



Aplicaciones y potencialidades de los métodos de diseño computacional en estudios ambientales y farmacocinéticos

Juan Alberto Castillo-Garit1* <https://orcid.org/0000-0003-0896-9484>

Humberto González-Díaz2 <https://orcid.org/0000-0002-9392-2797>

Yudith Cañizares-Carmenate3 <https://orcid.org/0000-0003-3972-1318>

Francisco Torrens4 <https://orcid.org/0000-0002-8534-208X>

Hai Pham-The5 <https://orcid.org/0000-0003-4531-7223>

Yoan Martínez-López6 <https://orcid.org/0000-0002-1950-567X>

Karel Diéguez-Santana7 <https://orcid.org/0000-0003-4064-0566>

¹ Unidad de Toxicología Experimental, Universidad de Ciencias Médicas de Villa Clara. Santa Clara, Villa Clara, Cuba

² Departamento de Química Orgánica II, Universidad del País Vasco (UPV/EHU). España

³ CAMD-BIR Unit, Facultad de Química-Farmacía, Universidad Central Marta Abreu de Las Villas. Santa Clara, Cuba

⁴ Instituto de Ciencia Molecular, Universidad de Valencia. Paterna, Valencia, España

⁵ Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hoan Kiem. Hanoi, Viet Nam

⁶ Departamento de Ciencias de la Computación, Facultad Ciencias de la Computación, Universidad de Camagüey. Camagüey, Cuba

⁷ Facultad de Ingeniería Ambiental, Universidad Estatal Amazónica. Puyo, Ecuador

*Autor para la correspondencia: jacgarit@yahoo.es o juancgarit@infomed.sld.cu

RESUMEN

Palabras clave

diseño computacional; ecotoxicología; estudios ambientales; farmacocinética; quiralidad

Introducción: La evaluación de los efectos adversos de los contaminantes sobre los ecosistemas y la predicción de propiedades farmacocinéticas de compuestos, son campos de investigación de gran interés. Objetivos: desarrollar modelos computacionales que permitan la predicción, lo más acertada posible, del potencial ecotoxicológico de sustancias químicas de interés; así como la predicción de propiedades farmacocinéticas. **Métodos:** Para el desarrollo de los modelos predictivos se recopilaron bases de datos de toxicidad, mecanismos de acción tóxica, permeabilidad a través de barreras biológicas y propiedades quirales. Se emplearon técnicas de estadística, técnicas de inteligencia artificial o de *machine learning* según el objetivo. Asimismo, se emplearon programas reconocidos internacionalmente y programas propios. **Resultados:** Como resultados más importantes se tuvieron nuevos modelos computacionales que permiten predecir mejor los efectos adversos de los compuestos sobre el medio ambiente y nuevos modelos para predecir el paso a través de membranas biológicas, así como la definición de nuevos descriptores quirales. Los resultados han sido publicados en revistas de alto impacto, presentados en eventos científicos y han obtenidos varios premios tanto nacionales como internacionales. **Conclusiones:** El trabajo permitió profundizar la colaboración con varias universidades nacionales y extranjeras. Los nuevos modelos desarrollados están a disposición de la comunidad científica internacional.



Potentialities and applications of computational design methods in environmental and pharmacokinetic studies

ABSTRACT

Keywords

computational design; ecotoxicology; environmental studies; pharmacokinetic; chirality

Introduction: Predictions of environmental adverse effects and pharmacokinetic properties of compounds are areas of great interest. Objectives: The aim of this work is to develop computational models to predict ecotoxicological risk and to predict pharmacokinetic properties. **Methods:** In order to develop the computational models, several databases of toxicity, mode of toxic action, permeability across human barriers and chiral properties were collected. According to the specific study, statistical techniques as well as artificial intelligence and/or machine learning techniques were employed. We used several international software as well as software developed by our groups. **Results:** The main results are the development of new QSAR models to predict environmental toxicity of chemicals, new models to predict the permeability of compounds as well as the computational definition of new chiral descriptors. The results have been published in internationally recognized journals, presented in conferences and have obtained several national and international awards. **Conclusions:** The work permits increasing the collaboration with several universities from Cuba and other countries. The new developed models are available for the international scientific community.

INTRODUCCIÓN

El método tradicional de búsqueda de nuevos principios activos basado en el sistema de 'prueba y error' a través de ensayos masivos de gran número de sustancias químicas, es cada vez más ineficiente, siendo necesario ensayar sobre 10 000 compuestos para encontrar el deseado. A estos datos se unen que el costo del desarrollo de un nuevo medicamento desde su obtención en el laboratorio hasta su uso en terapéutica es mayor de 2600 millones de USD ⁽¹⁾ y superior a los 12 años ⁽²⁾. Nótese que este tipo de ensayos experimentales tienen altos costos en términos de recursos materiales, humanos y de tiempo, asimismo el aspecto de tipo ético que conlleva la investigación con animales y su posterior sacrificio.

Las vías alternativas en que la industria farmacéutica ha depositado su confianza, y también de recursos, son la síntesis combinatoria y el diseño computacional de fármacos.⁽³⁻⁵⁾ De ahí surgieron las famosas siglas QSAR (del inglés *Quantitative Structure-Activity Relationships*), en la actualidad todas las compañías farmacéuticas y centros de investigación importantes hacen uso de algunas de estas técnicas para mejorar sus expectativas de obtener una nueva entidad molecular que pueda ser explotada comercialmente. En nuestro país se comenzaron a desarrollar estudios QSAR a partir de la década de 1990, pero no fue hasta los primeros años del nuevo siglo que se experimentó un desarrollo sustancial en esta temática.

El Centro de Bioactivos Químicos de la Universidad Central Marta Abreu de Las Villas (UCLV) desarrolló varios estudios computacionales con el fin de estimar una actividad o propiedad deseada con una capacidad predictiva que posibilitaran recomendar la síntesis selectiva y el ensayo experimental de los compuestos más prometedores, lo que permitió en el 2005 obtener un premio nacional de la Academia de Ciencias de Cuba (ACC)⁽⁶⁾. En ese compendio se definieron, implementaron computacionalmente y aplicaron nuevas familias de descriptores moleculares y macromoleculares: basados en cadenas de Markov y en la aplicación de conceptos de la matemática discreta y el álgebra lineal, así como nuevas aplicaciones de los momentos espectrales. Los nuevos descriptores moleculares fueron empleados en la modelación de propiedades físicas, químico-físicas, farmacológicas, toxicológicas, etc., así como en estudios de proteómica y bioinformática, demostrándose que los mismos eran útiles para el diseño molecular y permitían obtener modelos matemáticos más sencillos, interpretables y robustos que los descritos en la literatura hasta ese momento ⁽⁶⁾.

También se desarrollaron estudios QSAR en el Departamento de Farmacia, Facultad de Química y Farmacia (UCLV) orientados principalmente a la identificación de nuevos compuestos activos contra diversas enfermedades infecciosas de gran impacto social. Los resultados de estos trabajos recibieron un premio nacional ACC en el año 2006 ⁽⁷⁾. Los modelos

desarrollados en esos trabajos permitieron el descubrimiento (selección/identificación), mediante el tamizaje virtual, de nuevas familias de compuestos químicos con marcada acción frente bacterias, cáncer y enfermedades parasitarias tales como la malaria, trichomoniasis, tripanosomiasis, entre otras. También se definieron nuevas familias de descriptores moleculares útiles para el desarrollo de nuevos fármacos. Este sistema computacional propuesto representó un paso de avance para el descubrimiento de fármacos activos contra enfermedades microbianas de gran impacto social ⁽⁷⁾.

Posteriormente en el año 2011 se presentó otro trabajo a premio nacional ACC, también de modelación computacional, coordinado por la Universidad de Camagüey⁽⁸⁾. El mismo tuvo como objetivo desarrollar e implementar un enfoque *integral* novedoso de diseño racional de fármacos para optimizar el proceso de descubrimiento y posterior desarrollo de nuevas entidades moleculares activas eficaces y seguras, considerando el ciclo de diseño-síntesis-evaluación-patente y la transdisciplinariedad e internacionalización de la investigación. Entre las actividades estudiadas podemos destacar antitumoral, antituberculosis, neurodegenerativa, así como estudios de seguridad, enfocados a los perfiles de toxicidad-farmacocinética-farmacovigilancia de los nuevos compuestos identificados. Como resultado se diseñaron/descubrieron y se establecieron nuevos procedimientos de síntesis y caracterización espectroscópica de los nuevos compuestos; se corroboraron los perfiles de eficacia/seguridad por ensayos biológicos requeridos, y se patentaron nuevos bioactivos más promisorios ⁽⁸⁾.

Estos trabajos previos se enfocaron principalmente en la definición de nuevas herramientas (descriptores moleculares) o en la demostración de la utilidad del diseño computacional para obtener nuevos compuestos activos contra enfermedades específicas. Sin embargo, en ninguno de ellos se han explorado las enormes potencialidades del diseño computacional en para estudios ambientales (predicción del potencial ecotoxicológico, mecanismo de acción tóxica, etc.) y solamente en algunos casos han esbozado las posibilidades de aplicación en el campo de la farmacocinética.

Tomando en consideración lo antes expuesto, el presente trabajo tiene como objetivo general demostrar la aplicabilidad del diseño computacional en estudios ambientales (ecotoxicológicos) así como ampliar su aplicación a la predicción de propiedades farmacocinéticas. De forma particular, desarrollar nuevos modelos que permitan la predicción del potencial ecotoxicológico de sustancias químicas, así como de su mecanismo de acción tóxica; desarrollar modelos que permiten la predicción del paso de las sustancias a través de membranas biológicas (con énfasis en absorción intestinal y barrera hematoencefálica). Finalmente, también se incluye en esta

propuesta la extensión de índices moleculares (previamente definidos) para codificar la quiralidad, esta propiedad es muy importante (ver por ejemplo el caso de la talidomida) y debe ser tenida en cuenta en las modelaciones computacionales.

MÉTODOS

Bases de datos empleadas

Para el desarrollo de los estudios de modelación medioambiental se emplearon datos de toxicidad acuática de varias familias de compuestos químicos ⁽⁹⁻¹²⁾ (derivados bencénicos, derivados fenólicos, compuestos orgánicos diversos). Asimismo se emplearon bases de datos (BD) que contenían reportes del mecanismo de acción tóxica de los compuestos para realizar la predicción del mismo ^(13,14) y también datos de toxicidad ambiental de nanopartículas y nanomateriales ^(15,16). En los estudios de predicción de propiedades farmacocinéticas, utilizamos BD que contenían información acerca de: la permeabilidad de membrana celular ⁽¹⁷⁾, de permeabilidad en células de Caco-2 ⁽¹⁸⁻²⁰⁾ (modelo para la absorción intestinal) y datos de permeabilidad de compuestos a la barrera hematoencefálica ⁽²¹⁾. Para la definición de los descriptores quirales ⁽²²⁻²⁴⁾ y sus aplicaciones ^(25, 26) empleamos tres BD con compuestos que presentan diferentes tipos/cantidades de centros quirales en su estructura.

Programas empleados

Para el cálculo de los descriptores moleculares (DM) fueron empleados varios programas. Entre ellos uno el denominado TOMOCOMD-CARDD ⁽²⁷⁾, desarrollado íntegramente por el grupo del departamento de Farmacia (FQF, UCLV). Otro disponible comercialmente, el DRAGON ⁽²⁸⁾ uno de los más utilizado en este campo de investigaciones y del cual poseemos licencia a través de los colegas de la Universidad de Valencia que colaboran con nosotros. Además, empleamos el software libre CDK ⁽²⁹⁾ el cual es parte de una iniciativa de software libres y el programa MD-LOVIS ⁽³⁰⁾ desarrollado por uno de los autores. Todos y cada uno de estos programas permiten el cálculo de varias familias de DM que han sido implementados en ellos, permitiendo la descripción de cada compuesto de manera independiente. Para desarrollar los diferentes modelos computacionales empleamos los programas STATISTICA ⁽³¹⁾, WEKA ⁽³²⁾, MOBYDIGS ⁽³³⁾ y QSARINs ⁽³⁴⁾. Estos permitieron el desarrollo de técnicas desde las más sencillas hasta otras más complejas como son técnicas de inteligencia artificial o machine learning (ML) para obtener los diferentes modelos computacionales.

Métodos estadísticos y de análisis de datos

Para correlacionar los datos experimentales de las diferentes propiedades con los DM calculados, se utilizaron las

siguientes técnicas: análisis discriminante lineal (ADL), regresión lineal múltiple (RLM), árboles de clasificación (CT) ⁽³⁵⁾, *k-esimos* vecinos más cercanos (IBk) ⁽³⁶⁾, redes neuronales tipo perceptron-multicapa (MLP) ⁽³⁷⁾, máquinas de vectores soporte (SVM) ⁽³⁸⁾ y modelos basados en la teoría de las perturbaciones (PT). Para determinar los mejores modelos QSAR se emplearon los siguientes estadísticos: para el caso de los estudios de clasificación: coeficiente de correlación de Matthews (CCM), exactitud (Q), sensibilidad, especificidad, relación de falsos positivos (*fpr*) y área bajo la curva ROC (AURC) ⁽³⁹⁾. Mientras para los estudios de regresión empleamos los valores de: R^2 , s , F , q^2 , s_{cv} entre otros. A los mejores modelos se les realizaron los procedimientos de validación correspondientes.

RESULTADOS Y DISCUSIÓN

Modelación computacional de los efectos de compuestos químicos sobre el medio ambiente

En este epígrafe, aplicamos la modelación computacional a estudios medioambientales para predecir el potencial ecotoxicológico de sustancias químicas de interés, así como de su mecanismo de acción tóxica. Para lo cual se recopiló BD de reportes previos y se desarrollaron varios tipos de modelos para predecir efectos sobre el medio ambiente tales como: la toxicidad acuática⁽⁹⁻¹¹⁾, mecanismos de acción tóxica^(13, 14), fitotoxicidad⁽¹²⁾ y evaluación de riesgo de ambiental de nanopartículas⁽¹⁵⁾ y nanomateriales⁽¹⁶⁾. Cada una de estas temáticas será desarrollada con más detalles en los sub-epígrafes siguientes.

Predicción de la toxicidad acuática sobre *Tetrahymena pyriformis*

Entre las especies utilizadas en bioensayos ecotoxicológicos, los ciliados, ocupan un lugar destacado siendo el *Te-*

trahymena pyriformis el modelo más empleado. El ensayo *in vitro* de inhibición del crecimiento del 50 % de la población de *T. pyriformis* ⁽⁴⁰⁾ es uno de los más empleados para predecir la toxicidad de un compuesto y generalmente se expresa como IGC₅₀ (por sus siglas en inglés). Este ensayo fue estandarizado por comparación con el 'estándar dorado' de los test de toxicidad acuática, el test sobre *Pimephales promelas* ⁽⁴⁰⁾. Desde entonces, ha sido considerada una de las BD preferidas pues el ensayo es económico, rápido y fácil de realizar; los resultados son sensibles, realistas y permiten establecer relaciones cuantitativas concentración-respuesta ⁽⁴¹⁾. Por esto fue el biomarcador seleccionado para ser utilizado en la modelación computacional de la toxicidad acuática.

Se desarrollaron dos estudios de predicción de la toxicidad acuática de derivados bencénicos sobre *T. pyriformis* ^(9, 11). Se empleó una BD de 392 derivados del benceno de los cuales se conocía su valor de IGC₅₀ sobre *T. pyriformis*; esta BD es una de las más extensas empleadas para este tipo de compuestos y desde su publicación inicial ha sido empleada por varios investigadores. Utilizando la técnica de análisis de clúster dividimos la BD en serie de entrenamiento (SE) y serie de predicción (SP), con 313 y 79 compuestos respectivamente; como se muestra en la figura 1.

En el primer estudio ⁽⁹⁾ se desarrolló una extensa revisión sobre estudios QSAR desarrollados, sobre varios biomarcadores (*Pimephales promelas*, *Daphnia magna* y *Tetrahymena pyriformis*), para predecir la toxicidad acuática identificándose que faltaban estudios sobre *T. pyriformis* que cumplieran con los cinco principios de la OECD ⁽⁴²⁾. Por tanto, desarrollamos la predicción de la toxicidad acuática sobre *T. pyriformis* cumpliendo los cinco principios establecidos para que un modelo QSAR pueda ser empleado con fines regulatorios. Empleamos la BD, antes mencionada, de 392 derivados del benceno (*endpoint* definido). Empleando, como DM los índices

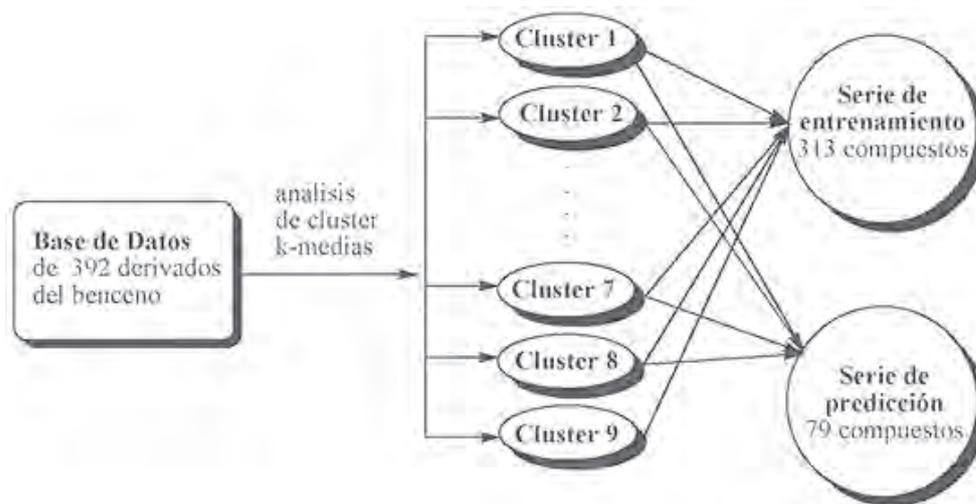


Fig. 1. Algoritmo utilizado para dividir la base de datos en SE y SP.

cuadráticos y la RLM como técnica estadística, (algoritmo no ambiguo). Además se definió el dominio de aplicación (DA) de los modelos (tercer principio) ⁽⁹⁾. Los modelos presentaron buenos parámetros de ajuste y validación (principio 4); el modelo no-estocástico ($R^2 = 0,807$ y $s = 0,334$) y el estocástico ($R^2 = 0,817$ y $s = 0,321$). Ambos fueron validados internamente empleando las técnicas *leave-one-out*, *bootstrapping* y *Y-scrambling* obteniéndose buenos resultados. Además, realizamos una validación empleando la SP, obteniendo valores de R^2_{pred} de 0,754 y 0,760, respectivamente. El principio 5 referido no es obligatorio por lo cual no fue abordado aquí. Los modelos desarrollados cumplieron los principios de la OECD convirtiéndose en excelentes herramientas para predecir la toxicidad de derivados bencénicos ⁽⁹⁾.

El segundo trabajo fue desarrollado para demostrar la utilidad, de una nueva familia de DM, definidos recientemente, para estudios de predicción medioambiental⁽¹¹⁾. Estos DM fueron nombrados como vectores atómicos ponderados (AWV:Atomic Weighted Vectors) y fueron implementados en el programa MD-LOVIS ⁽³⁰⁾. Empleamos la misma BD, descrita previamente, y se le calcularon los AWV empleando diferentes operadores de agregación. Se desarrollaron varios modelos QSAR para predecir la toxicidad los cuales presentaron adecuados parámetros estadísticos para la SE y SP. Además, los modelos fueron estables y robustos en la validación interna; la tabla 1 se muestra parte de la comparación que se hace en la tabla 5 de otro artículo ⁽¹¹⁾. Aquí solo se muestran aquellos modelos que emplearon exactamente la misma cantidad de compuestos (313) y presentaron un valor de $R^2 > 0,7$ en el modelo. Como puede verse los resultados de los modelos obtenidos con los AWV son superiores a los obtenidos previamente por otros autores. Los resultados sugieren que los descriptores AWV son una buena herramienta para codificar la información estructural de las moléculas y una alternativa

interesante para desarrollar modelos de predicción de la toxicidad acuática de derivados bencénicos sobre *T. pyriformis*.

También desarrollamos un estudio de predicción de toxicidad acuática de derivados fenólicos sobre *T. pyriformis* ⁽¹⁰⁾. Los derivados fenólicos existen comúnmente en el medio ambiente, siendo utilizados como componentes de tintes, polímeros, productos farmacéuticos y otras sustancias orgánicas. En este trabajo nos propusimos como objetivo, actualizar y ampliar la BD de derivados fenólicos contra *T. pyriformis* para desarrollar el modelo QSAR. Hasta donde sabemos es la BD más extensa recopilada hasta la fecha (358 derivados fenólicos) superando en más de 100 compuestos la reportada anteriormente. Se emplearon los DM del DRAGON ⁽²⁸⁾ y la RLM empleando STATISTICA⁽³¹⁾. El mejor modelo obtenido se muestra a continuación ⁽¹⁰⁾:

$$\begin{aligned} \text{Log}(1/\text{IGC}_{50}) = & -17,537 + 12,384*\text{MWC02} - 16,530*\text{MWC09} \\ & + 12,794*\text{MWC10} + 9,150*\text{piPC03} \\ & + 0,070*\text{piPC08} - 4,408*\text{TPC} + 24,910*\text{X3A} - 0,083*\text{nC-} \\ & \text{conj} + 0,743*\text{nR=Cs} \\ & - 0,487*\text{nRCN} - 0,624*\text{nCXr} - 0,505*\text{O-059} - 0,877*\text{LTD48} \end{aligned}$$

N = 240 $R^2 = 0,740$ $s = 0,439$ F = 45,91
p < 0,0001
 $Q^2_{boot} = 0,686$ SDEP = 0,455 $R^2_{pred} = 0,699$

Como puede verse los parámetros de ajuste del modelo son buenos para SE y SP; asimismo como resultado de los procesos de validación (interna y externa) podemos decir que el modelo es estable, robusto y que tiene un adecuado poder predictivo. El modelo fue comparado con otros previamente reportados en la literatura internacional, demostrándose que los parámetros estadísticos de nuestro modelo fueron superiores o comparables a los de los modelos previos⁽¹⁰⁾. Además, se brinda una explicación estructural de los DM del mo-

Tabla 1. Comparación entre los descriptores AWV y otros enfoques publicados previamente

Index	N	n	R ²	Q ² _{loo}	Q ² _{boot}	Q ² _{ext}	SDEC	F	S	Método estadístico
MD-AWV	313	6	0,837	0,830	0,827	0,789	0,305	262,63	0,309	RLM
MD-AWV	313	5	0,820	0,813	0,811	0,742	0,321	279,49	0,324	RLM
CoMSIA	313	7	0,844	0,514	-	-	0,295	-	-	PLS
CoMFA	313	7	0,816	0,611	-	-	0,319	-	-	PLS
VolSurf	313	7	0,806	0,758	-	-	0,322	-	-	PLS
ERM-VolSurf	313	7	0,797	0,783	-	-	0,330	-	-	RLM
Stochastic linear indices	313	6	0,733	0,704	-	-	0,411	139,94	0,394	RLM
Non-Stochastic linear indices	313	6	0,721	0,687	-	-	0,421	131,79	0,403	RLM
Topological descriptors	313	6	0,716	0,682	-	-	0,423	128,70	0,406	RLM

delo y su relación con la toxicidad de los compuestos. La idea principal de cualquier modelo es su uso posterior; para esto se descargó un conjunto de datos compuesto por 7932 compuestos del ChEMBL. Se eliminaron los compuestos que no estaban en el DA y se realizaron las predicciones para los 600 que cumplieran con este requisito. Se describieron las alertas estructurales más importantes y se ordenaron los compuestos en orden decreciente de toxicidad. Los resultados de este trabajo son de gran interés para la predicción del potencial ecotoxicológico de derivados bencénicos.

Predicción del mecanismo de acción tóxica

Aquí analizaremos en paralelo dos trabajos desarrollados para predecir el mecanismo de acción tóxica (MOA, siglas en inglés de *Mode of toxic Action*) de derivados fenólicos sobre *T. pyriformis*. En ambos empleamos la misma BD y también coinciden algunas técnicas empleadas para desarrollar los modelos. Los derivados fenólicos son contaminantes ambientales que ejercen su toxicidad mediante diferentes MOA⁽⁴³⁾. Muchos de los fenoles sintéticos son tóxicos y la mayoría de ellos están clasificados como contaminantes peligrosos. La BD seleccionada fue obtenida de la literatura⁽⁴⁴⁾ y consta de 221 derivados fenólicos con sus MOA. En el primer artículo⁽¹³⁾ empleamos DM del TOMOCOMD-CARDD⁽²⁷⁾ mientras en el segundo⁽¹⁴⁾ utilizamos los del programa CDK⁽²⁹⁾.

En el primer trabajo⁽¹³⁾ desarrollamos modelos con ADL y cuatro técnicas de ML (CT, IBK, MLP y SVM); mientras en el

segundo trabajo⁽¹⁴⁾ desarrollamos solamente las cuatro técnicas de ML. Los modelos desarrollados suponen un paso de avance en la modelación computacional del mecanismo de acción tóxica. Los resultados obtenidos con el ADL (93,21 %) están entre los tres mejores resultados alcanzados. Los modelos desarrollados con las técnicas de ML⁽¹³⁾ clasificaban entre los mejores resultados obtenidos hasta el momento, pero fueron superados posteriormente⁽¹⁴⁾ porque los nuevos modelos presentan un nivel de calidad superior. Todos los modelos obtenidos con técnicas de ML fueron validados con la SP de 22 derivados fenólicos mostrando un buen poder predictivo^(13, 14).

Predicción de efectos tóxicos de sustancias químicas sobre la germinación

Entre los diferentes biomarcadores empleados para evaluar el impacto ecotoxicológico por exposición a compuestos químicos en organismos terrestres se encuentra el ensayo de germinación en semillas de lechuga (*Lactuca sativa*)⁽⁴⁵⁾. Desarrollamos una modelación computación para predecir los efectos tóxicos de compuestos químicos en la germinación de semillas de *L. sativa*⁽¹²⁾. La BD recopilada está compuesta por 73 compuestos, ensayados en *L. sativa*, fue dividida en SE y SP, con 56 y 18 compuestos respectivamente. Se emplearon los DM del programa DRAGON⁽²⁸⁾ y el modelo fue obtenido con el programa QSARINS⁽³⁴⁾. La calidad del modelo se determinó examinando los parámetros estadísticos de la regresión para SE y SP.

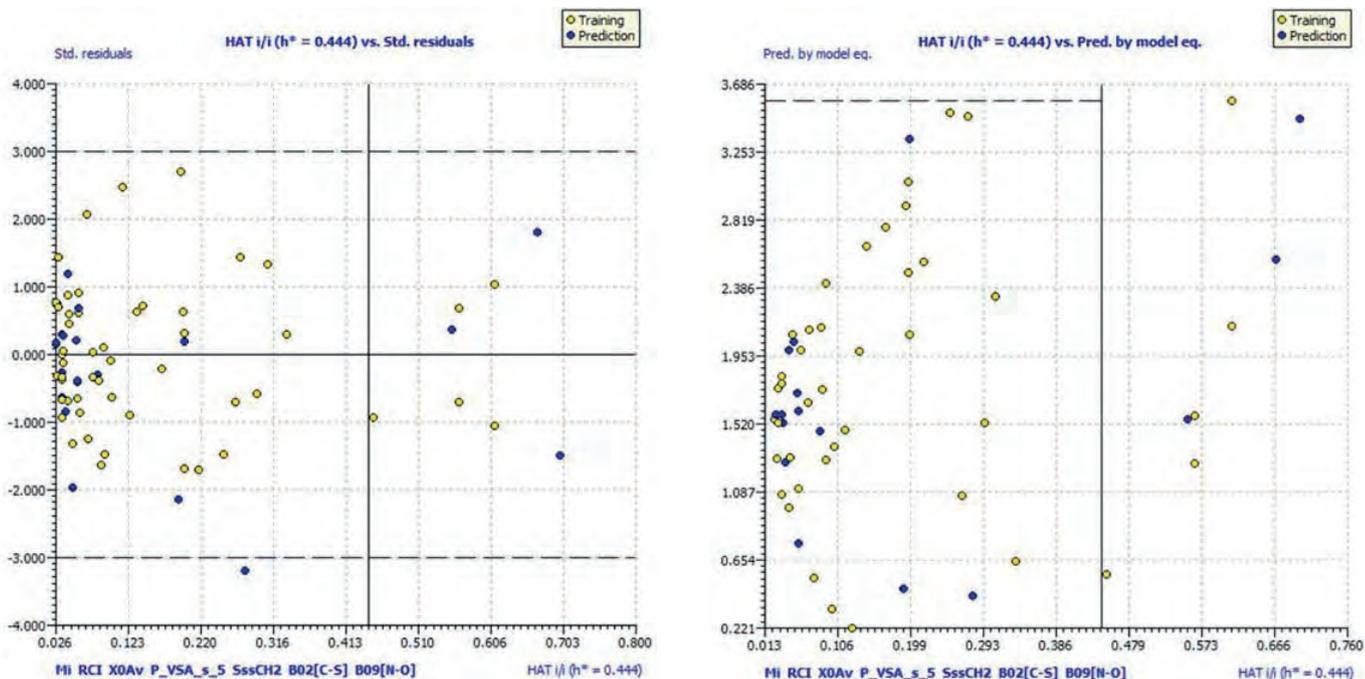


Fig. 2. Dominio de aplicación. Gráfico de Williams a la izquierda; gráfico de Insubria a la derecha.

El modelo desarrollado explica casi el 92% de la varianza experimental, la correlación entre DM es pequeña $K_{xx} = 0,315$ (muy poca información redundante). En este trabajo ⁽¹²⁾ realizamos validación interna de varios tipos: *dejando-uno-fuera*, *dejando-varios-fuera* y experimento de aleatorización de la variable respuesta (*Y-scrambling*) el cual demuestra que el modelo no es producto de una correlación casual. Adicionalmente se estableció el DA del modelo por dos métodos (figura 2): mediante el gráfico de Williams y mediante el gráfico de Insubria, que es un nuevo método propuesto en el programa QSARINS.

Predicción del riesgo ambiental de nanopartículas y nanomateriales

La nanotecnología ha traído grandes avances en muchos campos de la ciencia moderna por tanto los efectos tóxicos potenciales de las nanopartículas en diferentes ecosistemas son de especial interés. En el primer trabajo⁽¹⁵⁾ se desarrolló un modelo basado en la PT para abordar el problema de las nanopartículas; el modelo nano-QSAR se emplea para predecir simultáneamente la ecotoxicidad de diferentes nanopartículas ensayadas contra diferentes biomarcadores, considerando múltiples medidas de ecotoxicidad así como diferentes composiciones químicas, tamaños, condiciones de medición y el tiempo durante el cual los organismos de ensayo fueron expuestos a las nanopartículas. El modelo QSAR se desarrolló con una BD de 5520 casos (parejas de nanopartículas) mostrando exactitudes alrededor del 99 % para SE y SP. Para demostrar la aplicabilidad del modelo escogimos tres nanopartículas basadas en níquel (Ni), con valores experimentales reportados en la literatura. Las predicciones del modelo coincidían en gran medida con las evidencias experimentales confirmando que las Ni-nanopartículas eran menos ecotóxicas que otras nanopartículas. Los resultados obtenidos constituyen una herramienta de gran valor para realizar la predicción eficiente de la ecotoxicidad de nanopartículas bajo diversas condiciones experimentales ⁽¹⁵⁾.

Los nanomateriales han revolucionado la ciencia y tecnología modernas; sin embargo su uso y manipulación puede provocar serios daños a organismos vivos y ecosistemas. En el segundo trabajo ⁽¹⁶⁾ se desarrolló un modelo QSAR unificado basado en la PT para predecir simultáneamente la citotoxicidad y ecotoxicidad de los nanomateriales bajo diferentes condiciones experimentales. El modelo fue desarrollado con una BD conformada por 36488 parejas de nanomateriales. Divididos en 27 347 parejas como SE y 9141 como SP. El modelo mostró exactitudes por encima del 98 % para la SE y SP; siendo empleado para predecir la toxicidad de nanomateriales no incluidos en la BD original. Los resultados obtenidos del *screening* virtual realizado a diferentes nanomateriales

para predecir su potencial de citotoxicidad y ecotoxicidad fueron consistentes con los obtenidos experimentalmente. El modelo desarrollado en este trabajo puede ser considerado un enfoque alternativo para brindar nueva información acerca del comportamiento de los nanomateriales en relación a su citotoxicidad y ecotoxicidad ⁽¹⁶⁾.

Modelación computacional de propiedades farmacocinéticas. Transporte a través de membranas biológicas

En este epígrafe, se aplica la modelación computacional a la predicción de propiedades farmacocinéticas de sustancias químicas de interés. Entre las propiedades farmacocinéticas más importantes encontramos la permeabilidad a través de membranas biológicas. Los enfoques computacionales o *in silico* son métodos alternativos para el cribado (eficiente y rentable) de bibliotecas y en la estimación de los parámetros de permeabilidad para nuevas entidades químicas potencialmente activas. ⁽¹⁷⁻²¹⁾

Predicción de permeabilidad a través de la membrana celular

En el primer estudio, se realizaron mediciones de Permeabilidad (P) en la membrana de microbios ruminales en diferentes niveles iniciales de tensión superficial (ST) y área de superficie específica (ASE) ⁽¹⁷⁾. La BD empleada en este trabajo provino de dos fuentes fundamentales, datos previamente publicados por Liu y col. ⁽⁴⁶⁾ y de resultados experimentales de permeabilidad celular obtenidos en este trabajo siguiendo la metodología descrita⁽¹⁷⁾. Se propone un nuevo método para la predicción de P basado en los métodos de PT los cuales comienzan con una solución exacta conocida de un problema al agregar correcciones debido a las variaciones de diferentes condiciones experimentales (C_c). El mejor modelo presentó sensibilidad, especificidad y exactitud superiores a 0,89, y CCM mayor que 0,78 para 77 781 casos que conformaron la BD. Además, también se refiere una simulación de diagrama de fase ternaria con valores predichos de permeabilidad celular en diversas condiciones experimentales. En consecuencia, el modelo propuesto brinda una oportunidad para investigar el efecto de las perturbaciones en la permeabilidad de la membrana celular de los microbios ruminales involucrados en las variables de fermentación *in vitro*.

Predicción de permeabilidad a través de células de Caco-2

La administración oral de fármacos es la vía más apropiada en muchos casos, ya que es la ruta preferida por los pacientes y además presenta menor costo de producción. La mono-capa de células de Caco-2 es el modelo *in vitro* más avanzado porque

esta línea celular expresa una serie de propiedades biológicas de la membrana. Mimetizando las propiedades morfológicas y funcionales de la barrera de células del epitelio intestinal *in vivo* ⁽²⁰⁾. Primero desarrollamos modelos de clasificación usando ADL para predecir el paso de compuestos a través de Caco-2 ⁽¹⁸⁾. Se empleó una BD compuesta por 157 compuestos obtenida de la literatura ⁽⁴⁷⁾. La BD es diversa y tiene representación de todos los tipos de mecanismos de absorción intestinal. Los mejores modelos presentaron valores de exactitud, sensibilidad y especificidad por encima de 90 %; también fueron robustos y estables a diferentes perturbaciones. Los modelos mostraron además adecuado poder predictivo en la SP con valores sobre 90 %; asimismo, fueron comparados con otros enfoques previamente reportados mostrando resultados superiores. Finalmente, simulamos una búsqueda virtual para predecir el perfil de absorción de 241 compuestos obteniendo buenos resultados ⁽²⁵⁾.

Posteriormente, desarrollamos modelos de RLM en Caco-2 ⁽¹⁹⁾ empleando una BD previamente conformada ⁽⁴⁷⁾. Los mejores modelos mostraron valores de $R^2 = 0,72$ ($s = 0,435$) en el caso del modelo no-estocástico y de $R^2 = 0,66$ ($s = 0,464$) para el modelo estocástico. Los valores de q^2 fueron de 0,66 y 0,63, respectivamente. El poder predictivo y la robustez de los modelos fue estimada usando estrategias de validación interna y externa (usando SP). También desarrollamos una comparación con los modelos previamente publicados con la misma DB (idéntica división en SE y SP). Nuestros modelos mostraron el mejor poder predictivo en comparación modelos desarrollados previamente; ya que el R^2_{ext} es superior a 0,70 y anteriormente solo se habían alcanzado 0,64 como valor máximo ^(48,49).

Una revisión anterior examina el uso del modelo de Caco-2 para la predicción de la permeabilidad intestinal ⁽²⁰⁾. Estos modelos computacionales se han venido desarrollando por más de una década y se han aplicado diferentes técnicas en el desarrollo de ellos; sin embargo, la extrapolación a absorción intestinal se encuentra aún de un 80 % a un 85 %. Los principales problemas identificados en este trabajo fueron: conformación de BD, variabilidades inter-laboratorios, mecanismos metabólicos del Caco-2 y la inconsistencia entre los resultados (*in vitro-in vivo*). En este trabajo se hizo un resumen de los avances y limitaciones de los modelos desarrollados con Caco-2 y se exponen además algunas consideraciones a tener en cuenta con vistas a mejorar la aplicabilidad de estudios computacionales en esta línea celular ⁽²⁰⁾.

Predicción de la permeabilidad a través de la barrera hematoencefálica

La barrera hematoencefálica (BHE) es una barrera fisiológica compleja que separa al sistema nervioso central

(SNC) del flujo sanguíneo y ayuda en el mantenimiento de la homeostasis cerebral ⁽⁵⁰⁾. Aunque la BHE es de naturaleza protectora, la baja permeabilidad o incapacidad de los compuestos de atravesarla es un impedimento significativo para candidatos a fármacos del SNC ⁽⁵¹⁾. Desarrollamos un modelo QSAR ⁽²¹⁾ empleando la mayor base de datos disponible hasta este momento (497 compuestos) y DM del DRAGON ⁽²⁸⁾. La BD fue separada en SE (381) y SP (116) y el modelo fue obtenido empleando CT. Aunque nuestro modelo no es complejo su rendimiento es bastante bueno; obteniéndose para la SE y SP los siguientes resultados: exactitud (87,66 % y 87,93 %), sensibilidad (87 % y 86,67 %), especificidad (89,23 % y 89,66 %) y RFP (11,6 % y 10,7 %), respectivamente. Se realizó una comparación con otros modelos previamente desarrollados obteniéndose el mejor rendimiento. Nuestro modelo es una poderosa herramienta de predicción, que es capaz de modelar el paso a través de la BHE ⁽²¹⁾.

Definición de nuevos índices quirales y sus aplicaciones, definición de nuevos descriptores basados en vectores atómicos ponderados

Definición de nuevos índices quirales. En este epígrafe se aborda la definición de índices 3D-quirales (2,5) basados en la generalización y extensión de tres familias de índices previamente definidos en el programa TOMOCOMD-CARDD. Estas familias de índices son: los índices bilineales basados en relaciones atómicas ⁽²²⁾, los índices lineales basados en relaciones de enlaces ⁽²³⁾ y los cuadráticos basados en relaciones de enlaces ⁽²⁴⁾. De esta manera se crearon nuevos índices moleculares que permiten codificar la quiralidad central y otros aspectos estructurales como la isomería Z-E y axial/ecuatorial.

Las familias de DM implementados en TOMOCOMD-CARDD, se basan en dos tipos de representaciones de las moléculas: una vectorial y otra matricial. La forma matricial está asociada al pseudografo y resulta en una matriz de adyacencia de átomos/enlaces (según corresponda). Por otro lado la forma vectorial consiste en un vector de R^n donde n será el número de átomos/enlaces de la molécula y los componentes de este vector propiedades de esos átomos/enlaces. Finalmente para derivar las familias de DM se aplican diferentes transformaciones algebraicas: lineal, cuadrática y bilineal. Estas son incapaces de describir las características tridimensionales de las moléculas. En este sentido y con el propósito de resolver este problema, se introduce un factor trigonométrico de corrección de quiralidad 3D en el vector molecular, \bar{x} . En

este enfoque un vector molecular quiral, $*\bar{x}$, es obtenido; en

donde cada componente de \bar{x} (por ejemplo la electronegati-

vidad de Mulliken, x_A) se sustituye por el siguiente término: $[x_A + \text{sen}((\omega_A + 4\Delta)\pi/2)]$. Este factor trigonométrico de corrección de quiralidad 3D utiliza una variable *dummy* (ficticia), ω_A y un parámetro entero, Δ ⁽²²⁻²⁴⁾.

Con el objetivo de validar los nuevos índices quirales se seleccionaron tres bases de datos. Estas han sido empleadas en estudios previos y son muy útiles pues permiten la comparación entre los nuevos índices propuestos y los ya existentes. La primera está compuesta por 32 isómeros del perindoprilato, un inhibidor de la ECA. La segunda está compuesta por siete pares de enantiómeros de 3-(3-hydroxifenil) piperidinas-N-alquiladas agonistas del receptor σ . La tercera y última es la conocida como base de Cramer, son 31 esteroides que han sido empleados ampliamente para la validación de índices 3D, quirales, etc. Los resultados obtenidos en los tres trabajos ⁽²²⁻²⁴⁾ permitieron demostrar la validez de los nuevos descriptores quirales. Este aporte tiene importancia pues no existe una gran batería de descriptores quirales de utilidad para los investigadores y muchas de las actividades biológicas dependen en último término de la estereoespecificidad y estereoselectividad en la interacción ligando-receptor. Además, la mayoría de los métodos 3D que existen en la actualidad necesitan de alineamiento por lo que este nuevo enfoque también constituye una adecuada alternativa cuando no sea posible lograr dicho alineamiento.

En una revisión sobre el desarrollo de índices quirales en estudios QSAR, ⁽⁵³⁾ se hizo especial énfasis en los DM del programa TOMOCOMD-CARDD que habían sido extendidos para codificar al quiralidad. Se emplearon las tres bases de datos para hacer más fácil que la comparación con enfoques previamente publicados en la literatura y entre ellos ⁽⁵³⁾.

Aplicación de los nuevos índices quirales. El Fenoterol, 5-[1-hydroxy-2-[[2-(4-hydroxyphenyl)-1-methylethyl]-amino]ethyl]-1,3-benzenediol, es un agonista del adrenergico receptor β_2 (β_2 -AR) que se emplea en el tratamiento del asma y puede ser utilizado también en el fallo cardiaco congestivo. Este compuesto presenta dos centros quirales y existe como cuatro estereoisómeros. Desarrollamos dos trabajos donde aplicamos los descriptores quirales antes definidos; se empleó como BD los 26 posibles estereoisómeros del fenoterol ^(25,26). Los modelos fueron desarrollados empleando RLM y mostraron buenos parámetros estadísticos. Los índices lineales ⁽²⁶⁾ describieron más del 92 % de la varianza mientras los índices cuadráticos ⁽²⁵⁾ tuvieron un desempeño superior al 94 %. En ambos estudios se comparó con modelos previamente publicados mostrando un desempeño superior de nuestros modelos ^(25,26).

Definición de nuevos descriptores basados en vectores atómicos ponderados. En esta sección se abordara brevemente la definición matemática de una nueva familia de DM

conocidos como AWWs. En este trabajo ⁽⁵²⁾ se exponen aspectos de la definición matemática y se validan los descriptores. En la introducción se hace una revisión sobre los diferentes tipos de descriptores existentes y los programas en los cuales están implementados. También se hace mención a un grupo de BD que se consideran como BD modelo, estas son las denominadas *bases de datos de Sutherland* las que fueron empleadas para validar los AWW y comparar su desempeño con otros descriptores conocidos. Fueron obtenidos modelos QSAR para cada una de estas ocho BD, todos los modelos fueron validados y además comparados con otros previamente publicados mostrando buen desempeño los nuevos AWW ⁽⁵²⁾. Los resultados obtenidos corroboran la idea de que los nuevos descriptores AWW son una excelente herramienta para estudios QSAR a pesar de su simplicidad.

Conclusiones

En este trabajo han quedado evidenciadas las potencialidades de aplicación del diseño computacional en estudios ambientales y farmacocinéticos. En este sentido, se obtuvieron *nuevos modelos computacionales* que permiten predecir el potencial ecotoxicológico o el mecanismo de acción tóxica de compuestos químicos. Asimismo, se desarrollaron *nuevos modelos* que permiten la predicción de propiedades farmacocinéticas (como la permeabilidad a través de varias membranas biológicas) y se extendieron varias familias de DM para codificar la quiralidad.

Los modelos obtenidos resultaron interpretables, predictivos, estadísticamente robustos y estables; los nuevos modelos fueron comparados con modelos similares, previamente publicados, mostrando un mejor desempeño en casi todos los casos. Lográndose un gran impacto en la esfera científico-tecnológica ya que la aplicación de los resultados alcanzados ha posibilitado y permitirá introducir en las investigaciones tanto a nivel nacional como internacional herramientas capaces predecir el efecto ambiental de compuestos de interés, así como la estimación de propiedades farmacocinéticas. Posibilitando un ahorro sustancial del uso de animales de laboratorio, recursos humanos y materiales, los cuales son muy importantes dentro del desarrollo científico cubano e internacional actual. Los resultados integrales aquí reportados son expresión de la colaboración entre instituciones cubanas y foráneas, la publicación de los resultados alcanzados en el presente trabajo ha proporcionado un reconocimiento de la comunidad científica nacional e internacional.

REFERENCIAS BIBLIOGRÁFICAS

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econom. 2016;47:20-33.

2. Lombardino JG, Lowe III JA. The role of the medicinal chemist in drug discovery – THEN and NOW. *Nature Rev Drug Discov.* 2004;3:853-62.
3. Basak SC, Grunwald GD. Molecular Similarity and Estimation of Molecular Properties. *J Chem Inf Comput Sci.* 1995;35(3):366-72.
4. Estrada E, Vilar S, Uriarte E, Gutierrez Y. In Silico Studies toward the Discovery of New Anti-HIV Nucleoside Compounds with the Use of TOPS-MODE and 2D/3D Connectivity Indices. 1. Pyrimidyl Derivatives. *J Chem Inf Comput Sci.* 2002;42(5):1194-203.
5. García-Domenech R, Julián-Ortiz Jvd. Antimicrobial Activity Characterization in a Heterogeneous Group of Compounds. *J Chem Inf Comput Sci.* 1998;38(3):445-9.
6. Premio-ACC. "De la estructura molecular a la actividad biológica: Aplicaciones y potencialidades de novedosos métodos de diseño computacional en estudios QSAR, QSPR, QSPkR, QSTR, Proteómica y Bioinformática"2005.
7. Premio-ACC. "Bases Teóricas para el Desarrollo e Integración de un Sistema Computacional en el Descubrimiento de Fármacos Activos Contra Enfermedades Microbianas de Gran Impacto Social"2006.
8. Premio-ACC. "Diseño-síntesis-evaluación-propiedad intelectual: descubrimiento y desarrollo de nuevos compuestos bioactivos más eficaces y seguros utilizando un enfoque integral in silico"2011.
9. Castillo-Garit JA, Abad C, Casañola-Martin GM, Barigye SJ, Torrens F, Torreblanca A. Prediction of Aquatic Toxicity of Benzene Derivatives to *Tetrahymena pyriformis* According to OECD Principles. *Curr Pharm Des.* 2016;22(33):5085-94.
10. Dieguez-Santana K, Pham-The H, Villegas-Aguilar PJ, Le-Thi-Thu H, Castillo-Garit JA, Casañola-Martin GM. Prediction of acute toxicity of phenol derivatives using multiple linear regression approach for *Tetrahymena pyriformis* contaminant identification in a median-size database. *Chemosphere.* 2016;165:434-41.
11. Martínez-López Y, Barigye SJ, Martínez-Santiago O, Marrero-Ponce Y, Green J, Castillo-Garit JA. Prediction of aquatic toxicity of benzene derivatives using molecular descriptor from atomic weighted vectors. *Environ Toxicol Pharm.* 2017;56:314-21.
12. Castillo-Garit JA, González Y, Albear EM, Rodríguez E, Pérez-Doñate V, Pérez-Giménez F. Environmental toxicity prediction using computational tools: prediction of potential hazardous effects of chemicals in *Lactuca sativa* seed germination. *Nereis.* 2019;11:15-30.
13. Brito-Sánchez Y, Castillo-Garit JA, Le-Thi-Thu H, González-Madariaga Y, Torrens F, Marrero-Ponce Y, et al. Comparative study to predict toxic modes of action of phenols from molecular structures. *SAR QSAR Environ Res.* 2013;24(3):235-51.
14. Castillo-Garit JA, Casañola-Martin GM, Barigye SJ, Pham-The H, Torrens F, Torreblanca A. Machine learning-based models to predict modes of toxic action of phenols to *Tetrahymena pyriformis*. *SAR QSAR Environ Res.* 2017;28(9):735-47.
15. Kleandrova VV, Luan F, González-Díaz H, Ruso JM, Melo A, Speck-Planche A, et al. Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environ Int.* 2014;73:288-94.
16. Kleandrova VV, Luan F, González-Díaz H, Ruso JM, Speck-Planche A, Cordeiro MNDS. Computational Tool for Risk Assessment of Nanomaterials: Novel QSTR-Perturbation Model for Simultaneous Prediction of Ecotoxicity and Cytotoxicity of Uncoated and Coated Nanoparticles under Multiple Experimental Conditions. *Environ Sci Tech.* 2014;48(24):14686-94.
17. Liu Y, Ran T, Tenorio-Borroto E, Tang S, Pazos A, Tan Z, et al. Experimental and chemometric studies of cell membrane permeability. *Chem Intel Lab Syst.* 2016;154:1-6.
18. Castillo-Garit JA, Cañizares-Carmenate Y, Marrero-Ponce Y, Torrens F, Abad C. Prediction of ADME properties, Part 1: Classification models to predict Caco-2 cell permeability using atom-based bilinear indices. *Afinidad.* 2014;71(566):129-38.
19. Le-Thi-Thu H, Canizares-Carmenate Y, Marrero-Ponce Y, Torrens F, Castillo-Garit JA. Prediction of Caco-2 Cell Permeability Using Bilinear Indices and Multiple Linear Regression. *Lett Drug Des Discov.* 2016;13(2):161-9.
20. Pham-The H, Cabrera-Pérez MÁ, Nam N-H, Castillo-Garit JA, Rasulev B, Le-Thi-Thu H, et al. In Silico Assessment of ADME Properties: Advances in Caco-2 Cell Monolayer Permeability Modeling. *Curr Top Med Chem.* 2018;18(26):2209-29.
21. Castillo-Garit JA, Casañola-Martin GM, Le-Thi-Thu H, Pham-The H, Barigye SJ. A Simple Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Classification Trees. *Med Chem.* 2017;13(7):664-9.
22. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, Rotondo R. Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification. *J Mol Graphics Model.* 2007;26:32-47.
23. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, García-Domenech R, Romero-Zaldivar V. Bond-Based 3D-Chiral Linear Indices: Theory and QSAR Applications to Central Chirality Codification. *J Comput Chem.* 2008;29:2500-12.
24. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, García-Domenech R, Rodríguez-Borges JE. Applications of Bond-Based 3D-Chiral Quadratic Indices in QSAR Studies Related to Central Chirality Codification. *QSAR & Comb Sci.* 2009;28:1465-77.
25. Castillo-Garit JA, García-Domenech R, Marrero-Ponce Y, Torrens F, Abad C. Atom-based 3D-chiral quadratic indices. Part 3: Prediction of the binding affinity of the stereoisomers of fenoterol to the β_2 adrenergic receptor. *Afinidad.* 2011;68(555):334-40.
26. Castillo-Garit JA, García-Domenech R, Marrero-Ponce Y, Torrens F, Abad C. Prediction of the Binding Affinity between Fenoterol Derivatives and the β_2 -Adrenergic Receptor Using Atom-Based 3D-Chiral Linear Indices. *Nereis.* 2012;4.
27. Marrero-Ponce Y, Romero V. TOMOCOMD software (TOPOlogical MOlecular COMputer Design) for Windows. 1.0 ed. Central University of Las Villas. Villa Clara. Cuba 2002.
28. Todeschini R, Consonni V, Pavan M. Dragon Software v. 2.1, 2002.
29. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J Chem Inf Comput Sci.* 2003;43(2):493-500.
30. Martínez-López Y, Marrero-Ponce Y, SJ. B, inventors; CAMD-BIR Unit, assignee. MD-LOVIs for windows: Software for molecular descriptors calculator from LOVIs. Version 1.0. 2010.
31. STATISTICA version. 6.0. 2001;StatSoft.;Tulsa.
32. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques.* Second ed. Gray J, editor: Morgan Kaufmann Publishers; 2005.

33. Todeschini R, Consonni V, Mauri A, Pavan M. MOBYDIGS version 1.0. Milano 2005.
34. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: A new software for the development, analysis and validation of QSAR MLR models. *J Comput Chem*. 2013;34(24):2121-32.
35. Quinlan R. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers; 1993.
36. Aha D, Kibler D. Instance-based learning algorithms. *Machine Learning*. 1991; 6:37-66.
37. Zupan J, Gasteiger J. *Neural Networks for Chemistry and Drug Design*. Weinheim (Germany): Wiley-VCH Publishers; 1999.
38. Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges CJC, Smola AJ, editors. *Advances in kernel methods*: MIT Press; 1999. p. 185-208.
39. Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*. 2000;16(5):412-24.
40. Schultz TW. TETRATOX: *Tetrahymena pyriformis* population growth impairment endpoint- A surrogate for fish lethality. *Toxicol Methods*. 1997;7:289-309.
41. Bradbury SP, Russom CL, Ankley GT, Schultz TW, Walker JD. Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals. *Environ Toxicol Chem*. 2003;22(8):1789-98.
42. Guidance document on the validation of quantitative structure-activity relationship [QSAR] models., (2007).
43. Garg R, Kurup A, Hansch C. Comparative QSAR: on the toxicology of the phenolic OH moiety. *Crit Rev Toxicol*. 2001;31(2):223-45.
44. Aptula AO, Netzeva TI, Valkova IV, Cronin MTD, Schultz TW. Multivariate discrimination between modes of toxic action of phenols. *Quant Struct Act Relat*. 2002;21:12-22.
45. Wang W. Root elongation method for toxicity testing of organic and inorganic pollutants. *Environ Toxicol Chem*. 1987;6(5):409-14.
46. Liu Y, Ran T, Tan Z-l, Tang S-x, Wang P-p. Effects of surface tension and specific surface areas on in vitro fermentation of fiber. *Acta Vet Zootech Sin*. 2013;44 901-10.
47. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, García-Domenech R. Estimation of ADME properties in drug discovery: Predicting Caco-2 cell permeability using atom-based stochastic and non-stochastic linear indices. *J Pharm Sci*. 2008;97:1946-76.
48. Guangli M, Yiyu C. Predicting Caco-2 permeability using support vector machine and chemistry development kit. *J Pharm Pharmacol*. 2006;9(2):210-21.
49. Hou TJ, Zhang W, Xia K, Qiao XB, Xu XJ. ADME evaluation in drug discovery. 5. Correlation of Caco-2 permeation with simple molecular properties. *J Chem Inf Comput Sci*. 2004;44(5):1585-600.
50. Bujak R, Struck-Lewicka W, Kaliszan M, Kaliszan R, Markuszewski MJ. Blood-brain barrier permeability mechanisms in view of quantitative structure-activity relationships (QSAR). *J Pharm Biomed Anal*. 2015;108:29-37.
51. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov*. 2007;6(7):521-32.
52. Martínez-López Y, Caballero Y, Barigye SJ, Marrero-Ponce Y, Millán-Cabrera R, Madera J, et al. State of the Art Review and Report of New Tool for Drug Discovery. *Curr Top Med Chem*. 2017;17(26):2957-76.
53. Marrero-Ponce Y, Castillo-Garit JA, Castro EA, Torrens F, Rotondo R. 3D-Chiral (2.5) Atom-Based TOMOCOMD-CARDD Descriptors: Theory and QSAR Applications to Central Chirality Codification. *J Math Chem*. 2008;44:755-86.

Recibido: 22/04/2020

Aprobado: 18/07/2020

Agradecimientos. Los autores quieren agradecer al colectivo de colaboradores siguiente: Alejandro Pazos, Alejandro Speck-Planche, Amparo Torreblanca, André Melo, Bakhtiyor Rasulev, Concepcion Abad, Eberts M. Albear, Eduardo Alberto Castro, Efrain Edgar Chaluisa Quishpe, Elizabeth Rodríguez, Esvieta Tenorio-Borroto, Facundo Pérez-Giménez, Feng Luan, Gerardo M. Casañola-Martin, Huong Le-Thi-Thu, James Green, Jose E. Rodríguez-Borges, Juan M. Ruso, Julio Madera, Maria Natália D.S. Cordeiro, Miguel Á. Cabrera-Pérez, Nguyen-Hai Nam, Oscar Martínez Santiago, Pedro J. Villegas-Aguilar, Ramón García-Domenech, Reisel Millan-Cabrera, Richard Rotondo, Shaoxun Tang, Stephen J. Barigye, Tao Ran, Valeria V. Kleandrova, Vicente Romero Zaldívar, Virginia Pérez-Doñate, Yaile Caballero, Yisel González Madariaga, Yoan Brito Sánchez, Yong Liu, Yovani Marrero Ponce, Yuleidis González Pérez, Zhiliang Tan.

Contribución de autoría

1. Conceptualización: J.A. Castillo-Garit, H. González-Díaz, Y. Cañizares-Carmenate, H. Pham-The, Y. Martínez-López y K. Diéguez-Santana.
2. Curación de datos: J.A. Castillo-Garit, Y. Cañizares-Carmenate, H. Pham-The, Y. Martínez-López y K. Diéguez-Santana.
3. Análisis formal: -
4. Adquisición de fondos: F. Torrens Zaragoza y H. González-Díaz.
5. Investigación: J.A. Castillo-Garit, H. González-Díaz, Y. Cañizares-Carmenate, H. Pham-The, Y. Martínez-López y K. Diéguez-Santana.
6. Metodología: H. Pham-The, F. Torrens Zaragoza, J.A. Castillo-Garit y H. González-Díaz.
7. Administración del proyecto: -
8. Recursos: -
9. Software: F. Torrens Zaragoza, Y. Martínez-López y H. González-Díaz.
10. Supervisión: J.A. Castillo-Garit, F. Torrens Zaragoza y H. González-Díaz.
11. Validación: Y. Cañizares-Carmenate, Y. Martínez-López y K. Diéguez-Santana.
12. Visualización: -
13. Redacción – borrador original: J.A. Castillo-Garit y Y. Cañizares-Carmenate.
14. Redacción – revisión y edición: J.A. Castillo-Garit, F. Torrens Zaragoza y H. González-Díaz.

Conflictos de interés. Los autores no declaran conflictos de interés.

Financiamiento. Los trabajos que conforman este compendio han recibido financiamiento de las siguientes Instituciones/proyectos:

- **España:** Universidad de Valencia bajo el programa "Estades Temporals per a Investigadors Convidats"; Ministerio de Economía y Competitividad a través de los proyectos (CTQ2004-07768-C02-01/BQU, CCT-005-07-00365 y BFU2013-41648-P); Ministerio de Sanidad (SAF2005-PI052128); Ministerio de Ciencia e Innovación a través de los proyectos (BFU 2010-19118, SAF2009-10399 y MAT2011-25501); Universidad Católica de Valencia San Vicente Mártir; "Galician Network for Colorectal Cancer Research (REGICC)" (Ref. R2014/039), apoyado por la Xunta de Galicia, mediante el proyecto "Collaborative Project on Medical Informatics (CIMED)" P113/00280.
- **Unión Europea:** (FEDER funds) bajo el programa QREN a través del proyecto NORTE-07-0124-FEDER-000067-NANOCHEMISTRY y FEDER a través de COMPETE.
- **Portugal:** FCT, Fundação para a Ciência e Tecnologia a través del proyecto Pest-C/EQB/LA0006/2013.
- **Cuba:** a través del programa VLIR-UCLV titulado "Strengthening postgraduate education and research in Pharmaceutical Sciences";
- **Viet-Nam:** National Vietnam National University y Hanoi University of Pharmacy.
- **Ecuador:** Universidad Estatal Amazónica a través de proyectos internos.
- **China:** Fundación Nacional para las Ciencias Naturales (Grant No. 31172234), Academia de Ciencias de China (Grant No. XDA05020700) y Hunan Provincial Natural Science Foundation of China (Grant No. 12JJ3015).

