza mediante los apéndices del insecto. Aunque, diversas observaciones demostraron que el engrosamiento del revestimiento de la cutícula del mosquito podría reducir su permeabilidad, y eventualmente, aumentar la biodisponibilidad del agente químico sobre otros organismos no objetivos y su entorno.¹⁴⁴

(IV). *Perturbación del comportamiento*: Como último tipo de resistencia, la perturbación del comportamiento en los insectos, puntualmente en los mosquitos, se basa en los cambios de conducta atribuidos a la exposición prolongada de un insecticida. Aunque, este modo de resistencia no tiene el mismo impacto que la resistencia metabólica, puede contribuir en la modificación del efecto tóxico de un insecticida.¹⁴⁵

La resistencia por perturbación del comportamiento se divide en dos principales respuestas, la irritación por contacto directo o la repelencia espacial. Esta última respuesta, se origina particularmente en las áreas fumigadas con agentes insecticidas, donde la acción tóxica del compuesto, promueve una rápida reacción del insecto, que busca alejarse del área tratada.¹⁴⁶

Una de las principales consecuencias del fenómeno de resistencia, es su persistencia en las generaciones de insectos. Por ello, resulta importante encontrar alternativas químicas eficientes sobre los vectores de enfermedades, como el *A. aegypti* L., que puedan eludir los casos de resistencia y que sean inocuos para la salud humana y el ambiente.

4.11. Estrategias de gestión y prevención de la resistencia

En la actualidad, la fuerte dependencia a los insecticidas de tipo organofosforado, organoclorado, carbamato y piretroide ha promovido el desarrollo de diferentes modos de resistencia en insectos, basadas principalmente en alteraciones genéticas de los sitios polimórficos de las proteínas, que limitan las

interacciones entre el ligando (molécula insecticida) y el receptor diana (proteína objetivo).^{137, 146}

Para el mosquito *A. aegypti* L. se conocen diversas mutaciones en los canales de sodio controlados por voltaje, los cuales son el objetivo de los organoclorados, tal como el Dicloro-Difenil-Tricloroetano (DDT); agente químico ampliamente utilizado en tratamiento de plagas del sector agrícola durante el siglo pasado.¹⁴¹

En la actualidad, existen distintas estrategias para prevenir y controlar la resistencia de los insectos, que proponen la regulación de las concentraciones y dosis aplicadas, la aleatorización constante de los insecticidas y la diversificación de sus modos de acción. De esta forma, se estima que la resistencia a un determinado grupo químico podría reducirse, y eventualmente desaparecer de una población de insectos.

Sin embargo, el éxito de cualquier estrategia de prevención o manejo de la resistencia, ya sea por rotaciones, mosaicos, mezclas o combinaciones, reside en conocer el modo de acción, las propiedades fisicoquímicas y la vida residual de los sustancias químicas empleadas. Además, es necesario integrar las distintas técnicas biológicas, físicas y genéticas destinadas al control de vectores, con el objetivo de minimizar el uso de insecticidas y evadir la resistencia en los insectos tratados.

4.12. Estrategias *in silico* en el control del *Aedes aegypti* L.

En términos generales, los productos químicos empleados en el control de mosquitos se dividen en dos categorías de acción: los agentes larvicidas o pupicidas y los agentes insecticidas, también conocidos como adulticidas. Estas sustancias difieren en su modo de aplicación, propiedades fisicoquímicas y características estructurales.

Los compuestos de acción larvicida son usados para reducir la población de los mosquitos durante su etapa inmadura, es decir, cuando son más susceptibles y se pueden localizar en un medio limitado. Por el contrario, los agentes insecticidas están destinados a erradicar rápidamente el mosquito adulto, con amplia dispersión geográfica y alta capacidad de transmitir algún virus.

Si bien estas sustancias tienen por objetivo diferentes etapas del ciclo de vida del vector, su actuación en conjunto representa mayor control sobre los mosquitos transmisores de patógenos a los humanos. Es así como la participación de diversas técnicas prospectivas orientadas en la predicción de las propiedades biológicas de interés en términos de salud pública, como la actividad larvicida o insecticida, han presentado un incremento paulatino durante los últimos años.¹⁴⁷⁻¹⁵⁰

Por lo anterior, en esta sección de Tesis se recopilaron de la literatura especializada distintas herramientas cualitativas y cuantitativas desarrolladas durante la última década (2009-2019) para predecir la actividad larvicida (LC₅₀ o IC₅₀) de distintas moléculas naturales, semisintéticas y sintéticas sobre el mosquito vector *A.*

aegypti.^{11, 13, 16} Todos los estudios computacionales extraídos de la literatura se agruparon de acuerdo al origen del conjunto molecular evaluado.

4.13. Compuestos naturales

4.13.1. Benzoquinonas

Inicialmente, Sousa y colaboradores, 2010;¹⁵¹ desarrollaron un ensayo de la actividad larvicida contenida en compuestos benzoquinónicos, donde se destacaron seis benzoquinonas conocidas como: *p*-benzoquinona, 2-metil-*p*-benzoquinona, 2-isopropil-*p*-benzoquinona, 2,6-dimetil-*p*-benzoquinona, 2,5-dimetil-*p*-benzoquinona y timoquinona.

Estas moléculas provenientes de la naturaleza fueron objeto de estudio en un análisis SAR, que buscaba identificar el potencial tóxico de cada sustancia sobre larvas de *A. aegypti* L. en III *instar*. Para este fin, los compuestos fueron evaluados en un rango de concentraciones de 33–90 ppm.

A partir de los resultados obtenidos, se determinó que el número, la posición y el tamaño de los sustituyentes están fuertemente asociados a la bioactividad de los compuestos.

4.13.2. Monoterpenos

A partir del estudio propuesto por Santos *et al.*, 2011, se realizó un análisis SAR para identificar las características estructurales asociadas con la actividad larvicida de 14 monoterpenos sobre el *A. aegypti* L.¹⁵²

Con este análisis, los autores determinaron que la presencia de heteroátomos en la estructura base de isopreno y la sustitución de dobles enlaces por epóxidos más reactivos, reducen el potencial larvicida. Asimismo, identificaron que los dobles enlaces conjugados *exo*, favorecen la actividad.

Otro análisis basado en monoterpenos y sus análogos, fue conducido por Alencar-Filho *et al.*, 2016. En este caso, los autores se enfocaron en el diseño y aplicación un modelo matemático para predecir la actividad larvicida de 31 monoterpenos, incluidos en la Figura 4.8.¹⁵³

La digitalización y el refinamiento de las estructuras moleculares se realizó con el programa Gaussian, donde se aplicó un nivel de teoría basado en el método de la Teoría del Funcional de Densidad (DFT), con funcional B3LYP y base 6-31+G (d, p).

A partir del conjunto molecular optimizado se calcularon 1666 descriptores moleculares provistos por la plataforma libre E-Dragon, distribuidos entre descriptores constitucionales, topológicos, electrónicos y geométricos. Para seleccionar los

descriptores moleculares más apropiados se utilizó la técnica lineal de selección de variables PLS acoplada con el algoritmo de Selección de Predictores Ordenados (OPS).

Los autores obtuvieron el modelo QSAR expresado en la Ec. 4.1, conformado por un descriptor geométrico GETAWAY ($R8e+$) de autocorrelación máxima en una distancia de 8 enlaces, ponderado por electronegatividad de Sanderson.

El descriptor $MAXDN$, es un índice topológico que describe los valores de máxima variación negativa electrotopológica. Por último, nCt detalla el número total de Carbonos terciarios (sp^3) presentes en cada compuesto.

$$\text{Log}(LC_{50}) = -18.28(R8e+) + 0.57(MAXDN) + 0.18(nCt) + 1.62 \quad (4.1)$$

La combinación de estos tres descriptores moleculares del programa E-Dragon, proporciona predicciones acertadas de la LC_{50} sobre larvas *A. aegypti*. De tal forma que los parámetros estadísticos considerados exhibieron valores aceptables del coeficiente de determinación (R^2) y la desviación estándar (S), tanto para el conjunto de calibración ($R_{cal}^2 = 0.83$, $S_{cal} = 0.19$) como en el conjunto de predicción ($R_{pred}^2 = 0.89$, $S_{pred} = 0.34$); además de superar la prueba de CV-LOO con un valor de $R_{LOO}^2 = 0.77$.

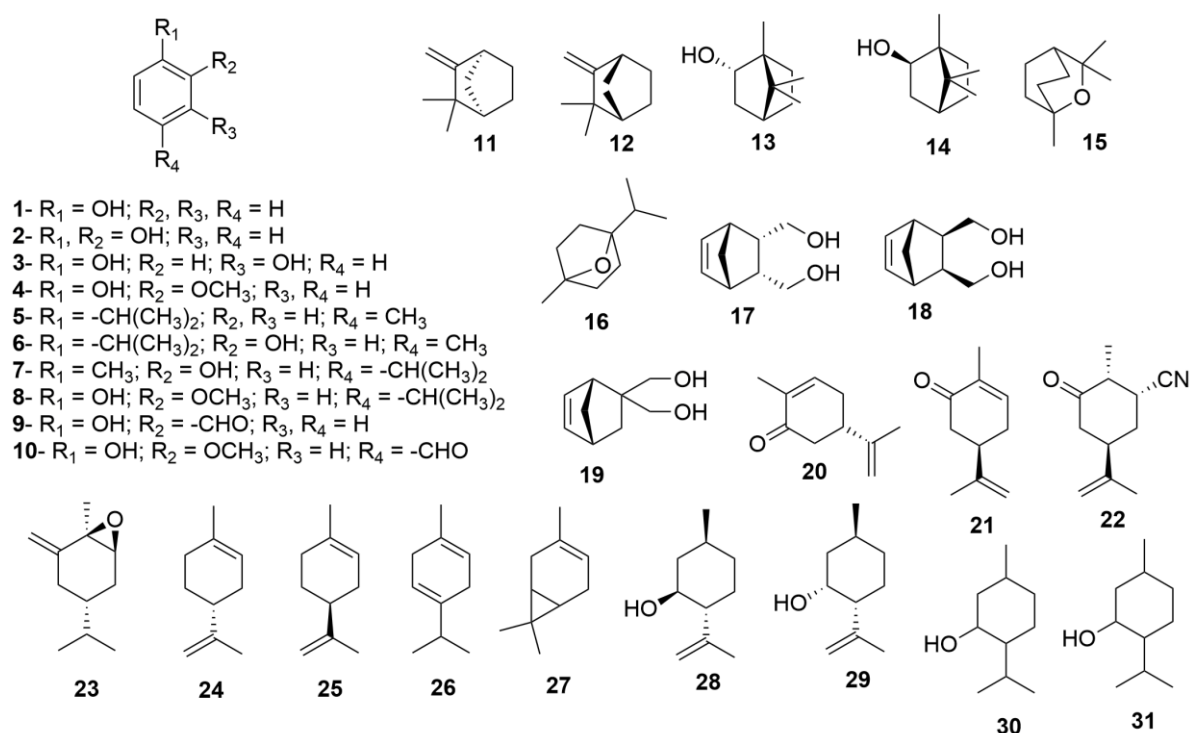


Figura 4.8. Estructuras moleculares de 31 monoterpenos evaluados por Alencar-Filho *et al.*, 2016.¹⁵³

4.14. Compuestos modificados químicamente

4.14.1. Compuestos naturales y sus derivados

En el campo de los compuestos naturales y sus análogos semisintéticos, se han desarrollado diferentes estudios *in silico* para determinar la toxicidad aguda LC₅₀ sobre larvas *A. aegypti* L. Bajo esta consideración, un conjunto molecular de 55 agentes naturales y semisintéticos de tipo oxigenados, terpenos, terpenoides, fenilpropanoides y cetonas de la Figura 4.9; fue el objeto de estudio en el trabajo químico-informático desarrollado por Scotti *et al.*, 2014.¹⁵⁴

Las técnicas PCA, PCA-consensuado y PLS fueron empleadas para examinar 128 descriptores moleculares 3D de Campos de Interacción Molecular (MIFs) acoplados a descriptores de campos de fuerza GRID, contenidos en el programa Volsurf+.

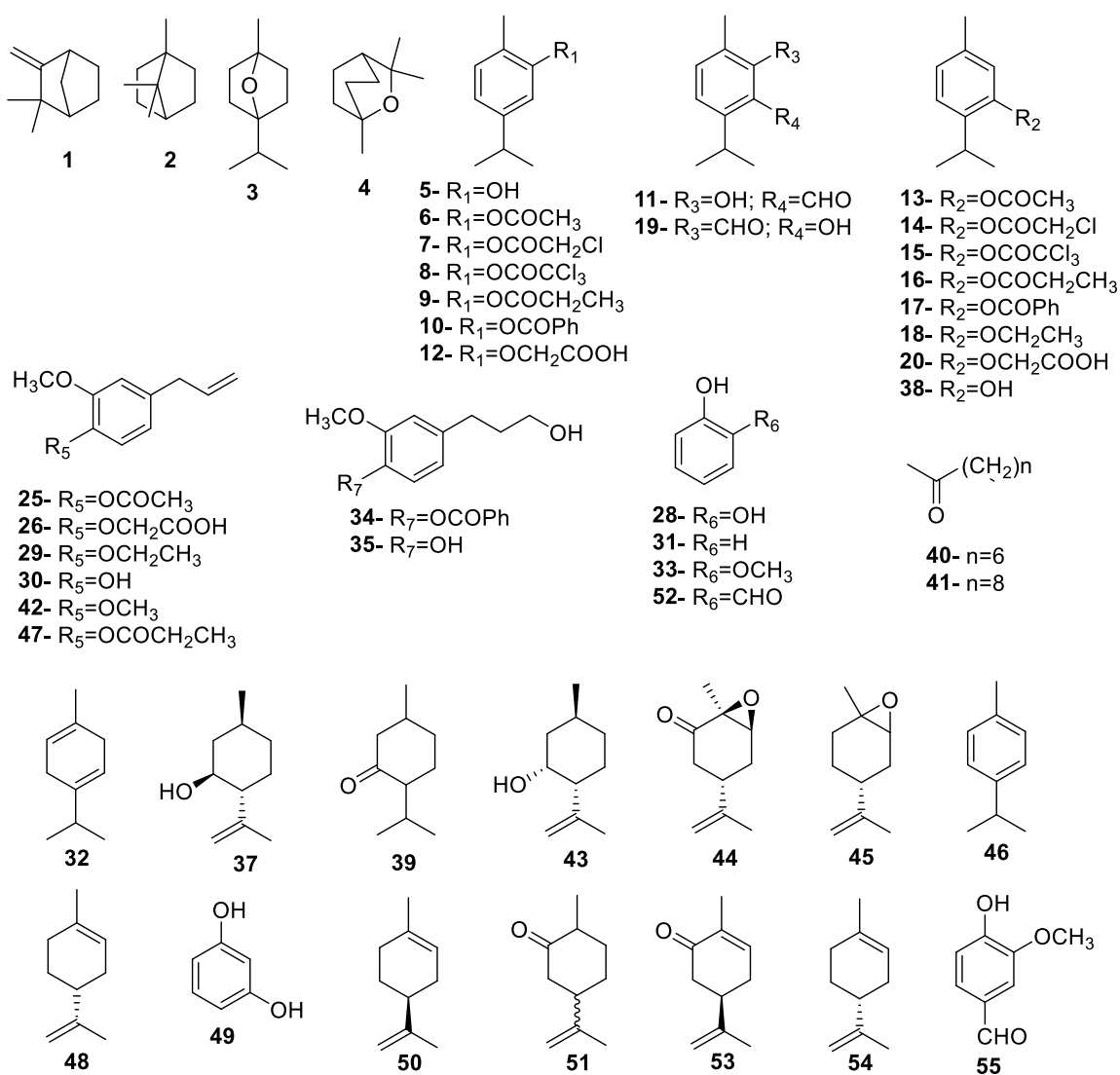


Figura 4.9. Estructuras moleculares evaluadas por Scotti *et al.*, 2014.¹⁵⁴

Los resultados obtenidos revelaron que la técnica PCA proporciona mejores predicciones de la actividad larvicida expresada como LC₅₀, en relación a las otras técnicas de selección de variables contempladas. Los parámetros estadísticos presentaron valores apropiados en el conjunto de calibración ($R_{cal}^2 = 0.71$), predicción ($R_{pred}^2 = 0.68$) y en CV-LOO ($R_{LOO}^2 = 0.67$). Entre las conclusiones del trabajo, se destacó que el potencial larvicida sobre *A. aegypti* L. está asociado con la hidrofobicidad de las moléculas.

A partir del trabajo anterior, De *et al.*, 2018,¹⁵⁵ propusieron un modelo QSAR para predecir la actividad larvicida de 61 compuestos derivados de las plantas, con base en la estrategia estadística PLS acoplada a Algoritmos Genéticos (GA). Esta técnica de selección de variables fue programada por los autores y se encuentra disponible junto con una herramienta de Validación Cruzada Doble (DCV), de acceso libre en: (http://teqip.jdvu.ac.in/QSAR_Tools/).

Particularmente, este estudio logró establecer una correlación entre la actividad larvicida y las características estructurales determinadas por descriptores topológicos simples, denominados índice de Átomo Topoquímico Extendido (ETA), calculados a través del programa de acceso libre PaDEL.¹⁵⁶

En este caso, no fue necesario realizar un análisis exhaustivo de la conformación de las estructuras moleculares. Después de realizar el filtrado de los descriptores, se obtuvieron 42 descriptores ETA con intercorrelación absoluta menor a 0,95 para diseñar el modelo.

Los resultados de esta investigación exhibieron un modelo lineal conformado por de 5 descriptores topológicos detallados en la Ec. 4.2, para estimar la LC₅₀ sobre larvas *A. aegypti* L.

$$\text{pLC}_{50} = 5.9 + 1.2\text{ETA_EtaP_F} - 5.3\text{ETA_dEpsilon_D} - 13.4\text{ETA_dAlpha_B} - 6.1\text{ETA_BetaP_s} - 2.1\text{ETA_dEpsilon_C} \quad (4.2)$$

$$N_{cal} = 41, R^2 = 0.73, Q^2 = 0.63, R_{LOO}^2 = 0.52, RMS_{cal} = 0.25, S_{cal} = 0.27, F = 32.7$$

$$N_{pred} = 20, Q_{F1}^2 = 0.67, Q_{F2}^2 = 0.65, r_{m(pred)}^2 = 0.53, RMS_{pred} = 0.29, S_{pred} = 0.33$$

Donde Q^2 el coeficiente de correlación de la técnica CV-LOO; RMS el error cuadrático medio; S desviación estándar; F radio de varianza; Q_{F1}^2 y Q_{F2}^2 los coeficientes de correlación del conjunto de predicción.

El modelo obtenido exhibió valores aceptables de los parámetro estadísticos que evalúan el conjunto de calibración ($R_{cal}^2 = 0.73$, $RMS_{cal} = 0.25$, $S_{cal} = 0.27$) y de predicción ($Q_{F1}^2 = 0.67$, $Q_{F2}^2 = 0.65$, $RMS_{pred} = 0.29$, $S_{pred} = 0.33$). Además, la prueba de CV interna presentó un valor admisible ($Q^2 = 0.63$).

Por medio del análisis de los coeficientes de los descriptores topológicos ETA, se determinó que el descriptor *ETA_dEpsilon_D* tiene mayor contribución sobre la respuesta biológica evaluada; esto revela que la presencia de grupos químicos donantes de electrones favorece la bioactividad de las moléculas.

El segundo descriptor en términos de contribución es el *ETA_Etap_F*, definido como índice de funcionalidad relativa al tamaño molecular, con la cual se obtiene una medida de presencia de heteroátomos y enlaces múltiples.

El descriptor *ETA_dAlpha_B*, tiene el tercer grado de importancia sobre la actividad. Esta variable provee información acerca del área de la superficie polar, y determinó que el aumento de la superficie polar, disminuye el poder larvicida de la molécula.

El descriptor *ETA_BetaP_s* se basa en los valores β sumados para todos los enlaces sigma de acuerdo con el número de vértices, como lo indica la Ec. 4.3. Por ello, *ETA_BetaP_s* realiza el recuento de los átomos electronegativos de la molécula en relación a su tamaño molecular. Según el coeficiente de este descriptor, a medida que aumenta su valor, disminuye drásticamente la actividad larvicida.

$$\sum \beta'_s = \sum \beta_s / N_v \quad (4.3.)$$

El grado de importancia del descriptor *ETA_dEpsilon_C* es inferior a los anteriores. Esta variable topológica mide la electronegatividad la molécula e indica que a menor electronegatividad, mayor será el poder larvicida.

Por último, los autores concluyeron que la actividad larvicida sobre *A. aegypti* L. está fuertemente asociada a la presencia de: (i) grupos electrodonadores como –OH, –NH₂ y –SH (Figura 4.10.A); (ii) compuestos con heteroátomos o múltiples enlaces (Figura 4.10.B); (iii) grupos electroatrayentes, como halogenuros de alquilo (Figura 4.10.C). Además, determinaron que una extensa superficie polar, reduce el efecto larvicida de las moléculas.

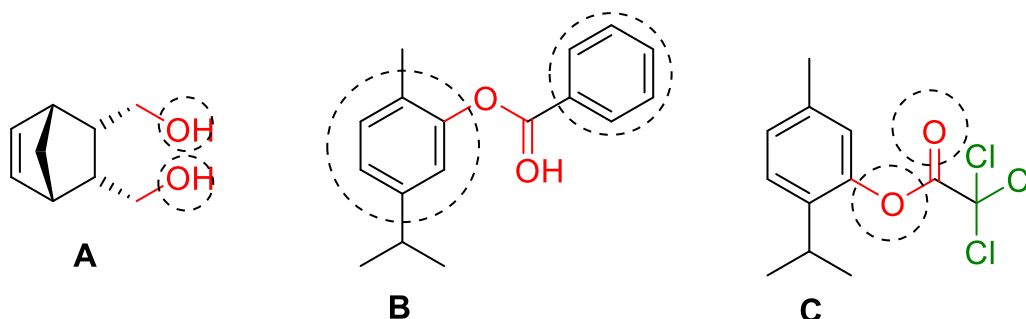


Figura 4.10. Compuestos bioactivos sobre larvas *A. aegypti*. Fuente: De *et al.*, 2018,¹⁵⁵

Un estudio más reciente reportado por Dos Santos *et al.*, 2018,¹⁵⁷ utilizó el conjunto molecular reportado por Scotti *et al.*, 2014,¹⁵⁴ para establecer un modelo

QSAR con capacidad de predecir la actividad larvica sobre el vector de múltiples *flavivirus*, el *A. aegypti* L.

Aquí, los autores analizaron el poder predictivo de dos modelos QSAR definidos como clásico y multidimensional, que difieren en el tipo de descriptores moleculares analizados y el método aplicado para diseñar el modelo. La optimización de la geometría molecular de cada compuesto se ejecutó bajo el método semiempírico Hamiltoniano PM6, seguido de un refinamiento basado en el nivel de teoría DFT-B3LYP/6-31+G(d), incorporado en el paquete Gaussian09W®.

En el diseño del modelo QSAR clásico, se empleó el programa Dragon6 para calcular 5.000 descriptores moleculares de tipo constitucional, topológico, electrónico y geométrico. Los cuales fueron sometidos a un procedimiento de selección de variables a través del algoritmo OPS, basado en vectores informativos iniciales, que provee las mejores relaciones entre variables dependientes e independientes. Adicionalmente, emplearon la técnica PLS para la selección y validación del mejor modelo QSAR, mediante el programa BuildQSAR.

Para determinar el poder predictivo del modelo, los autores dividieron el conjunto molecular total ($N=55$ moléculas) con el procedimiento de arranque para obtener dos subconjuntos de análisis; un conjunto de calibración ($N=45$ moléculas) y uno de predicción externa ($N=10$ moléculas), utilizado para evaluar el poder predictivo.

Como resultado, fue seleccionado un modelo QSAR clásico que vincula 5 descriptores moleculares teóricos, expresados por la Ec. 4.4, para predecir el efecto tóxico IC_{50} sobre larvas *A. aegypti* L.

$$pIC_{50} = -0.60MATS7e - 1.47VE1Sing_A - 0.11nCt + 0.29MlogP + 0.34R8u + 2.32 \quad (4.4)$$

$$N = 45, R_{cal}^2 = 0.75, S_{cal} = 0.284, R_{LOO}^2 = 0.68, R_{pred}^2 = 0.85$$

El modelo QSAR clásico, involucra 5 descriptores moleculares obtenidos mediante el programa comercial Dragon6 definidos de la siguiente manera: *MATS7e* detalla un descriptor de autocorrelación 2D de Moran, que expresa los valores de electronegatividad de los átomos separados por una distancia topológica de 7 enlaces. El descriptor *VE1Sing_A* se basa en una matriz 2D, donde la matriz de adyacencia representa el conjunto completo de conexiones entre pares de átomos adyacentes.

La variable *nCt* indica la importancia del número total de carbonos terciarios en la estructura molecular de los larvicidas naturales. Acorde con el coeficiente de regresión de este descriptor, a medida que aumenta el número de carbonos *sp3* se reduce el poder activo de las moléculas.

Seguidamente, *MlogP* es un descriptor basado en la medida experimental del coeficiente de partición octanol-agua de Moriguchi,¹⁵⁸ que expresa el carácter hidrófilo

o hidrófobo del compuesto. La última variable del modelo vincula un descriptor geométrico GETAWAY, que revela la influencia de los átomos a una distancia topológica de 8 enlaces en el sistema molecular.

Posterior a esto, Dos Santos y colaboradores desarrollaron un modelo QSAR utilizando el mismo conjunto molecular, donde se consideraron descriptores moleculares de tipo geométrico para determinar las características moleculares estéricas y electrostáticas asociadas a la respuesta larvicida. Este modelo multidimensional exhibió una clara interpretación del sentido físico de los campos involucrados en el modelo clásico reportado previamente.

Entre las conclusiones de esta investigación se determinó que las moléculas que presenten mayor número de vértices, una cadena hidrofóbica enlazada a un ciclohexano y menor volumen en el centro de estructura molecular, como ocurre con el compuesto ciclohexano *orto, para* sustituido, altamente lipofílico; presentan mejor perfil tóxico sobre la especie *A. aegypti* L. en etapa larval.

4.15 Compuestos de síntesis en el laboratorio

4.15.1. Compuestos organoestánicos

El uso de compuestos organoestánicos en el control de insectos no representa una novedad. Sin embargo, los efectos que estas sustancias puedan generar sobre los mosquitos, son aún desconocidos. Por esta razón, Song *et al.*, 2007,¹⁵⁹ sintetizaron una serie de compuestos triorganoestánicos, denominados 2,2,3,3-tetrametilciclopropanocarboxilatos, incluidos en la Figura 4.11; donde R corresponde a los sustituyentes -CH₃, -C₂H₅, -*n*-C₃H₇, -*n*-C₄H₉, -C₆H₅ o -C₆H₁₁.

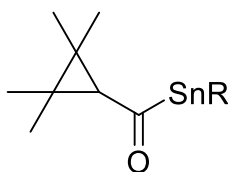


Figura 4.11. Estructura molecular base de los agentes organoestánicos bajo estudio. Autores: Hansch y Verma.¹⁶⁰

Los compuestos de la Figura 4.11, fueron empleados en un estudio QSAR desarrollado por Hansch y Verma en 2009,¹⁶⁰ para calcular IC₅₀. En este caso, el momento principal de inercia a lo largo del eje z (π_R) se incluyó como descriptor molecular. Los resultados brindaron un paralelismo racional entre el descriptor π_R y la actividad larvicida de los compuestos organoestánicos, expresado en la Ec.4.5.

$$\log 1 / IC_{50} = 0.54 \pi_R + 5.19 \quad (4.5)$$

$$N = 6, R_{cal}^2 = 0.81, S_{cal} = 0.24, R_{LOO}^2 = 0.57, Q = 3.80, F = 17.16$$

La Ec. 4.5 presentó valores aceptables en cada parámetro estadístico analizado. En este caso se dedujo que la toxicidad aguda IC_{50} se asocia con la hidrofobicidad de los grupos sustituyentes. De modo que, a mayor hidrofobicidad del grupo R, mayor actividad larvicida.

Seguidamente, fueron seleccionadas aleatoriamente dos moléculas organoestánicas (Figura 4.11 con $R=n-C_3H_7$ y $R=C_6H_5$) para conformar el conjunto de predicción externa; y los compuestos restantes formaron el conjunto de calibración del modelo. Esta prueba dio como resultado la Ec. 4.6.

$$\log 1 / IC_{50} = 0.51\pi_R + 5.12 \quad (4.6)$$

$$N = 4, R_{cal}^2 = 0.92, S_{cal} = 0.18, R_{pred}^2 = 0.74, Q = 5.2, F = 24.67$$

Por medio del modelo 4.5, se alcanzaron valores residuales para las moléculas **1** y **2** de 0.24 y 0.48, respectivamente. Por otra parte, los autores encontraron una importante correlación ($R^2=0.94$) entre el momento principal de inercia a lo largo del eje z (π_R) y la Refractividad Molar (MR) del grupo R, expresada en la Ec. 4.7.

$$\log 1 / IC_{50} = 0.55MR_R + 5.15 \quad (4.7)$$

$$N = 6, R_{cal}^2 = 0.89, S_{cal} = 0.18, R_{LOO}^2 = 0.78, Q = 5.24, F = 32.69$$

Posteriormente, los científicos Duong *et al.*, 2006,¹⁶¹ sintetizaron 11 compuestos definidos como trifenilestaño *p*-sustituidos benzoatos, con estructura base **A**, ilustrada en la Figura 4.12. Adicionalmente, diseñaron 11 triciclohexilestaño *p*-sustituidos benzoatos, los cuales tienen por estructura base **B**. El sustituyente R corresponde a los grupos: -H, -F, -Cl, -Br, -I, -OCH₃, -OH, -NO₂, -NH₂, -CH₃ y -C(CH₃)₃.

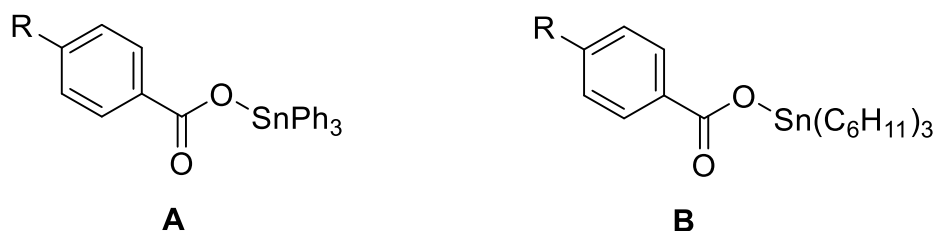


Figura 4.12. Estructuras moleculares base de trifenilestaño *p*-sustituidos benzoatos y triciclohexilestaño *p*-sustituidos benzoatos. Fuente Duong *et al.*, 2006.¹⁶¹

Posteriormente, se realizaron pruebas del potencial larvicida sobre la especie *A. aegypti* L. de cada compuesto. Los autores Hansch y Verma utilizaron los datos experimentales de la IC_{50} de los derivados de **A** para establecer un modelo de predicción de la actividad larvicida, expresado por la Ec. 4.8. En este caso, el modelo se calibró en ausencia de los compuestos cloro- sustituidos.

$$\log 1 / IC_{50} = -0.21\pi_R + 0.32MR_R + 5.73 \quad (4.8)$$

$$N = 10, R_{cal}^2 = 0.79, S_{cal} = 0.08, R_{LOO}^2 = 0.74, Q = 11.38, F = 13.09$$

Posterior a esto, seleccionaron aleatoriamente 3 análogos de **A** con R= -I, -NO₂, C(CH₃)₃, para constituir el conjunto de predicción externa. Las moléculas restantes conformaron el conjunto de calibración del modelo expresado en la Ec. 4.9.

$$\log 1 / IC_{50} = -0.17\pi_R + 0.40MR_R + 5.69 \quad (4.9)$$

$$N = 7, R_{cal}^2 = 0.89, S_{cal} = 0.06, R_{pred}^2 = 0.75, Q = 15.02, F = 17.23$$

A partir de la Ec. 4.8, obtuvieron valores residuales de 0.20, 0.08 y 0.19, para los análogos de R= -I, -NO₂ y -C(CH₃)₃, respectivamente. Posterior a esto, desarrollaron un modelo para estimar la IC₅₀ de los derivados de **B**, excluyendo el análogo R= -OCH₃. Así, lograron el modelo de la Ec. 4.10, que vincula el descriptor Es_{S-R} , denominado parámetro estérico de *Taft's* para los sustituyentes R.

$$\log 1 / IC_{50} = 0.22\pi_R + 0.13Es_{S-R} + 5.74 \quad (4.10)$$

$$N = 10, R_{cal}^2 = 0.85, S_{cal} = 0.08, R_{LOO}^2 = 0.59, Q = 11.51, F = 19.68$$

Seguidamente, fueron seleccionados aleatoriamente 3 análogos de **B** con R= -F, -NO₂, -C(CH₃)₃, para conformar el conjunto de predicción externa. Los 7 compuestos restantes se utilizaron para calibrar el modelo expresado en la Ec. 4.11.

$$\log 1 / IC_{50} = 0.17\pi_X + 0.12Es_{S-X} + 5.71 \quad (4.11)$$

$$N = 7, R_{cal}^2 = 0.94, S_{cal} = 0.04, R_{pred}^2 = 0.65, Q = 24.25, F = 31.89$$

La Ec. 4.10 presentó datos de residuos de 0.15, 0.24, 0.08 para las 3 moléculas incluidas en el conjunto de predicción. Por otra parte, los científicos Song *et al.*, 2004;¹⁶² realizaron un trabajo enfocado en la síntesis, caracterización estructural y medida de la actividad larvica sobre *A. aegypti* L. de 15 compuestos (X(p)-C₆H₄)₃SnY, con X= H, -Cl, -F, -CH₃ y -SCH₃, y Y= -Cl, -OH y -OAc.

Los datos de toxicidad aguda obtenidos del conjunto molecular previo, fueron empleados por Hansch y Verma en el desarrollo de un modelo matemático para predecir el potencial activo sobre larvas *A. aegypti* L.¹⁶⁰ Este estudio involucró el análisis de descriptores topológicos, como Índice de forma Kappa y una forma modificada de este.

Los resultados obtenidos exhibieron un modelo QSAR con capacidad de predicción de la IC_{50} aceptable, representado en la Ec. 4.12. Aunque, los derivados con $X = -SCH_3$ y $Y = -Cl, -OH$ y $-OAc$ fueron omitidos por no contar con datos experimentales de actividad larvicida.

$$\log 1 / IC_{50} = 6.46 \pi_x - 8.91 \pi_x^2 + 1.49 \sigma_x^+ + 5.28 \quad (4.12)$$

$$N = 12, R^2 = 0.85, S = 0.11, q^2 = 0.66, Q = 8.02, F = 12.23$$

La relación parabólica de π_x en la Ec. 4.12, sugiere que la actividad letal de los compuestos sobre el mosquito *A. aegypti* L. incrementa a medida que aumenta la hidrofobicidad de los sustituyente X, hasta un valor óptimo de $\pi_x = 0.36$ y luego decrece.

Asimismo, se destaca que el desarrollo de un modelo QSAR mono y biparamétrico con π_x y π_x, π_x^2 , conduce a valores sobresalientes del parámetro r^2 . El signo positivo en σ_x^+ , involucra que los sustituyentes en X con fuerte potencial electroatrayente, tal como $-NO_2, -CN, -CHO, -CF_3$ y $-COCF_3$, podrían aumentar el poder toxico de los compuestos destinados al control de larvas.¹⁶⁰

Seguido a esto, los autores desarrollaron un nuevo modelo que considera un conjunto de validación externa conformado por los compuestos con sustituyente $X = -CH_3, -F$ o $-Cl$ y $Y = OH$; los cuales fueron elegidos aleatoriamente del conjunto molecular total. Los 9 compuestos restantes fueron utilizados para calibrar el modelo expresado en la Ec. 4.13.

$$\log 1 / IC_{50} = 6.37 \pi_x - 8.88 \pi_x^2 + 1.48 \sigma_x^+ + 5.28 \quad (4.13)$$

$$N = 9, R^2 = 0.86, S = 0.12, q^2 = 0.50, Q = 7.47, F = 9.98$$

Los valores residuales determinados por la Ec. 4.13 para cada compuesto del validación sustituido en X por grupos $-CH_3, -F$ o $-Cl$ fueron de 0.12, 0.04 y 0.16, consecutivamente.

Posteriormente, la Ec. 4.11 y 4.12 fueron utilizadas para predecir el potencial larvicida de los compuestos sustituidos en $X = -C \equiv CH, -COOC_2H_5$ o -4 -piridinil, y en $Y = -Cl, -OH$ o OAc . De este modo, se identificó que la estructura molecular más activa sobre larvas *A. aegypti* L. es el derivado sustituido en $X = 4$ -pirinidil.

En orden de agrupar toda la información experimental disponible de los agentes organoestánicos con actividad larvicida sobre *A. aegypti* L, los autores Hansch y Verma desarrollaron un modelo QSAR basado en los datos reportados por Song *et al.*, 2004, Duong *et al.*, 2006 y Song *et al.*, 2007,^{159, 161, 162} a excepción de tres sustancias sin valores de medida experimental.

$$\log 1 / IC_{50} = 0.15 \log P - 0.33 I_3 - 0.46 I_4 + 5.28 \quad (4.14)$$

$$N = 37, R^2 = 0.71, S = 0.18, q^2 = 0.64, Q = 4.58, F = 27.06$$

Los resultados obtenidos, indican que la Ec. 4.14 constituye un paralelismo racional entre la actividad larvicida y los descriptores moleculares semiempíricos y teóricos, como la variable del coeficiente de partición octanol-agua ($\log P$); el descriptor I_3 , representado por valores binarios de 0 o 1, que indican la presencia de compuestos triciclohexilestaño *p*-sustituidos benzoatos. Asimismo, el descriptor I_4 , que codifica la presencia de compuestos organoestánicos tris-(fenil-*p*-sustituidos).

Además, los autores indicaron que el diseño del modelo QSAR monoparámetro vinculó al descriptor $\log P$ con $R^2 = 0.32$; es decir, carente de información para describir el fenómeno asociado a la toxicidad sobre larvas *A. aegypti* L.

Por último, para validar la relación estructura-actividad expresada en la Ec. 4.13, se estableció un conjunto de validación externo conformado por 9 compuestos seleccionados aleatoriamente del conjunto molecular total ($N=37$). Los compuestos químicos restantes se usaron para definir el modelo expresado por la Ec. 4.15. La cual se utilizó para estimar el poder larvicida del conjunto de validación externo, logrando valores residuales en un rango de 0.05–0.43.

$$\log 1 / IC_{50} = 0.15 \log P - 0.30 I_3 - 0.40 I_4 + 5.19 \quad (4.15)$$

$$N = 28, R^2 = 0.67, S = 0.19, q^2 = 0.55, Q = 4.32, F = 16.02$$

Por otra parte, Freitas *et al.*, 2010 desarrollaron un modelo matemático para predecir la poder larvicida de los compuestos organoestánicos mencionados anteriormente sobre el mosquito *Anopheles stephensi*.¹⁵⁷⁻¹⁶²

Este enfoque vinculo descriptores de Análisis de Imagen Multivariante (MIA) para establecer una correlación racional entre la estructura molecular y la propiedad biológica. Ante la necesidad de seleccionar los descriptores más representativos para la respuesta final considerada, se empleó la técnica basada en regresión PLS.

Como resultado de este estudio se estableció un modelo QSAR conformado por 5 componentes PLS, con valores aceptables de la estadística evaluada, con confiabilidad de $R^2 = 0.80$ y $R_{LOO}^2 = 0.52$, y capacidad predictiva de $R_{pred}^2 = 0.66$.¹⁶³

4.16. Referencias

1. World Health Organization (WHO). A global brief on vector-borne diseases. WHO Press: Ginebra, 2014. [Citado el 20 de Abril de **2019**]. Disponible en: <http://www.who.int/campaigns/world-health-day/2014/global-brief/>
2. World Health Organization (WHO). Investing to overcome the global impact of neglected tropical diseases: third report. WHO Press: Ginebra, 2015. [Citado el 20 de Abril de **2019**]. Disponible en: http://www.who.int/neglected_diseases/Third_report_2015/
3. World Health Organization (WHO). Sixth meeting of the vector control advisory group (VCAG). Drexler, A., Ed.; WHO Press: Geneva, **2017**, p. 4-24. Licence: CC BY-NC-SA 3.0 IGO. Disponible en: https://www.who.int/neglected_diseases/vector_ecology/resources/WHO_HTM_NTD_VEM_2017.05/en/
4. Organización Mundial de la Salud (OMS). Enfermedades transmitidas por vectores. Centro de Prensa OMS en línea: *Respuesta Mundial para el Control de Vectores 2017-2030*. Vol 5.4., 2017. [Citado el 25 de Abril de **2019**]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/vector-borne-diseases>
5. Manjarres-Suarez, A., Olivero-Verbel, J. Chemical control of *Aedes aegypti*: a historical perspective. *Revista Costarricense de Salud Pública*. **2013**, 22, p. 68–75.
6. Roy, K., Kar, S., Nas, R.N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier: London. **2015**.
7. Hansch, C and Leo, A., Exploring QSAR, Fundamentals and Applications in Chemistry and Biology. Vol 1. American Chemical Society (ACS) Professional Reference Book: Washington DC. **1995**, p. 139–205.
8. Hansch, C and Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal American Chemical Society*. **1964**, 86 (8), 1616–1626.
9. Ramsden, C.A. *Quantitative drug design - Comprehensive medicinal chemistry*. In: Hansch, C., Sammes, P.G., Taylor, J.B., Eds. The rational design, mechanistic study; therapeutic application of chemical compounds. Pergamon Press: Oxford, **1990**.
10. Hansch C.L.A. *Substituent constants for correlation analysis in chemistry and biology*. John Wiley and Sons: New York. **1979**.
11. Muhammad, U., Uzairu, A., Arthur, D.E., Review on: Quantitative Structure Activity Relationship (QSAR) modelling. *Journal Analytical & Pharmaceutical Research*. **2018**, 7(2), p.240–242.

12. Charrel, R., Leparc-Goffart, I., Gallian, P., de Lamballerie, X. Globalization of chikungunya: 10 years to invade the world. *Clinical Microbiology and Infection*. **2014**, 20, p. 662–663.
13. Peter, S.C., Dhanja, J.K., Malik, V., Radhakrishnan, N., Jayakanthan, M., Sundar, D. *Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications*. In: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds. *Encyclopedia of Bioinformatics and Computational Biology*. Vol 2, Elsevier: Amsterdam. **2019**, p. 661–676.
14. Dearden, J.C. The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relationships*. **2016**, 1(1), p.1–44.
15. Devillers, J., Lagneau, C., Lattes, A., Garrigues, J.C., Clémenté, M.M., Yébakima, A. In silico models for predicting vector control chemicals targeting *Aedes aegypti*. *SAR and QSAR Environmental Research*. **2014**, 25(10), p. 805–835.
16. Illing, P. General Overview of the Safety Evaluation of Chemicals. In Hester, R.E and Harrison, M.R (Eds.), *Alternatives to Animal Testing*. Environmental Science and Technology, 23ed. The Royal Society of Chemistry: Cambridge. **2006**, p. 3–24.
17. World Health Organization (WHO). Pesticides and their application for the control of vectors and pests of public health importance. 6th ed. World Health Organization, Department of Control of Neglected Tropical Diseases, Pesticides Evaluation Scheme, Geneva. 2006. [Citado el 27 de Abril de **2019**] Disponible en: http://whqlibdoc.who.int/hq/2006/WHO_CDS_NTD_WHOPES_GCDPP_2006.1_eng.pdf
18. Charrel, R., Leparc-Goffart, I., Gallian, P., de Lamballerie, X. Globalization of chikungunya: 10 years to invade the world. *Clinical Microbiology and Infection*. **2014**, 20, p. 662–663.
19. Chouin-Carneiro, T., Barreto dos Santos, F. *Transmission of Major Arboviruses in Brazil: The Role of Aedes aegypti and Aedes albopictus Vectors*. In: Shields, B.D.C., Ed. *Biological Control of Pest and Vector Insects*. IntechOpen, **2017**, p. 231–255.
20. Velandia, M.L., Castellanos, J.E. Virus del dengue: estructura y ciclo viral. Asociación Colombiana de Infectología. *Infectio*. **2011**, 15(1), p. 33–43.
21. Westaway, E.G., Blok, J. *Taxonomy and evolutionary relationships of flaviviruses*. In: Gubler, D.J., Kuno, C., Eds. *Dengue and Dengue Hemorrhagic Fever*. CAB International: London. **1997**, p. 147–173.
22. Centro para el Control y la Prevención de Enfermedades los Estados Unidos de Norteamérica (CDC). Hoja de Datos sobre el Dengue. 2015. [Citado el 5 de Mayo de **2019**]. Disponible en: <https://www.cdc.gov/spanish/enfermedades/dengue/hojadosdatos.htm>

23. World Health Organization (WHO). *Dengue guidelines for diagnosis, treatment, prevention and control*. In: Special Programme for Research and Training in Tropical Diseases (TDR). UNICEF, UNDP, World Bank and WHO., Eds. WHO Press: Geneva. **2009**, p. 3–17.
24. World Health Organization (WHO). *Epidemiology*. WHO Press: Ginebra, 2017. [Citado el 5 de Mayo de **2019**]. Disponible en: <https://www.who.int/denguecontrol/epidemiology/en/>
25. Weaver, S.C. Arrival of Chikungunya virus in the New World: prospects for spread and impact on public health. *PLoS Neglected Tropical Diseases*. **2014**, 8 (6), p. 2921–2925.
26. Organización Mundial de la Salud (OMS). Chikungunya. Centro de Prensa- Notas descriptivas OMS en línea: Washington, 2017. [Citado el 5 de Mayo de **2019**]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/chikungunya>
27. Centers for Disease Control and Prevention (CDC)., Organización Panamericana de la Salud (OPS) y Organización Mundial de la Salud (OMS). Preparación y respuesta ante la eventual introducción del virus chikungunya en las Américas. OPS en línea: Washington. **2011**, p. 43–67.
28. Weaver, S.C., Lecuit, M.D. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. *The new England Journal of Medicine*. **2015**, 372, p. 1231–1239.
29. Laras, K., Sukri, N.C., Larasati, R.P., Bangs, M.J., Kosim, R., Djauzi, S., Wandra, T., Master, J., Kosasih, H., Hartati, S., Beckett, C., Sedyaningsih, E. R., Beecham, J. III., Corwin, A.L. Tracking the re-emergence of epidemic chikungunya virus in Indonesia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **2005**, 99 (2), p. 128–141.
30. Pastorino, B., Muyembe-Tamfum, J.J., Bessaud, M., Tock, F., Tolou, H., Durand, J.P., Peyrefitte, C.N. Epidemic resurgence of Chikungunya virus in Democratic Republic of the Congo: identification of a new Central African strain. *Journal of Medical Virology*. **2004**, 74, p. 277–282.
31. Organización Panamericana de la Salud (OPS)., Organización Mundial de la Salud (OMS). Chikungunya. Boletín informativo Marzo. OPS-OMS en línea: Washington, 2014. [Citado el 7 de Mayo de **2019**]. Disponible en: www.paho.org/chikungunya
32. Organización Panamericana de la Salud (OPS)., Organización Mundial de la Salud (OMS). Chikungunya. Número de Casos reportados de fiebre de chikungunya en las Américas – SE 52 (30 de Diciembre de 2016). OPS-OMS en línea: Washington, 2016. [Citado el 7 de Mayo de **2019**]. Disponible en: https://www.paho.org/hq/index.php?option=com_topics&view=article&id=343&Itemid=40931&lang=es

33. Organización Panamericana de la Salud (OPS)., Organización Mundial de la Salud (OMS). Chikungunya. Número de Casos reportados de fiebre de chikungunya en las Américas – SE 51 (22 de Diciembre de 2017). OPS-OMS en línea: Washington, 2017. [Citado el 7 de Mayo de **2019**]. Disponible en: https://www.paho.org/hq/index.php?option=com_topics&view=article&id=343&Itemid=40931&lang=es
34. Liu-Helmersson, L., Quam, M., Wilder-Smith, A., Stenlund, H., Ebia, K., Massad, E., Rocklöv, J. Climate Change and *Aedes* Vectors: 21st Century Projections for Dengue Transmission in Europe. *EBio Medicine*. **2016**, 7, p. 267–277.
35. Liu-Helmersson, J., Rocklöv, J., Sewe, M., Brännström, A. Climate change may enable *Aedes aegypti* infestation in major European cities by 2100. *Environmental Research*. **2019**, 172, p. 693–699.
36. Guerbois, M., Fernandez-Salas, I., Azar, S.R., Danis-Lozano, R., Alpuche-Aranda, C.M., Leal, G., Garcia-Malo, I.R., Diaz-Gonzalez, E.E., Casas-Martinez, M., Rossi, S.L., Del Río-Galván, S.L., Sanchez-Casas, R.M., Roundy C.M., Wood, T.G., Widen, S.G., Vasilakis, N., Weaver, S.C. Outbreak of Zika virus infection, Chiapas State, Mexico, 2015, and first confirmed transmission by *Aedes aegypti* mosquitoes in the Americas. *Journal Infectious Diseases*. **2016**, 214 (9), p. 1349–1356.
37. Boyer, S., Calvez, E., Chouin-Carneiro, T., Diallo, D., Anna-Bella, F. An overview of mosquito vectors of Zika virus. *Microbes and Infection*. **2018**, 20 (11-12), p. 646–660.
38. Kauffman, E.B., Kramer, L.D. Zika Virus Mosquito Vectors: Competence, Biology, and Vector Control. *The Journal of Infectious Diseases*. **2017**, 216, p. 976–990.
39. Organización Panamericana de la Salud (OPS)., Organización Mundial de la Salud (OMS). Zika., Plataforma de Información en Salud para las Américas (PLISA), SE 53. OPS-OMS en línea: Washington, 2016. [Citado el 7 de Mayo de **2019**]. Disponible en: <http://www.paho.org/data/index.php/es/>
40. Weger-Lucarelli, J., Rückert, C., Chotiwan, N., Nguyen, C., Luna, S.M.G., Fauver, J.R., Foy, B.D., Perera, R., Black, W.C., Kading, R.C., Ebel, G.D. Vector competence of American mosquitoes for three strains of Zika virus. *PLOS Neglected Tropical Diseases*. **2016**, 10 (10), p. 1–16.
41. Chouin-Carneiro, T., Vega-Rua, A., Vazeille, M., Yebakima, A., Girod, R., Goindin, D., Dupont-Rouzeyrol, M., Lourenço-de-Oliveira, R., Failloux, A.B. Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika virus. *PLoS Neglected Tropical Diseases*. **2016**, 10 (3), p.1–11.
42. Organización Mundial de la Salud (OMS). Información General sobre el virus del zika (ZIKV) y sus complicaciones. OMS en línea: Washington, 2017. [Citado el 8 de Mayo de **2019**]. Disponible en: <https://www.who.int/features/qa/zika/es/>

43. Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman, R.K., Vesnaver-Vipotnik, T., Fabjan, V.V., Vizjak, A., Pižem, J., Petrovec, M., Avšič, Ž.T. Zika virus associated with microcephaly. *The New England Journal of Medicine*. **2016**, 374, p.951–958.
44. Driggers, R.W., Cheng-Ying Ho, M.D., Korhonen, E.M., Kuivanen, S., Jääskeläinen, A.J., Smura, T., Rosenberg, A., Hill, D.A., DeBiasi, R.L., Vezina, G., Timofeev, J., Rodriguez, F.J. Zika virus infection with prolonged maternal viremia and fetal brain abnormalities. *The New England of Journal Medicine*. **2016**, 374, p. 2142–2151.
45. World Health Organization (WHO). Zika casualty statement: zika virus infection: update on the evidence for a causal link to congenital brain abnormalities and Guillain-Barré syndrome. WHO online: Washington, 2016. [Citado el 8 de Mayo de **2019**]. Disponible en: <https://www.who.int/emergencies/zika-virus/causality/en/>
46. World Health Organization (WHO). WHO Zika virus research agenda. WHO reference number: WHO/ZIKV/PHR/16.1. WHO Press: Geneva, 2016. [Citado el 8 de Mayo de **2019**]. Disponible en: <https://www.who.int/reproductivehealth/zika/zika-virus-research-agenda/>
47. Service, M. *Medical Entomology for Students*, 5 Ed., Cambridge University Press: New York. **2012**, p. 303–310.
48. Powell, J.R., Tabachnick, W.J., Arnold, A. Genetics and the origin of a vector population – *Aedes aegypti*, a case-study. *Science*. **1980**, 208, p. 1385–1387.
49. Moore, M., Sylla, M., Goss, L., Burugu, M.W., Sang, R., Kamau, L.W., Kenya, E.U., Bosio, C., Munoz, M. de L., Sharakova, M., Black, W.C. Dual African origins of global *Aedes aegypti* s.l. populations revealed by mitochondrial DNA. *PLoS Neglected Tropical Diseases*. **2013**, 7 (4), p. 2175–2182.
50. Powell, J.R. Tabachnick, W.J. History of domestication and spread of *Aedes aegypti* – A review. *Memórias do Instituto Oswaldo Cruz*. **2013**, 108, p. 11–17.
51. Integrated Taxonomic Information System (ITIS). *Aedes aegypti* L. Mosquito Classification Comparison. The Walter Reed Biosystematics Unit, 2013. [Citado el 15 de Mayo de **2019**]. Disponible en: [www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=126240; WRBU \(2014\)](http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=126240; WRBU (2014))
52. Service, M. *Medical Entomology for Students*, 5 Ed. Cambridge University Press: New York. **2012**, p. 303.
53. Clements, A.N. The Biology of Mosquitoes, Vol. 1. In *Development, Nutrition and Reproduction*. 2 Ed. CABI Publishing: Oxford. **2000**.

54. García, I., Araujo-Muñoz, B., Amaya-Aguirre, I., Roldán, I.P., García-Moreno, A., Refoyo, P.R. Manual de Laboratorio de Parasitología: 12 Insectos Dípteros. *Reduca Biología-Serie Parasitología*. **2009**, 2 (5), p. 80–100.
55. Organization for Economic Co-operation and Development (OECD). Safety Assessment of Transgenic Organisms in the Environment, Vol 8: OECD Consensus Document of the Biology of Mosquito *Aedes aegypti*, Harmonisation of Regulatory Oversight in Biotechnology, OECD Publishing: Paris. **2018**.
56. Bar, A. Andrew, J. Morphology and morphometry of *Aedes aegypti* adult mosquito. *Annual Review and Research in Biology*, **2013**, 3, p. 52–69.
57. National Environment Agency (NEA). *Wolbachia-Aedes Mosquito Suppression Strategy: male mosquitoes do not bite*. NEA copyright. [Citado el 18 de Mayo de **2019**]. Disponible en: <https://www.nea.gov.sg/corporate-functions/resources/research/wolbachia-aedes-mosquito-suppression-strategy/male-mosquitoes-do-not-bite>
58. Farnesi, L.C., Martins, A.J., Valle, D., Rezende, G.L. Embryonic development of *Aedes aegypti* (Diptera: Culicidae): influence of different constant temperatures. *Memórias do Instituto Oswaldo Cruz*. **2009**, 104, p. 124–126.
59. Juliano, S.A., O’Meara, G.F., Morrill, J.R., Cutwa, M.M. Desiccation and thermal tolerance of eggs and the coexistence of competing mosquitoes. *Oecologia*. **2002**, 130, p. 458–469.
60. Couret, J., Dotson, E.M., Benedict, M. Temperature, larval diet, and density effects on development rates and survival of *Aedes aegypti* (Diptera: Culicidae). *PLoS One*. **2014**, 9 (2), doi: 10.1371/journal.pone.0087468.
61. Ponlawat, A., Harrington, L.C. Factors associated with male mating success of the dengue vector mosquito, *Aedes aegypti*. *The American Society of Tropical Medicine and Hygiene*. 2009, 79(3), p. 312–318.
62. Bar, A. Andrew, J. Morphology and morphometry of *Aedes aegypti* larvae. *Annual Review and Research in Biology*. **2013**, 3, p. 1–21.
63. Brady, O.J. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasite Vectors*. **2013**, 6 (351), doi:10.1186/1756-3305-6-351.
64. Rubio, A., Cardo, M.V., Vezzani, D. Tire-breeding mosquitoes of public health importance along an urbanization gradient in Buenos Aires, Argentina. *Memórias do Instituto Oswaldo Cruz*. **2011**, 106 (6), p. 678–684.
65. Sirot, L.K., Poulson, R.L., Caitlin McKenna, M., Girnary, H., Wolfner, M.F., Harrington, L.C. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female

- feeding and reproduction. *Insect Biochemistry and Molecular Biology*. **2008**, 38, p. 176–189.
66. Boyer, S., Toty, C., Jacquet, M., Lemperiere, G., Fontenille, D. Evidence of multiple inseminations in the field in *Aedes albopictus*. *PLoS One*. **2012**, 7, doi:10.1371/journal.pone.0042040
 67. Conde-Osorio, A.M. Estudio de la longevidad y el ciclo gonotrófico del *Aedes* (*Stegomyia*) *aegypti* (LINNAEUS, 1762), cepa Girardot (Cundinamarca) en condiciones de laboratorio. Pontificia Universidad Javeriana, Facultad de Ciencias, Carrera de Biología. Trabajo de grado: Bogotá, **2003**, p. 7–67.
 68. Nyasembe, V.O., Tchouassi, D.P., Mbogo, C.M., Sole, C.L., Pirk, C., Torto, B. Linalool oxide: Generalist plant-based lure for mosquito disease vectors. *Parasites and Vectors*, **2015**, 8 (581), p. 1–8.
 69. Ponlawat, A., Harrington, L.C. Blood feeding patterns of *Aedes aegypti* and *Aedes albopictus* in Thailand. *Journal of Medical Entomology*. **2005**, 42, p.844–849.
 70. World Health Organization (WHO). Global strategy for dengue prevention and control 2012–2020. WHO press: Geneva. **2012**.
 71. Cummins, B., Cortez, R., Foppa, I.M., Walbeck, J., Hyman, J.M. A spatial model of mosquito host-seeking behavior. *PLoS Computational Biology*. **2012**, 8 (5), doi:10.1371/journal.pcbi.1002500.
 72. Smallegange, R.C., Verhulst, N.O., Takken, W., Sweaty skin: an invitation to bite?. *Trends in Parasitology*. **2011**, 27, p. 143–148.
 73. Reiter, P. Oviposition, dispersal, and survival in *Aedes aegypti*: implications for the efficacy of control strategies. *Vector-borne Zoonotic Diseases*. **2007**, 7, p. 261–273.
 74. David, M.R., Lourenco-de-Oliveira. R., Freitas, R.M. Container productivity, daily survival rates and dispersal of *Aedes aegypti* mosquitoes in a high income dengue epidemic neighbourhood of Rio de Janeiro: presumed influence of differential urban structure on mosquito biology. *Memórias do Instituto Oswaldo Cruz*. **2009**, 104, p. 927–932.
 75. Bagny, L., Delatte, H., Elissa, N., Quilici, S., Fontenille, D. *Aedes* (Diptera: Culicidae) vectors of arboviruses in Mayotte (Indian Ocean): distribution area and larval habitats. *Journal of Medical Entomology*. **2009**, 46, p. 198–207.
 76. Hardy, J.L., Houkm E.J., Kramer, L.D., Reeves, W.C. Intrinsic factors affecting vector competence of mosquitoes for arboviruses. *Annual Review of Entomology*. **1983**, 28, p. 229–262.
 77. Beerntsen, B.T., James, A.A., Christensen, B.M. Genetics of mosquito vector competence. *Microbiology and Molecular Biology Review*. **2000**, 64, p. 115–137.

78. Ye, Y.H., Chenoweth, S.F., Carrasco, A.M., Allen, S.L., Frentiu, F.D., van den Hurk, A.F., Beebe, N.W., McGraw, E.A. Evolutionary potential of the extrinsic incubation period of dengue virus in *Aedes aegypti*. *Evolution*. **2016**.
79. Clements, A.N. The physiology of mosquitoes. In: *International Series of Monographs on pure and applied biology*. The Macmillan Company: New York. **1963**. p. 393.
80. Richards, A.G., Richards, P.A. The peritrophic membranes of insects. *Annual Review Entomology*. **1977**, 22, p. 219–240
81. Kato, N., Mueller, C.R., Fuchs, J.F., McElroy, K., Wessely, V., Higgs, S., Christensen, B.M. Evaluation of the function of a type I peritrophic matrix as a physical barrier for midgut epithelium invasion by mosquito-borne pathogens in *Aedes aegypti*. *Vector Borne Zoonotic Diseases*. **2008**, 8, p. 701–712.
82. Salazar, M.I., Richardson, J.H., Sánchez-Vargas, I., Olson, K.E., Beaty, B.J. Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes. *BMC Microbiology*. **2007**, 7 (9), doi: 10.1186/1471-2180-7-9.
83. Forrester, N.L., Coffey, L.L., Weaver, S.C. Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. *Viruses*. **2014**, 6, p. 3991–4004.
84. Kramer, L.D., Ebel, G.D. Dynamics of flavivirus infection in mosquitoes. *Advances in Virus Research*. **2003**, 60, p. 187–232
85. Schmid, M.A., Glasner, D.R., Shah, S., Michlmayr, D., Kramer, L.D., Harris, E. Mosquito saliva increases endothelial permeability in the skin, immune cell migration, and dengue pathogenesis during antibody-dependent enhancement. *PLoS Pathog*. **2016**;12:e1005676
86. Black, W.C., Bennett, K.E., Gorrochotegui-Escalante, N., Barillas-Mury, C., Fernandez-Salas, I., Munoz, M.L., Farfan, J.A., Olson, K.E., Beaty, B.J. Flavivirus susceptibility in *Aedes aegypti*. *Archives of Medical Research*. **2002**, 33, p. 379–388.
87. Tabachnick, W.J. Nature, nurture and evolution of intra-species variation in mosquito arbovirus transmission competence. *International Journal of Environment Research and Public Health*. **2013**, 10 (1), p. 249–277.
88. Failloux, A.B., Vazeille, M., Rodhain, F. Geographic genetic variation in populations of the dengue virus vector *Aedes aegypti*. *Journal of Molecular Evolution*. **2002**, 55, p.653–663.
89. Gonçalves, C.M., Melo, F.F., Bezerra, J.M.T., Chaves, B.A., Silva, B.M., Silva, L.D., Pessanha, J.E., Arias, J.R., Secundino, N.F., Norris, D.E., Pimenta, P.F. Distinct variation in vector competence among nine field populations of *Aedes aegypti*

- from a Brazilian dengue-endemic risk city. *Parasite Vector*. **2014**, 7 (320), doi:10.1186/1756-3305-7-320.
90. Foster, W.A., Walker, E.D. Mosquitoes (Culicidae)", in Mullen, G. & Durden, L. eds., *Medical and Veterinary Entomology*. Academic Press: San Diego. **2002**, p. 203-262.
 91. World Health Organization (WHO). Control of dengue vectors: Tools and strategies. In: *Report of the Scientific Working Group Meeting on Dengue*. WHO: Geneva. **2007**, p. 110–119.
 92. World Health Organization (WHO). Guidelines for Dengue Surveillance and Mosquito Control. Regional Office for the Western Pacific. WHO Press: Manila. **1995**.
 93. Lacey, L.A. *Bacillus thuringiensis* serovariety israelensis and *Bacillus sphaericus* for mosquito control. *Journal of the American Mosquito Control Association*, **2007**, 23 (2), p. 133–163.
 94. Boyce, R. Lenhart, A., Kroeger, A., Velayudhan, R., Roberts, B., Horstick, O. *Bacillus thuringiensis israelensis* (Bti) for the control of dengue vectors: Systematic literature review. *Tropical Medicine & International Health*, **2013**, 18 (5), p. 564–577.
 95. World Health Organization (WHO). Space Spray Application of Insecticides for Vector and Public Health Pest Control A Practitioner’s Guide, Communicable Disease Control, Prevention and Eradication WHO Pesticide Evaluation Scheme (WHOPES). World Health Organization: Geneva. **2003**
http://apps.who.int/iris/bitstream/handle/10665/68057/WHO_CDS_WHOPES_GCDPP_2003.5.pdf?sequence=1&isAllowed=y
 96. Venkatesh, A. Tyagi, B.K. Predatory potential of *Bradinopyga* 118 ocalize and *Ceriagrion coromandelianum* larvae ON dengue vector *Aedes aegypti* under controlled conditions (Anisoptera: Libellulidae; Zygoptera: Coenagrionidae; Diptera: Culicidae). *Odonatologica*. **2013**, 42 (2), p. 139–149.
 97. Kay, B.H., Nam, V.S., Tien, T.V., Yen, N.T., Phong, T.V., Diep, V.T., Ninh, T.U., Bektas, A., Aaskov, J.G. Control of *Aedes* vectors of dengue in three provinces of Vietnam by use of *Mesocyclops* (Copepoda) and community-based methods validated by entomologic, clinical, and serological surveillance. *The American Journal of Tropical Medicine and Hygiene*. **2013**, 66, p. 40–48.
 98. McMeniman, C.J., Lane, R.V., Cass, B.N., Fong, A.W., Sidhu, M., Wang, Y.F., O'Neill, S.L. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*". *Science*. **2009**, 323, p. 141–144.
 99. McMeniman, C.J. Lane, A.M., Fong, A.W.C., Voronin, D.A., Iturbe-Ormaetxe, I., Yamada, R., McGraw, E.A., O'Neill, S.L. Host adaptation of a *Wolbachia* strain

- after long-term serial passage in mosquito cell lines. *Applied and Environmental Microbiology*. **2008**, 74, p. 6963–6969.
100. Ricci, I., Cancrini, G., Gabrielli, S., D'amelio, S., Favia, G. Searching for *Wolbachia* (Rickettsiales: Rickettsiaceae) in mosquitoes (Diptera: Culicidae): Large polymerase chain reaction survey and new identifications. *Journal of Medical Entomology*, **2002**, 39, p. 562–567.
 101. Moreira, L.A., Iturbe-Ormaetxe, I., Jeffery, J.A., Lu, G., Pyke, A.T., Hedges, L.M., Rocha, B.C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L.E., Johnson, K.N., Kay, B.H., McGraw, E.A., van den Hurk, A.F., Ryan, P.A., O'Neill, S.L. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya and Plasmodium. *Cell*, **2009**, 139, p. 1268–1278.
 102. Nguyen, T.H., Nguyen, H.L., Nguyen, T.Y., Vu, S.N., Tran, N.D., Le, T.N., Vien, Q.M., Bui, T.C., Le, H.T., Kutcher, S., Hurst, T.P., Duong, T.T.H., Jeffery, J.A.L., Darbro, J.M. Field evaluation of the establishment potential of wMelPop *Wolbachia* in Australia and Vietnam for dengue control. *Parasites and Vectors*. **2015**, 8 (1), p. 1–14.
 103. Dicko, A.H. Using species distribution models to optimize vector control: The tsetse eradication campaign in Senegal. *Proceedings of the National Academy of Sciences*. **2014**, 111, p. 10149–10154.
 104. Alphey, L. Genetic Control of Mosquitoes. *Annual Review of Entomology*. **2014**, 59, p. 205–224.
 105. Harris, A.F. Lancelot, R., Seck, M.T., Guerrini, L., Sall, B., Lo, M., Vreysen, M.J., Lefrançois, T., Fonta, W.M., Peck, S.L., Bouyer, J. Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology*, **2012**, 30, p. 828–830.
 106. Carvalho, D.O., McKemey, A.R., Garziera, L., Lacroix, R., Donnelly, C.A., Alphey, L., Malavasi, A., Capurro, M.L. Suppression of a field population of *Aedes aegypti* in Brazil by sustained release of transgenic male mosquitoes. *PLoS Neglected Tropical Diseases*, **2015**, 9 (7), doi:10.1371/journal.pntd.0003864.
 107. Gorman, K., Young, J., Pineda, L., Márquez, R., Sosa, N., Bernal, D., Torres, R., Soto, Y., Lacroix, R., Naish, N., Kaiser, P., Tepedino, K., Philips, G., Kosmann, C., Cáceres, L. Short-term suppression of *Aedes aegypti* using genetic control does not facilitate *Aedes albopictus*. *Pest Management Sciences*. **2016**, 72 (3), p. 618–628.
 108. World Health Organization (WHO). Manual on Environmental Management for Mosquito Control with Special Emphasis on Malaria Vectors, World Health Organization: Geneva. **1982**.

109. Pan American Health Organization (PAHO). Dengue and Dengue Hemorrhagic Fever in the Americas: Guidelines for Prevention and Control, PAHO Scientific Publication 548, Pan American Health Organization: Washington DC. **1994**.
110. Erlanger, T., Keiser, L., Utzinger, J. Effect of dengue vector control interventions on entomological parameters in developing countries: A systematic review and meta-analysis. *Medical and Veterinary Entomology*. **2008**, 22, p. 203–221.
111. Bouyer, J., Lefrançois, T. Boosting the sterile insect technique to control mosquitoes. *Trends in Parasitology*. **2014**, 30, p. 271–273.
112. Mains, J.W., Brelsfoard, C.L., Dobson, S.L. Male mosquitoes as vehicles for insecticide. *PLoS Neglected Tropical Diseases*. **2015**, 9, doi:10.1371/journal.pntd.0003406.
113. Vazquez-Prokopec, G.M., Montgomery, B.L., Horne, P., Clennon, J.A., Ritchie, S.A. Combining contact tracing with targeted indoor residual spraying significantly reduces dengue transmission. *Science Advances*. **2017**, 3 (2), doi: 10.1126/sciadv.1602024.
114. Esu, E., Lenhart, A., Smith, L., Horstick, O. Effectiveness of peridomestic space spraying with insecticide on dengue transmission; Systematic review. *Tropical Medicine and International Health*. **2010**, 15, p. 619–631.
115. Ritchie, S.A. Devine, D.J. Confusion, knock-down and kill of *Aedes aegypti* using metofluthrin in domestic settings: A powerful tool to prevent dengue transmission?. *Parasites & Vectors*. **2013**, 6, p. 1–9.
116. World Health Organization (WHO). WHO Recommended Insecticides for Space Spraying Against Mosquitoes. WHO Press: Geneva. **2012**. www.who.int/whopes/Insecticides_for_space_spraying_Jul_2012.pdf.
117. World Health Organization (WHO). *Global Plan for Insecticide Resistance Management in Malaria Vectors (GPIRM)*, WHO: Geneva. **2012**.
118. Brogdon, W.G., McAllister, J.C. Insecticide resistance and vector control. *Emerging Infectious Diseases*. **1998**, 4, p. 605–613.
119. Hemingway, J., Ranson, H. Insecticide resistance in insect vectors of human disease. *Annual Review of Entomology*. **2000**, 45, p. 371–391.
120. Liu, N., Zhu, F., Qiang, X., Zhang, L. Pyrethroid resistance in mosquitoes. *Insect Science*. **2006**, 13, p. 159–166.
121. WHO Pesticide Evaluation Scheme (WHOPES). Guidelines for Laboratory and Field Testing of Mosquito Larvicides, World Health Organization Pesticide Evaluation Scheme: Geneva. **2005**.

122. Horstick, O., Runge-Ranzinger, S., Nathan, M.B., Kroeger, A. Dengue vector-control services: how do they work? A systematic literature review and country case studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **2010**, 104, p. 379–386.
123. Eto, M. Organophosphorus Pesticides: Organic and Biological Chemistry, CRC Press: Boca Raton. **1974**.
124. Khambay, B.P.S. Pyrethroid insecticides. *Pesticide Outlook*. **2012**, 13, p. 49-54
125. Nauen, R. Insecticide mode of action: Return of the ryanodine receptor". *Pest Management Science*. **2006**, 62, p. 690–692.
126. Despres, L., David, J.P., Gallet, C. The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*. **2007**, 22, p. 298–307.
127. Flores, A.E., Fernandez-Salas, I., Albeldaño-Vázquez, W., Badii, M.H. Elevated alfa-esterases levels associated with permethrin tolerance in *Aedes aegypti* (L.) from Baja California, Mexico. *Pesticide Biochemistry and Physiology*. **2005**, 82, p. 66–78.
128. Flores, A.E., Grajales, J.S., Salas, I.F., Garcia, G.P., Becerra, M.H., Lozano, S., Brogdon, W.G., Black, W.C., Beaty, B. Mechanisms of insecticide resistance in field populations of *Aedes aegypti* (L.) from Quintana Roo, Southern Mexico. *Journal of the American Mosquito Control Association*, **2006**, 22 (4), p. 672–677.
129. Hemingway, J., Hawkes, N.J., McCarroll, L., Ranson, H. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochemistry and Molecular Biology*. **2004**, 34, p. 653–665.
130. Ranson, H., Hemingway, J. Mosquito glutathione transferases. *Methods in Enzymology*, **2005**, 401, p. 226–241.
131. Sieglaff, D.H., Duncan, K.A., Brown, M.R. Expression of genes encoding proteins involved in ecdysteroidogenesis in the female mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology*. **2005**, 35, p. 471–490.
132. Fournier, D. Mutations of acetylcholinesterase which confer insecticide resistance in insect populations. *Chemico-Biological Interactions*. **2005**, 15, p. 15, 257–261.
133. Gupta, R.C. Toxicology of Organophosphate & Carbamate Compound. 1 Ed. Elsevier Academic Press: Oxford. **2006**.
134. Davies, T.G.E., Field, L.M., Usherwood, P.N., Williamson, M.S. DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life*. **2007**, 59, p. 151–162.
135. Rinkevich, F.D., Du, Y.Z., Dong, K. Diversity and convergence of sodium channel mutations involved in resistance to pyrethroids", *Pesticide Biochemistry and Physiology*. **2013**, 106, p. 93–100.

136. Vontas, J., Kioulo, E., Pavlidi, N., Morou, E., della Torre, A., Ranson, H. Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*. *Pesticide Biochemistry and Physiology*. **2012**, 104, p. 126–131.
137. Saavedra-Rodriguez, K., Urdaneta-Marquez, L., Rajatileka, S., Moulton, M., Flores, E.A., Fernandez-Salas, I., Bisset, J., Rodriguez, M., McCall, P.J., Donnelly, M.J., Ranson, H., Hemingway, J., Black, W.C. A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American *Aedes aegypti*. *Insect Molecular Biology*. **2007**, 16 (6), p. 785–798.
138. McAllister, J.C., Godsey, M.S., Scott, M.L. Pyrethroid resistance in *Aedes aegypti* and *Aedes albopictus* from Port-au-Prince, Haiti. *Journal of Vector Ecology*. **2012**, 37, pp. 325–332.
139. Maestre-Serrano, R., Gomez-Camargo, D., Ponce-Garcia, G., Flores, A.E. Susceptibility to insecticides and resistance mechanisms in *Aedes aegypti* from the Colombian Caribbean Region. *Pesticide Biochemistry and Physiology*. **2014**, 116, p. 63–73.
140. Alvarez, L.C., Ponce, G., Saavedra-Rodriguez, K., Lopez, B., Flores, A.E. Frequency of V1016I and F1534C mutations in the voltage-gated sodium channel gene in *Aedes aegypti* in Venezuela. *Pest Management Science*, **2015**, 71 (6), p. 863–869.
141. Brengues, C., Hawkes, N.J., Chandre, F., McCarroll, L., Duchon, S., Guillet, P., Manguin, S., Morgan, J.C., Hemingway, J. Pyrethroid and DDT cross-resistance in *Aedes aegypti* is correlated with novel mutations in the voltage-gated sodium channel gene. *Medical and Veterinary Entomology*. **2013**, 17 (1), p. 87–94.
142. Valles, S.M., Dong, K., Brenner, R.J. Mechanisms responsible for cypermethrin resistance in a strain of German cockroach, *Blattella germanica*. *Pesticide Biochemistry and Physiology*. **2000**, 66, p. 195–205.
143. Ahmad, M., Denholm, I., Bromilow, R.H. Delayed cuticular penetration and enhanced metabolism of deltamethrin in pyrethroid-resistant strains of *Helicoverpa armigera* from China and Pakistan. *Pest Management Science*. **2006**, 62, p. 805–810.
144. Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., Field, L.M., Millar, N.S., Williamson, M.S., Bass, C. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet*. **2010**, 6, doi:10.1371/journal.pgen.1000999.
145. Grieco, J.P., Achee, N.L., Chareonviriyaphap, T., Suwonkerd, W., Chauhan, K., Sardelis, M.R., Roberts, D.R. A new classification system for the actions of IRS chemicals traditionally used for malaria control. *PLoS ONE*. **2007**, 2 (8), doi:10.1371/journal.pone.0000716.

146. Rajatileka, S., Black, W.C., Saavedra-Rodriguez, K., Trongtokit, Y., Apiwathnasorn, C., McCall, P.J., Ranson, H. Development and application of a simple colorimetric assay reveals widespread distribution of sodium channel mutations in Thai populations of *Aedes aegypti*. *Acta Tropica*. **2008**, 108 (1), p. 54–57.
147. Anastas, P.T., Warner, J.C. *Green Chemistry: Theory and Practice*. Oxford University Press: New York, **1998**, p.30.
148. Scotti, L., Ishiki, H.M., Cavalcante Duarte, M., Branquinho Oliveira, T., Scotti, M.T. *Computational Approaches in Multitarget Drug Discovery*. In Nicolotti, H., Ed. *Computational Toxicology: Methods and Protocols*. Humana Press: Bari. **2018**, p. 327–345.
149. Demchuk, E., Ruiz, P., Chou, S., Fowler, B.A. SAR/QSAR methods in public health practice. *Toxicology and Applied Pharmacology*. **2011**, 254, p. 192–197.
150. Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To?. *Journal of Medicinal Chemistry*. **2014**, 57, p. 4977–5010.
151. De Sousa, D.P., Viera, Y.W., Uliana, M.P., Melo, M.A., Brocksom, T.J., Cavalcanti, S.C.H. Larvicidal activity of para-benzoquinone. *Parasitology Research*. **2010**, 107, p. 741–745.
152. Santos, S.R.L., Melo, M.A., Cardoso, A.V., Santos, R.L.C., de Sousa, D.P., Cavalcanti, S.C.H. Structure-activity relationships of larvicidal monoterpenes and derivatives against *Aedes aegypti* Linn. *Chemosphere*. **2011**, 84, p. 150–153.
153. Alencar Filho, E.B., Castro Silva JW and Cavalcanti SCH, Quantitative structure-toxicity relationships and molecular highlights about *Aedes aegypti* larvicidal activity of monoterpenes and related compounds. *Medical Chemistry Research*. **2016**, 25, p. 2171–2178.
154. Scotti, L., M.T. Scotti, V.B. Silva, S.R.L. Santos, S.C.H. Cavalcanti, and F.J.B. Mendonça Jr, Chemometric studies on potential larvicidal compounds against *Aedes aegypti*. *Medical Chemistry*. **2014**, 10, p. 201–210.
155. De, P., Aher, R.B., Roy, K. Chemometric modeling of larvicidal activity of plant derived compounds against zika virus vector *Aedes aegypti*: application of ETA indices. *RSC Advances*. **2018**, 8, p. 4662–4670.
156. Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*. **2011**, 32 (7), 1466–1474.
157. Dos Santos, I.M., Gomes-Agra, J.P., Cavalcante de Carvalho, T. G., Azevedo Maia, G.L., Alencar Filho, E.B. Classical and 3D QSAR studies of larvicidal monoterpenes

- against *Aedes aegypti*: new molecular insights for the rational design of more active compounds. *Structural Chemistry*. **2018**, 29(5), p. 1287–1297.
158. Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., Matsushita, Y. Simple method of calculating octanol/water partition coefficient. *Chemical and Pharmaceutical Bulletin*. **1992**, 40 (1), p. 127–130.
159. Song, X., Zapata, A., Hoerner, J., de Dios, A.C., Casabianca, L., Eng, G. Synthesis, larvicidal, QSAR and structural studies of some triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates. *Applied Organometallic Chemistry*. **2007**, 21, p. 545–550.
160. Hansch, C., Verma, R.P. Larvicidal activities of some organotin compounds on mosquito larvae: A QSAR study. *European Journal of Medicinal Chemistry*. **2009**, 44 (1), p. 260–273.
161. Duong, Q., Song, X., Mitrojorgji, E., Gordon, S., Eng, G. Larvicidal and structural studies of some triphenyl- and tricyclohexyltin para-substituted benzoates. *Journal of Organometallic Chemistry*. **2006**, 691, p. 1775–1779.
162. Song, X., Duong, Q., Mitrojorgji, E., Zapata, A., Nguyen, N., Strickman, D., Glass, J., Eng, G. Synthesis, structure characterization and larvicidal activity of some tris-(para-substitutedphenyl) tins. *Applied Organometallic Chemistry*. **2004**, 18 p. 363–368.
163. Freitas, M.P., Ramalho, T.C., da Cunha, E.F.F., Cormanich, R.A. *Multivariate image analysis applied to QSAR as a tool for mosquitoes control: Dengue and yellow Fever*. In Ramalho, T.C., Freitas, M.P., da Cunha, E.F.F. Eds. *Chemoinformatics: Directions Toward Combating Neglected Diseases*. Bentham Science Publishers Ltd: Rio de Janeiro, **2010**, p. 172–182.

CAPITULO 5: APLICACIONES BIOLÓGICAS

QSAR DE INTERÉS EN EL CONTROL DEL *Aedes aegypti* L.

5.1 Introducción

Conforme a la extensa diversidad estructural y funcional de los productos químicos utilizados en la industria farmacéutica, textil, agrícola, alimentaria, cosmética, sanitaria y de cuidado personal, se puede deducir que el campo de análisis de la teoría QSAR es amplio e involucra numerosas áreas de aplicación. Generalmente, las aplicaciones QSAR se encuentran clasificadas según el tipo de efecto final evaluado, ya sea una respuesta final activa, tóxica o fisicoquímica.^{1,2}

En este contexto, las investigaciones QSAR orientadas en predecir la respuesta biológica (actividad), se enfocan sobre todas las sustancias químicas destinadas a la producción de fármacos, suplementos, alimentos, aditivos, conservantes, cosméticos y fitosanitarios. En general, estos compuestos se caracterizan por presentar diversas respuestas activas, tal como antioxidante, repelente, antiparasitaria, anti-HIV, analgésica, antibacterial, antidepresiva, insecticida, anticolinérgica, antimicótico, inhibición del crecimiento u ovicida.

Por otra parte, las propiedades fisicoquímicas han sido supervisadas a través del enfoque de las Relaciones Cuantitativas Estructura–Propiedad (QSPR). Los modelos QSPR evalúan cualquier tipo de producto químico que produzca un efecto sobre los recursos naturales y el entorno.³

El modelado QSPR permite estimar el nivel de peligrosidad o biodisponibilidad de los agentes antropogénicos, a partir de la observación o predicción de diferentes parámetros fisicoquímicos, como el punto ebullición, presión de vapor, coeficiente de partición octanol/agua, refractividad molar, solubilidad acuosa, punto de fusión, entre otros.

El nivel de peligrosidad de los compuestos químicos depende directamente de la dosis, la concentración y el tiempo exposición a los mismos. Por lo tanto, el esfuerzo por anticipar el grado de toxicidad de cualquier tipo de sustancia, involucra diferentes medidas experimentales *in vivo*, *in vitro* e *in situ*, que en general son complejas, costosas y a largo plazo; en especial si se trata de evaluar amplios conjuntos moleculares.⁴

Ante estas limitaciones, las respuestas toxicológicas de las moléculas son supervisadas de forma rápida y acertada por medio de las Relaciones Cuantitativas Estructura–Toxicidad (QSTR). Los estudios QSTR más comunes buscan medir teóricamente la toxicidad aguda en peces, toxicidad acuosa, toxicidad oral, cutánea e inhalatoria aguda, genotoxicidad, citotoxicidad, carcinogenicidad, sensibilidad cutánea, cardiotoxicidad y toxicidad en modelos animales (peces, ratas, ratones, conejos o perros).^{5,6}

En particular, diversas investigaciones orientadas en el modelado QSAR han permitido reducir de forma significativa el tiempo, los costos y el número de ensayos *in vivo* o *in vitro* requeridos para realizar la evaluación del potencial de riesgo o medir la respuesta activa de cualquier sustancia química comercial, nueva o hipotética.

Cabe destacar que los modelos basados en QSAR tienen la capacidad de proveer datos a priori de la actividad contenida en amplios conjuntos de moléculas sin requerir la síntesis y caracterización de las mismas.⁷

Conjuntamente, el creciente interés de los gobiernos y los organismos reguladores por controlar el almacenamiento, distribución y administración de diferentes productos comerciales de uso común clasificados como peligrosos para la salud humana y los ecosistemas, promovió el diseño y aplicación de numerosas herramientas computacionales, como alternativas rápidas y coherentes para obtener información específica y suficiente en la toma de decisiones regulatorias; especialmente ante la ausencia de datos o evidencia experimental.⁸

Una de las características más novedosas de la teoría QSAR es su participación como estrategia computacional asociada al concepto de las 3Rs; el cual busca Reducir el número de animales utilizados para obtener resultados precisos, Reemplazar el uso de mamíferos adultos de gran tamaño por embriones animales; y Replantear los protocolos basados en pruebas *in vivo*, con el fin de excluir la crueldad, el sacrificio o el maltrato animal.⁹⁻¹³

Con base en lo anterior, el presente capítulo de Tesis contiene diferentes estrategias QSAR basadas principalmente en los aspectos no conformacionales de las estructuras moleculares. Cada modelo fue diseñado, validado y aplicado a partir de varios conjuntos moleculares heterogéneos extraídos de la literatura especializada, constituidos por moléculas provenientes de la naturaleza o de modificaciones químicas realizadas en el laboratorio.

Las características estructurales de cada compuesto fueron codificadas por miles de descriptores moleculares generados con diferentes programas de cálculo de descriptores disponibles y de acceso libre. En cada estudio, la búsqueda de la relación estructura–actividad más apropiada se logró a través del enfoque de MLR, supervisado por la técnica RM.

Todos los modelos QSAR cumplieron con diferentes criterios de validación interna y externa. Asimismo, se consideraron los 5 principios propuestos por la Organización para la Cooperación y el Desarrollo Económico (OCDE) para su pertinente aceptación.

Puntualmente, las herramientas QSAR incluidas en el presente capítulo de Tesis fueron desarrolladas con el objetivo de proveer una guía computacional para el diseño racional, selectivo y sustentable de nuevas entidades químicas menos tóxicas, con actividades específicas y efectos favorables para la salud pública mundial y la conservación del medio ambiente.

Resulta importante destacar que entre los fármacos más efectivos para tratar los trastornos neurodegenerativos, como la enfermedad de Alzheimer (AD), se destacan los inhibidores de la enzima AChE (IsAChE). Este tipo de sustancias químicas actúan en el CNS, incrementando la actividad colinérgica cerebral. Este modo de acción también se implementa en el diseño de los pesticidas más novedosos y eficientes para controlar las plagas y los vectores de enfermedades, como el mosquito *A. aegypti* L.

5.2. Análisis QSAR en inhibidores de la enzima acetilcolinesterasa (AChE) derivados de Tacrina

La Enfermedad de Alzheimer (AD), la forma más común de demencia, es un trastorno neurodegenerativo crónico caracterizado por el deterioro cognitivo progresivo y la pérdida de memoria. Acorde con la OMS, en el mundo existen 50 millones de personas que padecen esta enfermedad y cada año se registran aproximadamente 10 millones de casos nuevos, con mayor incidencia en adultos de la tercera edad.

A partir de esto, se estima que el número de pacientes con Alzheimer aumente a 82 millones en 2030 y 152 millones en 2050.¹⁴ En Argentina, esta enfermedad representa la quinta causa de mortalidad en adultos mayores, perjudicando alrededor de 300 mil argentinos hasta la fecha. Asimismo, según la organización Alzheimer's Association, Argentina es el primer país en sur América donde la población mayor de 65 años tiene más probabilidades de desarrollar Alzheimer.^{15,16}

De acuerdo con la hipótesis colinérgica, la AD se atribuye principalmente a la disminución gradual del neurotransmisor acetilcolina (ACh), encargado de transmitir las señales en el proceso sináptico de las células nerviosas.¹⁷ La descomposición de este neurotransmisor se atribuye principalmente a la acción de la AChE. Esta enzima está codificada para hidrolizar el ACh, y así regular su concentración en diferentes regiones del Sistema Nervioso Central (CNS). Con base en esto, todos los procedimientos terapéuticos desarrollados hasta el momento para tratar la AD, se enfocan en la inhibición de esta colinesterasa.^{18,19}

Entre los diferentes fármacos inhibidores de colinesterasa, se destacan la Rivastigmina, Tacrina, Galantamina y Donepezilo, por su eficacia en el aumento de la actividad colinérgica cerebral; estos compuestos presentan sobresaliente acción anticolinesterasa, con valores de Concentración Inhibitoria media (IC₅₀) entre 0.2–22 μM para AChE.

No obstante, diversos estudios han demostrado que además de su alto valor comercial, estos fármacos presentan efectos adversos en el organismo, como náusea, vómito y malestar estomacal; asimismo, exhiben baja selectividad y tendencia a disminuir su actividad inhibitoria a largo plazo.^{20, 21} Por lo tanto, el interés por descubrir nuevos inhibidores selectivos y de fácil acceso económico continúa vigente.

Particularmente, el compuesto Tacrina con estructura molecular exhibida en la Figura 5.1, fue diseñado en la década de los 70s, y aprobado en 1993 por la FDA US como el primer fármaco para tratar la AD.²² Aunque, Tacrina es el fármaco paliativo más antiguo diseñado en función de la hipótesis colinérgica; existe un amplio interés por encontrar nuevos agentes químicos más seguros y accesibles, que superen la actividad anticolinérgica de Tacrina. De hecho, se ha demostrado experimentalmente que los congéneres de bis-tacrina presentan mayor actividad inhibitoria *in vitro* de la AChE con respecto a Tacrina.^{23, 24}

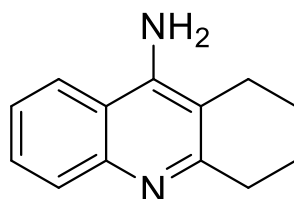


Figura 5.1. Estructura molecular 2D de Tacrina. Fuente: Pubchem CID 1935.

Ante este panorama, la posibilidad de emplear estrategias matemáticas como la teoría QSAR,²⁵ que permitan vincular de forma cuantitativa las estructuras moleculares de los compuestos químicos con una delimitada respuesta biológica final, es una gran ventaja para los investigadores al momento de seleccionar las estructuras más prometedoras de un amplio conjunto molecular donde todos los compuestos son potencialmente activos. Y así, promover que el diseño de medicamentos sea más racional; es decir, que requieran menos tiempo y bajos costos de síntesis, caracterización y ensayo de decenas de moléculas.²⁶

En este contexto, resulta importante destacar que la mayoría de las investigaciones QSAR enfocadas en predecir la actividad inhibitoria de las AChE analizan un gran número de derivados de Tacrina, que involucran medidas de inhibición *in vitro* de la AChE extraída de diferentes fuentes.

Puntualmente, este estudio QSAR fue diseñado para establecer las relaciones cuantitativas de estructura-actividad para diez conjuntos moleculares basados en análogos de Tacrina, medidos bajo diferentes condiciones bioquímicas.

5.2.1. Datos de la actividad inhibitoria de la AChE

Se recolectó de la literatura especializada diez conjuntos de datos compuestos por diferentes clases de nuevo derivados de Tacrina como inhibidores de la enzima AChE proveniente de tres diferentes fuentes animales.

El ensayo *in vitro* de la AChE se basó en el método colorimétrico propuesto por Ellman *et al.*, 1961,²⁷ en la AChE extraída de bovino (bAChE) y de eritrocitos humanos (hAChE); se empleó yoduro de acetiltiocolina de 0.53 o 0.13 mM como sustrado de bAChE y hAChE, respectivamente. Asimismo, se evaluó el poder de inhibición de la AChE proveniente del pez *Electrophorus electricus* (EeAChE), con cloruro de acetilcolina de 29.5 mM como sustrado, a través del método de espectrometría propuesto por Rappaport *et al.*, 1959.

Los conjuntos de datos se identificaron como **a – g**. El conjunto **a** está compuesto por 7 pirano[2,3-b]quinolinas, 6 benzonaftiridinas y Tacrina, como sustancia de referencia,²⁹ con datos experimentales de la inhibición *in vitro* de bAChE.

El conjunto **b** contiene 2 hidroclouros de 6-clorotacrina y ésteres tricíclicos, yoduro de propidio, 10 dihidroclouros de la pirano [3,2-c] quinolina-6-clorotacrina y 6-clorotacrina, con valores de experimentales inhibición de bAChE y hAChE.³⁰

El conjunto **c** consiste en 17 híbridos de tacrina-dihidropiridina, Tacrina y Propidio, que presentan valores de inhibición *in vitro* de EeAChE y hAChE.³¹

El conjunto **d** involucra 8 derivados relacionados con híbridos de donepezil-Tacrina, Tacrina y Donepezilo, con valores experimentales de inhibición de bAChE.³²

El conjunto **e** contiene 12 dímeros basados en Tacrina, bis-(7)-Tacrina y Tacrina con información experimental de inhibición de hAChE.³³

El conjunto **f** está conformado por 12 análogos de Tacrina de 2-aminopiridina-3-carbonitrilos altamente sustituidos y Tacrina, con datos experimental de inhibición de bAChE y EeAChE.³⁴

El conjunto **g** incorpora 18 inhibidores de AChE de sitio de unión doble basados en Tacrina inhibidores, Tacrina, 6-cloro-Tacrina, yoduro de propidio y Donepezilo, con datos experimentales de inhibición de la hAChE³⁵

Todos los valores experimentales de la actividad inhibitoria *in vitro* de la enzima AChE, expresados como $\log_{10} IC_{50}$ para cada conjunto de datos (**a–g**) se encuentran disponibles en el apartado anexo, Tablas 5.2.1 a 5.2.7.

5.2.2. Descriptores moleculares

Las moléculas estudiadas fueron digitalizadas a través del programa HyperChem 6.03.³⁶ Las estructuras moleculares de los compuestos se optimizan en primer lugar con el procedimiento Campo de fuerza de Mecánica Molecular (MM+) incluido en el paquete Hyperchem.

Las geometrías resultantes se refinan aún más por medio del Método Semiempírico PM3 (Método Paramétrico-3), mediante la aplicación del algoritmo de Polak-Ribiere y un límite de gradiente de 0.01 kcal.Å⁻¹.

Se calcularon 4885 descriptores moleculares por medio del programa comercial Dragon versión 6.0,³⁷ el cual genera 19 clases de descriptores 0D-3D, tal como los Constitucionales, Topológicos, Geométricos, Electrónicos, tipo 3D – Morse, WHIM, GETAWAY, RDF, recuentos de caminos molecular, descriptores BCUT, Autocorrelación 2D, funciones de distribución radial, grupos funcionales y fragmentos centrados en el átomo, entre otras clases.

Además, se incorporaron descriptores químico-cuánticos externos al programa Dragon, tal como el momento dipolar molecular, la energía total, la energía HOMO y LUMO y el GAP de energía HOMO-LUMO ($\Delta_{Homo-Lumo}$).³⁸

Luego, se eliminaron los descriptores con valores constantes, casi constantes (solo 1 valor diferente) y datos ausentes; además, se verificó la dependencia lineal de los mismos. Para finalmente obtener una matriz con 1502 descriptores linealmente independientes.

5.2.3. Selección de los modelos de regresión QSAR

La búsqueda de las mejores regresiones lineales se realizó mediante la técnica RM basada MLR, la cual fue programa por nuestro grupo a través del programa Matlab 7.0.^{39, 40} La estrategia RM se utilizó para explorar las características estructurales más representativas de cada subconjunto de calibración de los conjuntos **a-g** para la actividad inhibitoria *in vitro* de la bACHe, hAChE y EeAChE.

De esta manera, los mejores modelos de MLR conformados por $d=1-4$ descriptores moleculares, fueron encontrados a partir un grupo de descriptores con 1502 variables independientes. Siempre aplicando el criterio común de mantener la dimensión (d) del modelo lo más pequeña posible.

5.2.4. Validación interna y externa de los modelos

Las estrategias de validación teóricas utilizadas para evaluar los modelos QSAR, consiste en la Validación Cruzada Interna (CV) Dejar-uno-afuera (LOO), regulada a través de los parámetros R_{LOO}^2 y S_{LOO} .⁴¹ Además, se considera la prueba de Aleatorización-Y,⁴² para verificar la veracidad de los modelos lineales QSAR.

Finalmente, también se aplica la práctica estándar que consiste en apartar del conjunto molecular inicial algunos compuestos para conformar el subconjunto de predicción (N_{pred}). El cual será utilizado para medir la capacidad que tiene cada modelo QSAR obtenido para estimar los valores de actividad de los compuestos del subconjunto *pred*, los cuales no están involucrados durante el ajuste del modelo.

La construcción de cada modelo QSAR se basa únicamente en la información química del subconjunto de calibración (N_{cal}). El criterio de selección de las moléculas que conforman los subconjuntos *cal* y *pred*, consiste en generar subconjuntos balanceados en relación estructura-actividad, a partir de la similaridad estructural de las sustancias derivadas de Tacrina.

5.2.5. Resultados y Discusión

La aplicación del método de selección de variables RM acoplado a MLR sobre la matriz de descriptores con $d=1502$ variables, condujo a los mejores modelos lineales para cada conjunto de datos constituidos con inhibidores de colinesterasa.

Para determinar el número apropiado de descriptores incluidos en la relación estructura-actividad, se consideró el razonamiento de Occam o el concepto de reglas de pulgar, que sugieren incorporar un descriptor por cada 6 o 7 datos del conjunto de calibración (N_{cal}). A continuación, se exhiben los mejores modelos QSAR encontrados para predecir la IC_{50} de cada conjunto molecular (**a-g**).

- Conjunto de datos **a** (bAChE):

$$\log_{10} IC_{50} = 1.34(\pm 2) - 1.05(\pm 0.3)Mor06v - 44.82(\pm 10)G1u \quad (5.1)$$

$$N_{cal} = 12, R = 0.90, R^2 = 0.81, S = 0.29, F = 18.36, p < 10^{-4}, R_{ij\max} = 0.43, o(> 2.S) = 0$$

$$R_{LOO} = 0.84, R_{LOO}^2 = 0.70, S_{LOO} = 0.39, S^{aleat} = 0.31$$

- Conjunto de datos **b** (bAChE):

$$\log_{10} IC_{50} = -7.24(\pm 0.4) + 6.74(\pm 0.4)GATS2e - 4.47(\pm 0.4)R5e \quad (5.2)$$

$$N_{cal} = 12, R = 0.98, R^2 = 0.96, S = 0.23, F = 143.36, p < 10^{-4}, R_{ij\max} = 0.46, o(> 2.S) = 0$$

$$R_{LOO} = 0.96, R_{LOO}^2 = 0.92, S_{LOO} = 0.37, S^{aleat} = 0.30$$

- Conjunto de datos **b** (hACHe):

$$\log_{10} IC_{50} = -18.16(\pm 0.8) + 101.91(\pm 6)Qmean + 35.56(\pm 6)R3m^+ \quad (5.3)$$

$$N_{cal} = 11, R = 0.99, R^2 = 0.98, S = 0.22, F = 179.5, p < 10^{-4}, R_{ijmax} = 0.61, o(> 2.S) = 0$$

$$R_{LOO} = 0.98, R_{LOO}^2 = 0.96, S_{LOO} = 0.27, S^{aleat} = 0.27$$

- Conjunto de datos **c** (EeACHe):

$$\log_{10} IC_{50} = -2.61(\pm 0.3) - 83.23(\pm 8)R5e^+ - 1.59(\pm 0.1)nNHR \quad (5.4)$$

$$N_{cal} = 16, R = 0.97, R^2 = 0.94, S = 0.24, F = 109, p < 10^{-4}, R_{ijmax} = 0.25, o(> 2.S) = 0$$

$$R_{LOO} = 0.95, R_{LOO}^2 = 0.90, S_{LOO} = 0.32, S^{aleat} = 0.50$$

- Conjunto de datos **c** (hACHe):

$$\log_{10} IC_{50} = 16.02(\pm 4) - 52.45(\pm 10)X1A \quad (5.5)$$

$$N_{cal} = 9, R = 0.90, R^2 = 0.81, S = 0.24, F = 28.44, p < 10^{-4}, R_{ijmax} = 0, o(> 2.S) = 0$$

$$R_{LOO} = 0.84, R_{LOO}^2 = 0.70, S_{LOO} = 0.30, S^{aleat} = 0.25$$

- Conjunto de datos **d** (bACHe):

$$\log_{10} IC_{50} = -5.51(\pm 0.1) - 1.26(\pm 0.1)Mor09u - 32.16(\pm 2.0)R2m^+ \quad (5.6)$$

$$N_{cal} = 9, R = 0.98, R^2 = 0.96, S = 0.18, F = 98.81, p < 10^{-4}, R_{ijmax} = 0.30, o(> 2.S) = 0$$

$$R_{LOO} = 0.95, R_{LOO}^2 = 0.90, S_{LOO} = 0.33, S^{aleat} = 0.26$$

- Conjunto de datos **e** (hACHe):

$$\log_{10} IC_{50} = 5.38(\pm 2) - 6.66(\pm 1)BELm2 - 2.69(\pm 0.3)Mor27e \quad (5.7)$$

$$N_{cal} = 12, R = 0.96, R^2 = 0.92, S = 0.29, F = 51.70, p < 10^{-4}, R_{ijmax} = 0.04, o(> 2.S) = 0$$

$$R_{LOO} = 0.92, R_{LOO}^2 = 0.87, S_{LOO} = 0.39, S^{aleat} = 0.46$$

- Conjunto de datos **f** (bACHe):

$$\log_{10} IC_{50} = -17.54(\pm 2) + 1.51(\pm 0.2)AMW - 0.25(\pm 0.2)MEcc \quad (5.8)$$

$$N_{cal} = 11, R = 0.98, R^2 = 0.96, S = 0.25, F = 21.33, p < 10^{-4}, R_{ijmax} = 0.136, o(> 2.S) = 0$$

$$R_{LOO} = 0.96, R_{LOO}^2 = 0.92, S_{LOO} = 0.36, S^{aleat} = 0.39$$

- Conjunto de datos *f* (EeAChE):

$$\log_{10} IC_{50} = 15.37(\pm 1) + 13.38(\pm 2)E2u + 18.30(\pm 1)HATS6m \quad (5.9)$$

$$N_{cal} = 11, R = 0.99, R^2 = 0.98, S = 0.19, F = 150.88, p < 10^{-4}, R_{ijmax} = 0.66, o(> 2.S) = 0$$

$$R_{LOO} = 0.97, R_{LOO}^2 = 0.94, S_{LOO} = 0.28, S^{aleat} = 0.34$$

- Conjunto de datos *g* (hAChE):

$$\log_{10} IC_{50} = 66.88(\pm 6) + 0.01(\pm 0.1)MPC09 - 78.66(\pm 6)MATS1m - 6.09(\pm 0.3)RDF020m \quad (5.10)$$

$$N_{cal} = 19, R = 0.98, R^2 = 0.96, S = 0.21, F = 131.81, p < 10^{-4}, R_{ijmax} = 0.65, o(> 2.S) = 3$$

$$R_{LOO} = 0.97, R_{LOO}^2 = 0.94, S_{LOO} = 0.29, S^{aleat} = 0.53$$

Los parámetros estadísticos R , R^2 y S expresados para el conjunto de ecuaciones (5.1 a 5.10) presentan una bondad de ajuste, capacidad predictiva y veracidad adecuada en cada modelo obtenido. El parámetro R_{ijmax} define el coeficiente de correlación máximo entre pares de descriptores, y $o(> 2.S)$ es el número de moléculas con valor de residuo mayor o igual a dos veces la desviación estándar del modelo.

Todos los modelos QSAR propuestos superaron diferentes criterios de validación interna y externa detallados en la sección 5.2.4. En particular, los valores de los parámetros R_{LOO} y S_{LOO} demuestran la estabilidad de los modelos ante la variación sucesiva de los coeficientes de regresión de los descriptores. Además, la prueba de Aleatorización-Y demostró que los modelos obtenidos son el resultado de una verdadera relación estructura-actividad anticolinesterasa con un valor de $S_{cal} < S^{aleat}$, después de analizar 1.000 casos se tomó el mejor resultado de S^{aleat} .

En la mayoría de los casos, se evidenció que un solo descriptor molecular no tiene la capacidad suficiente para brindar predicciones acertadas del potencial inhibitorio de la AChE contenido en las moléculas. Por el contrario, los modelos con 2 o más descriptores presentan mejor correlación con los datos experimentales de la actividad anticolinesterasa de los análogos de Tacrina.

Las ecuaciones 5.1-5.10 exhibieron una aceptable capacidad predictiva de la actividad inhibitoria de la AChE contenida en las moléculas de los subconjuntos de predicción. De modo que, cada paralelismo matemático establecido está capacitado para cuantificar la capacidad inhibitoria de moléculas derivadas de Tacrina sobre la enzima AChE extraída de diferentes fuentes animales (bAChE, hAChE, EeAChE).

Conforme a lo anterior, se buscó verificar el comportamiento predictivo de cada modelo QSAR graficando la actividad $\log_{10} IC_{50}$ predicha como una función de los datos experimentales de cada conjunto molecular (a–g).

A partir de la Figura 5.2 hasta la Figura 5.11 (A) se observó una marcada tendencia lineal de entre los valores del eje de ordenadas (datos de predicción) y el eje de abscisas (datos observados).

Asimismo, los gráficos de dispersión de los valores de residuo de cada modelo, exhibidos en las Figuras 5.2 a 5.11 (B), permiten verificar la tendencia de los datos a obedecer un patrón aleatorio alrededor de la línea cero del eje Y, lo cual indica que se cumple el postulado de la técnica MLR dentro del rango de actividad considerado en cada ecuación.

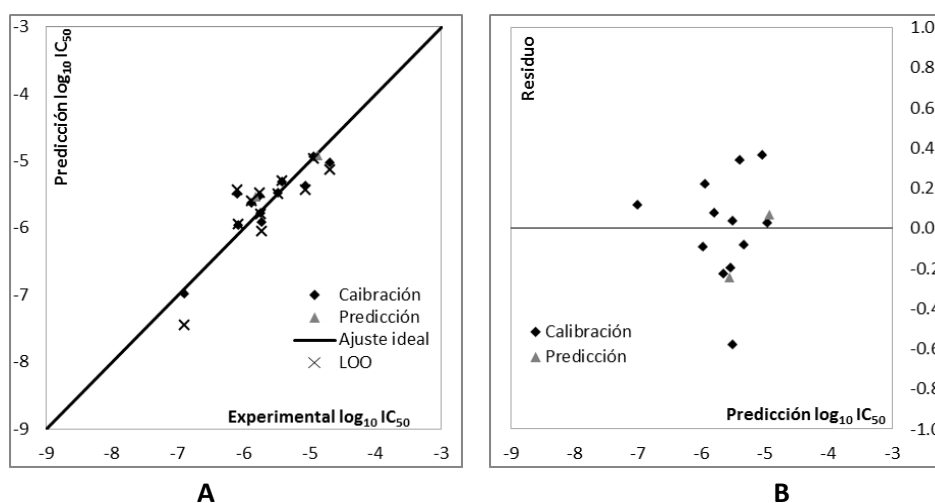


Figura 5.2. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto a (bAChE). **B).** Dispersión de los valores de residuo.

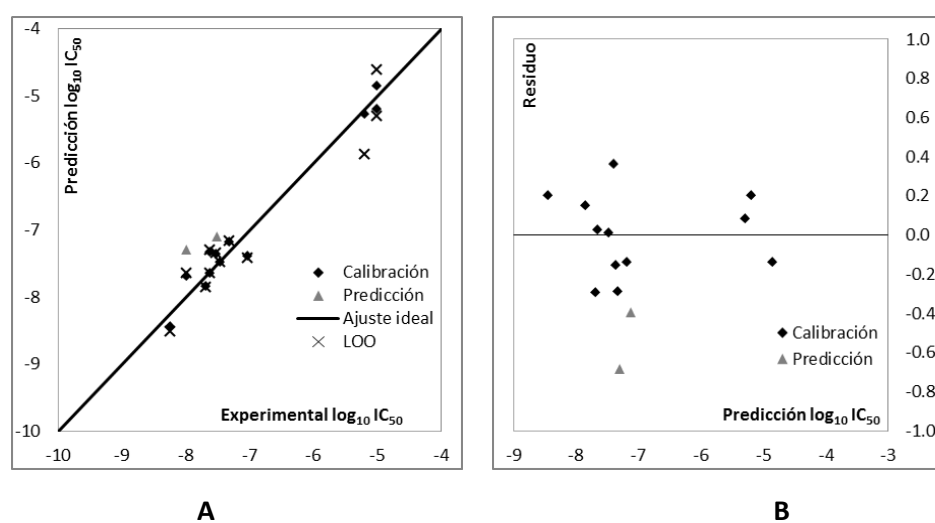


Figura 5.3. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto b (bAChE). **B).** Dispersión de los valores de residuo.

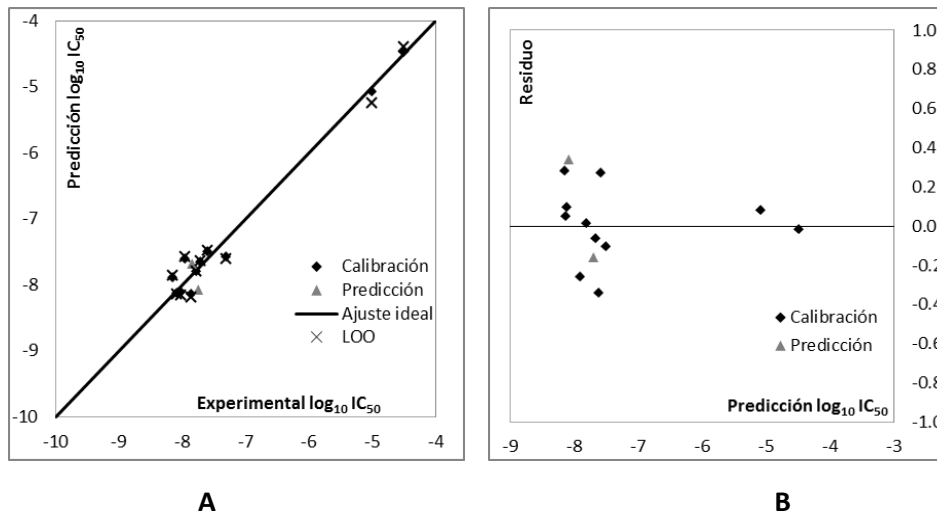


Figura 5.4. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **b** (hAChE). **B).** Dispersión de los valores de residuo.

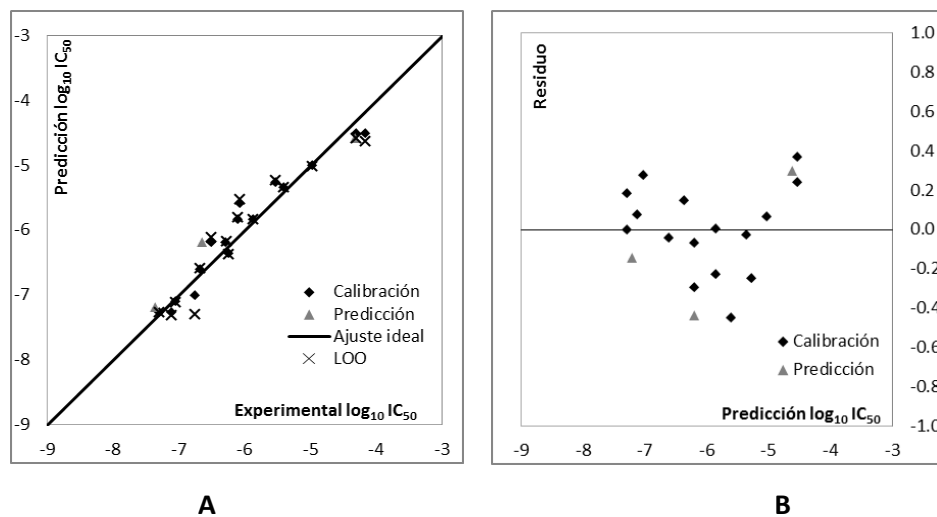


Figura 5.5. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **c** (EeAChE). **B).** Dispersión de los valores de residuo.

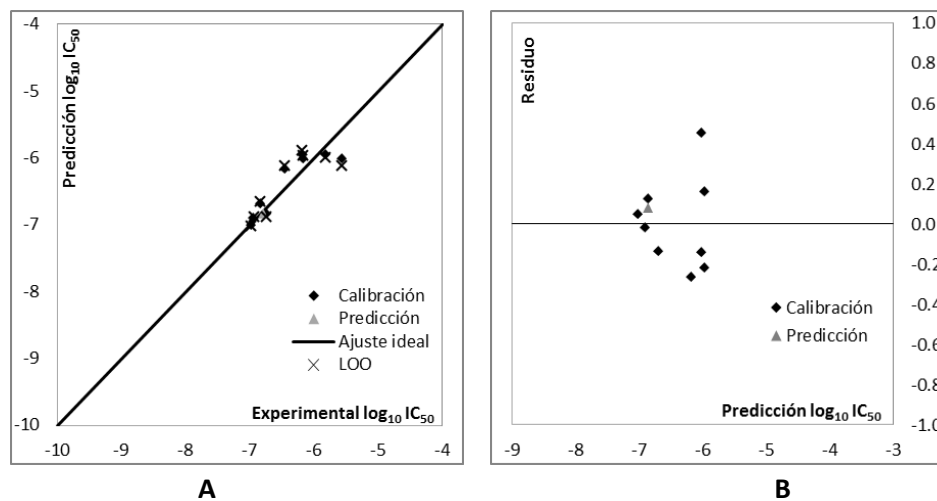


Figura 5.6. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **c** (hAChE). **B).** Dispersión de los valores de residuo.

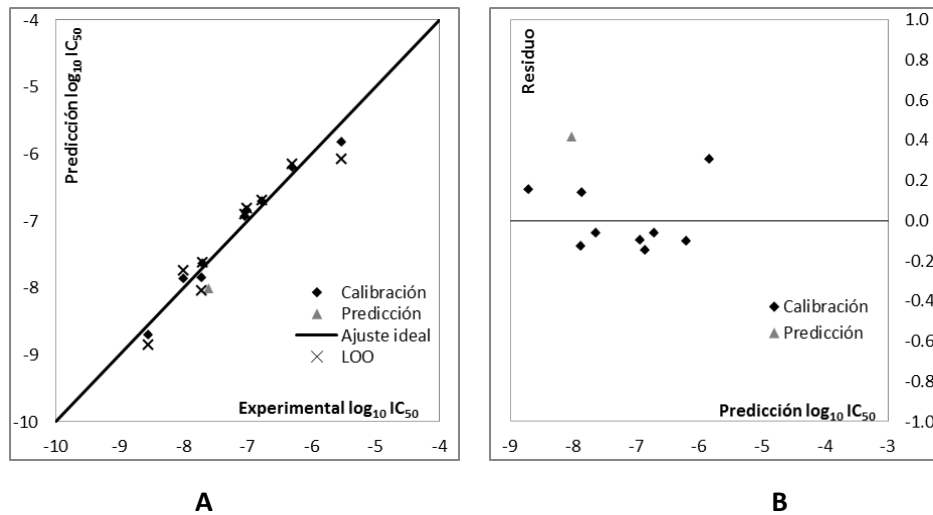


Figura 5.7. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **d** (bAChE). **B).** Dispersión de los valores de residuo.

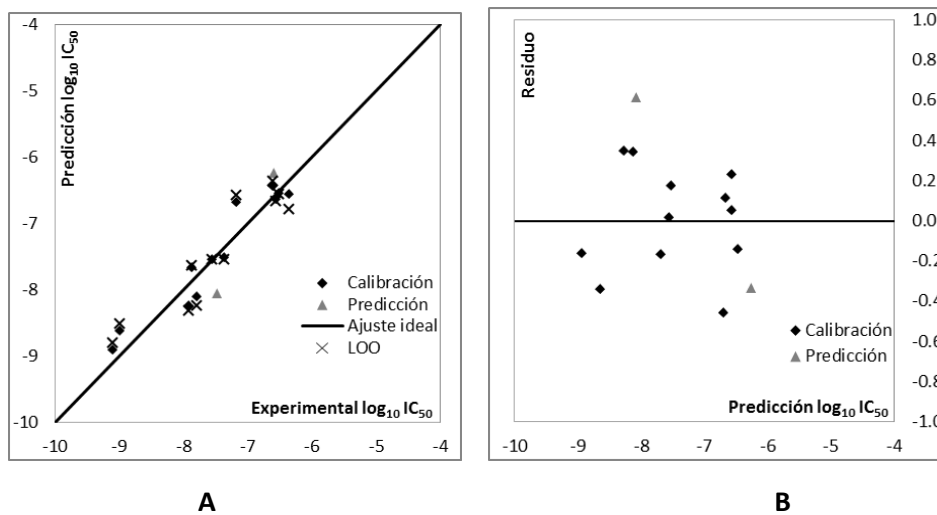


Figura 5.8. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **e** (hAChE). **B).** Dispersión de los valores de residuo.

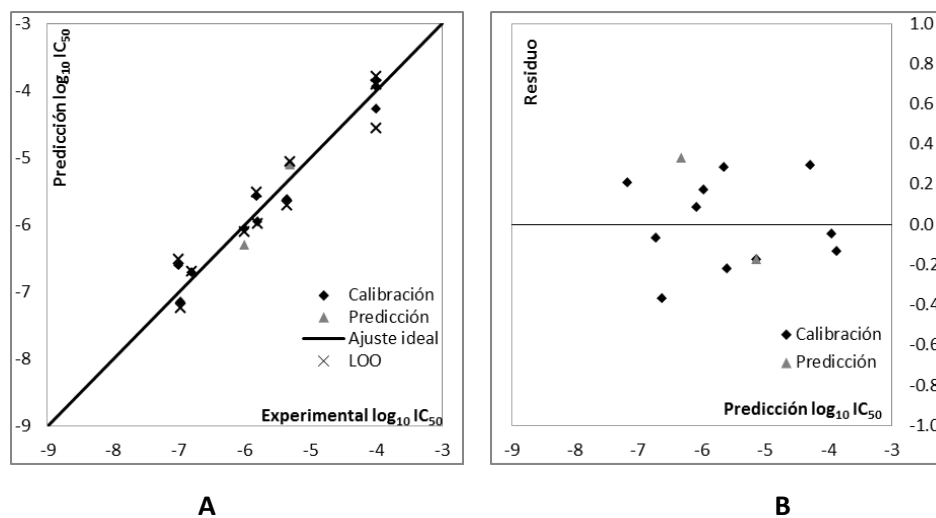


Figura 5.9. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **f** (bAChE). **B).** Dispersión de los valores de residuo.

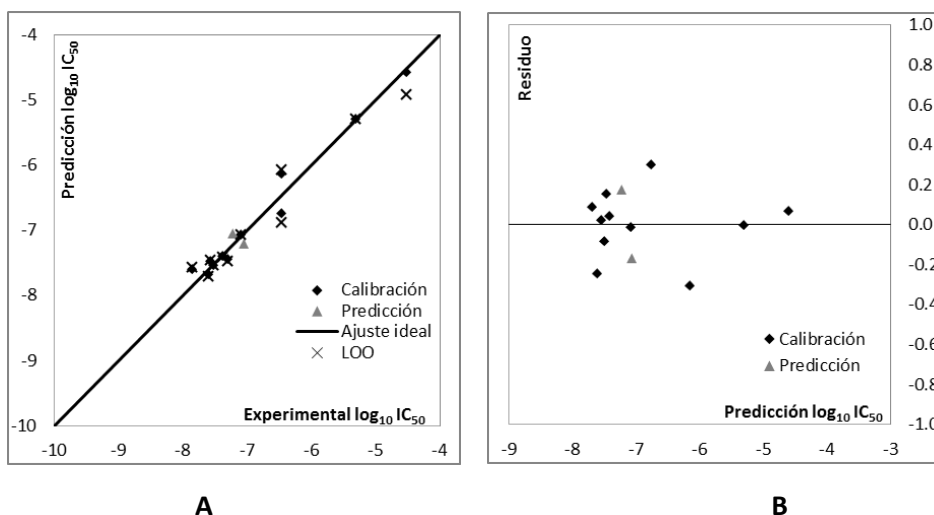


Figura 5.10. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **f** (EeChE). **B).** Dispersión de los valores de residuo.

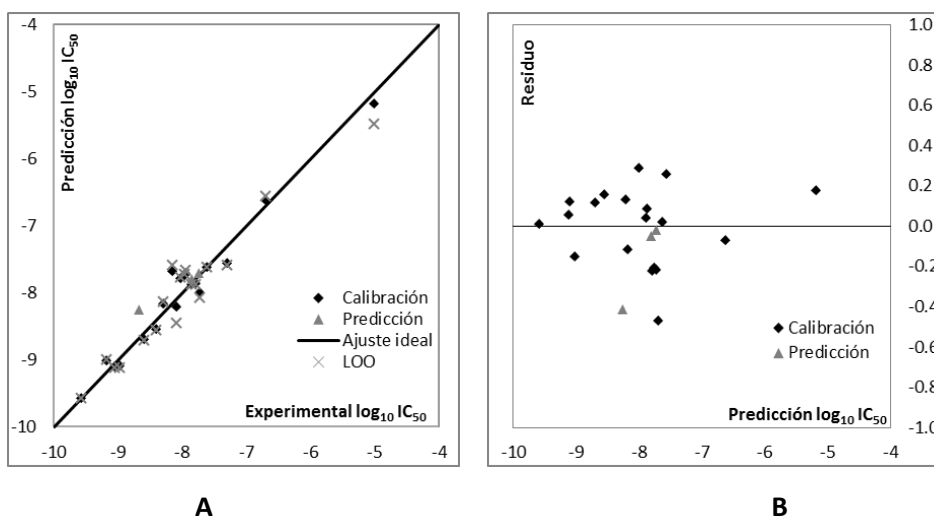


Figura 5.11. A). Valores experimentales y predichos del $\log_{10} IC_{50}$ para el conjunto **g** (hAChE). **B).** Dispersión de los valores de residuo.

Aunque, se detectaron algunos valores atípicos en la Ec. 5.10, que superan el valor del parámetro $\sigma(2.5)$; el principal objetivo de este estudio consiste en proponer modelos generales que tengan aplicabilidad a cualquier biomolécula sin restricciones; por lo tanto, no se eliminaron datos de los subconjuntos *cal*.

Los valores de los descriptores moleculares involucrados en las ecuaciones 5.1-5.10, así como los datos experimentales y de predicción del $\log_{10} IC_{50}$ se encuentran detallados en el apartado anexo, Tablas 5.2.1 a 5.2.7.

En general, los modelos QSAR establecidos involucraron diferentes descriptores conformacionales y no conformacionales, definidos brevemente en la Tabla 5.2.8 del material anexo.

Particularmente, los descriptores moleculares que resultaron ser más representativos para predecir la actividad anticolinesterasa son los descriptores geométricos de tipo *3D-Morse*, *WHIM*, *GETAWAY*, *RDF*; así como los descriptores topológicos de tipo Autocorrelación 2D junto con diferentes índices de conectividad y los descriptores 0D, tal como el promedio de peso molecular o el número de amina secundarias.

Si bien los modelos QSAR obtenidos fueron diseñados para predecir la capacidad inhibitoria *in vitro* de la AChE de diez conjuntos moleculares específicos, basados en compuestos derivados de Tacrina; inicialmente se buscó desarrollar un modelo QSAR que analizara todas las moléculas derivadas del fármaco Tacrina ($N=106$).

Sin embargo, los resultados exhibieron un importante deterioro en la calidad de los parámetros estadísticos del modelo QSAR resultante, y en consecuencia la capacidad predictiva y credibilidad del mismo se vieron afectadas. Una explicación racional de este comportamiento, se asocia con la diversidad experimental contemplada para medir la IC_{50} , y la implementación de enzimas AChE provenientes de diferentes fuentes animales.

Conforme a lo anterior, se puede destacar la importancia de diseñar modelos QSAR generales, a partir de conjuntos de datos con valores concretos, obtenidos por medio de un protocolo de ensayo estandarizado, con bajo error experimental. Ya que de esto depende la calidad, veracidad, coherencia y capacidad predictiva del modelo.

5.2.6. Conclusiones

En el presente estudio se desarrollaron diez diferentes modelos basados en las relaciones cuantitativas estructura-actividad de nuevos inhibidores de la enzima acetilcolinesterasa (AChE), encargada de regular la producción de impulsos nerviosos en el CNS. Cada modelo QSAR propuesto representa la relación matemática que existe entre las estructuras moleculares y la actividad anti-AChE exhibida por las mismas.

Los modelos QSAR obtenidos tienen una capacidad predictiva adecuada, y la escasa cantidad de observaciones disponibles en cada conjunto de datos se analiza con éxito mediante la estrategia estadística MLR acoplada al RM. La importancia de este estudio radica en la necesidad de encontrar nuevos inhibidores de AChE inspirados en el fármaco Tacrina, con menos efectos secundarios y de fácil acceso económico.

No obstante, ante la continua innovación de las entidades químicas destinadas al tratamiento de diversos trastornos neurodegenerativos que afectan a la población más vulnerable, tal como la enfermedad de Alzheimer; es necesario proponer constantemente diferentes herramientas computacionales con habilidad de evaluar a priori la respuesta anticolinesterasa final exhibida por amplios conjuntos moleculares basados en estructuras moleculares nuevas o hipotéticas derivadas de Tacrina u otros fármacos comerciales.

5.3. Las Relaciones Cuantitativas Estructura–Actividad antifúngica (QSAR) de compuestos derivados de cinamato.

Los ésteres de ácido cinámico o cinamatos son una familia de ésteres aromáticos con una estructura como se muestra en la Figura 5.12. Los ácidos cinámicos y los correspondientes cinamatos (ésteres α,β insaturados) son compuestos naturales que pertenecen a la familia de los fenilpropanoides.

Los cinamatos naturales más importantes se derivan de los siguientes ácidos cinámicos, como el ácido ferúlico (ácido 4-hidroxi-3-metoxicinámico), el ácido cumárico (ácido p-hidroxicinámico), el ácido cafeico (ácido 3,4-dihidroxicinámico) y el ácido sináptico (4-hidroxi-3,5-dimetoxicinámico).⁴³ En general, son sustancias químicas bioactivas que exhiben una amplia variedad de propiedades biológicas específicas.

En ese contexto, los cinamatos son compuestos orgánicos relevantes en el diseño de una amplia gama de productos industriales como lubricantes y plastificantes. En particular los cinamatos de alquilo se utilizan en la elaboración de productos cosméticos y saborizantes alimentarios, ya que presentan pronunciados aromas de frutas o flores.⁴⁴

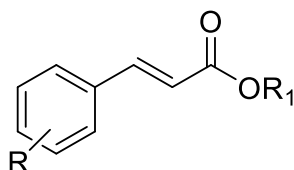


Figura 5.12. Estructura molecular de compuestos cinamatos.

De hecho, los cinamatos con diversas estructuras poseen una variedad de propiedades biológicas como: antioxidante, antimicrobiano, anticancerígeno, antiinflamatorio, entre otros.

Otro compuesto análogo de cinamato de gran utilidad en el tratamiento de diferentes patologías neurodegenerativas es el imidoalquil cinamato, 2-(N- β -ftalimido) etil cinamato, un potente inhibidor de la enzima 17-hidroxiesteroide deshidrogenasa asociada con enfermedades como el cáncer de mama, la enfermedad de Alzheimer y la hiperplasia prostática benigna.⁴⁵⁻⁴⁸

Entre los plaguicidas sustentables y seguros para la salud humana se encuentran los compuestos provenientes de la naturaleza; entre este grupo, los derivados de cinamato han contribuido en el control de diversas plagas que causan importantes pérdidas en el sector agrícola.

Específicamente, el cinamato de metilo tiene un efecto insecticida contra *Sitophylus oryzae* (Coleoptera: Curculionidae) y *Musca domestica* (Diptera: Muscidae);^{49, 50} el cinamato de etilo tiene un efecto antialimentario contra *Spodoptera littoralis* (Lepidoptera: Noctuidae).⁵¹

A partir de esta breve revisión de las propiedades biológicas que contienen los compuestos inspirados en la estructura molecular de los cinamatos, se propuso aplicar un estudio de la teoría QSAR enfocado en los aspectos conformacionales y no conformacionales de la estructura molecular, con el objetivo de cuantificar la actividad fungicida de una serie de derivados de cinamato, que exhiben varios patrones de sustitución que resultan efectivos contra las especies de hongos *Phytium sp* y *Corticium rolfsii*; estos parásitos se destacan por generar importantes pérdidas en la producción de alimentos primarios, se propagan a través de las fuentes hídricas naturales y son persistentes en el entorno.⁵²

5.3.1. Datos de la actividad antifúngica (GI)

Para iniciar el modelado QSAR, se recopilaron de la literatura especializada los valores de la actividad antifúngica experimental de una serie de 27 derivados del cinamato, expresada como la Inhibición del Crecimiento (GI 100 ppm) contra las especies *Pythium sp* y *Corticium rolfsii*.⁵³ Tanto los valores observados de la actividad GI como las estructuras moleculares de los 27 derivados semi-sintéticos se encuentran disponibles en la Tabla 5.3.1 del apartado anexo.

El principal objetivo de este estudio fue establecer paralelismos matemáticos racionales entre las estructuras químicas de los análogos de cinamato y la actividad antifúngica exhibida por los mismos. Seguidamente, aplicar el mejor modelo lineal obtenido para predecir la actividad fungicida de un nuevo conjunto molecular conformado por 21 derivados de alquil y aril cinamatos, sintetizados en el laboratorio dirigido por el Dr. Romanelli, ubicado en el Centro de Investigación y Desarrollo en Ciencias Aplicadas “Dr. Jorge J. Ronco” (CINDECA) de la UNLP. Las estructuras moleculares de estos compuestos encuentran disponibles en la Tabla 5.3.2 del apartado anexo.

5.3.2. Descriptores moleculares

La digitalización y optimización de las 27 estructuras moleculares se realizó a través del programa HyperChem 6.03,³⁶ empleando en primer lugar un procedimiento de Campo de fuerza de Mecánica Molecular (MM +). Posteriormente, las geometrías resultantes se refinan aún más por medio del Método Semiempírico PM3 (Método Paramétrico-3), con la aplicación del algoritmo de Polak-Ribiere y un límite de gradiente de 0.01 kcal.Å^{-1} .

Las características estructurales de las estructuras moleculares optimizadas fueron codificadas por 4885 descriptores moleculares conformacionales y no conformacionales generados por el programa comercial Dragon versión 6.³⁷ El cual incorpora 19 clases de descriptores de dimensionalidad 0D, 1D, 2D y 3D.

Adicionalmente, se contempló la información codificada por diferentes descriptores químico-cuánticos como el número de electrones, el momento dipolar molecular, la energía de orbital HOMO y de orbital LUMO y la diferencia de energía entre orbitales HOMO-LUMO ($\Delta_{Homo-Lumo}$).³⁸

Para excluir los descriptores “colineales” (idénticos), los que presentan datos ausentes y los pares linealmente dependientes fueron identificados dentro del RM, y solo se conserva una variable de cada par para continuar con el modelado QSAR. Este proceso de filtrado generó un conjunto con 1502 descriptores linealmente independientes.

5.3.3. Selección del modelo lineal QSAR

Se recurrió a la técnica de selección de descriptores RM,³⁹ como estrategia de selección de los descriptores moleculares más relevantes para estimar la respuesta antifúngica final, a partir de la matriz de descriptores con 1502 variables independientes elegibles.

Este algoritmo fue propuesto por nuestro grupo de investigación hace algunos años para generar modelos QSAR de regresión lineal multivariable con error minimizado, moderado a través de la desviación estándar (S) o el error cuadrático medio (RMS) del subconjunto de calibración.

5.3.4. Validación interna y externa del modelo

Entre las estrategias más comunes para validar teóricamente los modelos QSAR se destacan las pruebas de CV-LOO y la Aleatorización-Y,^{41, 42} ejecutando más de 150 mil casos por modelo. Como complemento de la validación teórica del modelo se calcularon los parámetros propuesto por Golbraikh *et al.*, 2003.⁵⁴

Por otra parte, la estrategia más eficiente para determinar la capacidad predictiva de un modelo QSAR es a través de la validación externa basada en el subconjunto de predicción (N_{pred}), conformado por moléculas extraídas del conjunto de datos inicial; las cuales no se consideran durante la calibración de los valores de los parámetros estadísticos de referencia.

En este caso, fueron elegidos dos subconjuntos moleculares de magnitud $N_{cal} = 23$ y $N_{pred} = 4$ (15% de los datos iniciales). Esta partición se enfocó en cada bioactividad, de tal forma que ambos subconjuntos compartan características cualitativas de relación estructura-actividad. Cabe resaltar, que el subconjunto *cal* debe ser representativo en términos de diversidad molecular aportada por el conjunto de datos inicial, y cubrir de manera uniforme todo el rango de valores de cada actividad antifúngica evaluada.

Un criterio importante para considerar las predicciones de un modelo QSAR validas, es la apropiada definición del AD del mismo; este análisis define el alcance químico de sus predicciones. En este caso, se definió el AD por medio de la técnica de influencia (h),⁵⁵ basada en distancias. La información detallada de esta técnica se encuentra en el capítulo 3, sección 7 de la presente Tesis.

5.3.5. Resultados y Discusión

El análisis QSAR consiste en buscar diversas relaciones lineales predictivas para las actividades antifúngicas de las estructuras químicas basadas en cinamato. Para fines de modelado, se transformaron los valores experimentales en $\log_{10}(GI + 50)$. Las ecuaciones de regresión lineal óptimas se establecen mediante el enfoque RM, el cual encontró los descriptores moleculares más representativos para la actividad, examinando $d = 1 - 4$ descriptores.

El tamaño de tales regresiones lineales (d) tiene que obedecer el criterio de relación entre las variables dependientes e independientes del modelo; donde por cada 6 o 7 datos del subconjunto *cal* se incluye 1 descriptor al modelo, a fin de reducir los problemas de sobreajuste ($N_{train}/d \geq 6$).

Modelo QSAR para predecir el $\log_{10}(GI + 50)$ sobre la especie *Pythium sp*

La exploración de una matriz de descriptores con $d = 1502$ variables fue realizada por la técnica RM, con el fin de encontrar modelos de regresión más apropiados para estimar la inhibición del crecimiento de 27 derivados de cinamato sobre la especie *Pythium sp*.

La Tabla 5.1 detalla los parámetros estadísticos contemplados como criterio de selección de los mejores modelos lineales QSAR, donde el modelo con tres descriptores (QSAR-3) presenta mejor calidad estadística en ambos subconjuntos moleculares, conservando la relación óptima descriptor/moléculas $N_{train}/d = 7.7$, lo que promueve su habilidad para estimar la bioactividad bajo estudio.

Con el fin de verificar el comportamiento predictivo del modelo QSAR-3, se graficaron los valores de la actividad antifúngica estimados como una función de los valores experimentales, ilustrado en la Figura 5.13.A.

Adicionalmente, el gráfico de la Figura 5.13.B exhibe la dispersión de los valores de residuo del modelo, donde se puede apreciar un comportamiento aleatorio de los puntos en torno al valor 0 del eje de ordenadas, y la ausencia de algún tipo de patrón infrecuente que indique la existencia de factores no modelados.

Tabla 5.1. Mejores modelos lineales QSAR para predecir GI de 27 derivados de cinamato sobre *Pythium sp* (100 ppm).

Coeficiente de regresión (error estándar)		Calidad estadística
QSAR-1		
<i>Intercepto</i>	2.184 (0.09)	$N_{cal}/d = 23, R_{cal} = 0.52, R_{cal}^2 = 0.27, S_{cal} = 0.13$ $\sigma \geq 2S = 0, R_{LOO}^2 = 0.13, S_{LOO} = 0.15,$ $R_{pred}^2 = 0.91, S_{pred} = 0.13, RMS_{pred} = 0.09$
<i>RDF020m</i>	-0.660 (0.2)	
QSAR-2		
<i>Intercepto</i>	2.659 (0.2)	$N_{cal}/d = 11.5, R_{cal} = 0.71, R_{cal}^2 = 0.50, S_{cal} = 0.11,$ $R_{ij\max}^2 = 0.41, \sigma(2.S) = 0, R_{LOO}^2 = 0.31, S_{LOO} = 0.13,$ $R_{pred}^2 = 0.96, S_{pred} = 0.17, RMS_{pred} = 0.08$
<i>RDF040e</i>	-0.0327 (0.008)	
<i>R6u⁺</i>	-9.139 (2)	
QSAR-3		
<i>Intercepto</i>	2.564 (0.09)	$N_{cal}/d = 7.7, R_{cal} = 0.84, R_{cal}^2 = 0.71, S_{cal} = 0.08,$ $R_{ij\max}^2 = 0.33, \sigma(2.S) = 0, R_{LOO}^2 = 0.55, S_{LOO} = 0.11,$ $R_{pred}^2 = 0.82, S_{pred} = ND^*, RMS_{pred} = 0.07$
<i>RDF020u</i>	-0.141 (0.02)	
<i>RDF150m</i>	0.238 (0.05)	
<i>HATS2m</i>	-1.719 (0.4)	
QSAR-4		
<i>Intercepto</i>	1.049 (0.2)	$N_{cal}/d = 5.7, R_{cal} = 0.91, R_{cal}^2 = 0.82, S_{cal} = 0.07,$ $R_{ij\max}^2 = 0.71, \sigma(2.S) = 0, R_{LOO}^2 = 0.72, S_{LOO} = 0.08,$ $R_{pred}^2 = 0.94, S_{pred} = ND^*, RMS_{pred} = 0.07$
<i>DISPv</i>	0.0534 (0.007)	
<i>RDF065m</i>	-0.106 (0.01)	
<i>Mor27v</i>	-2.753 (0.3)	
<i>E2e</i>	1.473 (0.3)	

*-No Disponible

A partir del gráfico de dispersión de los valores de residuo (Figura 5.12.B) se logró determinar que el compuesto **22** presenta un valor de residuo cercano al límite $\sigma(2S)$. La exclusión de dicho compuesto permitiría mejorar la calidad estadística del subconjunto *cal* para el modelo QSAR-3, pasando de $R_{cal} = 0.84, S_{cal} = 0.08$ a $R_{cal} = 0.88, S_{cal} = 0.07$. Este comportamiento se atribuye a que el derivado **22** presenta elementos estructurales diferentes en comparación con los otros compuestos del subconjunto *cal*, y esto modifica la capacidad del modelo para predecirlo de forma acertada.

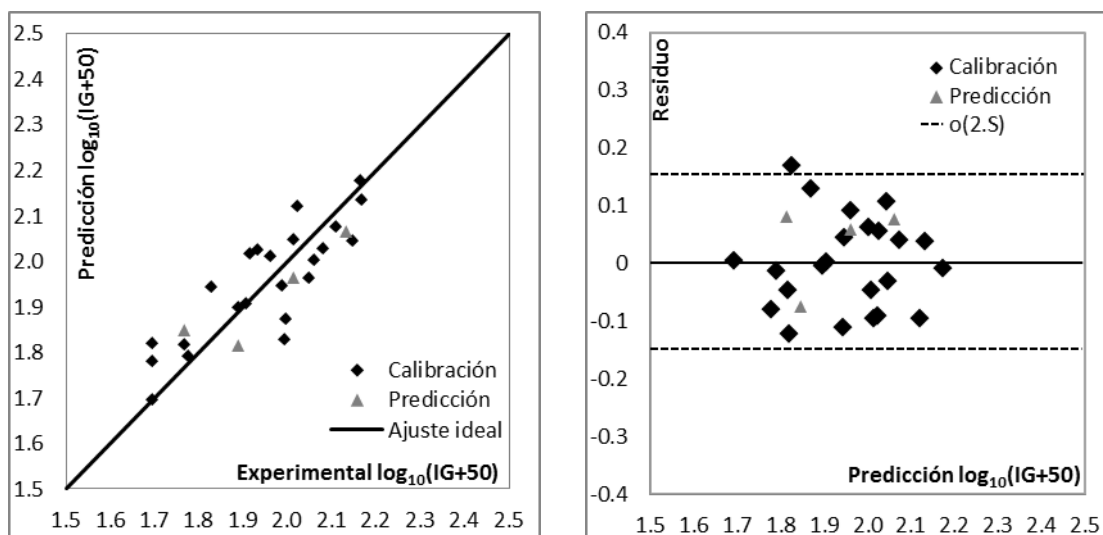


Figura 5.13. A). Gráfico de los valores de $\log_{10}(IG + 50)$ sobre *Pythium sp* predichos como una función de los experimentales. **B).** Dispersión de los valores de residuo del QSAR-3.

Sin embargo, el error de predicción del modelo mejorado, medido a través del subconjunto *pred* ($RMS_{pred} = 0.076$, $S_{pred} = 0.10$) y los resultados la prueba CV-LOO no son muy distintos con respecto al modelo QSAR-3 original. Por lo tanto, no hay razones suficientes para excluir dicha molécula. Además, se promueve el desarrollo de modelos QSAR generales, que logren abarcar un espacio químico importante.

Los valores experimentales de la actividad antifúngica ($\log_{10}(IG + 50)$), y los estimados por el modelo QSAR-3 para cada subconjunto molecular se encuentran disponibles en la Tabla 5.3.3 del apartado anexo.

El análisis del dominio de aplicación del modelo QSAR-3 revela que los 4 compuestos del subconjunto *pred* se encuentran dentro del área de influencia del modelo ($h_i < h^* = 0.52$); por lo tanto, los valores estimados de la actividad antifúngica de los derivados de cinamato pueden considerarse confiables.

En el caso del subconjunto de *cal*, solo el compuesto **26** excede el límite de advertencia del modelo (h^*), y el **24** se localiza cercano al h^* . Aunque, presentan un valor de residuo por debajo del $o(2S)$. Los valores de influencia (h_i) para los subconjuntos *cal* y *pred* se registran en la Tabla 5.3.4 del material anexo.

El resultado de la prueba de Aleatorización-Y revela que el modelo QSAR-3 no es producto de una correlación circunstancial, ya que el error calculado para el subconjunto *cal* ($S_{cal} < S^{aleat} = 0.09$) es inferior a la variación calculada por medio de esta prueba.

Para verificar el comportamiento predictivo del modelo en conjuntos moleculares externos, se estudiaron y aplicaron los parámetros de validación externa propuestos por Golbraikh *et al.*, 2003,⁵⁴ los cuales entregan los siguientes valores:

$$k = 1.01 \text{ y } k' = 0.98$$

$$R_0^2 = 0.79 \text{ y } R_0'^2 = 0.60$$

$$1 - R_0^2 / R_{pred}^2 = 0.035 \text{ y } 1 - R_0'^2 / R_{pred}'^2 = 0.26$$

$$R_m^2 = 0.68$$

Estos resultados promueven la aceptación del modelo lineal QSAR-3 como una herramienta predictiva de la actividad antifúngica sobre *Pythium sp.* Como una forma de interpretar los modelos expuestos en la Tabla 1, se brinda una breve descripción de los descriptores involucrados en cada expresión matemática en la Tabla 5.3.5 del apartado anexo. En particular el mejor modelo QSAR conformado por tres descriptores se basa en dos clases de descriptores diferentes:

Dos descriptores geométricos basados en la Función de Distribución Radial (RDF), un RDF número 2.0 sin ponderar (*RDF020u*) y un RDF número 15.0 ponderado por masas atómicas (*RDF150m*).

Un descriptor geométrico de Ensamblado de Geometría, Topología y Pesos Atómicos (GETAWAY) representado como *HATS2m*, que codifica la autocorrelación ponderada por influencia molecular de retraso 6 / y una doble ponderación por masas atómicas.

La matriz de correlación de estos descriptores geométricos se expresa en la Tabla 5.2; donde se destaca la baja intercorrelación entre pares de descriptores del modelo. Los valores numéricos de los descriptores del modelo QSAR-3 se disponen en la Tabla 5.3.6 del material anexo.

Tabla 5.2. Matriz de correlación de los descriptores incluidos en el modelo QSAR-3.

QSAR-3	<i>RDF020u</i>	<i>RDF150m</i>	<i>HATS2m</i>
<i>RDF020u</i>	1.00	0.57	0.44
<i>RDF150m</i>		1.00	0.10
<i>HATS2m</i>			1.00

A partir del análisis del parámetro b_j^2 , calculado a partir de los coeficientes estandarizados de los descriptores, se puede verificar el grado de importancia que tienen los descriptores del modelo sobre la actividad antifúngica (IG). De modo que, el descriptor que tenga mayor coeficiente estandarizado será el más importante para estimar la actividad, tal como se muestra a continuación:

$$RDF020u (1.1) > RDF150m (0.7) > HATS2m (0.6)$$

En este caso particular, el descriptor *RDF020u* basado en la distribución de distancias en la representación geométrica de una molécula, presenta valores numéricos que varían de acuerdo con las variaciones numéricas de la actividad experimental. Por lo tanto, es el más representativo para estimar la respuesta antifúngica final de los derivados de cinamato sobre la especie *Pythium sp.*

Sin embargo, las magnitudes relativas de los coeficientes estandarizados (1.1, 0.7 y 0.6) sugieren que los descriptores se complementan entre sí dentro de la ecuación lineal y que tienen una importancia comparable para predecir la actividad estudiada.

Otra forma de interpretar el modelo QSAR-3 es a través de los signos de los coeficientes de regresión de los descriptores. Para este modelo en particular, los descriptores *RDF020u* y *HATS2m* tienen coeficiente de regresión negativos en el modelo lineal, y sus valores numéricos son positivos (Tabla 5.3.6 del material anexo); mientras que *RDF150m* presenta un coeficiente de regresión positivo, al igual que sus valores numéricos.

Por lo tanto, se puede deducir que a medida que aumenta el valor del descriptor *RDF150m*, y de forma simultánea disminuyen los valores de los descriptores *RDF020u* y *HATS2m*, se pueden obtener compuestos de alta efectividad contra el hongo de las plantas *Pythium sp.* Tal como ocurre con los compuestos **9**, **10**, **11**, **13** y **25** detallados en la Tabla 5.3.1 del apartado anexo.

Modelo QSAR para predecir el $\log_{10}(GI + 50)$ sobre la especie *Corticium rolfsii*

Por otra parte, el análisis QSAR de la actividad antifúngica sobre la especie *Corticium rolfsii* del conjunto de datos analizado previamente, condujo a obtener los mejores modelos lineales expresados en la Tabla 5.3, con valores mínimos del parámetro S_{cal} según lo estipulado por la técnica RM.

Tabla 5.3. Mejores modelos lineales QSAR para predecir GI de 27 derivados de cinamato sobre *Corticium rolfsii* (100 ppm).

Coeficiente de regresión (error estándar)		Calidad estadística
QSAR-5		
<i>Intercepto</i>	2.181 (0.07)	$N_{cal}/d = 23, R_{cal} = 0.57, R_{cal}^2 = 0.32, S_{cal} = 0.10, ,$ $\alpha(2.S) = 0, R_{LOO}^2 = 0.18, S_{LOO} = 0.11,$ $R_{pred}^2 = 0.82, S_{pred} = 0.08, RMS_{pred} = 0.05$
<i>PHI</i>	-0.0425 (0.01)	
QSAR-6		
<i>Intercepto</i>	2.247 (0.05)	$N_{cal}/d = 11.5, R_{cal} = 0.79, R_{cal}^2 = 0.62, S_{cal} = 0.08,$ $R_{ijmax}^2 = 0.80, \alpha(2.S) = 0, R_{LOO}^2 = 0.52, S_{LOO} = 0.08,$ $R_{pred}^2 = 0.74, S_{pred} = 0.19, RMS_{pred} = 0.09$
<i>BELm8</i>	-0.700 (0.1)	
<i>RDF070v</i>	0.0794 (0.02)	

QSAR-7		
<i>Intercepto</i>	2.700 (0.1)	$N_{cal}/d = 7.7, R_{cal} = 0.87, R_{cal}^2 = 0.76, S_{cal} = 0.06,$ $R_{ijmax}^2 = 0.77, \sigma(2.S) = 1, R_{LOO}^2 = 0.60, S_{LOO} = 0.09,$ $R_{pred}^2 = 0.98, S_{pred} = ND^*, RMS_{pred} = 0.04$
<i>BELv5</i>	-0.709 (0.1)	
<i>RDF080u</i>	-0.0247 (0.006)	
<i>L3e</i>	0.338 (0.06)	
QSAR-8		
<i>Intercepto</i>	4.645 (0.5)	$N_{cal}/d = 5.7, R_{cal} = 0.94, R_{cal}^2 = 0.88, S_{cal} = 0.05,$ $R_{ijmax}^2 = 0.76, \sigma(2.S) = 0, R_{LOO}^2 = 0.80, S_{LOO} = 0.06,$ $R_{pred}^2 = 0.97, S_{pred} = ND^*, RMS_{pred} = 0.04$
<i>Mv</i>	-2.585 (0.6)	
<i>BELv5</i>	-1.063 (0.1)	
<i>RDF080u</i>	-0.0255 (0.004)	
<i>L3u</i>	0.432 (0.05)	

*-No Disponible

Claramente, el modelo QSAR-7 con tres descriptores moleculares representar la mejor relación estructura–actividad antifúngica encontrada para la especie *Corticium rolfsii*, en un radio favorable de $N_{train}/d = 7.7$.

La representación gráfica del comportamiento predictivo del modelo QSAR-7, se encuentra en la Figura 5.14.A, exhibiendo una tendencia lineal entre los datos estimados y los observados para la respuesta antifúngica final. Del mismo modo, la Figura 5.14.B muestra la dispersión que tienen los valores de residuo de los subconjuntos *cal* y *pred*; donde se destaca la presencia de un comportamiento atípico del compuesto **12** con valor de residuo superior a $\sigma(2S)$ y cercano a $\sigma(2.5S)$.

Seguidamente, el análisis de Aleatorización-Y reportó un valor de $S^{aleat} = 0.07$, superando el S_{cal} del modelo QSAR-7. De esta forma, se comprueba la estabilidad del modelo ante la variación aleatoria de los valores de la actividad. A partir del análisis de los parámetros de Golbraikh *et al.*, 2003; se comprobó la habilidad predictiva del QSAR-7 con los siguientes valores:

$$k = 0.98 \text{ y } k' = 1.01$$

$$R_0^2 = 0.89 \text{ y } R_0'^2 = 0.77$$

$$1 - R_0^2 / R_{pred}^2 = 0.08 \text{ y } 1 - R_0'^2 / R_{pred}^2 = 0.21$$

$$R_m^2 = 0.69$$

Con respecto a la definición del AD,⁵⁵ los cuatro compuestos del subconjunto *pred* se ubican dentro del dominio del modelo QSAR-7, con valores de influencia $h_i < h^* = 0.52$; de modo que, sus predicciones son correctas y coinciden con la información experimental.

Para el subconjunto *cal* solo dos sustancias exceden en límite crítico de influencia h^* ; aunque esto refuerza el espacio químico habilitado para usar el modelo con baja

incertidumbre. Los valores de influencia h_i del modelo QSAR-7 para las moléculas de *cal* y *pred* se registran en la Tabla 5.3.7 del apartado anexo.

El paralelismo matemático definido por QSAR-7 involucra una mezcla de descriptores moleculares 2D y 3D de tres diferentes clases, especificadas a continuación:

- Un descriptor topológico BCUT representado como *BELv5*, el auto valor más bajo no. 5 de matriz de carga, ponderado por volúmenes atómicos de van der Waals.
- Un descriptor geométrico RDF representando por *RDF080u*, indica un número RDF 8.0 no ponderado.
- Un descriptor geométrico WHIM dado como *L3e*, índice WHIM direccional de tercer componente, ponderado por las electronegatividades atómicas de Sanderson.

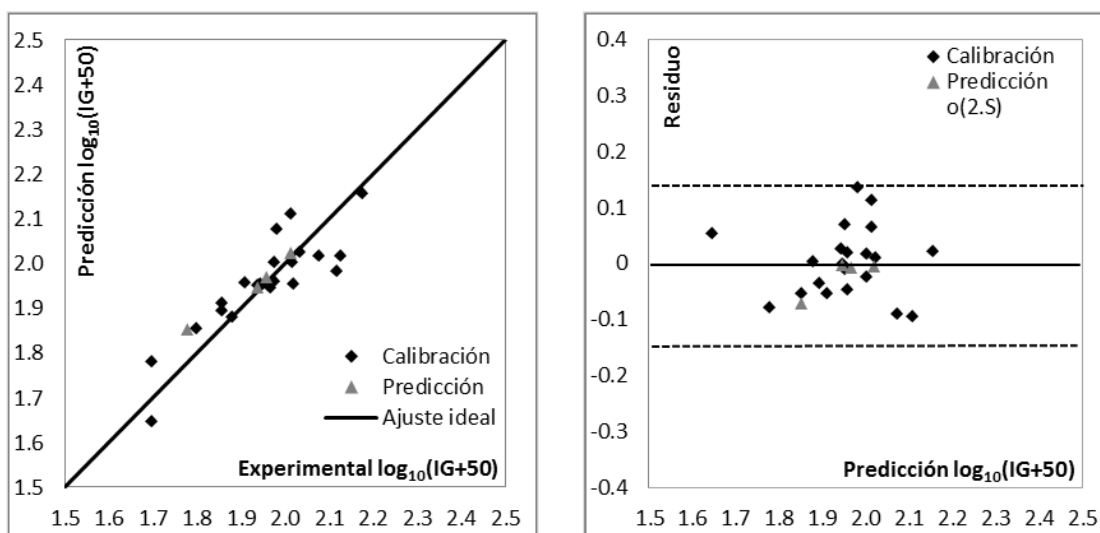


Figura 5.14. A). Gráfico de la actividad $\log_{10}(IG + 50)$ sobre *Corticium rolfsii* estimadas como una función de las observables. **B).** Dispersión de los valores de residuo del QSAR-7.

Los descriptores involucrados en el modelo QSAR-7 no presentan importantes valores de intercorrelación entre pares de descriptores, tal como se muestra en la Tabla 5.4. La información numérica de los tres descriptores del modelo QSAR-7 se encuentra registrada en la Tabla 5.3.6 del material anexo. Después de verificar el grado de importancia de los descriptores de QSAR-7 sobre la respuesta antifúngica analizada, se obtuvo el siguiente orden:

$$L3e (1.6) > BELv5 (1.2) > RDF080u (1.0)$$

Nuevamente, los tres descriptores toman valores numéricos positivos (Tabla anexa 5.3.6). Conforme a los signos de los coeficientes de regresión, el aumento de los

valores del descriptor *L3e*, junto con la disminución de los valores de los descriptores *BELv5* y *RDF080u* conducirían a estructuras que tienen mayor actividad de inhibición del crecimiento contra la especie *Corticium rolfssii*.

Tabla 5.4. Matriz de correlación de los descriptores incluidos en el modelo QSAR-7.

QSAR-7	<i>BELv5</i>	<i>RDF080u</i>	<i>L3e</i>
<i>BELv5</i>	1.00	0.76	0.83
<i>RDF080u</i>		1.00	0.88
<i>L3e</i>			1.00

Aplicaciones de los modelos QSAR-3 y QSAR-7 para predecir la actividad fungicida de un nuevo conjunto molecular de derivados de cinamato.

Como paso final del presente análisis QSAR, se aplicaron los modelos QSAR-3 y QSAR-7 para predecir la actividad antifúngica ($\log_{10}(GI + 50)$) de 21 derivados de cinamato de arilo y alquilo sintetizados, detallados en la Tabla 5.3.2 del apartado anexo. Los cuales no cuentan con información experimental de las actividades fungicidas sobre las especies *Pythium sp* y *Corticium rolfssii*.

El primer paso necesario consiste en calcular los valores de influencia (h_i) de los nuevos compuestos. Los resultados estimados para esta nueva librería molecular se exponen en la Tabla anexa 5.3.8. Los derivados **37**, **45-48**, y de otros análogos de cinamatos se predicen como activos contra la especie *Pythium sp*; así mismo, los compuestos **36**, **40**, **44** y **48** son previstos con alto poder de inhibición sobre *Corticium rolfssii*.

Sin embargo, los valores de influencia calculados para este conjunto de datos, revelan que tales moléculas tienen altos niveles de influencia y se localizan afuera del AD de cada modelo. Por lo tanto, los valores de predicción obtenidos presentan alta incertidumbre, y pueden ser el producto de una sustancial extrapolación de los modelos lineales.

Otra posible causa de este comportamiento, se asocia a la diferencia estructural de los compuestos **45-48** con respecto a las moléculas utilizadas para calibrar los modelos QSAR-3 y QSAR-7. En particular, el grupo funcional Ftalimidoetilo, no fue considerado en la construcción de las relaciones lineales estructura-actividad fungicida.

Entre los nuevos compuestos sintetizados **28-48**, se encuentran algunas estructuras prometedoras, que caen dentro del dominio de aplicación de los modelos QSAR-3 y QSAR-7 con $h_i < h^*$; por lo tanto, merecen ser analizadas experimentalmente a través de bioensayos estandarizados que permitan elucidar la actividad antifúngica contenida en dichas sustancias, especialmente en las moléculas **28**, **38** y **42** contra la especie *Pythium sp*, y **28** y **34** contra la especie *Corticium rolfssii*.

5.3.6. Conclusiones

Con el presente estudio QSAR se logró establecer dos modelos de regresión basados en descriptores 2D y 3D para predecir la capacidad de inhibición del crecimiento (GI) de 27 compuestos derivados de cinamato sobre las especies de parásito *Corticium rolfii* y *Pythium sp*, catalogados como hongos de diversas plantas, con importantes efectos negativos sobre los cultivos, ya que son persistentes y suele propagarse de forma rápida por medio de las aguas superficiales de riego.

Las herramientas QSAR-3 y QSAR-7 exhiben la habilidad de anticipar la actividad fungicidas GI, y en consecuencia, representan una importante guía computacional para generar conocimiento químico asociado a una determinada cantidad de estructuras moleculares requeridas para diseñar novedosos fungidas de alta efectividad, biodegradables y creados por medio de procedimiento de síntesis ecocompatible.

5.4. Análisis QSAR de compuestos derivados de las plantas con actividad larvicida sobre el *Aedes aegypti* L., vector del dengue, chikungunya y el zika

Actualmente, existen tres principales estrategias de control y prevención de las Enfermedades Transmitidas por Vectores (VBDs): las vacunas, los antivirus y los programas de control de mosquitos. Sin embargo, estas estrategias no han sido altamente exitosas, ya que los casos de infección y reinfección por *flavivirus* siguen ocurriendo en diferentes zonas del planeta.

Dentro de los programas de control de mosquitos, se han implementado masivamente diversos larvicidas de síntesis como el organofosforado temefos por su efectividad a bajas concentraciones. Sin embargo, el uso repetitivo e indiscriminado de este tipo de productos químicos ha provocado impactos ambientales negativos, el desarrollo de resistencia en las poblaciones de mosquitos y la bioacumulación en los organismos no objetivos.^{56,57}

En consecuencia, muchos estudios se han centrado en desarrollar nuevas estrategias basadas en larvicidas derivados de las plantas y sus análogos semisintéticos, que permitan controlar las poblaciones de mosquitos en las etapas inmaduras de forma selectiva, con menor impacto sobre la salud humana y los recursos naturales, a partir de materias primas renovables y biodegradables.^{58,59}

No obstante, el diseño de nuevos agentes larvicidas efectivos es limitado, ya que requiere tiempo y recursos económicos para realizar la síntesis, caracterización y los ensayos biológicos y clínicos. Por ello, la búsqueda, identificación y desarrollo de nuevas biomoléculas selectivas y potentes contra vectores han sido generalmente asistidas por técnicas *in silico*.

En este contexto, el modelado QSAR ha demostrado ser una herramienta exitosa para predecir las propiedades biológicas, organolépticas y fisicoquímicas de un compuesto químico a través del conocimiento de su estructura molecular.⁶⁰

Hasta el momento, la cantidad de investigaciones QSAR orientadas en predecir la actividad larvica sobre el mosquito *A. aegypti* L. es reducida. Por ello, el principal objetivo de los trabajos realizados en la presente Tesis fue estudiar, desarrollar, aplicar e interpretar mediante el modelado QSAR los conjuntos de datos disponibles de moléculas naturales o inspiradas en la naturaleza con prominente poder larvica sobre el vector la fiebre de chikungunya, el dengue y el zika.

5.4.1. Datos de la actividad larvica (LC_{50})

Se recolectaron de la literatura especializada 62 estructuras moleculares de sustancias derivadas de las plantas con actividad larvica sobre el mosquito *A. aegypti* L, exhibidas en la Figura 5.15.⁶¹⁻⁶⁴

Este conjunto de datos está conformado por estructuras moleculares heterogéneas de terpenos, cetonas, compuestos oxigenados y fenilpropanoides.

Las medidas experimentales de la actividad larvica fueron realizadas acorde con el protocolo propuesto por la OMS para determinar la toxicidad aguda media (LC_{50}) sobre larvas *A. aegypti* de III–IV *instar* de tamaño, en un periodo de prueba de 24 a 48 horas posteriores al tratamiento químico.

Cabe destacar que este conjunto molecular de sustancias naturales y sus derivados, representó hasta 2017 la única fuente de información química concreta, segura y menos tóxica para iniciar la búsqueda de nuevos fitosanitarios sustentables mediante un enfoque QSAR.

Los 62 datos de actividad larvica expresados como LC_{50} ($\mu\text{g mL}^{-1}$) se transformaron a escala logarítmica ($\log_{10} LC_{50}$) para reducir el rango de valores de la actividad y facilitar el modelado. Los valores de la actividad larvica $\log_{10} LC_{50}$ de los 62 compuestos derivados de las plantas se disponen en la Tabla 5.4.1.

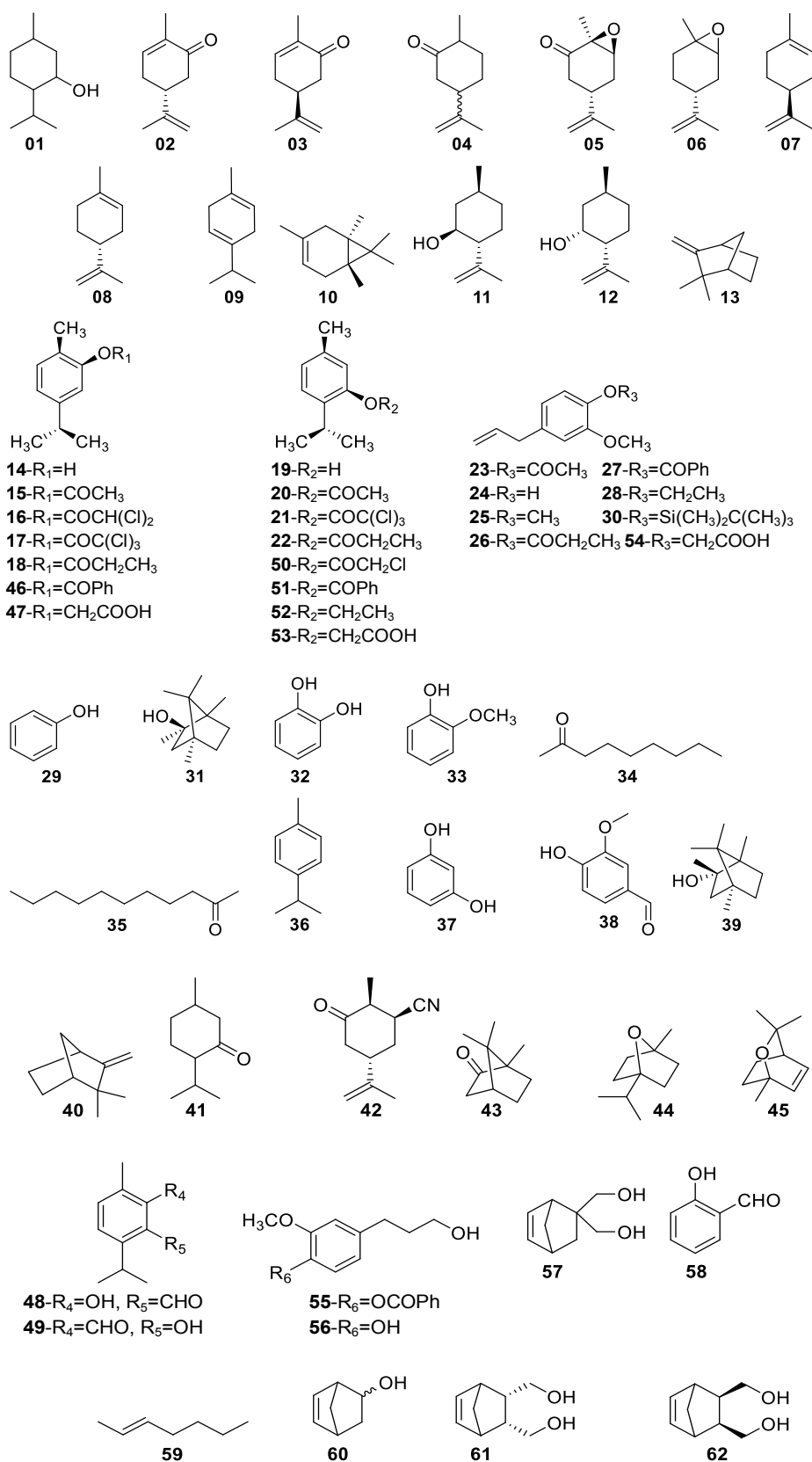


Figura 5.15. Estructuras moleculares con datos experimentales de actividad larvica (LC_{50}) sobre el *A. aegypti* L.

5.4.2. *Descriptores moleculares*

Para digitalizar las estructuras moleculares de los 62 análogos larvicidas se acudió al programa HyperChem 6.03.³⁶ Para contemplar la información contenida en los descriptores 3D, fue necesario optimizar las estructuras moleculares bajo un nivel de teoría normal. En este caso, se empleó un procedimiento de Campo de fuerza de Mecánica Molecular (MM +); luego, las geometrías resultantes fueron refinadas con el Método Semiempírico PM3 (Método Paramétrico-3), con la aplicación del algoritmo de Polak-Ribiere y un límite de gradiente de 0.01 kcal.Å⁻¹.

Posterior a esto, se calcularon los descriptores moleculares con el programa Dragon 6.³⁷ Este programa de descriptores genera un conjunto con 4885 variables independientes e incluye varios tipos de descriptores que caracterizan los aspectos multidimensionales de la estructura química.

Antes de iniciar la construcción del modelo lineal, se excluyeron los descriptores con valores constantes o casi constantes, y aquellos con al menos un valor faltante; luego, se aplicó el análisis de dependencia lineal entre pares de descriptores, donde considera solo uno descriptor por par. Con este proceso se logra un conjunto conformado por 1738 descriptores linealmente independientes.

5.4.3. *Selección del modelo QSAR*

La técnica estadística MLR ha demostrado ser de gran utilidad en muchas disciplinas para desarrollar modelos predictivos QSAR. Los modelos lineales son generales y muestran claramente el efecto de incluir/excluir descriptores en la ecuación; por lo tanto, es posible sugerir relaciones de causa/efecto a través de paralelismos tan simples.

La principal ventaja de desarrollar modelos de regresión lineal es que tienen menos problemas de sobreajuste, porque el método MLR no requiere demasiados parámetros optimizados durante el diseño del modelo, solo un coeficiente de regresión por descriptor. Por ello, en este estudio QSAR se consideró que la técnica MLR acoplada a la técnica RM,³⁹ es la mejor opción de búsqueda de modelos predictivos QSAR, especialmente cuando el conjunto molecular es reducido y se cuenta con pocos datos experimentales disponibles.⁶⁵

5.4.4. *Validación interna y externa del modelo*

Una partición racional de los subconjuntos de calibración y predicción deben tener relaciones estructura-actividad similares. Ya que la división aleatoria de los compuestos no conduce a resultados de predicción adecuados, fue necesario emplear una división racional del conjunto molecular inicial (N=62) a través del Método de Subconjuntos Balanceados (BSM).⁶⁶

Esta estrategia fue desarrollada por nuestro grupo y se basa en el método de análisis de agrupamientos k -medias (k -MCA) ejecutado en Matlab.⁴⁰ La idea principal detrás de la técnica BSM es crear grupos k o grupos de compuestos de tal forma que los compuestos en el mismo grupo sean muy similares en términos de distancias métricas (distancia euclídeas), y los compuestos en diferentes grupos sean distintos. De este modo, el procedimiento BMS asegura que el subconjunto de calibración sea representativo del subconjunto predicción.

Al igual que en los estudios QSAR previos, para verificar la calidad interna de los modelos obtenidos se consideraron los criterios de validación teóricos CV-LOO y Aleatorización-Y.^{41, 42} Además, se analizaron los parámetros postulados por Golbraikh *et al.*, 2003,⁵⁴ para confirmar la capacidad predictiva de los modelos lineales.

5.4.5. Resultados y Discusión

En los últimos años, ha incrementado el número de investigaciones enfocadas en identificar nuevas sustancias larvicidas de origen natural. En ese contexto, las plantas constituyen un laboratorio natural de nuevas entidades químicas, caracterizadas por presentar diferentes propiedades biológicas de interés en el control de insectos plaga o vectores de enfermedades, como el mosquito *Aedes aegypti* L.

Para iniciar el procesamiento del modelo QSAR, se empleó la técnica BSM para dividir el conjunto de datos ($N=62$) en dos subconjuntos equitativos en relación estructura-actividad; denominados subconjunto de calibración ($N_{cal}=52$), requerido para diseñar el modelo, y subconjunto de predicción ($N_{pred}=10$), constituido por el 20% de la información del conjunto de datos inicial (moléculas **3, 8, 13, 17, 19, 20, 22, 39, 46 y 62**), y se utiliza para verificar la capacidad predictiva del modelo resultante.

Seguidamente, se establecieron las mejores regresiones lineales a través de la estrategia RM, empleada para explorar 1738 descriptores moleculares linealmente independientes generados por medio del programa comercial Dragon6. El método RM minimiza el parámetro S_{cal} y selecciona los mejores modelos conformados por $d=1-6$ descriptores más representativos para la actividad. Asimismo, la selección del mejor modelo se evalúa en función de los parámetros R^2 y R_{ijmax}^2 .

A partir de Tabla 5.5 se logró determinar que el parámetro S_{cal} mejora a medida que aumenta el número de descriptores del modelo, y que para el modelo QSAR con $d=6$ descriptores el parámetro S_{pred} es superior en relación al resto de los modelos con $d=1-5$. La definición de los descriptores moleculares incluidos en cada modelo mencionado se exhibe en la Tabla 5.4.2 del apartado anexo.

Conforme a lo anterior, se determinó que el mejor modelo lineal es el que incorpora cinco descriptores moleculares independientes de la conformación expresado en la Ec. 5.11. En este caso, el modelo seleccionado presenta valores

aceptables de los parámetros estadísticos R^2 , S y $R_{ij\max}^2$ considerados como criterio de selección del modelo.

Tabla 5.5. Modelos QSAR con $d=1-6$ descriptores más representativos en 62 moléculas derivadas de las plantas. El mejor modelo se encuentra marcado en negrita.

d	R_{cal}^2	S_{cal}	R_{pred}^2	S_{pred}	$R_{ij\max}^2$	Descriptores Moleculares
1	0.30	0.41	0.39	0.43	0.00	CATS2D_05_LL
2	0.52	0.34	0.80	0.33	0.04	X5v, BLTF96
3	0.60	0.31	0.74	0.40	0.08	X4v, MATS5e, BLTF96
4	0.64	0.30	0.74	0.39	0.13	X4sol, SpMAD_B(m), GASTS7i, BLTF96
5	0.69	0.28	0.78	0.39	0.26	J.Dz(i), ATSC5S, JGI7, SpMax2.Bh(m), H.052
6	0.74	0.26	0.76	0.55	0.40	JGI3, chi1_EA(bo), Eig08_EA(dm), CATS2D_08_DA, F10[C-C], BLTF96

La capacidad predictiva de la Ec. 5.11 fue definida por medio de las moléculas del subconjunto $pred$, el cual tiene un porcentaje de varianza explicada de 78% y la dispersión de los datos es $S_{pred} = 0.39$. Además, se obtuvieron valores apropiados de los parámetros de validación interna ($R_{LOO}^2 = 0.60$ y $S_{LOO} = 0.32$), que garantizan la estabilidad del modelo $d=5$, ante la adición o eliminación consecutiva de una molécula de calibración por vez.

La estrategia de Aleatorización-Y verificó que el modelo obtenido no es producto de una correlación fortuita, y por el contrario, define una relación consistente entre la estructura y la propiedad biológica ($S^{aleat} = 0.36$), luego de analizar 30.000 casos posibles se elige el mejor resultado.

$$\log_{10} LC_{50} = 8.45 - 0.45(\pm 0.1)J.DZ(i) + 0.02(\pm 0.003)ATSC5s + 19.89(\pm 6)JGI7 - 1.65(\pm 0.2)SpMax2.Bh(m) + 0.07(\pm 0.01)H.052 \quad (5.11)$$

$$N_{cal} = 52, d = 5, R_{cal}^2 = 0.69, S_{cal} = 0.28, N_{cal} / d = 10, F = 20, R_{ij\max}^2 = 0.26, o(2.5S) = 1,$$

$$R_{LOO}^2 = 0.60, S_{LOO} = 0.32, S^{aleat} = 0.36$$

$$N_{pred} = 10, R_{pred}^2 = 0.78, S_{pred} = 0.39$$

Además, la Ec. 5.11 cumple con los parámetros de validación externa propuestos por Golbraikh *et al.*, 2003. Los valores resultantes de este análisis se exhiben a continuación:

$$k = 0.95 \text{ y } k' = 1.04$$

$$R_0^2 = 0.76 \text{ y } R_0'^2 = 0.61$$

$$1 - R_0^2 / R_{pred}^2 = 0.014 \text{ y } 1 - R_0'^2 / R_{pred}^2 = 0.21$$

$$R_m^2 = 0.69$$

Ahora para examinar la contribución de los cinco descriptores moleculares involucrados en el modelo QSAR se brindará una breve descripción de cada descriptor no conformacional a continuación:

- Un descriptor basado en la matriz 2D representado por $J.DZ(i)$, se obtiene a partir de un índice de conectividad J de Balaban obtenido de la matriz Barysz ponderada por el potencial de ionización.
- Dos descriptores de autocorrelación 2D denominados $ATSC5s$ y $JGI7$, proporcionan información originada por la autocorrelación centrada de Broto-Moreau de distancia 5 ponderada por el estado-I, y el índice de carga topológica media de orden 7, respectivamente.
- Un descriptor de valores propios de carga, representado por $pMax2.Bh(m)$, que indica el mayor auto valor n. 2 de la matriz de carga ponderada por la masa atómica.
- Un descriptor constitucional de fragmento centrado en el átomo de H catalogado como $H.052$ que aporta información sobre los átomos de H enlazados a un grupo sustituyente CO (sp3) enlazado a un átomos de Carbono.

La matriz de correlación de la Tabla 5.6 exhibe la ausencia de altas intercorrelaciones entre los cinco descriptores mencionados anteriormente. Los valores de cada descriptor molecular y los datos de la actividad larvícida $\log_{10} LC_{50}$ experimental y predicha por la Ec. 5.11 se encuentran disponibles en las Tablas 5.4.3 y 5.4.4 del material anexo.

Tabla 5.6. Matriz de correlación de los descriptores moleculares incluidos en la Ec. 5.11.

Ec. 5.11	$J.DZ(i)$	$ATSC5s$	$JGI7$	$pMax2.Bh(m)$	$H.052$
$J.DZ(i)$	1	0.154	0.243	0.034	0.413
$ATSC5s$		1	0.101	0.389	0.184
$JGI7$			1	0.514	0.122
$pMax2.Bh(m)$				1	0.027
$H.052$					1

La Figura 5.16.A corresponde al gráfico de predicción de los 62 compuestos semi-sintéticos como una función de la información experimental (LC_{50}) reportada para cada subconjunto molecular. Asimismo, la figura 5.16.B exhibe la distribución aleatoria de los valores de residuo dentro del rango de toxicidad aguda delimitado por las predicciones del modelo. Aunque, se detectó un valor atípico de la propiedad predicha por la Ec. 5.11. La molécula **55** del subconjunto *caI* presenta un comportamiento irregular con respecto a las demás moléculas evaluadas.

Después de verificar la estructura molecular digitalizada, los valores de los descriptores moleculares y la fuente de los datos experimentales, se encontró que el modelo propuesto no tiene la habilidad de predecir acertadamente esta molécula.

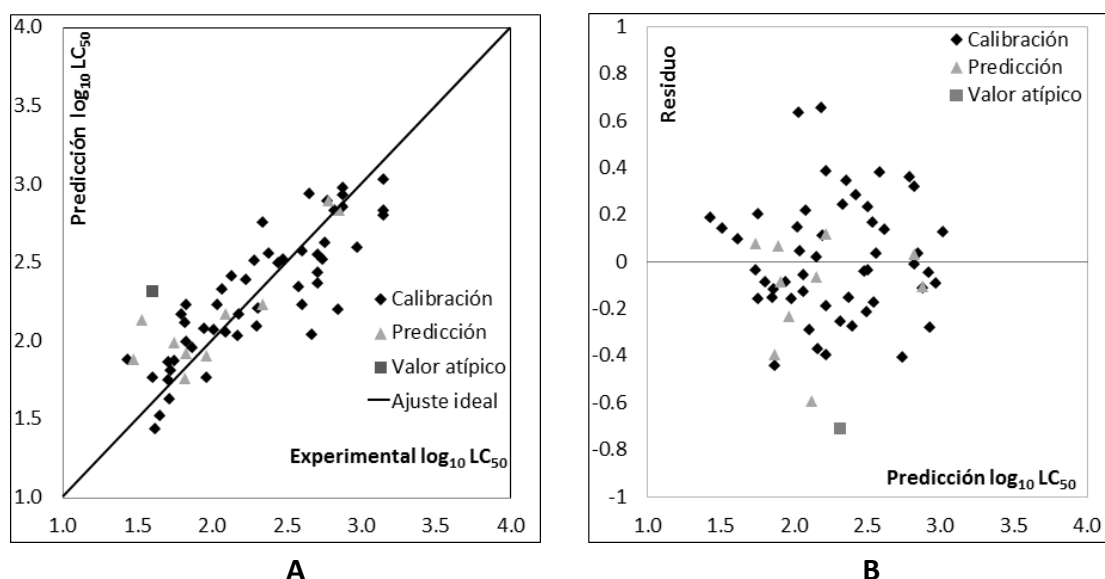


Figura 5.16. A). Gráfico de la actividad larvicida $\log_{10} LC_{50}$ experimental y calculada por la Ec. 5.11. **B).** Dispersión de los valores de residuo.

Conforme al análisis del AD de la Ec. 5.11, se obtuvo un diagrama de Williams exhibido en la Figura 5.17. Donde se puede apreciar que el poder predictivo del modelo no se ve afectado por la valor atípico del compuesto **55**, ya que las moléculas que comprenden el subconjunto *pred* se encuentran dentro del AD del modelo; lo que indica que las predicciones de $N_{pred} = 10$ son confiables y provienen de una verdadera relación estructura-actividad.

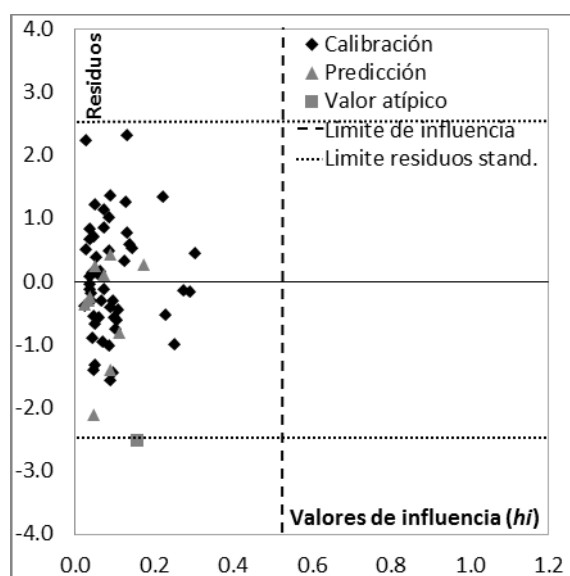


Figura 5.17. Valores de influencia (h_i), límite de influencia ($h^* = 0.53$) y residuo estandarizado $\sigma(2.5\sigma)$ de la Ec. 5.11.

Después de verificar el comportamiento de los valores estimados por el modelo QSAR, se logró apreciar que diferentes moléculas de los subconjuntos *cal* y *pred* exhiben alta capacidad larvicida sobre el *A. aegypti* L. Los compuestos **7** y **8** son hidrocarburos cíclicos insaturados con enlaces dobles *endo* y *exo* que exhiben un potencial larvicida efectivo en concentraciones menores a $50 \mu\text{g} / \text{mL}$ ($\log_{10} < 1.7$). Los valores de la actividad larvicida observada y estimada por la Ec. 5.11, así como los valores de influencia (h_i) del modelo se encuentran exhibidos en la Tabla 5.4.4 del material anexo

Además, los compuestos **55**, **19**, **21** y **50** derivados de la naturaleza son análogos funcionales y estructurales, conformados por un anillo aromático unido a un grupo éster activador con sustituyente de tipo hidroxilo, cloro o benceno, y presentan un poder larvicida similar.

Por último, se encontró que la molécula **59** con estructura molecular alifática ($\text{CO}(\text{CH}_2)_6$) tiene un importante valor de toxicidad aguda media (LC_{50}), análogo al poder larvicida de la molécula **55**, aunque no son análogos estructurales.

Conforme a lo anterior, se investigó el rendimiento del modelado QSAR cuando se evalúa un conjunto de datos más específico, conformado por moléculas de igual clase químicas. Se buscaron las mejores regresiones lineales para dos conjuntos moleculares más específicos, basados en el conjunto de datos inicial.

El primer conjunto definido como **I**, contiene 34 moléculas con un anillo aromático (**14-30**, **32**, **33**, **36-38**, **46-56** y **58**). Para iniciar el modelado se aplicó la técnica BSM, generando los subconjuntos moleculares $N_{cal} = 27$ (**14**, **16**, **19-24**, **26-30**, **32**, **33**, **36**, **38**, **46**, **48-50**, **52-56** y **58**) y $N_{pred} = 7$ (**15**, **17**, **18**, **25**, **37**, **47**, **51**).

El segundo conjunto denominado **II**, comprende 28 compuestos de cadena carbonada alifática, alicíclica y bicíclica (**1-13**, **31**, **34**, **35**, **39**, **40-45**, **57**, **59-62**); en este caso, la técnica BSM generó dos subconjuntos, $N_{cal} = 22$ (**1**, **2**, **4-8**, **10**, **11**, **13**, **31**, **34**, **35**, **41-45**, **57** y **59-61**) y $N_{pred} = 6$ (**3**, **9**, **12**, **39**, **40** y **62**).

La Tabla 5.7 exhibe los mejores modelos de regresión QSAR de $d=1-5$ descriptores definidos para los conjuntos **I** y **II**, mediante la exploración coherente de $d=1738$ descriptores realizada por la técnica RM.

Para el conjunto **I** se encontró el mejor modelo QSAR con cuatro descriptores moleculares; mientras que para el conjunto **II**, tan solo dos descriptores simples fueron requeridos para establecer una correcta relación estructura-actividad.

Tabla 5.7. Mejores modelos QSAR con $d = 1 - 5$ descriptores más representativos para los conjuntos de datos I y II. El mejor modelo se encuentra marcado en negrita.

Conjunto I con N = 34						
d	R^2_{cal}	S_{cal}	R^2_{pred}	S_{pred}	$R^2_{ij\max}$	Descriptors moleculares
1	0.27	0.40	0.71	0.20	0	<i>GATS7p</i>
2	0.44	0.36	0.71	0.24	0.01	<i>GATS7p, CATS2D_08_DD</i>
3	0.56	0.33	0.53	0.32	0.16	<i>MATS1v, MATS7i, Eig13_AEA(dm)</i>
4	0.66	0.29	0.60	0.38	0.35	<i>GATS5i, Eig13_AEA(dm), C.026, CATS2D_03_DA</i>
5	0.77	0.25	0.71	0.49	0.94	<i>MATS5i, GATS5i, Eig13_AEA(dm), CATS2D_03_DA, CATS2D_09_AL</i>
Conjunto II con N = 28						
1	0.62	0.32	0.75	0.27	0	<i>X4</i>
2	0.83	0.22	0.80	0.33	0.25	<i>ATS2m, Eig10_EA(bo)</i>
3	0.93	0.14	0.70	0.50	0.19	<i>nCt, CATS2D_06_LL, B01[C-O]</i>
4	0.96	0.11	0.89	0.59	0.18	<i>ICR, Eig10_EA(dm), nCt, LLS:_02</i>
5	0.97	0.10	0.90	ND *	0.46	<i>ICR, Eig10_EA(dm), nCt, F02[C-O], LLS_02</i>

*No Disponible

De acuerdo con los resultados de la Tabla 5.7, los modelos QSAR más específicos no logran superar el poder predictivo del modelo general definido en la Ec. 5.11. De este modo, se comprobó que la Ec. 5.11 define la mejor relación cuantitativa estructura-actividad establecida para el conjunto de datos con 62 moléculas larvicidas.

Los valores de los descriptores moleculares involucrados en los modelos QSAR propuestos para los conjunto I y II se encuentran disponibles en las Tabla 5.4.5 y 5.4.6 del material anexo. Además, se pueden observar los gráficos de predicción y residuo para cada conjunto molecular (I y II) en las Figura anexas 5.4.A y 5.4.B.

Otro enfoque considerado durante este estudio fue la búsqueda de un modelo QSAR basado en descriptores moleculares de fácil interpretación y con capacidad de anticipar el efecto tóxico agudo de las moléculas sobre las etapas inmaduras del mosquito *A. aegypti* L., puntualmente en larvas de III y IV *instar*.

Para esto, se buscaron los descriptores moleculares más simples entre los 4885 descriptores del programa comercial Dragón 6. Seguidamente, se empleó el método RM para buscar los mejores modelos QSAR con $d = 1 - 6$, explorando un conjunto con 223 descriptores de fácil interpretación exhibidos en la Tabla 5.8.

Tabla 5.8. Los mejores modelos QSAR con $d=1-6$ descriptores a partir de un conjunto de 233 descriptores de fácil interpretación. El mejor modelo se marca en negrita.

d	R_{cal}^2	S_{cal}	R_{pred}^2	S_{pred}	R_{ijmax}^2	Descriptores moleculares
1	0.26	0.42	0.79	0.41	0.00	$B06[C-C]$
2	0.49	0.35	0.79	0.33	0.03	$Rbrind, BLTF96$
3	0.54	0.34	0.85	0.32	0.42	$Se, Rbrind, BLTF96$
4	0.60	0.32	0.87	0.33	0.21	$TRS, TIE, Ui, BLTF96$
5	0.62	0.31	0.87	0.35	0.21	$TRS, TIE, nR=Ct, BLTF96$
6	0.69	0.29	0.82	0.40	0.56	$nRCOOR, nHAcc, B06[C-C], B10[C-O], F05[C-O], Uc$

El modelo QSAR, definido en la Ec. 5.12, muestra una capacidad predictiva del $\log_{10} LC_{50}$ superior a los otros modelos, e incluye cuatro descriptores de fácil interpretación, asociados a la respuesta final bajo estudio.

$$\log_{10} LC_{50} = 3.38 + 0.07(\pm 0.02)TRS + 0.02(\pm 0.003)TIE - 0.3(\pm 0.06)Ui + 0.38(\pm 0.07)BLTF96 \quad (5.12)$$

$$N_{cal} = 52, d = 4, R_{cal}^2 = 0.60, S_{cal} = 0.32, N_{cal} / d = 13, F = 17, R_{ijmax}^2 = 0.21, o(2.5S) = 2, \\ R_{LOO}^2 = 0.50, S_{LOO} = 0.36, S^{aleat} = 0.40 \\ N_{pred} = 10, R_{pred}^2 = 0.87, S_{pred} = 0.33$$

Los descriptores moleculares simples involucrados en la Ec. 5.12 pertenecen a tres diferentes clases de descriptores, detalladas a continuación:

- Dos descriptores de propiedades moleculares representados como Ui y $BLTF96$, que contiene información acerca del índice de insaturación de los compuestos y la toxicidad Verhaar de referencia en peces y ecosistemas acuáticos considerando el parámetro MLOG P (mmol/l), respectivamente.
- Un descriptor constitucional de grupo aromático definido como TRS , con información del tamaño total del anillo aromático.
- Un descriptor basado en ITs descrito como TIE , que codifica información de los índices topológicos de estados-E.

A partir de la Tabla 5.9 es posible apreciar que los cuatro descriptores simples no presentan inconvenientes de intercorrelaciones demasiado altas. Los valores de los descriptores más simples se incluyen en la Tabla 5.4.7 del apartado anexo.

Tabla 5.9. Matriz de correlación de los descriptores simples incluidos en la Ec. 5.12.

Ec. 5.12	TRS	TIE	Ui	BLTF96
TRS	1	0.013	0.026	0.03
TIE		1	0.452	0.46
Ui			1	0.13
BLTF96				1

Los resultados obtenidos muestran una bondad de ajuste más baja para el modelo 5.12 ($R_{cal}^2 = 0.60$, $S_{cal} = 0.32$), y no logra superar los resultado obtenidos para el modelo 5.11 ($R_{cal}^2 = 0.69$, $S_{cal} = 0.28$). La Ec. 5.12 involucra dos valores atípicos con un valor de residuo mayor a 2.5 veces S_{cal} .

Con el fin de verificar la correlación encontrada entre los 62 datos observados y los predichos por la Ec. 5.12, se generó el gráfico exhibido en la Figura 5.18.A, y el grado de dispersión de los valores de residuo se dispone en la Figura 5.18.B.

Si bien la Ec. 5.12 presenta mayor capacidad predictiva de N_{pred} ($R_{pred}^2 = 0.87$, $S_{pred} = 0.33$) en comparación con la Ec. 5.11 ($R_{pred}^2 = 0.78$, $S_{pred} = 0.39$); este resultado puede ser casual, ya que la calidad de N_{cal} empeora para el modelo 5.12. Además, la prueba CV-LOO revelo la baja estabilidad de la Ec. 5.12 ($R_{LOO}^2 = 0.50$, $S_{LOO} = 0.36$) con respecto a la Ec. 5.11 ($R_{LOO}^2 = 0.60$, $S_{LOO} = 0.32$), y se ubica en el límite permitido ($R_{LOO}^2 > 0.5$) por dicha prueba.

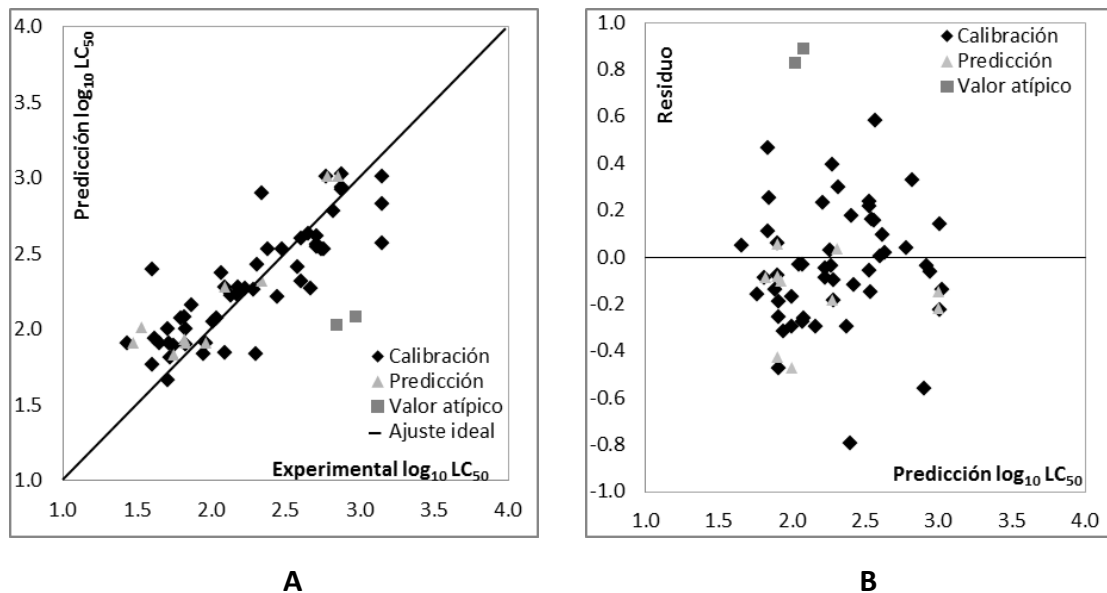


Figura 5.18. A). Valores de predicción del $\log_{10} LC_{50}$ en función de la información experimental. **B).** Dispersión de los valores de residuo.

Conforme a los valores de la prueba de Aleatorización-Y ($S_{cal} < S^{aleat}$ (0.40)) se determinó que el modelo de la Ec. 5.12 es estable. También, cumple con los criterios propuesto por Golbraikh *et al.*, 2003, obteniendo los siguientes valores:

$$k = 0.93 \text{ y } k' = 1.07$$

$$R_0^2 = 0.86 \text{ y } R_0'^2 = 0.82$$

$$1 - R_0^2 / R_{pred}^2 = 0.005 \text{ y } 1 - R_0'^2 / R_{pred}^2 = 0.05$$

$$R_m^2 = 0.81$$

A partir de los resultados alcanzados con este estudio se estableció que la mejor herramienta *in silico* obtenida es el modelo QSAR expresado en la Ec. 5.11, que involucra cinco descriptores independientes de la conformación con capacidad de predecir acertadamente la actividad larvicida sobre el *A. aegypti* L. de 62 sustancias químicas de origen natural o semisintético, requeridas en el diseño ecocompatible de nuevos larvicidas para reducir los casos clínicos asociados a los virus DENV, ZIKV o CHIKV.

5.4.6. Conclusiones

Un conjunto de datos conformado por 62 moléculas derivadas de las plantas fue estudiado a través de tres diferentes enfoques QSAR, donde se consideró un conjunto de datos general y dos conjuntos de datos específicos de magnitud $I = 34$ moléculas y $II = 28$ moléculas, seleccionados según los fragmentos subestructurales aromáticos o alifáticos, respectivamente.

Luego, cada conjunto de datos fue dividido de forma equitativa a través del Método de Subconjuntos Balanceados (BSM). Se buscaron los mejores modelos de regresión lineal por medio de la técnica RM, examinando 4885 descriptores moleculares del programa Dragon 6.

Los resultados obtenidos mediante la técnica MLR junto con RM fueron exitosos, con parámetros estadísticos de valores adecuados que corroboran la calidad, veracidad y robustez de las ecuaciones. En definitiva, el mejor modelo QSAR encontrado incluye cinco descriptores no conformacionales con capacidad de anticipar la respuesta final larvicida en moléculas nuevas, hipotéticas o carentes de esta información biológica.

5.5. Análisis QSAR de agentes derivados de las plantas con actividad larvicida sobre el *Aedes aegypti* L. mediante descriptores de acceso libre

En particular, la necesidad de diseñar herramientas computacionales aplicables, reproducibles, sencillas y acertadas para anticipar diversos efectos biológicos finales generados por una o varias sustancias químicas en un delimitado organismo o entorno, motivo el desarrollo del presente estudio QSAR enfocado en predecir el poder larvicida sobre el *A. aegypti* L. de un conjunto molecular compuesto por 60 metabolitos secundarios y sus análogos semisintéticos.

A través del modelado QSAR se establecen paralelismos matemáticos racionales que contemplan únicamente los aspectos más simples de la estructura molecular, codificados por miles de descriptores constitucionales y topológicos, generados por diferentes programas de acceso libre.

La exclusión de los descriptores moleculares geométricos, por ejemplo el descriptor de distribución de carga que requiere una representación 3D, evita las ambigüedades derivadas de la existencia de varios estados conformacionales exhibidos para una misma molécula, dependiendo de las rotaciones del enlace y el ángulo entre enlaces.^{2, 67, 68} Por lo tanto, los valores de los descriptores 3D generalmente cambian según el usuario, los programas y las aproximaciones de la estructura molecular activa.

5.5.1. Datos de la actividad larvicida (LC_{50})

Para desarrollar un nuevo modelo QSAR basado en los aspectos no conformacionales de las moléculas, se utilizó el mismo conjunto de datos evaluado previamente por los modelos lineales QSAR expresados por las ecuaciones 5.11 y 5.12.

En este caso, fue necesario excluir las moléculas **27** y **56** detalladas en la Figura 5.15, ya que presentaban un importante error de medida experimental de la actividad larvicida sobre el mosquito *A. aegypti* L. De este modo, se evaluó un conjunto de moléculas extraído de la literatura con 60 datos de toxicidad aguda media (LC_{50}) sobre larvas del vector del dengue y el zika.⁶¹⁻⁶⁴

5.5.2. Descriptores moleculares

Inicialmente, se dibujaron las estructuras químicas en un formato bidimensional (2D) por medio del programa de acceso libre ACDLabs ChemSketch.^{69, 70} Este recurso gratuito permite digitalizar las estructuras moleculares en diversos formatos de representación molecular.

Además, permite calcular algunas propiedades atómicas como polarizabilidad, volumen molar, peso molecular; o propiedades fisicoquímicas de los compuestos como refractividad molar, constante dieléctrica, índice de refracción, tensión superficial o el Log P. Incluso, cuenta con módulos para verificar las formas tautoméricas.

Las 60 estructuras químicas analizadas se generaron tanto en notación SMILES canónica como en representación molecular 2D, sin realizar la optimización de la geometría; luego se guardaron en formato MDL mol (V2000).

A partir del conjunto molecular conformado por 60 larvicidas inspirados en la naturaleza, se codificaron las características moleculares independientes de la conformación a través de los descriptores moleculares e indicadores de fragmentos incluidos en diferentes programas disponibles de acceso libre.

El primer programa utilizado es el PaDEL versión 2.20 con capacidad de generar 1444 descriptores 0D-2D y 12 tipos de indicadores de fragmentos (16.092 bits).^{71, 72} Seguidamente, se utilizó el programa gratuito Mold², programado por la FDA US para generar 777 descriptores que codifican la estructura molecular bidimensionalmente a partir de archivos de entrada en formato MLD sdf.⁷³

Además, se agregaron 14 descriptores semiempíricos generados por el programa libre EPI Suite;⁷⁴ basado en propiedades fisicoquímicas y estimaciones con aplicación ambiental, tal como la constante de la ley de Henry ($\log K_H EPI$), el coeficiente de sorción del suelo y los sedimentos ($\log K_{oc} EPI$) y el logaritmo del coeficiente de partición octanol/agua ($\log K_{ow} EPI$).

Con el procedimiento anterior, se obtuvo un amplio conjunto de descriptores moleculares con 18326 variables no conformacionales, que brinda más información acerca de las características estructurales más relevantes de este conjunto de datos, generando más posibilidades estructurales para establecer paralelismos matemáticos relacionados con la respuesta larvicida final sobre el *A. aegypti* L.

Finalmente, a este conjunto de descriptores se le aplicó el análisis de dependencia lineal entre pares de variables, así como la exclusión de los descriptores constantes, casi constantes o con valores ausentes. Para finalmente obtener un conjunto representativo con 1750 descriptores linealmente independientes.

5.5.3. Selección del modelo QSAR

La técnica de selección de descriptores RM basada en MLR,^{39, 40} se empleó para explorar un conjunto de $d=1705$ los descriptores moleculares más representativos para la toxicidad aguda media (LC_{50}) de un conjunto de 60 moléculas; con el objetivo de encontrar las mejores regresiones lineales con error mínimo, controlando la desviación estándar (S_{cal}) o error cuadrático medio (RMS_{cal}) del subconjunto *cal*.

La técnica RM ha resultado útil en diferentes estudios de la teoría QSAR para descubrir el paralelismo matemático existente entre los aspectos estructurales no conformacionales de las moléculas y un delimitado efecto biológico final generado por las mismas.⁶⁶⁻⁶⁸

Conforme a lo anterior, se buscaron los mejores modelos de predicción QSAR basados en los descriptores 0D, 1D y 2D más representativos para estimar la actividad larvícida sobre las etapas inmaduras del vector de múltiples enfermedades endémicas.

5.5.4. Validación interna y externa del modelo

Actualmente, el análisis de un conjunto de moléculas de prueba externo, no considerado por el modelo durante la calibración de sus parámetros, se conoce como el criterio de validación más confiable. En este sentido, el conjunto de datos inicial ($N=62$) fue dividido por medio del BSM;⁶⁶ esta técnica ha garantizado la división equitativa de cada subconjunto molecular en diversos estudios de la teoría QSAR.

Asimismo, se utilizaron distintos criterios teóricos para validar internamente el modelo QSAR, tal como la validación cruzada (CV) LOO que evalúa la estabilidad del modelo ante la exclusión/inclusión de una molécula.⁴¹

Del mismo modo, se empleó la técnica de validación cruzada de LMO (donde M debe ser el porcentaje de moléculas eliminadas del subconjunto *cal*). En este caso, se excluye el 15% de los datos de *cal* (8 moléculas). Después de ejecutar 100 mil casos de eliminación aleatoria de datos, los resultados se contemplan a través de los parámetros estadísticos R_{LMO}^2 y S_{LMO} .⁷⁵

Además, se recurrió al análisis de Aleatorización-Y para medir el nivel de veracidad del modelo de regresión.⁴² Al igual que en los estudios QSAR previos, se evaluaron los parámetros postulados por Golbraikh para regular la capacidad predictiva del modelo.

Seguidamente, se evaluó el parámetro *VIF* que indica cuánto se infla la varianza del coeficiente del descriptor en comparación con el caso en que los descriptores son completamente ortogonales entre sí. Idealmente, el valor *VIF* de un descriptor debe ser inferior a 10.^{76,77}

Por último se aplicó la técnica de influencia para definir el AD del modelo lineal QSAR, y determinar el nivel de incertidumbre de sus predicciones; especialmente del subconjunto *pred* u otros conjuntos moleculares externos.^{55,75}

5.5.5. Resultados y Discusión

Un conjunto de datos compuesto por 60 moléculas naturales y sus análogas semisintéticas con prominente capacidad tóxica sobre las larvas de *A. aegypti* L. en III-IV *instar* fue utilizado como objeto de estudio para diseñar un modelo QSAR.

En este caso, se inició con la división del conjunto molecular en dos subconjuntos equilibrados en relación estructura-actividad generados por la técnica BSM; el subconjunto de mayor magnitud de datos se utilizó para calibrar el modelo ($N_{cal} = 50$),

y las moléculas remanentes conformaron el subconjunto $pred$ ($N_{pred} = 10$), utilizado para medir el poder predictivo del modelo más adelante.

Seguidamente, se examinaron las características estructurales más representativas del subconjunto N_{cal} por medio de la técnica RM basada en MLR, con un criterio de selección adecuado para explorar 1750 descriptores 0D-2D linealmente independientes.

El criterio considerado para seleccionar los mejores modelos lineales se enfoca en los valores de los parámetros estadísticos R^2 , S y $R_{ij\max}^2$; también utilizamos la práctica común de mantener el tamaño del modelo lo más pequeño posible, para evitar cualquier posible correlación inusual

Con base en los resultados exhibidos en la Tabla 5.10 es posible apreciar que el parámetro S_{cal} disminuye ante el incremento del número de descriptores; mientras que el parámetro S_{pred} exhibe valores aceptables hasta $d = 5$.

Tabla 5.10. Los modelos QSAR más representativos de $d = 1 - 7$ descriptores no conformacionales de acceso libre. El mejor modelo QSAR se resalta en negrita.

d	R_{cal}^2	S_{cal}	R_{pred}^2	S_{pred}	$R_{ij\max}^2$	Descriptores moleculares
1	0.43	0.36	0.71	0.31	0.00	<i>C3SP2</i>
2	0.67	0.27	0.87	0.21	0.06	<i>ATSC3v, log K_{oc}EPI</i>
3	0.79	0.22	0.92	0.20	0.40	<i>SdssC, KR3224, KR3369</i>
4	0.82	0.20	0.91	0.22	0.40	<i>AATSC0v, SdssC, KR3224, KR3369</i>
5	0.84	0.20	0.92	0.23	0.75	<i>SP.6, VP.2, TIC0, PC2, log K_{ow}EPI</i>
6	0.87	0.18	0.88	0.33	0.44	<i>ATS3e, SpMAx7: Bhm, ETA_BetaP_s, PC544, PC581, D777</i>
7	0.90	0.16	0.88	0.37	0.48	<i>ATS4p, MATS1m, nHCsatu, MDEO.11, KR2, KR3682, log K_{ow}EPI</i>

Conforme a lo anterior, se eligió la relación estructura-actividad que contiene cinco descriptores con alto poder predictivo del conjunto externo de predicción, expresada en la Ec. 5.13. El significado de los descriptores moleculares elegidos entre las mejores regresiones lineales se exhibe en la Tabla 5.5.1 del apartado anexo.

La Ec. 5.13 exhibe una sobresaliente capacidad predictiva del subconjunto N_{pred} , donde los porcentajes de la varianza explicada son $R_{cal}^2 = 84\%$ y $R_{pred}^2 = 92\%$. Asimismo, se determinó que el parámetro S presenta un nivel bajo de dispersión de los datos de predicción de cada subconjunto con valores de $S_{cal} = 0.20$ y $S_{pred} = 0.23$.

$$\log_{10} LC_{50} = 3.14 + 0.25 \cdot SP.6 + 0.30 \cdot VP.2 - 0.03TIC_0 + 0.30 \cdot PC2 - 0.46 \cdot \log K_{ow}EPI \quad (5.13)$$

$$N_{cal} = 50, d = 5, R_{cal}^2 = 0.84, S_{cal} = 0.20, N_{cal} / d = 10, R_{ij\max}^2 = 0.74, o(2.5S) = 0, \\ R_{LOO}^2 = 0.80, S_{LOO} = 0.22, R_{LMO}^2 = 0.72, S_{LMO} = 0.26, S^{aleat} = 0.36 \\ N_{pred} = 10, R_{pred}^2 = 0.92, S_{pred} = 0.23$$

Seguidamente, el modelo QSAR de la Ec. 5.13 superó las pruebas de validación cruzada interna LOO y LMO, empleadas para verificar la estabilidad del modelo ante la ausencia de 1 ($R_{LOO}^2 = 0.80, S_{LOO} = 0.22$) u 8 datos ($R_{LMO}^2 = 0.72, S_{LMO} = 0.26$) en N_{cal} .

Los resultados de la prueba de Aleatorización-Y ($S_{cal} < S^{aleat}$ (0.36)) ratifican que la Ec. 5.13 es una relación estructura-actividad larvicida coherente y robusta después de analizar 100 mil casos de aleatorización, e incluso cumple con los criterios de validación de Golbraikh *et al.*, 2003 con los siguientes valores:

$$k = 1.01 \text{ y } k' = 0.98 \\ R_0^2 = 0.91 \text{ y } R_0'^2 = 0.92 \\ 1 - R_0^2 / R_{pred}^2 = 0.01 \text{ y } 1 - R_0'^2 / R_{pred}^2 = 2.3 \times 10^{-4} \\ R_m^2 = 0.85$$

A través de la Figura 5.19.A se exhiben las 60 actividades larvicidas predichas como una función de valores observados para las moléculas de los subconjuntos *cal* y *pred*. La información numérica del gráfico se encuentra en la Tabla 5.5.2 del material anexo. El diagrama de dispersión de los valores de residuo ilustrado por la Figura 5.19.B, revela que los residuos tienden a obedecer un patrón aleatorio alrededor de la línea cero, siguiendo el criterio de la técnica MLR.

Los cinco descriptores no conformacionales involucrados en la Ec. 5.13 pertenecen a 4 diferentes clases, detalladas a continuación:

- Dos descriptores de conectividad chi de PaDEL representados por *SP.6*, camino simple de orden 6; *VP.2* camino de valencia de orden 2.
- Un descriptor de contenido de información denominado TIC_0 , índice de contenido de información total (simetría de vecinos más cercanos de orden 0).
- Un descriptor indicador de fragmentos Pubchem definido por *PC2*, indicador de la presencia de átomos individuales interpretados por el símbolo H_{16} .
- Un descriptor semiempírico representado como $\log K_{ow}EPI$, codifica el logaritmo del coeficiente de partición octanol / agua calculado a partir del módulo KOWWIN del programa EPI Suite.

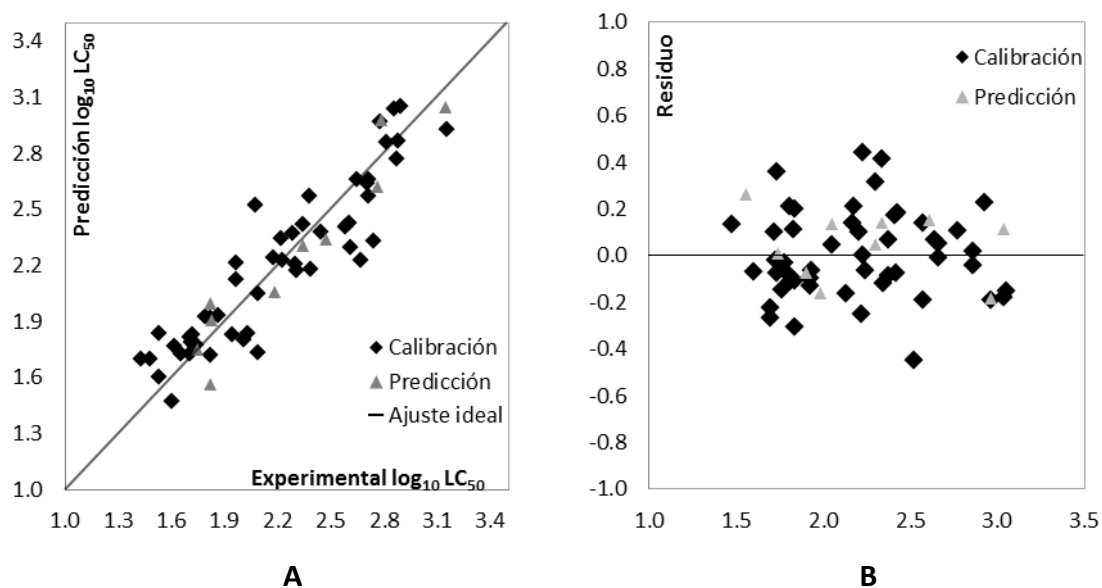


Figura 5.19. A). Valores de predicción del $\log_{10} LC_{50}$ en función de la información experimental. **B).** Dispersión de los valores de residuo.

Los descriptores del modelo QSAR obtenido presentan valores de intercorrelación bajos, acorde con la Tabla 5.11. Además, conforme a los resultados del parámetro *VIF*, los descriptores de la Ec. 5.13 tienen un valor *VIF* menor a 10 ($VIF \approx 2$); indicando que la multicolinealidad no es un problema en la selección de variable de la técnica RM.

Tabla 5.11. Matriz de los coeficientes de correlación de los descriptores incluidos en la Ec. 5.13 y sus valores *VIF*.

Ec. 5.13	<i>SP.6</i>	<i>VP.2</i>	<i>TIC₀</i>	<i>PC2</i>	$\log K_{ow} EPI$	<i>VIF</i>
<i>SP.6</i>	1	0.39	0.74	0.09	0.23	2.4
<i>VP.2</i>		1	0.51	0.15	0.45	1.8
<i>TIC₀</i>			1	0.05	0.24	2.8
<i>PC2</i>				1	0.12	1.0
$\log K_{ow} EPI$					1	1.3

Seguidamente, con la exploración del AD de la Ec. 5.13 se generó el diagrama de Williams expuesto en la Figura 5.20. En este caso, se puede observar que los compuestos del subconjunto *pred* se encuentran dentro del AD, con valores de influencia inferiores al límite ($h^* = 0.36$).

Además, la mayoría de las moléculas del subconjunto *cal* presentan valores de influencia altos ($h_i = 0.22 - 0.04$), otorgando la posibilidad de estimar la actividad larvicida de diversas moléculas nuevas o aún inexploradas. Y solo una molécula superó el límite de advertencia h^* , con $h_i = 0.48$. Los datos del análisis de influencia, así como los valores de los cinco descriptores están disponibles en las Tablas 5.5.2 y 5.5.3 del material anexo.

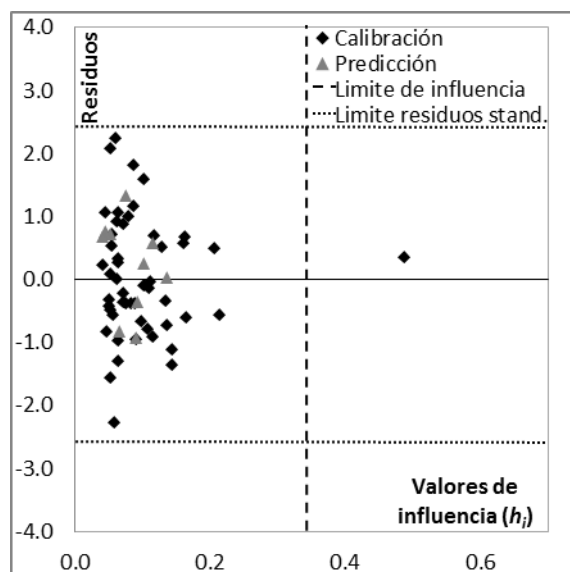


Figura 5.20. Valores de influencia (h_i), límite de influencia ($h^* = 0.36$) y residuo estandarizado $\sigma(2.5S)$ de la Ec. 5.13.

Por otra parte, el grado de contribución de los descriptores del modelo fue evaluado por el parámetro b_j^2 ; donde se determinó que los descriptores $\log K_{ow}EPI$, $VP.2$ y TIC_0 tienen mayor importancia para estimar la LC_{50} sobre larvas *A. aegypti* L, acorde al siguiente orden:

$$\log K_{ow}EPI (1.14) > VP.2 (0.97) > TIC_0 (0.61) > PC2 (0.33) > SP.6 (0.31)$$

En este contexto, los coeficientes de regresión de los descriptores pueden elucidar la influencia que tienen sobre la respuesta biológica final analizada.

Para el modelo 5.13 se realizó la interpretación expuesta en la Figura 5.21; donde la disminución de los valores de los descriptores $VP.2$ y $SP.6$, junto con el aumento de los descriptores $\log K_{ow}EPI$ y TIC_0 , y la ausencia del fragmento codificado por $PC2$, contribuye a la obtención de larvicidas naturales de efectividad plausible sobre el vector *A. aegypti* L.

$$\log_{10} LC_{50} = 3.14 + \boxed{0.25SP.6 + 0.29VP.2} \boxed{-0.03TIC_0} \boxed{+0.30PC2} \boxed{-0.46\log K_{ow}EPI}$$

se predice menor $\log_{10} LC_{50}$ solo si ➔ **Valores bajos de $VP.2$ y $SP.6$** ➔ **Valor = 0 del fragmento $PC2$** ➔ **Valores altos de $\log K_{ow}EPI$ y TIC_0**

Figura 5.21. Interpretación del grado de influencia de los descriptores de la Ec. 5.13.

Conforme a lo anterior, se comprobó que la interpretación dada para el modelo 5.13 se cumple en los compuestos **16**, **19** y **44** del subconjunto *cal* y **59** del subconjunto

pred, los cuales se predicen como activos o inactivos en coherencia con los valores de los cinco descriptores del modelo, tal como se muestra en la Figura 5.20.

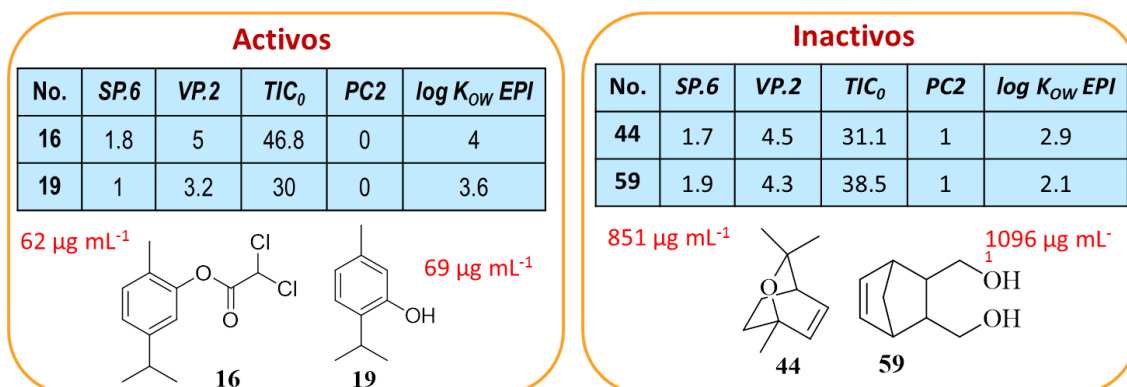


Figura 5.22. Moléculas con actividad asociada a las condiciones estructurales estipuladas por los descriptores de la Ec. 5.13.

Por último, se comprobó que la calidad estadística del modelo QSAR de la Ec. 5.13 es similar o supera varias herramientas computacionales reportadas hasta el momento, desarrolladas en el siguiente orden:

1. Scotti *et al.*, 2014:
2. Alencar-Filho *et al.*, 2016
3. De *et al.*, 2018
4. Dos Santo *et al.*, 2018.

La información completa de estos estudios computacionales se encuentra disponible en el capítulo 4, sección 14. Cabe resaltar que el modelo de la Ec. 5.13 también logró superar la calidad estadística y el poder predictivo de los modelos QSAR 5.11 y 5.12 detallados anteriormente.

5.5.6. Conclusiones

Con base en los resultados obtenidos en el presente estudio QSAR se encontró un modelo conformado por cinco descriptores independientes de la conformación, generados por diferentes programas de acceso libre. Este modelo exhibe un rendimiento predictivo aceptable del subconjunto *pred*, y puede cumplir con otras condiciones matemáticas necesarias, como la CV-LOO o CV-LMO, la técnica de Aleatorización-Y o el análisis del parámetro *VIF*. Además, el análisis del AD postuló que para predecir la actividad larvicida de nuevas moléculas con certeza, estas deben presentar un valor de influencia h_i menor al límite de advertencia (h^*) del modelo.

Asimismo, el grado de contribución de los descriptores moleculares detallado antes, permite elucidar las características estructurales que una sustancia química, ya sea de origen natural o de síntesis debe tener para exhibir un valor de toxicidad aguda

(LC₅₀) favorable en el control prematuro de los vectores transmisores de diferentes arbovirus.

Además, se logró proponer con éxito un modelo QSAR que no requiere la representación conformacional de las moléculas, ya que contempla únicamente los aspectos constitucionales y topológicos de las mismas, codificados por medio de miles de descriptores moleculares disponibles en diferentes programas de acceso libre.

5.6. Estudio QSAR no conformacional de 263 larvicidas derivados de las plantas sobre el vector del dengue y el zika, el *Aedes aegypti* L.

Ante el constante interés por desarrollar nuevos modelos QSAR para contribuir al control del vector *A. aegypti* L. se buscó proponer un novedoso modelo QSAR orientado en predecir un amplio conjunto molecular con 263 compuestos derivados de las plantas, recolectados de la literatura especializada para identificar las principales características estructurales que pueden influir en la actividad larvicida sobre el *A. aegypti* L.

El objetivo del presente estudio fue encontrar modelos simples basados únicamente en los aspectos estructurales 0D, 1D y 2D de las moléculas, codificados a través de los descriptores moleculares incluidos en los programas libres detallados en la sección anterior, junto con un nuevo programa de cálculo de descriptores denominado QuBiLs-MAS.^{78,79}

La principal ventaja de no analizar las conformaciones moleculares es que los únicos datos experimentales necesarios para desarrollar el modelo QSAR es la información experimental de la actividad bajo estudio. En este caso, los datos de la toxicidad aguda media (LC₅₀) reportados para las 263 sustancias químicas evaluadas.

5.6.1. Datos de la actividad larvicida (LC₅₀)

Un conjunto de datos con 263 compuestos larvicidas estructuralmente diversos, que incluyen quinonas, policétidos, fenilpropanoides, cumarinas, terpenoides, flavonoides, terpenos, alcaloides y diversos análogos semisintéticos, fue utilizado como objeto de estudio para desarrollar un nuevo modelo QSAR.⁸⁰⁻⁸²

En este caso, la actividad larvicida se define como la Concentración Letal media (LC₅₀), que indica la concentración a la que el 50% de la población de larvas entre III y IV *instar* muere después de 24 a 48 horas de tratamiento con la solución de prueba.

El conjunto de datos fue extraído de la literatura especializada y cada valor se convierte en escala logarítmica ($\log_{10} LC_{50}$) para fines de modelado. La lista completa del conjunto molecular estudiado y los valores experimentales de la actividad larvicida se proporcionan en el la Tabla 5.6.1 del apartado anexo.

5.6.2. Descriptores moleculares

El conjunto de descriptores teóricos y semiempíricos basados en los aspectos no conformacionales de la molécula se generó a través de los programas de acceso libre PaDEL,^{71, 72} Mold2,⁷³ y EPI Suite.⁷⁴ Además, se estudió y utilizó el programa QuBiLs para calcular otro tipo de descriptores bidimensionales calculados a partir de los mapas cuadráticos bilineales y *N*-lineales,^{78, 79} empleando las matrices de densidad electrónica teóricas de grafos y el módulo de ponderaciones atómicas MAS, del programa multiplataforma gratuito ToMoCoMD-CARDD.

El módulo algebraico QuBiLs-MAS tiene la capacidad de generar 8448 índices basados en tensor que pertenecen a 176 tipos de mapas algebraicos bilineales, cuadráticos y lineales, establecidos sobre índices métricos espaciales (matrices de disimilitud y ponderaciones atómicas).

Las matrices de *N*-tuplas se usan para representar las relaciones entre dos, tres y cuatro átomos, y también se pueden usar para codificar información relacionada con grupos o tipos de átomos que pertenecen a un fragmento molecular específico.⁷⁹

Para ejecutar el programa de acceso libre QuBiLs-MAS se deben seleccionar las siguientes opciones: formas algebraicas Bilineales (B), Cuadráticas (Q) y Lineales (F); restricciones dobles Basadas en Átomos (AB) y No Quirales (nCi). Formas de matrices de orden máximo 15 No Estocástica (NS), Estocástica Simple (SS), Estocástica Doble (DS) y Matriz de Probabilidad mutua (MP). Grupos totales Mantener Todo (KA) cortado.

Además, las propiedades incluidas son: el Log P (A) basado en Ghose-Crippen, Carga (C), Electronegatividad (E), Masa (M), Polarizabilidad (P), Área de Superficie Polar (PSA), Refractividad (R), Masa y volumen de van der Waals (V); Distancia euclídea (N2), Media Aritmética (alfa = 1) (AM) y Desviación Estándar (SD) invariantes con opción no estandarizada.

Posteriormente, se obtuvo una matriz de descriptores conformada por 26775 descriptores no conformacionales que codifican las estructuras moleculares del conjunto de datos para iniciar el estudio de la actividad larvicida. Luego, se identificaron los pares de descriptores linealmente dependientes, y solo se elimina una variable de cada par.

Con respecto a los descriptores irrelevantes; es decir, aquellos con valores constantes o casi constantes o con al menos un valor faltante, fueron excluidos de la matriz de descriptores original para minimizar la información redundante. De este modo, se obtiene una matriz con 10604 descriptores no conformacionales con alto contenido de información y linealmente independientes.

5.6.3. Selección del modelo QSAR

Se empleó la técnica de selección de variables RM basada en MLR, con el fin de generar modelos lineales explorando una matriz inicial conformada por $d=10604$ descriptores para obtener una matriz óptima con d descriptores ($d \ll D$) de menor desviación estándar (S_{cal}) o error cuadrático medio (RMS_{cal}) para el subconjunto *cal*.

Conforme a lo anterior, se buscaron los mejores modelos lineales basados en diferentes tipos de descriptores moleculares independientes de la conformación (0D, 1D y 2D) más representativos para predecir el efecto larvicida LC_{50} sobre el mosquito *A. aegypti* L de 263 compuestos derivados de los productos naturales.

5.6.4. Validación interna y externa del modelo

Como parte del procedimiento del modelado QSAR, se dividió el conjunto de datos con 263 compuestos derivados de las plantas mediante el método BSM en tres subconjuntos moleculares denominados calibración (*cal*), validación (*val*) y predicción (*pred*).

El subconjunto *cal* sirve para calibrar y obtener los parámetros del modelo QSAR, mediante la técnica RM; mientras que el subconjunto *val* ayuda a calibrar y validar de formar parcial el modelo al predecir la bioactividad de las moléculas no incluidas en *cal*. Seguidamente, el subconjunto *pred* contiene las moléculas no contempladas durante la construcción del modelo a través de los subconjuntos *cal* y *val*, y así puede demostrar el poder predictivo real del modelo QSAR obtenido.

Por otra parte, la validación teórica del modelo se ejecutó a través de pruebas comunes de validación como CV LOO y LMO. Se determinó la estabilidad del modelo ante la inclusión o exclusión de 1 molécula o del 30% de las moléculas de *cal* (40 moléculas), medida por los parámetros estadísticos R^2_{LOO} , RMS_{LOO} y $R^2_{L30\%O}$, $RMS_{L30\%O}$, respectivamente.

Además, se aplicaron otras pruebas de validación interna como Aleatorización-Y, ejecutando más 150 mil casos de aleatorización de los datos de la actividad. Se evaluó el parámetro *VIF* para los descriptores del modelo. La evaluación de los parámetros de validación teórica postulados por Golbraikh *et al.*, 2003,⁵⁴ siempre resulta necesaria para verificar la calidad predictiva del modelo QSAR obtenido.

Adicionalmente, acorde con los cinco postulados de la OCDE para definir la funcionalidad, aplicabilidad y reproducibilidad de un modelo QSAR, es necesario establecer la región teórica de aplicación del modelo con respecto a un nuevo conjunto molecular. Por ello, es indispensable definir el AD de cada modelo de predicción mediante un método adecuado para el conjunto de datos. En este caso, se ejecutó la técnica de influencia basada en criterios de distancia, detallada en el capítulo 3, sección 7, subsección 4 de la presente Tesis.

Finalmente, para verificar el grado de contribución de los descriptores moleculares del modelo se calculó el parámetro b_j^s basado en los valores estandarización de los coeficientes de regresión de los descriptores. Los valores numéricos se encuentran disponibles en la Tabla 5.6.5 del material anexo.

5.6.5. Resultados y Discusión

A partir de un amplio conjunto molecular constituido por 263 larvicidas derivados de las plantas contra el vector de *A. aegypti* L., se inició el análisis QSAR. Luego, se dividió el conjunto de datos inicial a través de la técnica BSM en tres subconjuntos equilibrados en términos de relación estructura-actividad, definidos como *cal* ($N_{cal} = 133$), *val* ($N_{val} = 65$) y *pred* ($N_{pred} = 65$); donde los compuestos de calibración incluidos en los subconjuntos *cal* y *val* constituyen el 75% de todo el conjunto de datos. Las estructuras moleculares de los 263 compuestos químicos, así como los valores de la actividad larvicida experimental y la identificación de los miembros de los subconjuntos *val* (*) y *pred* (^) se encuentran en la Tabla 5.6.1 del material anexo.

Las características estructurales más relevantes del subconjunto *cal* fueron examinadas por medio de la técnica de RM, que proporciona una forma de explorar los 10604 descriptores linealmente independientes. Los criterios de selección del modelo QSAR se enfocan en el valor mínimo de *RMS*, el máximo del R^2 y el mínimo de R_{ijmax}^2 .

De este modo se obtuvieron los mejores modelos de MLR con $d = 1 - 8$ descriptores, detallados en la Tabla 5.12. El significado de los descriptores moleculares más representativos para los 263 larvicidas analizados se encuentra en la Tabla 5.6.2 del material anexo.

A partir de los resultados de la Tabla 5.12 se aprecia claramente que los parámetros R_{cal}^2 y RMS_{cal} mejoran a medida que aumenta d en la ecuación lineal hasta $d = 8$. Por lo tanto, se logró verificar el funcionamiento adecuado de la técnica de RM.

Principalmente, se buscó mantener la dimensión del modelo lo más pequeña posible, lo que condujo a seleccionar el modelo con siete descriptores expresado en la Ec. 5.14, cuya calidad estadística es aceptable en los subconjuntos de *cal* y *val*. Además, es el único modelo de la Tabla 5.12 capaz de cumplir con las pruebas de validación interna y externa propuestas en la sección anterior.

Tabla 5.12. Los modelos QSAR basados en MLR más representativos para 263 larvicidas de *A. aegypti* L. El mejor modelo se resalta en negrita.

<i>d</i>	R_{cal}^2	S_{cal}	R_{val}^2	S_{val}	$R_{ij\max}^2$	Descriptorios moleculares
1	0.31	0.63	0.42	0.59	0	<i>D590</i>
2	0.42	0.58	0.64	0.46	0.01	<i>AATS5v, Sub282</i>
3	0.50	0.54	0.65	0.46	0.01	<i>AATS5i, KR1592, Sub282</i>
4	0.56	0.50	0.67	0.44	0.20	<i>PC34, PC777, Sub282, D589</i>
5	0.62	0.47	0.69	0.43	0.06	<i>VE2_Dze, PC34, KR1592, Sub282, D589</i>
6	0.65	0.45	0.70	0.42	0.47	<i>nRing, PC34, KR1592, KR3584, Sub282, D589</i>
7	0.68	0.43	0.72	0.41	0.19	<i>M16, PC34, PC199, KR1592, AP653, Sub282, D589</i>
8	0.71	0.43	0.69	0.46	0.25	<i>VE1_Dze, PC34, KR1592, AP653, Sub282, KRC4736, D178, D589</i>

Para demostrar que la Ec. 5.14 no es el resultado de una correlación casual, se empleó la técnica de Aleatorización-Y, obteniendo que $RMS_{cal} \ll RMS^{alet}$ (0.64); de esta manera, se propone una relación estructura-actividad larvicida válida.

$$\log_{10} LC_{50} = 5.6 + 0.8(\pm 0.3)M16 - 1.8(\pm 0.3)PC34 + 0.6(\pm 0.2)PC199 - 0.7(\pm 0.2)KR1592 + 1.1(\pm 0.3)AP653 - 1.2(\pm 0.2)Sub282 - 1.2(\pm 0.1)D589 \quad (5.14)$$

$$N_{cal} = 133, d = 7, R_{cal}^2 = 0.68, RMS_{cal} = 0.43, N_{cal} / d = 19, R_{ij\max}^2 = 0.19, VIF^{\max} = 1.1, \alpha(3S) = 0, R_{LOO}^2 = 0.64, RMS_{LOO} = 0.46, R_{L30\%O}^2 = 0.55, RMS_{L30\%O} = 0.5, RMS^{aleat} = 0.64$$

$$N_{val} = 65, R_{val}^2 = 0.72, RMS_{val} = 0.41$$

$$N_{pred} = 65, R_{pred}^2 = 0.58, RMS_{pred} = 0.37$$

Los criterios de validación de Golbraikh *et al.*, 2003,⁵⁴ fueron contemplados para el modelo QSAR de la Ec. 5.14 para garantizar su poder predictivo a través de los siguientes valores:

$$k = 0.97 \text{ y } k' = 0.97$$

$$1 - R_0^2 / R_{pred}^2 = 0.02 \text{ y } 1 - R_0'^2 / R_{pred}^2 = 0.2$$

$$R_m^2 = 0.52$$

Se verificó el comportamiento predictivo de la Ec. 5.14 a través del gráfico exhibido en la Figura 5.23.A, que contiene las actividades larvicidas calculadas por el modelo como una función de los valores experimentales. La figura 5.23.B representa la dispersión de los valores de residuo, con una marcada tendencia a obedecer un patrón aleatorio alrededor de la línea cero del eje Y; lo que sugiere que la Ec. 5.14 predice

todo el conjunto de datos sin errores sistemáticos o un sesgo residual amplio. Los valores usados para generar los gráficos se proporcionan en la Tabla 5.6.3 del apartado anexo.

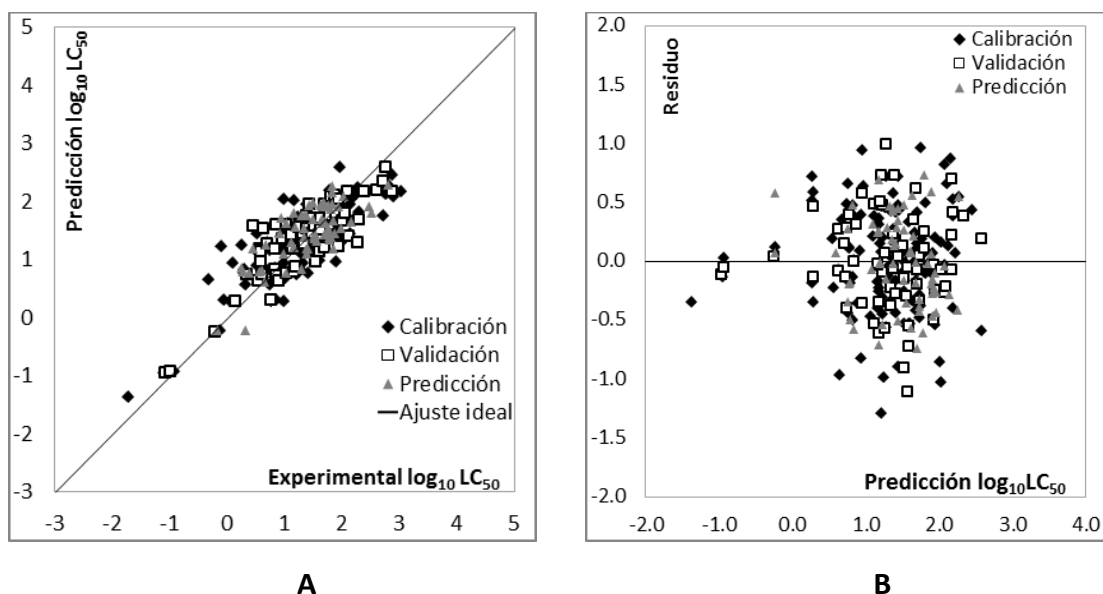


Figura 5.23. A. Valores de predicción del $\log_{10} LC_{50}$ en función de la información experimental. B. Dispersión de los valores de residuo.

También se analizó el parámetro $R_{jj\max}^2$ de la Ec. 5.14, que es el coeficiente de correlación máximo entre pares de descriptores. Un $R_{jj\max}^2 = 0.19$ indica que no hay problemas serios de superposición de información estructural.

Del mismo modo, la matriz de correlación del modelo proporcionada en la Tabla 5.13 muestra que los siete descriptores incluidos en el modelo QSAR tienen intercorrelaciones muy bajas; y el valor VIF^{\max} de todos los descriptores es mucho menor ($VIF^{\max} = 1.1$) que 10, lo que indica que no hay problemas de multicolinealidad entre los descriptores seleccionados por la técnica de RM.

Es de amplio conocimiento que un modelo QSAR es exitoso cuando supera el proceso de validación externa. Por esa razón, verificamos la capacidad del modelo para predecir la actividad larvicida de compuestos derivados de las plantas que no se consideraron durante la calibración del mismo.

Tabla 5.13. Matriz de correlación de los siete descriptores incluidos en la Ec. 5.14.

Ec. 5.14	<i>M16</i>	<i>PC34</i>	<i>PC199</i>	<i>KR1592</i>	<i>AP653</i>	<i>Sub282</i>	<i>D589</i>	<i>VIF</i>
<i>M16</i>	1	0.0004	0.0015	0.0008	0.0002	0.0763	0.0049	1.0
<i>PC34</i>		1	0.0023	0.0013	0.0004	0.0005	0.0062	1.0
<i>PC199</i>			1	0.0055	0.0015	0.0419	0.1852	1.1
<i>KR1592</i>				1	0.0008	0.0011	0.0072	1.0
<i>AP653</i>					1	0.0003	0.0157	1.0
<i>Sub282</i>						1	0.0148	1.0
<i>D589</i>							1	1.1

La Ec. 5.14 presenta un poder predictivo aceptable para el subconjunto de predicción conformado por 65 compuestos con información experimental no contemplada en la construcción del modelo. A partir de los valores de los parámetros $R^2_{pred} = 0.58$ y $RMS_{pred} = 0.37$, se corroboró que este modelo QSAR puede ser útil para predecir nuevos compuestos larvicidas carentes de información experimental sobre la toxicidad aguda media (LC_{50}) en vectores de enfermedades.

Ahora, es posible dar una breve descripción de cada descriptor no conformacional incluido en la Ec. 5.14, la cual involucra siete variables independientes que pertenecen a dos clases de descriptores moleculares: seis descriptores indicadores de fragmento y 1 descriptor topológico Mold², detallados a continuación:

- Un descriptor indicador de fragmento *MACCS*: *M16* se refiere a estructuras que tienen cicloalcanos de 5 o 6 miembros, unidos a un sustituyente hidroxilo (-OH), acetilo (-Ac) o acetoxi (-OAc).⁸³
- Dos descriptores indicadores de fragmento *PubChem*: *PC34* indica la presencia o el recuento de átomos químicos individuales representados por el símbolo atómico 2S; *PC199* denota cualquier anillo que no comparte tres átomos consecutivos con ningún otro anillo en la estructura química (ESSSR). En este caso *PC199* define la presencia de cualquier anillo de 6 miembros, que tiene un anillo ESSSR.⁸⁴
- Un descriptor indicador de fragmento *Klekota-Roth*: *KR1592* indica la presencia de una subestructura SMART que representa 3,3'-Dimetil-1,1'-bifenilo.
- Un descriptor indicador de fragmento de pares de átomos 2D: *AP653* denota la presencia de pares de átomos (O-Br) a una distancia topológica 9.
- Un descriptor indicador de subestructura: *Sub282* que simboliza la presencia de una subestructura química con un anillo de 5 o 6 miembros que contiene un átomo de oxígeno y un enlace acetal en la posición 2.
- Un descriptor topológico de Mold²: *D589* codifica el auto valor más alto de la matriz de simetría de Burden ponderado por la polarizabilidad de orden 2.

Una ventaja de los estudios QSAR radica en lograr relacionar de forma coherente las características moleculares de un conjunto de datos con la propiedad evaluada, para guiar la síntesis de nuevos compuestos inspirados en las plantas más activos. En

este contexto, se analizó grado de contribución de cada descriptor (b_j^s) de la Ec. 5.14, exhibiendo el siguiente orden de importancia:

$D589 (0.65) > PC34 (0.34) > Sub282 (0.32) > PC199 (0.24) > KR1592(0.20) > AP653 (0.17) > M16 (0.12)$

Los valores numéricos de estos descriptores se proporcionan en la Tabla 5.6.4 del material anexo. A partir del coeficiente de regresión de cada descriptor involucrado en la Ec. 5.14, fue posible plantear una posible interpretación de la relación matemática existente entre los aspectos estructurales codificados por los siete descriptores y la actividad larvicida sobre el *A. aegypti* L.

Por lo tanto, se propone una serie de condiciones útiles para guiar la síntesis química de nuevos larvicidas derivados de las plantas en la Figura 5.24.

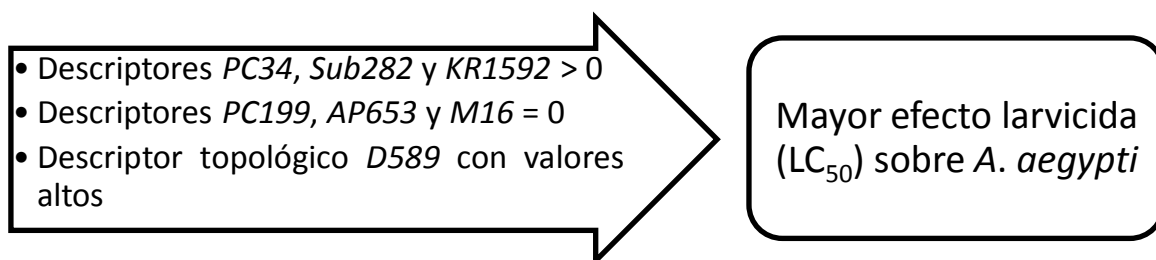


Figura 5.24. Valores adecuados de los descriptores de la Ec. 5.14 para obtener larvicidas efectivos.

Dichas características estructurales se encuentran en las moléculas **27-30** de calibración, **15, 26 y 31** de validación y **24** de predicción, las cuales exhiben valores de toxicidad aguda media ($\log_{10} LC_{50} = 4 \mu g mL^{-1}$) sobresalientes contra las larvas *A. aegypti* L.

Por otra parte, el análisis del dominio de aplicación del modelo basado en la técnica de influencia generó el diagrama de Williams ilustrado en la Figura 5.25. Dentro del enfoque de influencia, una molécula del subconjunto de *cal* o *val* con valor alto h_i reforzaría el modelo, pero si dicha molécula pertenece al subconjunto *pred* su actividad larvicida será predicha de forma poco confiable, como resultado una extrapolación sustancial del modelo.⁸⁵

Los valores de influencia (h) obtenidos para cada compuesto químico de los subconjuntos *cal*, *val* y *pred* se encuentran disponibles en la Tabla 5.6.2 del material anexo. Estos datos revelan que las moléculas **115 y 246** del subconjunto *pred* superan el límite de advertencia del modelo ($h^* = 0.18$). Cabe resaltar que dicho comportamiento puede atribuirse a la alta heterogeneidad del conjunto de datos modelado. No obstante, la Figura 5.25 exhibe que la mayoría de las moléculas de

predicción tienen valores $h_i < h^*$; por lo tanto los valores de $\log_{10} LC_{50}$ estimados para este subconjunto pueden considerarse confiables.

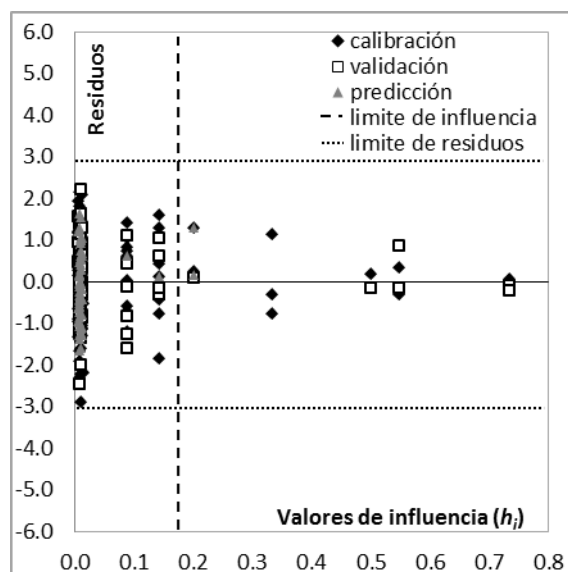


Figura 5.25. Valores de influencia (h_i) para la Ec. 5.14, el límite de influencia ($h^* = 0.18$) y residuo estandarizado $\sigma(3S)$.

El modelo de regresión expresado en la Ec. 5.14 ha demostrado ser capaz de cuantificar con éxito la actividad larvica de 263 moléculas derivadas de plantas. Por ello, se verificó la capacidad del modelo para clasificar las sustancias larvicidas en dos principales grupos activo e inactivo.

Para verificar la capacidad de clasificación del modelo se consideraron los datos de predicción para definir el umbral de actividad larvica, donde las moléculas con $LC_{50} \leq 40 \mu g mL^{-1}$ son consideradas activas y las moléculas con $LC_{50} > 40 \mu g mL^{-1}$ se toman como poco activas o inactivas. Debido a que la OMS no ha decretado un nivel de concentración que defina la efectividad larvica de las sustancias químicas. En este estudio, se utilizó el nivel de clasificación propuesto por Cheng *et al.*, 2013.⁸⁶

Luego, a través de las estadísticas de Cooper,⁸⁷ se calcularon tres parámetros estadísticos principales: precisión ($A\%$), referida al grado de concordancia de la clasificación; sensibilidad (SE), mide la cantidad de compuestos tóxico clasificados de forma correcta; y especificidad (SP), verifica la cantidad de sustancias no tóxicas clasificadas correctamente. Las ecuaciones consideradas para calcular estos parámetros son las siguientes:

$$A\% = \frac{VP + VN}{TOTAL} \quad (5.15)$$

$$SE = \frac{VP}{P} \quad (5.16)$$

$$SP = \frac{VN}{N} \quad (5.17)$$

Donde el parámetro VP indica el número de compuestos tóxicos estimados de forma correcta; VN es el número de compuestos no tóxicos estimados correctamente; P es el número total de compuestos identificados como tóxicos; N es el número total de compuestos identificados como no tóxicos.

Los resultados obtenidos de este análisis demuestran que la Ec. 5.14 tiene aceptable precisión, sensibilidad y especificidad para clasificar la actividad larvicida del subconjunto $pred$. Además, para corroborar la habilidad del modelo para clasificar las 65 moléculas de predicción, se dividieron en cuatro categorías según los datos de predicción de la LC_{50} .

De este modo, se evidenció que 27 moléculas tóxicas se predicen como tóxicas (VP) y 9 como no tóxicas (FN); además, 20 moléculas no tóxicas se predicen como no tóxicas (VN) y 9 como tóxicas (FP). Por lo tanto, el 73% de las moléculas de predicción se predice correctamente.

En definitiva, los productos de origen natural involucran una gran variedad de estructuras moleculares con diversas propiedades biológicas. Por ello, se buscó ampliar el espacio químico de la actividad larvicida, recolectando de la literatura un conjunto de datos con 237 moléculas derivadas de las plantas con actividad larvicida indeterminada sobre el *A. aegypti*.

Se aplicó el modelo QSAR expresado en la Ec. 5.14, basado en el protocolo de la OMS para predecir la actividad larvicida del conjunto molecular desconocido, denominado N_{desc} .

El primer paso para predecir la toxicidad aguda media (LC_{50}) consiste en verificar el AD del conjunto N_{desc} . Las 237 estructuras moleculares, los valores estimados de la actividad larvicida ($\log_{10} LC_{50}$) y valores de influencia obtenidos para las moléculas se encuentran en la Tabla 5.6.5 del material anexo.

El análisis del AD exhibió que solo 27 moléculas del conjunto N_{desc} tienen valores h_i superiores al límite de advertencia del modelo ($h^* = 0.18$); de modo que, las predicciones estas moléculas se consideran inconsistentes y poco confiables. No obstante, las 210 moléculas restantes se localizan dentro del AD, un rango de influencia inferior al límite h^* ; demostrando la sobresaliente capacidad de la Ec. 5.14 para cuantificar la actividad larvicida sobre el *A. aegypti* L. de este novedoso conjunto molecular nunca antes estudiado a través de la teoría QSAR.

5.6.6. Conclusiones

La presente investigación dio origen al primer modelo lineal QSAR basado en un conjunto de datos conformado por 263 larvicidas sobre el mosquito *A. aegypti*, altamente heterogéneo e inspirado en fitoquímicos. Con este modelo se logró predecir acertadamente el 78% del conjunto de molecular total en rango de actividad larvicida de $0.02 - 790 \mu\text{g mL}^{-1}$. Además, la relación estructura-actividad obtenida superó diferentes condiciones matemáticas necesarias para su aceptación, tal como las prueba de CV-LOO y LMO, Aleatorización-Y y el parámetro *VIF*.

Para estructuras químicas que caen dentro del AD de este modelo, se proporcionó una guía QSAR para la síntesis o identificación de nuevos larvicidas derivados de plantas. En este contexto, el modelo QSAR independiente de la conformación propuesto logró que el 89% de las moléculas del conjunto desconocido se localizara dentro del AD, suministrando predicciones acertada de la actividad larvicida contenida en las mismas.

Por lo tanto, la simplicidad del modelo lineal QSAR, la fácil interpretación y disponibilidad de los descriptores moleculares involucrados, así como el adecuado poder predictivo han hecho que este enfoque sea atractivo para guiar el diseño de nuevas moléculas potencialmente activas basadas en materias primas renovables.

5.7 Referencias

1. Li, L., Hu, J, Ho Y-S. Global performance and trend of QSAR/QSPR research: a bibliometric analysis. *Molecular Informatics*. **2014**, 33, p. 655–668.
2. Roy, K., Kar, S., Nas, R.N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier: London, **2015**.
3. Benfenati, E., Roncaglioni, A., Petoumenaou, M., Cappelli, C., Gini, G. Integrating QSAR and read across for environmental assessment. *SAR & QSAR in Environmental Research*. **2015**, 26, p. 605–618.
4. Demchuk, E., Ruiz, P., Chou, S., Fowler, B.A. SAR/QSAR methods in public health practice. *Toxicology and Applied Pharmacology*. **2011**, 254, p. 192-197.
5. Benfenati, E., Diana, R.G., Gini, G., Cardamone, L., Gocieva, M., Mancusi, M., Padovaní, R., Tamellini, L. *Introduction on in vivo, in vitro and in silico methods*. En Benfenati, E., Ed. The e-book on QSAR and REACH: theory, guidance and applications. Orchestra: Milan, **2012**, p. 5-37. Disponible en: http://ebook.insilico.eu/insilico-ebook-orchestrabenfenati-ed1_rev-June2013.pdf.
6. Illing, P. *General Overview of the Safety Evaluation of Chemicals*. In Hester, R.E., Harrison, M.R., Eds. *Alternatives to Animal Testing*. Environmental Science and Technology, vol 23. The Royal Society of Chemistry: Cambridge, **2006**, p.3-24.
7. Gini, G. Chapter 3 *QSAR: What else?*. En: Nicolotti, O., Ed. *Computational Toxicology: Methods and Protocols*, Methods in Molecular Biology, vol. 1800, Humana Press: New York, **2018**, p. 79–105.
8. Fjodorova, N., Novich, M., Vrachko, N. *et al.* Directions in QSAR modeling for regulatory uses in OECD member countries, EU and in Russia. *Journal of Environment Science and Health Part C*. **2008**, 26, p. 201–236.
9. Sanderson, H., Johnson, D., Reitsma, T. *et al.* Ranking and prioritization of environmental risks of pharmaceuticals in surface waters. *Regulatory Toxicology and Pharmacology*. **2004**, 39, p. 158–183.
10. Sanderson, H., Thomsen, M. Ecotoxicological quantitative structure–activity relationships for pharmaceuticals. *Bulletin of Environmental Contamination and Toxicology*. **2007**, 79, p. 331–335.
11. Kar, S., Roy, K. First report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. *Chemosphere*. **2010**, 81, p. 738–747.
12. Kar S, Roy K (**2012**) Risk assessment for ecotoxicity of pharmaceuticals - an emerging issue. *Expert Opin Drug Saf* 11:235–274
13. Kar, S., Roy, K., Leszczynski, J. *Impact of pharmaceutical on the environment: Risk assessment using QSAR modelling approach*. En: Nicolotti, O., Ed. *Computational Toxicology: Methods and Protocols*, Methods in Molecular Biology, vol. 1800, Humana Press: New York, **2018**, p. 395–473.

14. Organización Mundial de la Salud (OMS). Demencia, datos y cifras: centro de prensa. WHO press: Ginebra. **2019**. [Citado el 12 de Noviembre de 2019]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/dementia>
15. Alzheimer's Association. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. **2011**, 7, p. 208–44.
16. Alzheimer's association Argentina. Recursos para la enfermedad de Alzheimer y la demencia. [Citado el 12 de Noviembre de **2019**]. Disponible en: <http://www.alz.org/ar/demencia-alzheimer-argentina.asp>
17. Rosenberry, T.L. *Acetylcholinesterase*. Meister, A. Ed. Advances in Enzymology and Related Areas of Molecular Biology. Vol 43. John Wiley & Sons, Inc: X. **1975**.
18. Ballard, C.G. Advances in the treatment of Alzheimer's disease: Benefits of dual cholinesterase inhibition, European Neurology. **2002**, 47, p. 64–70.
19. Yu, I., Cao, R., Yi, W., Yan, Q., Chen, Z., Ma, L., Peng, W., Song, H. Synthesis of 4-[(diethylamino)methyl]-phenol derivatives as novel cholinesterase inhibitors with selectivity towards butyrylcholinesterase. Bioorganic & Medicinal Chemistry Letters. **2010**, 20, 3254.–3258.
20. Hossain, T., Saha, A., Mukherjee, A. Exploring molecular structural requirement for AChE inhibition through multi-chemometric and dynamics simulation analyses. Journal of Biomolecular Structure and Dynamics. **2017**, 36, p. 1274-1285.
21. Racchi, M., Mazzucchelli, M., Porrello, E., Lanni, C., Govoni, S. Acetylcholinesterase inhibitors: novel activities of old molecules. Pharmacology Research. **2004**, 50, p. 441–451.
22. Qizilbash, N., Birks, J., Lopez-Arrieta, J., Lewington, S., Szeto, S. Tacrine for Alzheimer's disease. Cochrane Database Syst. Rev. **2000**, 3, CD000202.
23. Carlier, P.R., Han, Y.F., Chow, E.S.H., Li, C.P.L., Wang, H., Lieu, T.X., Wong, H.S., Pang, Y.P: Evaluation of short-tether Bis-THA AChE inhibitors. A further test of the dual binding site hypothesis. Bioorganic & Medicinal Chemistry, **1999**, 7, p. 351–357.
24. Pang, Y.P., Quiram, P., Jelacic, T., Hong, F., Brimijoin, S. Highly potent, selective, and low cost bis-tetrahydroacminacrine Inhibitors of Acetylcholinesterase. Steps toward novel drugs for treating Alzheimer's disease. Journal of Biological Chemistry. **1996**, 271, p. 23646–23649.
25. Hansch, C & Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. Journal American Chemical Society. **1964**, 86 (8), 1616–1626.
26. Fernández, M., Carreiras, M.C., Marco, J.L., Caballero, J. Modeling of acetylcholinesterase inhibition by tacrine analogues using Bayesian-regularized Genetic Neural Networks and ensemble averaging. Journal of Enzyme Inhibitor and Medicinal Chemistry. **2006**, 21, p. 647–661.
27. Ellman, G.L., Courtney, K.D., Andres, V. Jr., Featherstone, R.M. A new and rapid colorimetric determination of acetylcholinesterase activity. Biochemical Pharmacology. **1961**, 7, p. 88–95.

28. Rappaport, F., Fischl, J., Pinto, N. An improved method for the estimation of cholinesterase activity in serum. *Clinical Chimica Acta*. **1959**, 4, p. 227–230.
29. Marco, J.L., de los Ríos, C., Carreiras, M.C., Baños, J.E., Badía, A., Vivas, N.M. Synthesis and acetylcholinesterase/butyrylcholinesterase inhibition activity of new tacrine-like analogues. *Bioorganic & Medicinal Chemistry*. **2001**, 9, p. 727–732.
30. Camps, P., Formosa, X., Galdeano, C., Muñoz-Torrero, D., Ramírez, L., Gómez, E., Isambert, N., Lavilla, R., Badia, A., Clos, M.V., Bartolini, M., Mancini, F., Andrisano, V., Arce, M.P., Rodríguez-Franco, M.I., Huertas, Ó., Dafni, T., Luque, F.J. Pyrano[3,2-c]quinoline-6-chlorotacrine hybrids as a novel family of acetylcholinesterase- and beta-amyloid-directed anti-Alzheimer compounds. *Journal of Medicinal Chemistry*. **2009**, 52, p. 5365–5379.
31. León, R., de los Ríos, C., Marco-Contelles, J., Huertas, O., Barril, X., Luque, F.J., López, M.G., García, A.G., Villarroya, M. New tacrine-dihydropyridine hybrids that inhibit acetylcholinesterase, calcium entry, and exhibit neuroprotection properties. *Bioorg Med Chem* **2008**, 16:7759–7769.
32. Alonso, D., Dorronsoro, I., Rubio, L., Muñoz, P., García-Palomero, E., Del Monte, M., Bidon-Chanal, A., Orozco, M., Luque, F.J., Castro, A., Medina, M., Martínez, A. Donepezil-tacrine hybrid related derivatives as new dual binding site inhibitors of AChE. *Bioorganic & Medicinal Chemistry*. **2005**, 13, p. 6588–6597.
33. Rizzo, S., Bisi, A., Bartolini, M., Mancini, F., Belluti, F., Gobbi, S., Andrisano, V., Rampa, A. Multi-target strategy to address Alzheimer's disease: Design, synthesis and biological evaluation of new tacrine-based dimers. *European Journal of Medicinal Chemistry*. **2011**, 46, p. 4336–4343.
34. Samadi, A., Valderas, C., de los Ríos, C., Bastida, A., Chioua, M., González-Lafuente, L., Colmena, I., Gandía, L., Romero, A., del Barrio, L., Martín-de-Saavedra, M.D., López, M.G., Villarroya, M., Marco-Contelles, J. Cholinergic and neuroprotective drugs for the treatment of Alzheimer and neuronal vascular diseases. II. Synthesis, biological assessment, and molecular modelling of new tacrine analogues from highly substituted 2-aminopyridine-3-carbonitriles. *Bioorganic & Medicinal Chemistry*. **2011**, 19, p. 122–133.
35. Camps, P., Formosa, X., Galdeano, C., Gómez, T., Muñoz-Torrero, D., Ramírez, L., Viayna, E., Gómez, E., Isambert, N., Lavilla, R., Badia, A., Clos, M.V., Bartolini, M., Mancini, F., Andrisano, V., Bidon-Chanal, A., Huertas, Ó., Dafni, T., Luque, F.J. Tacrine-based dual binding site acetylcholinesterase inhibitors as potential disease-modifying anti-Alzheimer drug candidates. *Chemico-biological Interactions*. **2010**, 187, 411–415.
36. Hyperchem, 6.03 (Hypercube). <http://www.hyper.com>.
37. Talete srl, d. Dragon (Software for Molecular Descriptor Calculation) Version 6.0–**2014**. Disponible en: <http://www.talete.mi.it>.
38. aKatritzky Group: www.ark.chem.ufl.edu., bJurs Research Group, Disponible en: <http://research.chem.psu.edu/pcjgroup/>.

39. Duchowicz, P.R., Castro, E.A. & Fernández, F.M. Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Communications in Mathematical and in Computer Chemistry*. **2006**, 55 (1), p. 179–192.
40. Matlab 7.0. Masachussetts, USA: The MathWorks, Inc., Disponible en: <http://www.mathworks.com>
41. Hawkins, D.M., Basak, S.C., Mills, D. Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*. **2003**, 43, p. 579–586.
42. Rücker, C., Rücker, G., Meringer, M. Y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information Modeling*. **2007**, 47, p. 2345-2357.
43. Clifford, M.N. Chlorogenic acids and other cinnamates – nature, occurrence, dietary burden, absorption and metabolism. *Journal of the Science of Food and Agriculture*. **2000**, 80, p. 1033–1043.
44. Shu, L., Hongjun, Y. An overview on synthetic methods of alkyl cinnamates. *European Chemical Bulletin*. **2003**, 2, p. 76–77.
45. Gobec, S., Sova, M., Kristan, K., Rižner, T.L., Cinnamic acid esters as potent inhibitors of fungal 17 β -hydroxysteroid dehydrogenase-a model enzyme of the short-chain dehydrogenase/reductase superfamily. *Bioorganic & Medicinal Chemistry Letters*. **2004**, 14, p. 3933–3936.
46. Kristan, K., Starčevic, S., Brunskole, M., Rižner, T.L. Gobec, S., Cinnamates and cinnamamides inhibit fungal 17 β -hydroxysteroid dehydrogenase. *Molecular and Cellular Endocrinology*. **2006**, 248, p. 239–241.
47. Zhao, Y., Wang, J., Balleve, O., Luo, H., Zhang, W. Antihypertensive effects and mechanisms of chlorogenic acids. *Hypertension Research*. **2011**, 35, p. 370–374.
48. Shankaran, K., Sloan, C.P., Sniekcus, V. Synthetic Connections to the aromatic directed metalation reaction. Radical-induced cyclization to substituted benzofurans, benzopyrans, and furopyridines. *Tetrahedron Letters*. **1985**, 26, p. 6001–6004.
49. Lee, E.J., Kim, J.R., Choi, D.R., Ahn, Y.J. Toxicity of cassia and cinnamon oil compounds and cinnamaldehyde-related compounds to *Sitophilus oryzae* (coleoptera: curculionidae). *Journal of Economic Entomology*. **2008**, 101 (6), p. 1960–1966.
50. Peterson, C.J., Tsao, R., Eggler, A.L., Coats, J.R. Insecticidal Activity of Cyanohydrin and Monoterpenoid Compounds. *Molecules*. **2000**, 5, p. 648–654.
51. Abdelgaleil, S.A.M., Abbassy, M.A., Belal, A.S.H., Abdel Rasoul, M.A.A. Bioactivity of two major constituents isolated from the essential oil of *Artemisia judaica* L. *Bioresource Technology*. **2008**, 99, p. 5947–5950.
52. Jarvis, W.R. *Managing Diseases in Greenhouse Crops*. APS Press: New York, **1992**.
53. Tawata, A., Taira, S., Kobamoto, N., Zhu, J., Ishihara, M., Toyama, S. Synthesis and antifungal activity of cinnamic acid esters. *Bioscience, Biotechnology, and Biochemistry*. **1996**, 60, p. 909–910.

54. Golbraikh, A., Tropsha, A. Beware of q^2 !. *Journal of Molecular Graphics and Modelling*. **2002**, 22, p. 269–276.
55. Gramatica, P., Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: a Clarification Concerning Metrics and Terminology. *Journal of Chemical Information and Modeling*. **2016**, 56 (6). p. 1127–1131.
56. World Health Organization (WHO). Vector resistance to pesticides: fifteenth report of the WHO expert committee on vector biology and control. WHO Technical Report Series 818, Geneva, **1992**.
57. Lima, T.C., Santos, S.R., Uliana, M.P., Santos, R.C., Brocksom, T.J., Cavalcanti, S.C.H., de Sousa, D.P. Oxime derivatives with larvicidal activity against *Aedes aegypti* L. *Parasitology Research*. **2015**, 114, p.2883–2891.
58. Kim, M.G., Jeon, J.H., Lee, H.S. Larvicidal activity of the active constituent isolated from *tabebuia avellanadae* bark and structurally related derivatives against three mosquito species. *Journal of Agriculture and Food Chemistry*. **2013**, 61, p. 10741–10745.
59. Yu, K.X., Wong, C.L., Ahmad, R., Jantan, I. Larvicidal activity, inhibition effect on development, histopathological alteration and morphological aberration induced by seaweed extracts in *Aedes aegypti* (Diptera: Culicidae). *Asian Pacific Journal of Tropical Medicine*. **2015**, 8, p. 1006–1012.
60. Hansch, C., Leo, A. Exploring QSAR, *Fundamentals and Applications in Chemistry and Biology*. American Chemical Society (ACS): Washington DC. **1995**, p. 139-205.
61. Santos, S.R.L., Silva, V.S., Melo, M.A., Barbosa, J.D.F., Santos, R.L.C., Sousa, D.P., Cavalcanti, S.C.H. Toxic Effects on and Structure-Toxicity Relationships of Phenylpropanoids, Terpenes and Related Compounds in *Aedes aegypti* Larvae. *Vector-Borne Zoonotic Diseases*. **2010**, 10, p. 1049–1054.
62. Santos, S.R.L., Melo, M.A., Cardoso, A.V., Santos, R.L.C., Sousa, D.P., Cavalcanti, S.C.H. Structure–activity relationships of larvicidal monoterpenes and derivatives against *Aedes aegypti* Linn. *Chemosphere*. **2011**, 84, p. 150–153.
63. Scotti, L., Scotti, M.T., Silva, V.B., Santos, S.R.L., Cavalcanti, S.C.H., Mendonça, F.J.B. Jr., Chemometric Studies on Potential Larvicidal Compounds Against *Aedes Aegypti*. *Med Chem*. **2014**, 10, p. 201–210.
64. Barbosa, J.D. F., Silva, V.B., Alves, P.B., Gumina, G., Santos, R.L.C., Sousa, D.P., Cavalcanti, S.C.H. Structure–activity relationships of eugenol derivatives against *Aedes aegypti* (Diptera: Culicidae) larvae. *Pest Management Science*. **2012**, 68, p. 1478–1483.
65. Duchowicz, P.R., Fioressi, S.E., Castro, E.A, Wróbel, K., Ibezim, N.E., Bacelo, D.E. Conformation-independent QSAR study on human epidermal growth factor receptor-2 (HER2) inhibitors. *Chemistry Select*. **2017**, 2, p. 3725–3731.
66. Duchowicz, P.R. Linear Regression QSAR Models for Polo-Like Kinase-1 Inhibitors. *Cells*. **2018**, 7 (2), p. 1–11.
67. Aranda J.F., Bacelo, D.E., Leguizamón Aparicio M.S., Ocsachoque, M.A., Castro, E.A. Duchowicz, P.R. Predicting the bioconcentration factor through a conformation-

- independent QSPR study. SAR & QSAR in Environmental Research. **2017**, 28, p. 749–763.
68. Duchowicz, P.R. QSPR studies on water solubility, octanol-water partition coefficient and vapour pressure of pesticides. SAR & QSAR in Environmental Research. **2019**, 31 (2), p. 135–148.
 69. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Computer Sciences. **1988**, 28 (1), p. 31–36.
 70. ACD/ChemSketch, Advanced Chemistry Development, Inc., Toronto, **2015**. Disponible en: www.acdlabs.com
 71. Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry. **2011**, 32, p. 1466–1474.
 72. PaDEL. www.yapcwsoft.com [Citado el 10 de Julio de 2018].
 73. Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., Su, Z., Perkins, R., Tong, W. Mold2, Molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. Journal of Computational Chemistry **2008**, 48, p. 1337–1344
 74. US EPA. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington DC. **2016**.
 75. Gramatica, P. Principles of QSAR models validation: Internal and external. QSAR in Combinatorial Science. **2007**, 26, p. 694–701.
 76. Stanton, D.T., Murray, W.J., Jurs, P.C. Comparison of QSAR and molecular similarity approaches for a structure-activity relationship study of DHFR inhibitors. Quantitative Structure -Activity Relationships. **1993**, 12, p. 239-245.
 77. Rafiei, H., Khanzadeh, M., Mozaffari, S., Bostanifar, M.H., Avval, Z.M., Aalizadeh, R., Pourbasheer, E. QSAR study of HCV NS5B polymerase inhibitors using the Genetic Algorithm-Multiple Linear Regression (GA-MLR). EXCLI Journal 2016, 15, p. 38–53.
 78. Valdes-Martini J.R., García Jacas, C.R., Marrero-Ponce, Y., Silveira Vaz d'Almeida, Y., Morrel, C. QuBiLS-MAS: free software for molecular descriptors calculator from quadratic, bilinear and linear maps based on graph-theoretic electronic-density matrices and atomic weightings; Version 1.0; CAMD-BIR Unit, CENDA Number of Register: 2373-2012. Central University of Las Villas, Villa Clara, Cuba. **2012**.
 79. Valdes-Martini JR, Marrero-Ponce Y, Garcia-Jacas CR, Martinez-Mayorga K, Barigye SJ, Silveira Vaz d'Almeida Y, Pham-The H, Perez-Gimenez F, Morell CA (2017) QuBiLS-MAS, open source multi-platform software for atom- and bond-based topological (2D) and chiral (2.5D) algebraic molecular descriptors computations. Aust J Chem 9:35–61.
 80. Geris, R., Ribeiro, P.R., Da Silva, M., Garcia, H.H., Garcia, I. Bioactive natural products as potential candidates to control *Aedes aegypti*, the vector of dengue: Atta-ur-Rahman (ed) Studies in Natural Products Chemistry, vol 37. Academic Press, Elsevier, London. **2012**.

81. Dias, C., Fernandes, D. Essential oils and their compounds as *Aedes aegypti* L. (Diptera: Culicidae) larvicides: review. *Parasitology Research*. **2014**, 113, p. 565–592.
82. Kishore, N., Mishra, M.M., Tiwari, V.K., Tripathi, V., Lall, N. Natural products as leads to potential mosquitocides. *Phytochemistry Reviews*. **2014**, 13, p. 587–627.

CONCLUSIONES GENERALES Y PERSPECTIVAS FUTURAS

En las estrategias de evaluación y determinación de la respuesta activa de las moléculas, se emplean diferentes pruebas *in vivo* e *in vitro* que proveen información acerca de los efectos nocivos o benignos que estas entidades químicas generan sobre un modelo animal o un cultivo celular. Este tipo de ensayos son considerados de alta complejidad, ya que al momento de evaluar grandes librerías moleculares, requieren mayor tiempo e insumos de alto costo. Además, la mayoría de los protocolos de ensayo *in vivo* son éticamente inaceptables ante la comunidad científica internacional por involucrar el sacrificio o maltrato animal.

Como alternativa a los ensayos *in vivo* e *in vitro* en la presente Tesis Doctoral se han estudiado, diseñado, validado y aplicado diferentes modelos lineales de la teoría QSAR enfocados en diversos conjuntos moleculares altamente heterogéneos conformados por sustancias químicas de origen natural, semisintético y de síntesis de amplio interés entre las estrategias públicas para controlar el mosquito *Aedes aegypti* L., principal transmisor de múltiples enfermedades endémicas como dengue, zika, fiebre amarilla y de chikungunya, que afectan la salud pública mundial.

Todos los modelos QSAR fueron diseñados y validados en coherencia con los 5 principios propuestos por la Organización para la Cooperación y el Desarrollo Económico (OCDE) para su pertinente aceptación ante la comunidad científica internacional.

Puntualmente, las herramientas QSAR incluidas en la presente Tesis fueron desarrolladas con el objetivo de proveer una guía computacional para el diseño rápido, racional y selectivo de nuevas entidades químicas seguras, específicas, con efectos favorables para la salud y el ambiente, con respuestas finales inéditas y provenientes de un recurso renovable. Especialmente, que puedan sustituir el uso indiscriminado de agrotóxicos comerciales o mejorar el poder farmacológico de los medicamentos actuales.

Con cada estudio QSAR propuesto en la presente Tesis se adquirieron conocimientos valiosos acerca de la funcionalidad, aplicabilidad, facilidad y certeza que tienen los modelos QSAR cuando son desarrollados a través de metodologías válidas,

que incorporen diferentes tipos de descriptores moleculares representativos, algoritmos inequívocos y definidos para seleccionar las mejores relaciones estructura-actividad, y estrategias de validación adecuadas, que garanticen la coherencia y certeza de las predicciones del modelo.

Los programas computacionales empleados para desarrollar cada modelo QSAR son de código abierto. Dichos programas sirven para digitalizar las estructuras moleculares y calcular diversos descriptores teóricos o semiempíricos, que codifican los aspectos constitucionales, topológicos, electrónicos y lipofílicos de las moléculas investigadas. En particular los programas ChemSketch, PaDEL, Mold², EPI Suite y QuBiLs-MAS, se encuentran disponibles en la web y son de acceso libre. Además utilizan un modo gráfico que facilita su manipulación.

Otro aspecto relevante de los modelos predictivos QSAR obtenidos se basa en considerar las características estructurales independientes de la conformación de diversos conjuntos moleculares para extraer información acerca de las respuestas biológicas o los efectos tóxicos finales generados por los mismos. Este enfoque representa un análisis alternativo de gran interés en el desarrollo de nuevas herramientas computacionales de predicción basadas en los aspectos estructurales invariantes de la molécula para anticipar el efecto final que produce sobre un organismo o entorno definido.

En la actualidad, existen muchos campos de aplicación en términos de salud pública, conservación de los ecosistemas, minimización de las pruebas en animales y obtención de productos naturales activos donde la teoría QSAR podría contribuir de forma favorecedora. Entre las investigaciones QSAR que continúan en desarrollo en el marco de la presente Tesis Doctoral se encuentran:

1. Obtener un modelo lineal QSAR para predecir la Concentración Letal media (LC₅₀) de un conjunto molecular heterogéneo de pesticidas comerciales con datos de toxicidad medidos a través del novedoso bioensayo del pez cebra *Danio rerio*, catalogado como una prueba de Toxicidad Embrionaria en Peces (FET) éticamente viable, sustentable y de bajo costo, que brinda información de la toxicidad *in vivo* e *in vitro* de los compuestos químicos.
2. Desarrollar modelos QSAR con capacidad de estimar la capacidad inhibitoria *in vitro* de la enzima acetilcolinesterasa (AChE) de un conjunto molecular proveniente de fuentes naturales, altamente requeridos en la producción de nuevos fitosanitarios más seguros, así como de fármacos más accesibles y con menores efectos secundarios para tratar distintos trastornos neurodegenerativos.

3. Analizar a través de la teoría QSAR diferentes conjuntos moleculares constituidos por mezclas complejas, como los aceites esenciales y extractos de plantas, reconocidos por contener una amplia variedad de propiedades biológicas, tal como actividad larvicida sobre el mosquito *Aedes aegypti* L. La principal ventajas de estas sustancias naturales es que presentan efectos sinérgicos que potencian el efecto tóxico sobre las larvas, y no requieren de procedimientos complejos de extracción y purificación.
4. La búsqueda, aplicación y difusión de nuevos programas disponibles de acceso libre para calcular nuevos descriptores moleculares, que puedan considerar más aspectos estructurales potencialmente asociados a las actividades de las moléculas.
5. La exploración y aplicación de nuevos criterios de validación interna y externa de los modelos de regresión QSAR, que promuevan la credibilidad, certeza y aplicabilidad de estas herramientas computacionales.

PUBLICACIONES Y TRABAJOS PRESENTADOS EN EVENTOS CIENTÍFICOS

La presente Tesis Doctoral dio lugar a las siguientes publicaciones:

1. K. Y. Wong, P. R. Duchowicz, A. G. Mercader, L. M. Saavedra, G. P. Romanelli, E. A. Castro. QSAR Analysis on Tacrine Inhibitors of Acetylcholinesterase. *Journal of Biomedical Science*. 2014, 21, p. 84-92.
(Detallado en la sección 5.2.)
2. L. M. Saavedra, D. Ruiz, G. P. Romanelli, P. R. Duchowicz. Quantitative Structure-Antifungal Activity Relationships for cinnamate derivatives. *Ecotoxicology and Environmental Safety*, 2015, 122, p. 521-527.
(Detallado en la sección 5.3)
3. L. M. Saavedra, G. P. Romanelli, C. E. Rozo, P. R. Duchowicz. The quantitative structure-insecticidal activity relationships from plant derived compounds against chikungunya and zika *Aedes aegypti* (Diptera:Culicidae) vector. *Science of the Total Environment*. 2018, 610-611, p. 937-943.
(Detallado en la sección 5.4)
4. L. M. Saavedra, G. P. Romanelli, P. R. Duchowicz. QSAR analysis of plant-derived compounds with larvicidal activity against zika *Aedes aegypti* (Diptera: Culicidae) vector using freely available descriptors. *Pest Management Science*. 2018, 74, p. 1608-1615.
(Detallado en la sección 5.5.)
5. L. M. Saavedra, G. P. Romanelli, P. R. Duchowicz. A non-conformational QSAR study for plant-derived larvicides against zika *Aedes aegypti* L. vector. *Environmental Science and Pollution Research*. 2020, 27, p. 6205-6214.
(Detallado en la sección 5.6.)

Los siguientes trabajos fueron realizados en colaboración:

6. P. R. Duchowicz, S. E. Fioressi, D. E. Bacelo, L. M. Saavedra, A. P. Toropova, A. A. Toropov. QSPR Studies on Refractive Indices of Structurally Heterogeneous Polymers. *Chemometrics and Intelligent Laboratory Systems*. 2014, 140, p. 86-91.
7. S. E. Fioressi, D. E. Bacelo, W. P. Cui, L. M. Saavedra, P. R. Duchowicz. QSPR study on refractive indices of solvents commonly used in polymer chemistry using flexible molecular descriptors. *SAR and QSAR in Environmental Research*. 2015, 26, p. 499-506.

8. L. Y. Vargas-Méndez, P. L. Sanabria-Florez, L. M. Saavedra-Reyes, D. R. Merchán-Arenas, V.V. Kouznetsov. Bioactivity of semisynthetic eugenol derivatives against *Spodoptera frugiperda* (Lepidoptera: Noctuidae) larvae infesting maize in Colombia. Saudi Journal of Biological Sciences. 2018, doi: 10.1016/j.sjbs.2018.09.010.
9. A. P. Toropova, P. R. Duchowicz, L. M. Saavedra, E. A. Castro, A. A. Toropov. The use of the index of ideality of correlation for improving the predictive potential of bioconcentration factor models. Molecular Informatics. 2020. <https://doi.org/10.1002/minf.201900070> (EN PRENSA)

Los siguientes trabajos fueron publicados en congresos:

1. Estudio QSPR en índices de refracción de polímeros mediante descriptores flexibles. P. R. Duchowicz, L. M. Saavedra, S. E. Fioressi, D. E. Bacelo, A. P. Toropova y A. A. Toropov. XXX Congreso Argentino de Química, Octubre 21 al 24 de 2014, Buenos Aires - Argentina.
2. Las Relaciones Cuantitativas Estructura-Propiedad en el estudio de Sustancia Químicas con Efectos Sobre el Medio Ambiente. L. M. Saavedra, G. P. Romanelli y P. R. Duchowicz. I Jornada de Tesis de la Facultad de la Facultad de Ciencias Exactas – UNLP, Octubre 28 al 30 de 2015, La Plata, Buenos Aires - Argentina.
3. Análisis QSAR de derivados de monoterpenos con actividad insecticida sobre el mosquito *Aedes aegypti* (Diptera: culicidae). L. M. Saavedra, C. E. Roza y P. R. Duchowicz. XX Simposio Nacional de Química Orgánica, 11 al 14 de Noviembre de 2015, Mar del Plata, Buenos Aires - Argentina.
4. Relaciones Cuantitativas Estructura-Actividad larvicida de compuestos natural en *Aedes aegypti* vector del virus Zika. L. M. Saavedra, G. P. Romanelli y P. R. Duchowicz. V Congreso Internacional Cambio Climático y Desarrollo Sostenible, Septiembre 14 al 16 de 2016, La Plata, Buenos Aires - Argentina.
5. Estudio de las Relaciones Cuantitativas Estructura-Actividad larvicida de compuestos naturales sobre *Aedes aegypti* L. vector del virus zika. L. M. Saavedra, G. P. Romanelli y P. R. Duchowicz. XXXI Congreso Argentino de Química (CAQ), Octubre 25 al 28 de 2016, Buenos Aires - Argentina.
6. Las Relaciones Cuantitativas Estructura-Actividad (QSAR) en el estudio de la toxicidad de pesticidas comerciales sobre embriones del pez cebrá (*Danio rerio*). L. M. Saavedra, G. P. Romanelli y P. R. Duchowicz. XXI Simposio Nacional de Química Orgánica (SINAQO), Noviembre 8 al 11 de 2017, Potrero de los Funes, San Luis - Argentina.

Las siguientes notas fueron publicadas como prensa escrita:

1. Científicos platenses buscan crear insecticidas contra el dengue y el zika. L.M. Saavedra, G. P. Romanelli, P. R. Duchowicz, Diario El Día, La Plata, Buenos Aires – Argentina. Febrero 2019.
2. Seleccionaron moléculas vegetales para hacer insecticidas contra el mosquito del dengue. L.M. Saavedra, G. P. Romanelli, P. R. Duchowicz, Agencia Nacional de Noticias, Argentina. Febrero 2019.
3. Desarrollan herramienta computacional contra el dengue y el Zika, L.M. Saavedra, G. P. Romanelli, P. R. Duchowicz, Agencia CyTA Noticias de Ciencia y Tecnología, Fundación Instituto Leloir, Febrero 2019.