

Predicción de Propiedades de Sustancias Químicas con la Teoría de las Relaciones Cuantitativas Estructura-Actividad y Estructura- Propiedad

Pablo R. Duchowicz

RESUMEN

Las Relaciones Cuantitativas Estructura-Actividad (QSAR) y Estructura-Propiedad (QSPR) permiten predecir un amplio espectro de propiedades fisicoquímicas y biológicas de las sustancias químicas que presenten interés en diferentes áreas de la Química, exhibidas en familias de compuestos químicos de baja y elevada masa molecular. Dicha teoría semiempírica formula modelos matemáticos predictivos derivados de la estructura molecular, y la finalidad de sus aplicaciones es conseguir mejorar la metodología general de trabajo para que represente una herramienta predictiva aplicable a cualquier problema fisicoquímico o biológico. Este objetivo puede considerarse un aporte científico realmente importante y significativo en el avance de la predicción de propiedades carentes de mediciones experimentales. Todo estudio QSAR/QSPR se enfoca en establecer paralelismos que seleccionen los factores estructurales microscópicos preponderantes que afectan la propiedad. La técnica QSAR/QSPR complementa estudios teóricos o experimentales que traten de elucidar racionalmente los interrogantes de tipo químico involucrados en los distintos problemas. Se sabe que el valor científico de estos esquemas usados de manera aislada es limitado, por lo que se busca mantener una permanente interrelación con grupos experimentales colaboradores de manera de resolver a la par los interrogantes planteados en estos estudios.

Hoy día encuentra verdadero interés la predicción teórica del gran número de propiedades fisicoquímicas y biológicas, en diversas áreas de la Química tales como la Bioquímica, la

Química Orgánica, Química Analítica, Química Medicinal, Química de los Alimentos, Biotecnología, Química Ambiental, y otras. En muchas circunstancias las medidas experimentales de tales propiedades perma-

necen desconocidas por trabajarse con compuestos que resultan ser nuevos, tóxicos o que demandan demasiado tiempo de medición experimental. A su vez, si la cuestión se resolviera con el mero procedimiento de síntesis y

testeo de sustancias sin otra guía que la posterior prueba y error, ello constituiría una metodología extremadamente laboriosa, costosa y nada científica.

Siempre que se busque predecir propiedades que dependan tanto de la reactividad química como de la estructura molecular, sin lugar a dudas deberán abordarse los métodos derivados de la Mecánica Cuántica con el fin de representar adecuadamente el fenómeno involucrado. Por otro lado, es bien sabido que los cálculos mecanocuánticos actuales sólo pueden resolverse con buena aproximación usando un nivel de teoría adecuado, siempre y cuando involucren moléculas de unos pocos átomos y se encuentren libres de interacciones intermoleculares en el estado gaseoso.

La mayoría de las aplicaciones del análisis de datos involucra intentos de ajustar un conjunto de observaciones experimentales a un modelo, usualmente de tipo cuantitativo. En muchas aplicaciones se supone que un modelo será utilizado de manera predictiva, pero las predicciones no tienen que ser necesariamente cuantitativas. Las formulaciones de la Teoría de las Relaciones

(QSPR) constituyen una hipótesis matemática que tiene su fundamento en el hecho de que la estructura molecular es responsable de las propiedades que exhibe la colección de moléculas interactuantes [1-5]. Los estudios pioneros en QSAR/QSPR fueron realizados por Corwin Hansch y Toshio Fujita en el año 1964 [6], y a partir de ese entonces los avances en esta materia no han cesado y las aportaciones significativas se suceden continuamente.

El hecho que los distintos compuestos químicos tienen diferentes efectos biológicos se conoce desde hace muchísimo tiempo. Uno de los primeros ejemplos de un compuesto con efectos medicinales fue el uso de Ma Huang, que contiene Efedrina, para tratar el asma y la fiebre del heno. Los ejemplos de sustancias bioactivas derivadas de distintas plantas son harto variados. Sin embargo, hasta el momento que la ciencia Química no estuvo suficientemente desarrollada para asignar estructuras a los compuestos no fue posible empezar a especular acerca de las causas de las propiedades biológicas. La habilidad para determinar estructuras permitió a los investigadores establecer relaciones estructura-actividad (SAR), que son simplemente las observaciones de que un cierto cambio en la estructura química de un determinado compuesto produce un cierto efecto sobre la actividad biológica.

Una verdadera demostración sobre la importancia que adquieren las predicciones QSAR/QSPR es la siguiente: el potencial de la Química Orgánica para la producción de nuevos compuestos es realmente enorme, ya sea que ellos estén destinados a usos farmacéuticos o agroquímicos, sean para generar fragancias, saborizantes o alimentos. En 1994, el Chemical Abstracts presentó más de 13 millones de compuestos, pero no obstante sólo una pequeña proporción de ellos se pudo llegar a sintetizar, al resultar la mayoría de las sustancias ser inestables. Por tanto, las relaciones estructura-propiedad permiten decidir a priori cuáles sustancias valdrían la pena sintetizar por presentar propiedades predichas buscadas.

La Teoría QSAR/QSPR también puede emplearse para describir a la propiedad en términos estructurales sugiriendo así paralelismos, para el descubrimiento, diseño y optimización molecular de nuevas drogas, y hasta para suministrar alguna información en cuanto a los mecanismos de las reacciones químicas. Una ventaja de poseer mejores descripciones de la estructura molecular es que resulta posible transferir información de una serie de moléculas a otra serie distinta. Este hecho constituye la real contribución de la teoría al mejoramiento de las propiedades fisicoquímicas, medicinales, alimentarias, etc. de las series de compuestos analizados.

Instituto de Investigaciones

Fisicoquímicas Teóricas y Aplicadas
(INIFTA), CONICET, UNLP, Diag. 113
y 64, C.C. 16, Sucursal 4, La Plata
(1900), Buenos Aires, Argentina
E-mail: prduchowicz@gmail.com

Cuantitativas Estructura Actividad
(QSAR) y Estructura-Propiedad

El diseño de nuevos métodos matemáticos eficientes que sean capaces de cuantificar mejor la hipotética relación estructura-propiedad/actividad, genera lugar a un amplio espectro de propiedades que pueden analizarse. Es importante señalar que la hipótesis fundamental de QSAR/QSPR no es predecir el mecanismo de acción molecular sino la propiedad, el resultado final del mecanismo que una estructura química produce. Sin embargo, la predicción acertada de la propiedad de una sustancia permite inferir alguna información del fenómeno involucrado, como cuando una estructura es predicha activa/inactiva.

Se busca establecer estudios QSPR en propiedades termodinámicas (entalpías de formación, cambios de energías libres y entropías estándares), por ejemplo en compuestos orgánicos acíclicos y aromáticos representativos con diferente tipo de heteroátomos en su estructura. Esto resulta importante en vista de que, si bien los métodos parametrizados de la Teoría de Orbitales Moleculares Semiempírica se diseñaron especialmente para obtener este tipo de parámetros termodinámicos, en muchas circunstancias sus predicciones presentan serias anomalías por tratarse con sistemas que no poseen información experimental alguna. Lo mismo sucede en cuanto a la determinación de puntos de ebullición normales,

una propiedad de amplia aplicación química que suele ser "complicada" de modelarse al estar determinada por las interacciones moleculares en el estado líquido y por la diferencia en la función de partición molecular interna entre las fases gaseosa y líquida a la temperatura de ebullición. El diseño de QSPR en estos estudios específicos es necesario en vista que los grupos experimentales necesitan conocer de antemano las predicciones de la propiedad fisicoquímica, para cumplimentar mejor sus trabajos experimentales en la identificación y caracterización de los compuestos químicos.

La búsqueda de moléculas aún no sintetizadas que puedan tener buena actividad y baja toxicidad para el medio ambiente, es uno de los objetivos de la Química Verde o Sustentable. Además, la predicción de nuevas estructuras químicas es lo que ciertas empresas farmacéuticas internacionales hacen para la búsqueda virtual de nuevos fármacos, por ejemplo. Tales predicciones constituyen el paso previo fundamental requerido por tales empresas, para que luego puedan realizar la síntesis experimental de dichos compuestos en el laboratorio mediante la guía de las estructuras predichas y probar su propiedad/actividad.

Se establecen estudios QSAR en moléculas que interactúen en

organismos biológicos, para predecir su nivel de actividad/potencia y asistir a la búsqueda de sustancias químicas con propiedades específicas deseables. Por ejemplo, resulta muy valiosa la propuesta de modelos in silico que sean capaces de cuantificar las relaciones estructura-actividad observadas durante el estudio de inhibidores no-nucleósidos de la enzima Transcriptasa Reversa (NNRTI) en el tratamiento clínico del SIDA. Cuando los NNRTI se utilizan en combinación con otros medicamentos anti-HIV, dicho cóctel de drogas bloquea la multiplicación del HIV en la sangre. Por tanto, resulta de interés la búsqueda de compuestos que posean un espectro de potencias favorables frente a la enzima, tales como los análogos estructurales de la droga Efavirenz. Otras enfermedades que se estudian involucran compuestos tipo-droga frente a Cáncer, Malaria, Alzheimer, Chagas, Tuberculosis, actividad antimotóptica, antifúngica, etc.

Se modela mediante QSAR el efecto contaminante de agentes químicos o sus mezclas en el medio ambiente, como toxicidad acuosa, persistencia en suelos, etc. La toxicidad acuosa exhibida por compuestos alifáticos heterogéneos, que inhiben el crecimiento de la especie ciliada *Tetrahymena pyriformis*, suele involucrar diferentes mecanismos experimentales de acción toxicológica y procesos de naturaleza narcó-

tica y electrofílica. Los estudios QSAR sobre compuestos que sigan un único mecanismo de acción obtendrán obviamente predicciones más confiables en comparación a los análisis que involucren sustancias con diferentes mecanismos. Se estudian también propiedades de agroquímicos: actividad antialimentaria frente a *Spodoptera litura*, inhibición de la actividad de la hormona juvenil de insectos, y otras.

El conocimiento de las propiedades de macromoléculas es crucial para la síntesis, desarrollo, y manufactura de nuevas sustancias, y el número de sus aplicaciones es innumerable. Por ejemplo, la predicción de la solubilidad acuosa de un compuesto químico, junto con su permeabilidad, constituye un desafío en la industria farmacéutica. Es posible incrementar la velocidad de disolución de una droga lipofílica, tal como Oxazepam o Griseofulvin, en base a reducir el tamaño de sus partículas mediante la formación de una dispersión sólida con un portador hidrofílico de peso molecular determinado, tal como Polietilenglicol (PEG), Polivinilpirrolidona (PVP) o Hidroxipropilcelulosa (HPC). La predicción QSPR de la liberación controlada de drogas inmersas en matrices sólidas inertes podría aportar datos interesantes en relación a las propiedades de disolución de las dispersiones sólidas, especialmente si logran simular las condiciones gastrointestinales, y además pueden aportar a la com-

prensión del mecanismo que conduce al aumento de solubilidad observado en dichos sistemas.

Varias de las propiedades fisicoquímicas de macromoléculas que se investigan son solubilidades acuosas y en otros solventes; temperaturas de transición vítrea; densidades; grado de cristalinidad; clasificación de polímeros en elastómeros, termoplásticos, y termoestables; módulo de elasticidad; fotoconductividad; luminiscencia de polímeros semiconductores; viscosidades relativas; y otras. Entre las propiedades biológicas de macromoléculas se estudian: actividad mutagénica; interacción de la droga *Paromomicina* con región RNA de HIV-1; actividad antiproliferativa; actividad antifúngica; clasificación de secuencias promotoras del RNA; y otras.

En la Teoría QSAR/QSPR la estructura química se representa con los denominados descriptores moleculares, que son cantidades numéricas (teóricas o empíricas) que reflejan diversas características constitucionales, topológicas, geométricas, electrónicas o lipofílicas de las moléculas estudiadas. Los descriptores generalmente son construidos mediante aproximaciones de la Química Cuántica, de la Teoría de la Información, o de la Teoría de Grafos [7,8]. El ejemplo más sencillo de descriptores es la cuenta de átomos y tipos de enlaces de la molécula. En la actualidad, se dispone de miles de definiciones de descriptores

asequibles de la literatura, y un problema principal a resolver en QSAR/QSPR es la adecuada selección de un conjunto reducido y representativo de descriptores moleculares, para diseñar un modelo que sea capaz de explicar y predecir lo mejor posible a la propiedad bajo estudio.

Es posible clasificar a los estudios QSAR/QSPR como tradicionales o semiempíricos. El trabajo fundamental que desarrollaron Hansch y Fujita se basó en realizar un modelado tipo tradicional, en el cual presenta especial interés establecer relaciones empíricas propiedad-propiedad. Por ejemplo, se pueden explicar propiedades "complejadas" como lo son los efectos biológicos de sustancias, en términos de propiedades físico-químicas más simples de entender, como es el caso de la solubilidad acuosa, el coeficiente de partición entre las fases *n*-octanol/agua, la refracción molar, puntos de ebullición, volúmenes molares o calores de vaporización. Todas estas propiedades que son medida del carácter lipofílico, forma molecular y propiedades electrónicas dependen de la estructura, si bien en un modo indirecto, y se pueden medir experimentalmente de manera más fácil que las propiedades que buscan modelarse. Sin embargo, en la práctica presenta mayor utilidad proponer modelos basados en descriptores derivados de la teoría. Con más de 100 años de aplicaciones en Química, la Teoría de Grafos [7] ha demostrado resultar de vital

importancia y su lenguaje más natural. Los descriptores topológicos que se derivan de esta teoría se obtienen a partir de la representación de la molécula por un grafo, es decir, su imagen en el plano, y proporciona sólo información de la constitución y conectividad de la estructura y nada acerca de los aspectos tridimensionales o estereoquímicos de la misma. El primer descriptor topológico se debe al trabajo realizado por Harry Wiener [9], en sus estudios de la variación de los puntos de ebullición de hidrocarburos acíclicos en función de la constitución molecular. Otros índices topológicos famosos que se utilizan ampliamente en los análisis QSAR/QSPR fueron posteriormente propuestos por Haruo Hosoya, Milan Randić, Lemont B. Kier y Lowell H. Hall, y otros autores [8].

A pesar de que los descriptores topológicos no tienen un significado físico preciso, son medida de la "forma topológica" molecular, representando su grado de ramificación, flexibilidad, forma geométrica, ciclicidad, y centricidad. Correlacionan bien con las propiedades cuando los aspectos geométricos y electrónicos tienen menor influencia; pe. volumen molar, calor de vaporización, presión crítica, tensión superficial, punto de fusión, energía de Gibbs, entalpía de formación, índice de refracción, o densidad.

El grupo de Alan R. Katritzky publicó correlaciones impresionantes para escalas de polaridad del solvente, temperaturas de transición vítrea de polímeros, y puntos de fusión de derivados bencénicos sustituidos [10,11]. Esta última propiedad depende mucho de la geometría y las fuerzas intermoleculares, por lo que puede correlacionarse con la estructura sólo en conjuntos de moléculas homogéneas. El grupo de Peter C. Jurs estableció estudios QSPR para la solubilidad acuosa de compuestos orgánicos y para valores de desplazamiento químico C13-NMR [12,13]. Se ha demostrado que en los estudios QSPR los índices topológicos están entre los mejores descriptores, pero en el caso de QSAR estas variables deben ser complementados con descriptores relacionados con la hidrofobicidad de un compuesto, concepto introducido por Hansch [14]. La hidrofobicidad se halla involucrada con el transporte molecular de una droga a través de membranas biológicas, y es indicadora de la forma en que el mismo se distribuye en el organismo.

La gran ventaja de los índices topológicos es su facilidad de cómputo para cualquier estructura imaginable con un relativamente bajo tiempo de cómputo. Cuando la propiedad fisicoquímica o biológica que se estudia se halla muy influenciada también por aspectos geométricos y elec-

trónicos de las moléculas, deben emplearse descriptores conformacionales. Es por ello que existen difundidos en la literatura miles de descriptores moleculares de todo tipo, y dos grupos de investigación ampliamente reconocidos, el de Alan R. Katritzky por un lado y el de Roberto Todeschini por otro, han elaborado programas comerciales (CODESSA [15], Dragon [16]) que permiten calcular diferente tipo de descriptores constitucionales, topológicos, geométricos, electrónicos y lipofílicos luego de la digitalización de las estructuras moleculares. Algunos programas de libre acceso ampliamente utilizados en el trabajo QSAR/QSPR son el PaDEL [17], Mold2 [18], Fragmentor [19], EPI Suite [20], CORAL [21], QuBiLS-MAS [22], LOVIs [23], y otros. En aquellas propiedades que así lo requieran, también se incorporan descriptores mecanocuánticos no provistos por estos programas.

Durante el diseño de descriptores moleculares eficientes para las investigaciones QSAR/QSPR, las aplicaciones concretas de los descriptores flexibles (específicos a la propiedad) del programa CORAL se han expandido en los últimos años, pues son variables que permiten ajustar los ambientes atómicos y distinguir mejor el rol de los heteroátomos. Es posible mejorar la capacidad predictiva de los modelos QSAR/QSPR a través de realizar combinaciones de descriptores

convencionales y descriptores flexibles, por lo que se investiga esta variante como metodología de trabajo.

El hecho de que un descriptor no sea un código sino un número que refleja alguna característica microscópica estructural impide reconstruir la estructura química a partir de los valores de sus descriptores. A pesar de ello es posible por ejemplo asistir al estudio de enfermedades como Tuberculosis o Malaria, pues el tratamiento QSAR/QSPR permite especificar los valores numéricos más favorables que deberían tener los descriptores moleculares participantes del modelo para que la propiedad predicha tenga un valor favorable. De esta manera es posible predecir el nivel de actividad/potencia de una sustancia y asistir a la búsqueda de sustancias con propiedades específicas deseables, evitando medir experimentalmente la propiedad si posee predicciones desalentadoras.

Los estudios QSAR/QSPR de mezclas químicas multicomponentes involucran la definición apropiada de descriptores de mezcla. Mediante el desarrollo de algoritmos que consigan este objetivo, es posible estudiar aceites esenciales, que poseen interés debido a sus diferentes bioactividades, como antibacterial, antifúngica, antiparásita, antimicobacteriana, etc.

En el tratamiento de macromoléculas, se sabe que los descriptores clásicos no fueron dise-

ñados para el tratamiento de moléculas con un gran número de átomos, por lo que resulta imprescindible extender su formulación a macromoléculas. Esto es en principio factible, ya que formalmente en un grafo no hay diferencia si un vértice (nodo) representa un átomo de carbono o si se trata de una base nitrogenada o de otra molécula [24]. Además, no es conveniente utilizar descriptores conformacionales si no se conoce en detalle la estructura macromolecular, por lo que suelen utilizarse descriptores independientes de los aspectos conformacionales de las macromoléculas.

Es factible desarrollar el trabajo QSAR/QSPR cotidiano gracias a la gran disponibilidad de métodos aproximados existentes en la literatura y que posibilitan establecer una relación matemática desconocida estructura-actividad/propiedad. El más simple y muy conocido de todos ellos es el análisis de Regresión Lineal Multivariable (MLR). Sin embargo, en aquellas situaciones en que la relación estructura-propiedad resulta más complicada y presenta un alto carácter no-lineal, como es el modelado de algunas bioactividades, es usual recurrir a otras técnicas estándares más elaboradas, tales como los Algoritmos Genéticos (GA) [25], o las Redes Neuronales Artificiales (ANN) [26]. Ambos métodos tienen analogías que permiten comprender fácilmente su funcionamiento: los AG se

basan en las reglas de evolución biológica de los organismos vivientes, mientras que las ANN fueron originalmente definidas como modelos de la actividad del cerebro humano: son capaces de reconocer relaciones altamente no-lineales entre la propiedad y la estructura durante el procesamiento de los datos. También es posible establecer modelos no-lineales y no-estadísticos por medio de la Teoría del Orden [27], que permite ordenar datos y realizar interpolaciones lineales sin necesidad de emplear los métodos clásicos dependientes de parámetros ajustables. La Teoría del Orden se basa en elementos básicos de la Matemática Discreta y no requiere conocer la función matemática del modelo en la relación estructura-propiedad.

Resulta necesario identificar los mejores descriptores contribuyentes y que expliquen de la mejor manera posible a la propiedad estudiada; esta selección de descriptores no es una tarea trivial. La búsqueda exacta (combinatorial) de los mejores descriptores moleculares de una propiedad implica realizar todas las combinaciones posibles entre los descriptores, impracticables cuando su número es elevado (por ejemplo 100.000 descriptores). Existen reportados en la literatura diferentes algoritmos matemáticos eficientes, cuyas soluciones son en algunos casos concordantes con las exactas y en otros casos muy próximas. La característica distintiva de estos

algoritmos es que permiten explorar miles de descriptores y alcanzar una solución optimizada plausible en un tiempo de cálculo relativamente corto. Algunos ejemplos de algoritmos de selección de descriptores son los famosos Método de Inclusión de Pasos (FSI) [28], los Algoritmos Genéticos [25], o el Método del Reemplazo (RM) [29].

Una etapa vital de todo estudio QSAR/QSPR es la validación de los modelos matemáticos establecidos, pues permite verificar el poder predictivo y validez general, de manera de revelar si resulta posible transferir la información adquirida con las moléculas de calibración (empleadas para calibrar el modelo) hacia moléculas no contempladas durante el ajuste de las relaciones QSAR/QSPR. La propuesta de métodos de validación es un área de investigación activa y en permanente desarrollo [30-35]. Recurre habitualmente al uso de conjuntos externos de moléculas, a la técnica de validación cruzada, técnica de aleatorización y, o a la definición del dominio de aplicación del modelo, aunque siempre deben investigarse nuevas estrategias para verificar la capacidad predictiva.

Es posible afirmar que la técnica de regresión lineal es la que mejor funciona para establecer modelos predictivos, es decir, que

no se limiten únicamente al ajuste de los datos que se usaron para su calibración sino que también sean de utilidad para predecir datos externos. En este contexto se cumple la hipótesis central en QSAR/QSPR: la estructura molecular origina las propiedades. Si bien los métodos no-lineales mejoran mucho la calidad de las predicciones, resulta difícil en la práctica calibrar dichos modelos para que posean general aplicación sobre otras sustancias de estructura análoga a las modeladas. Además, sabemos que el uso de esquemas lineales posibilita interpretar de manera directa y sencilla la contribución de cada descriptor molecular a la propiedad estudiada.

No es cierta la afirmación de que cuando se establecen los modelos QSAR/QSPR siempre se encuentra una buena correlación entre las propiedades de interés y los descriptores moleculares. Por el contrario, en muchísimas situaciones no hay buena correlación (coeficiente de correlación lineal menor a 0.5), y debe buscarse la manera de encontrar un resultado de utilidad: aquí es donde la resolución del problema QSAR/QSPR requiere una mayor audacia e ingenio en el tratamiento de los datos, puesto que no se trata de manera alguna de una tarea trivial. Algunas posibilidades son: definir de manera diferente al dominio de aplicación del modelo; introducir nuevos descriptores moleculares

distintos a los empleados inicialmente; analizar mejor las predicciones de moléculas que tengan mayor error; proponer nuevos métodos de predicción de la propiedad; transformar la propiedad de variable continua a variable discreta, etc.

La aproximación QSAR/QSPR ha ido evolucionando a lo largo de los años, desde un modelo de regresión simple con pocas variables empíricas, es decir, modelos tipo Hansch [14], hasta transformarse en una herramienta realmente importante y que es aplicable a un amplio rango de problemas químicos, biológicos, medicinales y farmacológicos [4]. En la actualidad muchos investigadores de QSAR/QSPR tratan en general de sugerir mejoras en la teoría, tales como (i) proponer descriptores moleculares noveles y más poderosos, (ii) indicar estrategias noveles para derivar los modelos matemáticos QSAR/QSPR. El resultado de lo primero es la ya mencionada sobrea-bundancia de descriptores. Lo segundo conduce a un continuo mejoramiento de la metodología.

En conclusión, todo estudio QSAR/QSPR busca lograr establecer relaciones estructura-actividad/propiedad que resulten adecuadas para predecir propiedades de sustancias químicas, y asistir a los grupos experimentales en sus estudios; aplicar los modelos desarrollados para su propósito fundamental, que es la

predicción de estructuras químicas sin datos experimentales de su propiedad; con la ayuda de los paralelismos QSAR/QSPR, poder derivar alguna clase de interpretación a nivel molecular de los fenómenos estudiados y si fuera posible proveer alguna información de los mecanismos involucrados; demostrar que el avance logrado en los últimos años en el diseño de modelos QSAR/QSPR permitiría diseñar estructuras moleculares que conduzcan a propiedades específicas deseadas; comprensión detallada de las metodologías matemáticas predictivas que puedan llegar a emplearse a la hora de modelar cualquier tipo de propiedad fisicoquímica o biológica de sustancias; desarrollar herramientas alternativas para el análisis de datos y que sean de fácil aplicación, y que además complementen las metodologías estándares existentes en la literatura. Una meta de investigación importante hoy día es estudiar la manera de modelar nuevos sistemas químicos, tales como nanosistemas, líquidos iónicos, mezclas complejas, o macromoléculas complejas. Existe poca investigación difundida en la literatura sobre el tratamiento de este tipo de problemas mediante QSAR/QSPR.

El trabajo de investigación QSAR/QSPR desarrollado en los últimos años involucra el estudio de diferentes propiedades fisicoquímicas y biológicas, en diferente tipo de familias de

compuestos químicos. Para ello, la información experimental precisa de la propiedad medida en conjuntos moleculares homogéneos o heterogéneos se extrae de la literatura reciente, o por otro lado se utiliza información suministrada por grupos experimentales con los cuales se colabora en estos estudios. En estas bases de datos multidisciplinarias, se establecen modelos predictivos de utilidad para los fines específicos de cada problema estudiado. No obstante, la finalidad de todas estas investigaciones desarrolladas es siempre conseguir mejorar la metodología general de trabajo que permita perfeccionar los modelos matemáticos obtenidos, para así poder predecir cualquier propiedad/actividad imaginable en conjuntos moleculares de interés.

REFERENCIAS

- [1] C. Hansch, A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington D. C., 1995
- [2] T. Puzyn, J. Leszczynski, M. T. D. Cronin, *Recent Advances in QSAR Studies: Methods and Applications, Challenges and Advances in Computational Chemistry and Physics*. Springer, Netherlands, 2010
- [3] E. Benfenati, *Theory, guidance and applications on QSAR and REACH*, Orchestra, 2012.
- http://ebook.insilico.eu/insilico-ebook-orchestra-benfenati-ed1_rev-June2013.pdf

- [4] K. Roy, *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*, IGI Global, 2015
- [5] J. C. Dearden, *Int. J. Quant. Struct. Prop. Relat.* 2016, 1(1), 1
- [6] C. Hansch, T. Fujita, *J. Am. Chem. Soc.* 1964, 86, 1616
- [7] O. Ivanciuc, *Graph Theory in Chemistry and Drug Design*, CRC Press, New York, 2009
- [8] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley-VCH, Weinheim, 2009.
- [9] H. Wiener, *J. Am. Chem. Soc.* 1947, 69(1), 17
- [10] A. R. Katritzky, L. Mu, M. Karelson, *J. Chem. Inf. Comput. Sci.* 1997, 37, 756
- [11] A. R. Katritzky, U. Maran, M. Karelson, V. S. Lobanov, *J. Chem. Inf. Comput. Sci.* 1997, 37, 913
- [12] T. M. Nelson, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* 1994, 34, 601
- [13] B. E. Mitchell, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* 1996, 36, 58
- [14] H. Kubinyi, *QSAR: Hansch Analysis and Related Approaches*, Wiley-Interscience, New York, 2008
- [15] CODESSA-QSPR/QSAR Software, <http://www.codessa-pro.com>.

- [16] Dragon, https://chm.kode-solutions.net/products_dragon.php
- [17] PaDEL 2.20 (Pharmaceutical Data Exploration Laboratory), <http://www.yapcsoft.com>
- [18] H. Hong, Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins, W. Tong, *J. Chem. Inf. Model.* 2008, 48, 1337
- [19] Fragmentor, Laboratoire de Chémoinformatique, Chimie de la Matière Complexe (SMS UMR 7140), Université de Strasbourg, France
- [20] EPI Suite, U.S. EPA. <https://www.epa.gov/tsca-screening-tools/download-epi-suite-estimation-program-interface-v411>
- [21] CORAL, <http://www.insilico.eu/coral>
- [22] J. R. Valdés-Martini, Y. Marrero-Ponce, C. R. García-Jacas, K. Martínez-Mayorga, S. J. Barigye, Y. Silveira Vazd'Almeida, H. Pham-The, F. Pérez Giménez, C. A. Morell, *J. Cheminform.* 2017, 9, 35
- [23] Y. Martínez-López, Y. Marrero-Ponce, S. J. Barigye, E. Teran, O. Martínez-Santiago, C. H. Zambrano, F. J. Torres, *Mol. Divers.* 2020, 24, 913
- [24] H. Gonzalez-Díaz, C. R. Munteanu, *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*, Transworld Research Network, New York, 2010
- [25] J. Devillers, *Genetic Algorithms in Molecular Modeling (Principles of QSAR and Drug Design)*, Academic Press, New York, 2012
- [26] D. J. Livingstone, *Artificial Neural Networks: Methods and Applications (Methods in Molecular Biology)*, Humana Press, New York, 2009b
- [27] R. Brüggemann, L. Carlsen, *Partial Order in Chemistry and Environmental Sciences*, Springer, Heidelberg, 2006
- [28] N. R. Draper, H. Smith, *Applied Regression Analysis*, Third Edition, John Wiley&Sons Inc., New York, 1998
- [29] P. R. Duchowicz, E. A. Castro, F. M. Fernández, *MATCH Commun. Math. Comput. Chem.* 2006, 55, 179
- [30] OECD, <http://www.oecd.org/fr/securitechimique/risques/38130292.pdf>
- [31] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 2002, 20, 269
- [32] P. Gramatica, *QSAR Comb. Sci.* 2007, 26, 694
- [33] C. Rücker, G. Rücker, M. Meringer, *J. Chem. Inf. Model.* 2007, 47, 2345
- [34] K. Roy, S. Kar, P. Ambure, *Chemom. Intell. Lab. Syst.* 2015, 145, 22
- [35] K. Roy, R. N. Das, P. Ambure, R. B. Aher, *Chemom. Intell. Lab. Syst.* 2016, 152, 18