



QSAR en Toxicología: predicción de toxicidad basada en propiedades fisicoquímicas y estructurales

QSAR in Toxicology: toxicity prediction based on physicochemical and structural properties

DANIEL BUVAT DE VIRGINI^{1*}, JOSÉ M DUARTE-VIVAS^{2*}, JOSÉ R TORRES^{*}

Resumen

Ante la creciente demanda de nuevos fármacos y la búsqueda de estudios más eficientes, las herramientas informáticas han experimentado un auge en el campo de la toxicología. La generación de amplias bases de datos de índole clínica, biológica y química ha dado lugar al desarrollo de herramientas computacionales para el estudio toxicológico de diversos compuestos. Las relaciones cuantitativas de estructura y actividad (QSAR) permiten estudiar, mediante modelos matemáticos, la relación entre las características fisicoquímicas y la actividad biológica de diversos xenobióticos. En el presente trabajo se contemplan los principios fundamentales de estos análisis, los descriptores moleculares más utilizados y su dimensionalidad, así como la aplicabilidad de los modelos QSAR y de metodologías estadísticas para su uso como mecanismo de extrapolación a modelos *in vivo*.

Palabras clave: QSAR, toxicidad, predicción, estructura fisicoquímica, modelado

Abstract

With the growing demand for new drugs and the need for more efficient studies of them, computational tools have seen a boom in toxicology. The generation of extensive clinical, biological, and chemical databases has led to the formulation of computational tools for the toxicological study of various compounds. Quantitative structure-activity relationships (QSAR) allow, through mathematical models, the study of the relationship between the physicochemical characteristics and the biological activity of various xenobiotics. The present work considers the fundamental principles of these analyses, the most commonly used molecular descriptors and their dimensionality, the applicability of QSAR models, and statistical methodologies for their extrapolation to *in vivo* models.

Keywords: QSAR, toxicity, prediction, physicochemical structure, modeling

* Maestría de Toxicología, Facultad de Farmacia, Universidad Central de Venezuela. Caracas-Venezuela Correspondencia: todoposfarucv@gmail.com

Orcid: ¹[0009-0008-5359-7126](https://orcid.org/0009-0008-5359-7126)
²[0009-0006-0788-9772](https://orcid.org/0009-0006-0788-9772)

DOI: [10.54305/RFFUCV.2026.89.1.9](https://doi.org/10.54305/RFFUCV.2026.89.1.9)
Disponible: http://saber.ucv.ve/ojs/index.php/rev_ff

Recepción: 06/05/2026

Aprobación: 04/06/2026

Rev. Fac. Farmacia 89(1): 122-145. 2026

Introducción

La demanda global de fármacos ha aumentado de forma sostenida en los últimos años, consolidando al mercado farmacéutico como uno de los más grandes del mundo. En 2021, la industria invirtió 276.000 millones de dólares en investigación y desarrollo (I+D), equivalente al 27% de sus ingresos mundiales, con el fin de desarrollar nuevos fármacos eficaces y seguros (Garthwaite, 2025). El estudio toxicológico constituye una etapa esencial en la evaluación de compuestos con potencial terapéutico, ya que todo nuevo fármaco debe cumplir criterios toxicológicos que garanticen la seguridad del consumidor.

La toxicidad de un compuesto depende de múltiples factores, entre ellos la vía de exposición, la dosis, la duración de la exposición y diversas características físicas, bioquímicas y estructurales propias de cada molécula o familia química. Actualmente se emplean numerosos marcadores toxicológicos —como la carcinogenicidad, la genotoxicidad y la teratogenicidad— que pueden clasificarse cuantitativamente (LD50, NOAEL, LOAEL, entre otros) o cualitativamente (tóxico/no tóxico; toxicidad alta, media o baja). Muchos de estos parámetros se obtienen mediante ensayos en animales. Sin embargo, aunque los estudios *in vivo* siguen siendo un requisito regulatorio, implican un elevado consumo de tiempo y recursos y requieren un número significativo de animales, lo que plantea consideraciones bioéticas relevantes (Raies y Bajic, 2016; Forest y col., 2018).

En este contexto, los modelos computacionales se han convertido en herramientas valiosas para reducir la carga

experimental y mejorar la eficiencia en la predicción de efectos tóxicos, lo que se traduce en alternativas costo-efectivas (Raies y Bajic, 2016; Myatt y col., 2018; Li y col., 2022). Organismos internacionales como la Organización para la Cooperación y el Desarrollo Económico (OECD) recomiendan integrar herramientas bioinformáticas y computacionales como complemento de las evaluaciones toxicológicas *in vitro* e *in vivo*, con el fin de disminuir costos, reducir el uso de animales, optimizar las pruebas biológicas y acelerar el desarrollo de nuevos fármacos (OECD, 2013; De y col., 2022a,b).

Durante el último siglo, el uso de técnicas computacionales predictivas ha crecido exponencialmente (De y col., 2022a,b; Lu y col., 2023). Su adopción se ha visto favorecida por la disponibilidad de extensas bases de datos clínicas, biológicas y toxicológicas, que incluyen información sobre dianas moleculares, mecanismos de acción, rutas metabólicas y otros aspectos relevantes (Raies y Bajic, 2016; Lu y col., 2023). Estas aproximaciones se fundamentan en la relación entre las propiedades físicas, químicas y estructurales de las moléculas y la actividad biológica que ejercen.

Los primeros antecedentes se remontan a Richet (1893), quien correlacionó los efectos citotóxicos de compuestos orgánicos con su solubilidad en agua. Posteriormente, se reconoció la existencia de una relación entre la estructura molecular y la actividad biológica (SAR). Meyer (1899) describió una relación lineal entre la lipofiliidad de los anestésicos y su potencia narcótica, basada en los coeficientes de partición. Más adelante, Hansch y Fujita (1964) desarrollaron el primer modelo lineal multivariado que correlacionaba parámetros fisicoquímicos con la actividad biológica,

introduciendo formalmente el concepto de relación cuantitativa estructura-actividad (QSAR). En la actualidad, la toxicología regulatoria busca transitar de un enfoque empírico, basado en ensayo y error, a uno predictivo, sustentado en mecanismos y estructuras (Danishuddin y Khan, 2016; De y col., 2022a,b).

QSAR se define como un método para construir modelos matemáticos o computacionales que establecen correlaciones estadísticas entre características fisicoquímicas de las moléculas y su actividad, generalmente expresada mediante marcadores toxicológicos o biológicos cuantitativos (Verma y col., 2010). Esta metodología permite predecir la actividad biológica y la toxicidad de fármacos y xenobióticos no evaluados experimentalmente. La presente revisión sintetiza los principios fundamentales del QSAR y examina sus aplicaciones en la predicción de diversos efectos toxicológicos, así como las metodologías empleadas para su estimación.

Materiales y métodos

Se realizó una búsqueda exhaustiva y sistemática de estudios en inglés con acceso a texto completo y de acceso abierto en la plataforma PubMed, de evidencia científica de 2008 a 2025, representativos del contexto QSAR en toxicología, utilizando la herramienta de almacenamiento de artículos científicos como Rayyan. Se utilizaron términos como "Quantitative Structure-Activity Relationship (QSAR)", "Toxicity QSAR", "*In silico* toxicology", "Machine learning" y "Toxicity prediction". Bioinformatics"; "Toxicity with

Computational"; "Environmental toxicology and QSAR"; "Predictive -QSAR", "Prediction of biological activity", entre otros. La investigación, redacción y revisión del contenido se realizaron entre mayo de 2025 y mayo de 2026.

DESCRIPTORES MOLECULARES

La relación estructura actividad constituye el principio central que explica cómo las características moleculares determinan la interacción de un compuesto con biomoléculas y, en consecuencia, su efecto biológico o tóxico. Este fundamento sustenta los enfoques computacionales basados en ligandos, ampliamente utilizados para identificar características químicas críticas en el diseño de fármacos y en la evaluación toxicológica. La capacidad predictiva derivada de esta relación ha permitido anticipar la bioactividad química, propiedades físicas (Muratov y col., 2020), actividad antiproliferativa en células cancerosas (Gandhi y col., 2021), actividad antifúngica (Andrade-Ochoa y col., 2023), propiedades farmacéuticas y toxicológicas de pesticidas potenciales (Janicka y Sliwinska, 2022) y efectos de sensibilización cutánea (Scheufen y col., 2024), lo que confirma su relevancia transversal.

El QSAR se fundamenta en la premisa de que existe una relación cuantificable entre la estructura molecular y la actividad biológica o toxicológica de un compuesto (Muratov y col., 2020). Bajo esta hipótesis, incluso modificaciones estructurales mínimas pueden generar cambios significativos en la función biológica. Para operacionalizar esta relación, se emplean descriptores moleculares, variables cuantitativas que

representan propiedades fisicoquímicas, topológicas, termodinámicas, constitucionales, cuánticas, geométricas o estructurales (Beltrán Pérez y col., 2022; Shahlaei, 2013). Estos descriptores actúan como variables predictoras en modelos matemáticos, cuya selección depende de su relevancia para explicar la actividad dentro de una familia de compuestos. En consecuencia, los modelos QSAR suelen construirse con un número reducido de descriptores pertinentes para evitar redundancias y mejorar la interpretabilidad (Danishuddin y Khan, 2016).

El uso indiscriminado de múltiples descriptores, especialmente en modelos de aprendizaje automático, puede conducir al sobreajuste, lo que reduce la capacidad predictiva y diluye la relación causal entre la estructura y la actividad (Shi y col., 2011). Por ello, la selección racional de descriptores es esencial para generar modelos robustos, reproducibles y de alto valor predictivo (Shahlaei, 2013). La literatura también destaca que los modelos QSAR, como herramientas *in silico*, constituyen alternativas valiosas a la experimentación animal y proporcionan principios y recursos aplicables en diversas áreas biomédicas y toxicológicas (Madden y col., 2020).

Los descriptores se clasifican según su dimensionalidad, que refleja el nivel de información estructural incorporada. Tradicionalmente abarcan desde 0D hasta 4D, aunque algunos autores proponen extensiones hacia 5D (ajuste inducido) y 6D (solvatación), aún no universalmente aceptadas (Verma y col., 2010). Esta clasificación permite estructurar modelos predictivos con distintos grados de complejidad, acordes con la naturaleza del fenómeno biológico o toxicológico que se desea modelar.

Las dimensiones de los descriptores moleculares reflejan el nivel de información estructural necesario para construir modelos QSAR y permiten capturar distintos aspectos de la arquitectura química de un compuesto.

Los descriptores 0D representan propiedades globales independientes de la geometría molecular, como el número de átomos, el peso molecular o el número de enlaces. Aunque son simples, constituyen la base para la caracterización inicial de cualquier sustancia en QSAR (Verma y col., 2010; Muratov y col., 2020).

Los descriptores 1D se relacionan con grupos funcionales, fragmentos estructurales o propiedades fisicoquímicas básicas. Entre ellos destaca el coeficiente de partición logP, un parámetro clave para la actividad antifúngica y la absorción y distribución de fármacos (Andrade Ochoa y col., 2023; Muratov y col., 2020). Propiedades como la constante de acidez (K_a) también influyen en los procesos de absorción (Verma y col., 2010), y los compuestos nitroaromáticos han sido modelados con éxito mediante 1D QSAR (Kuz'min y col., 2008).

Los descriptores 2D describen la conectividad atómica y la topología molecular sin considerar la conformación tridimensional. Incluyen índices de conectividad y *fingerprints* o farmacóforos 2D, ampliamente utilizados para evaluar la similitud estructural y predecir la actividad en estudios de diseño de fármacos, incluidas aplicaciones en enfermedades neurodegenerativas (Verma y col., 2010; Makhouri y Ghasemi, 2022).

Los descriptores 3D incorporan la geometría espacial y las propiedades estéricas y electrostáticas, fundamentales

para comprender la interacción con dianas biológicas. Enfoques como el mapeo de farmacóforos 3D y HQSAR han demostrado su utilidad en la investigación biomédica, especialmente cuando la conformación molecular determina la afinidad ligando receptor (Muratov y col., 2020; Makhouri y Ghasemi, 2022; Verma y col., 2010).

Los descriptores 4D extienden los descriptores 3D al integrar la variabilidad conformacional y temporal, lo que permite modelar interacciones dinámicas. Aunque no se describen descriptores específicos, técnicas como el *docking* molecular se emplean para evaluar la nanotoxicidad y diseñar fármacos, lo cual se alinea con el concepto de descriptores dependientes del tiempo (Huang y col., 2021; Makhouri y Ghasemi, 2022; Muratov y col., 2020).

En conjunto, estos descriptores cuantifican propiedades moleculares clave para predecir la actividad biológica y la toxicidad. Entre los más influyentes destacan:

- Lipofilicidad (logP), esencial para los procesos ADME (Absorción, Distribución, Metabolismo y Excreción) (Andrade Ochoa y col., 2023; Muratov y col., 2020).
- Efectos estéricos, que determinan la complementariedad espacial con receptores, especialmente relevantes en modelos 4D QSAR (Muratov y col., 2020).
- Efectos electrónicos, que modulan la reactividad y la formación de aductos con biomoléculas (Muratov y col., 2020).
- Farmacóforos y toxicóforos, que representan patrones estructurales asociados a la actividad o la toxicidad (Makhouri y Ghasemi, 2022; Muratov y col., 2020).

La integración de estos descriptores mediante algoritmos matemáticos permite desarrollar modelos QSAR robustos para predecir el comportamiento toxicológico y farmacológico de nuevas sustancias, lo que consolida su papel en el diseño racional de fármacos y en la toxicología predictiva (Huang y col., 2021; Makhouri y Ghasemi, 2022; Muratov y col., 2020).

El desarrollo de modelos QSAR representa un pilar fundamental en la toxicología y el diseño de fármacos modernos, al permitir transformar datos químicos y biológicos en predicciones robustas y confiables sobre el comportamiento de nuevas moléculas (Ren y col., 2023). Este enfoque computacional no solo agiliza la evaluación de la seguridad, sino que también contribuye a reducir la experimentación animal y a optimizar recursos, elementos cruciales en la investigación actual (Mei y Wu, 2022). La metodología para construir modelos QSAR es un proceso estructurado y riguroso cuyo objetivo es garantizar que las predicciones sean precisas, interpretables y aplicables a nuevos compuestos. Aun así, una alta predicción de los modelos puede lograr una alta precisión de interpretación, pero la alta capacidad predictiva de los modelos no garantiza una alta precisión de interpretación. Es aquí donde radica la importancia de mejorar las investigaciones, los métodos y los modelos, ya que este es el futuro como herramienta necesaria para la ciencia y la regulación (Matveieva y Polishchuk, 2021).

SELECCIÓN Y EXTRACCIÓN DE DESCRIPTORES

La calidad y el poder predictivo de un modelo QSAR dependen directamente de la representación adecuada de las moléculas mediante sus descriptores. La extracción de estos descriptores es un proceso que, mediante herramientas computacionales especializadas, puede generar miles de posibles características para cada compuesto. No obstante, la generación masiva de descriptores a menudo se traduce en un gran número de variables que pueden ser irrelevantes o altamente correlacionadas, lo que introduce ruido y redundancia en el conjunto de datos y podría comprometer la estabilidad y la interpretabilidad del modelo (Playe y Stoven, 2020). Por ello, la selección de descriptores constituye una etapa crítica y esencial. Su objetivo principal es identificar un subconjunto óptimo de descriptores relevante para la actividad o toxicidad estudiada y que, al mismo tiempo, evite la redundancia y capture la información esencial de la molécula (Dablander y col., 2023).

Para reducir la dimensionalidad de los datos, se emplean ampliamente técnicas como el Análisis de Componentes Principales (PCA), que transforman un elevado número de descriptores correlacionados en un conjunto menor de variables no correlacionadas, denominadas componentes principales (Wang y col., 2015). Asimismo, métodos iterativos como la selección hacia adelante/atrás añaden o eliminan descriptores en función de su impacto en la capacidad predictiva del modelo, mientras que los algoritmos genéticos exploran combinaciones complejas de descriptores para optimizar el rendimiento (Nha Tran y col., 2023). Así, como ya hemos mencionado, el fin último

de esta fase es maximizar la predictividad y la interpretabilidad del modelo, previniendo el sobrefit (Belfield y col., 2023).

MÉTODOS DE MODELADO

Una vez definido el conjunto óptimo de descriptores, el siguiente paso consiste en seleccionar y aplicar un algoritmo de modelado que establezca una relación predictiva entre estos descriptores y la propiedad biológica o molecular de interés (Cronin y col., 2025). Los métodos de modelado quimiométrico pueden clasificarse, a grosso modo, en dos vertientes: modelos lineales y no lineales. Algunos de los sistemas de modelado matemático son mencionados en la Tabla I; sin embargo, se mencionarán algunos otros:

- *Regresión lineal simple*: Este método predice la probabilidad de que una instancia pertenezca a una clase específica, transformando una combinación lineal de los descriptores en una probabilidad, a menudo mediante una función logística (Liu y col., 2024). Es simple y eficiente para problemas de clasificación binaria.
- *Regresión Lineal Múltiple (MLR)*: Este es un método sencillo y altamente interpretable que asume una relación lineal directa entre los descriptores y la propiedad a predecir. Su simplicidad facilita la identificación de los descriptores que ejercen un impacto directo y proporcional sobre la respuesta biológica (Tuenter y col., 2017).
- *Los mínimos cuadrados parciales (PLS)*: son una técnica de aprendizaje

Tabla I.
Sistemas de modelado matemático (Lin y Chou, 2022)

Redes neuronales artificiales	
Redes neuronales de escapatización	Todas las neuronas se dividen en tres capas, con información que fluye de la primera capa de neuronas de entrada a la segunda capa de neuronas ocultas, y luego a la tercera capa de neuronas de salida
Redes neuronales regularizadas en bayesiano	Aplicar métodos bayesianos para realizar la regularización de manera que la complejidad del modelo se equilibre con la precisión de reproducir los datos de entrenamiento
Redes neuronales asociativas	Aplicar ensamblaje aprendiendo a las redes neuronales de la propaganda
Redes neuronales profundas	Redes neuronales artificiales con múltiples capas ocultas (también llamadas aprendizaje profundo)
Métodos no supervisados	
Análisis de componentes de principio	Reducir la dimensionalidad de los datos a sólo los primeros componentes principales, preservando al mismo tiempo la mayor cantidad posible de la variación de los datos
Mapas de autoorganización de Kohonen	Mapea moléculas del espacio original del descriptor de una rejilla 2D de neuronas. Moléculas similares serán mapeadas a las mismas neuronas estrechamente localizadas en la red

ampliamente utilizada; combina las ventajas de la integración del análisis de componentes principales y de la regresión por mínimos cuadrados. Se utiliza para predecir una tabla de datos a partir de otra (de Souza y col., 2019).

- *Redes neuronales artificiales (ANN)*: Inspiradas en la estructura y el funcionamiento del cerebro humano, las ANN poseen la capacidad de capturar relaciones no lineales extremadamente complejas entre los descriptores y la actividad biológica (Fattouche y col., 2024). Sin embargo, su arquitectura multicapa y compleja las convierte en “cajas negras”, lo que dificulta interpretar cómo llegan a sus

predicciones (Matveieva y Polishchuk, 2021).

- *Máquinas de vectores de soporte (SVM)*: Las SVM son particularmente eficaces en problemas de alta dimensionalidad y pueden manejar relaciones no lineales al proyectar los datos a un espacio de mayor dimensión, donde la separación entre clases o la regresión es más sencilla. Son notablemente robustas frente al ruido y a la presencia de características irrelevantes (Hatmal y col., 2021).
- *Algoritmos de Clasificación Categórica*: Cuando el objetivo del modelo es predecir si un compuesto

pertenece a una categoría específica (por ejemplo, "tóxico" o "no tóxico", "activo" o "inactivo"), se emplean algoritmos de clasificación categórica (Bishop y col., 2024). Estos modelos no emplean metodologías avanzadas de análisis debido a su carácter categórico; sin embargo, sólo pueden emplearse cuando se tienen bien definidos los cortes predictivos por variables y los descriptores necesarios para ello, por lo que se requiere un procesamiento estadístico previo.

- *Máquinas de Vectores de Soporte para Clasificación (SVMs)*: Es la versión de SVM adaptada para problemas de clasificación, que busca el hiperplano óptimo que maximice el margen entre las diferentes clases de datos, lo que se traduce en una separación robusta (Liu y col., 2024).
- *Árboles de Decisión y Bosques Aleatorios (Random Forests)*: Los Árboles de Decisión son métodos no paramétricos que construyen modelos basados en reglas de decisión simples y son inherentemente interpretables. Los Bosques Aleatorios mejoran la robustez y la precisión al combinar múltiples árboles de decisión y promediar sus predicciones, lo que los hace eficientes para manejar grandes conjuntos de datos y revelar relaciones complejas de forma más interpretable que las ANN (Liu y col., 2024).

VALIDACIÓN DE MODELOS

La validación es la fase más crítica en el ciclo de vida de un modelo QSAR, ya que es el proceso que garantiza su fiabilidad, robustez y, crucialmente, su capacidad

para predecir con precisión la actividad de compuestos que no han sido vistos durante el entrenamiento (De y col., 2022a,b). Se deben realizar diversos tipos de validación.

Primero, se realiza una validación interna que consiste en evaluar la consistencia y la estabilidad del modelo con los datos de entrenamiento; es decir, su capacidad para reproducir los datos con los que fue formulado el modelo o con los que fue construido. El método más común para la validación interna es la Validación Cruzada (*Cross-validation*), que implica dividir el conjunto de datos original en subconjuntos. En el enfoque *Leave-One-Out (LOO)*, el modelo se entrena con todos los datos excepto uno, que se utiliza para la evaluación; este proceso se repite para cada punto de datos. Alternativamente, en la validación *k-fold*, los datos se dividen en "k" subconjuntos (*folds*); el modelo se entrena "k" veces, utilizando en cada iteración un subconjunto diferente como conjunto de prueba y los restantes para el entrenamiento. La premisa de esta validación interna consiste en evaluar la bondad de la aptitud y la robustez (Király y col., 2022).

El otro tipo de validación que se realiza es la validación externa. Esta se considera el "estándar de oro" en la evaluación de modelos QSAR. La validación externa es indispensable para evaluar la capacidad real de predicción y la capacidad de generalización del modelo en compuestos completamente nuevos que el modelo no ha "visto" en ninguna etapa previa. Para llevar a cabo esta validación, se reserva un conjunto de pruebas independiente que no se utiliza en ninguna fase de la construcción del modelo, incluida la selección de descriptores o el ajuste de parámetros.

Un modelo se considera verdaderamente robusto y confiable si mantiene un buen rendimiento no solo en la validación interna, sino, de manera crucial, también en la validación externa, lo que demuestra su aplicabilidad a nuevas estructuras químicas. Para la validación externa, tendríamos que evaluar la capacidad predictiva de los modelos. La OECD, con respecto a la implementación de buenas prácticas para la validez de los modelos, hace énfasis en 5 principios que deben cumplirse: 1) *Endpoints* o descriptores con valor predictivo, 2) algoritmo inequívoco definido, 3) dominio definido de aplicabilidad, 4) medidas apropiadas de bondad de ajuste, robustez y predictividad, 5) interpretación mecanicista, dentro de la posibilidad (Király y col., 2022).

MÉTRICAS DE VALIDACIÓN

La validación de un modelo estadístico es un requisito indispensable para garantizar su confiabilidad y aplicabilidad. En QSAR, esta validación se fundamenta en métricas que permiten evaluar tanto el ajuste interno como la capacidad predictiva externa del modelo. La regresión constituye una herramienta central para relacionar la estructura molecular con la actividad biológica o fisicoquímica, facilitando la predicción del comportamiento de nuevos compuestos y acelerando el desarrollo de fármacos y la evaluación de riesgos químicos (Wunsch y col., 2021).

Entre los indicadores más utilizados se encuentra el coeficiente de determinación (R^2), empleado en la validación externa para evaluar la capacidad predictiva del modelo. Aunque valores superiores a

0,60 se consideran aceptables, el R^2 por sí solo no es suficiente; la comparación entre los valores predichos y observados en un conjunto de datos independiente es esencial para robustecer y generalizar el modelo (Qin y col., 2017).

Los modelos de clasificación categórica también resultan relevantes para priorizar compuestos según su actividad biológica, al establecer correlaciones entre propiedades moleculares y efectos biológicos. Esta aproximación permite evaluar de manera eficiente grandes volúmenes de sustancias químicas (Kim y col., 2021).

Asimismo, métricas como la sensibilidad y la especificidad proporcionan información crítica sobre el rendimiento del modelo. La sensibilidad refleja la capacidad de identificar correctamente los compuestos activos (Khan y col., 2024), mientras que la especificidad indica la correcta clasificación de los compuestos inactivos. Un equilibrio adecuado entre ambas es fundamental para garantizar la precisión y la fiabilidad del modelo QSAR (Shiri y col., 2016). El valor predictivo positivo (VPP) también se utiliza, aunque puede mostrar variabilidad significativa (40–82%) según el modelo y los datos, lo que subraya la importancia de evaluar cuidadosamente la precisión global (Muto y col., 2024).

Otros indicadores complementarios fortalecen la validación. La exactitud (accuracy) es esencial para garantizar la reproducibilidad y la validez de las predicciones, especialmente en modelos aplicados a compuestos con potencial terapéutico. El área bajo la curva (AUC; *Area Under the Curve*) de la curva ROC (*Receiver Operating Characteristic*) es un índice que

mide qué tan bien un modelo es capaz de discriminar entre dos clases, por ejemplo: tóxico vs. no tóxico, activo vs. inactivo, etc. En los modelos QSAR, es uno de los criterios de desempeño más utilizados porque resume la capacidad predictiva global del modelo en un único valor. Permite evaluar la capacidad discriminativa del modelo, con valores entre 0,78 y 0,90 considerados sólidos en QSAR (Wei y col., 2019).

El Dominio de Aplicabilidad (AD) define el espacio químico en el que las predicciones del modelo son confiables. Su delimitación es crucial para evitar extrapolaciones imprecisas y asegurar la aceptación regulatoria de los modelos (Greco y col., 2025; Gadaleta y col., 2022). La validación externa mediante métricas como el coeficiente de correlación de concordancia (CCC), la detección de datos atípicos y el análisis de desviaciones estándar permite clasificar el desempeño del modelo como bueno, moderado o deficiente, garantizando su uso responsable y científicamente sólido (Shayanfar y Shayanfar, 2022).

En síntesis, sin una validación rigurosa, incluso los modelos QSAR más sofisticados carecerían de credibilidad científica y de aplicabilidad práctica.

Aplicaciones del QSAR para la determinación de parámetros toxicológicos

Los avances en toxicología predictiva han permitido desarrollar modelos QSAR altamente confiables para evaluar la toxicidad aguda, tanto en mamíferos como en anfibios (DL50). Para la DL50 oral, enfoques innovadores como el sistema de

acceso molecular con redes neuronales artificiales (MACCS-ANN) combinan huellas moleculares con redes neuronales, logrando una excelente capacidad predictiva ($Q^2 = 0,821$, $AUC > 0,9$) e identificando alertas estructurales clave (nitrilos, cloroalquilos) mediante la minería de subestructuras (MoSS). Estos modelos superan a los métodos tradicionales, como el modelo de algoritmo Genético-Regresión Lineal Múltiple (GA-MLR), al integrar parámetros moleculares fundamentales (polarizabilidad, potencial de ionización) y ofrecer una clasificación semicuantitativa, lo que reduce la dependencia de los ensayos en animales. En paralelo, para anfibios, donde los datos son escasos, modelos QSAR basados en regresión y validados mediante Monte Carlo han demostrado robustez ($R^2 = 0,96$ en validación) al predecir la DL50-12 h en Rana japónica, destacando el papel potenciador de los grupos nitro y halógenos en la toxicidad. Ambos casos subrayan el valor de los QSAR para priorizar compuestos en etapas tempranas, con aplicaciones que van desde el diseño de moléculas más seguras hasta la conservación de especies sensibles (Fan y col., 2018; Toropov y col., 2022).

TOXICIDAD CRÓNICA Y SUBCRÓNICA

Los modelos QSAR basados en técnicas avanzadas, como el *Deep Learning*, están revolucionando la estimación de umbrales toxicológicos clave, como el NOAEL (*No Observed Adverse Effect Level*) y el LOAEL (*Lowest Observed Adverse Effect Level*), al identificar patrones ocultos en las propiedades estructurales y biológicas de los compuestos. Estos enfoques predictivos permiten determinar dosis seguras (NOAEL) y los puntos de inicio de toxicidad (LOAEL) sin depender exclusivamente de ensayos

in vivo, optimizando así la evaluación de riesgos mediante un enfoque más rápido y ético. Al correlacionar características moleculares específicas con respuestas biológicas, los modelos no solo reducen la necesidad de experimentación animal, sino que también proporcionan insights mecanísticos sobre qué atributos químicos desencadenan efectos adversos, lo que facilita el diseño de compuestos más seguros y la priorización de sustancias en etapas tempranas del desarrollo (Selvestre y col., 2022).

MUTAGENICIDAD Y CARCINOGENICIDAD

Los modelos *in silico*, como las relaciones cuantitativas estructura actividad (QSAR), permiten predecir la actividad biológica de sustancias químicas a partir de su estructura molecular. Entre ellos, los modelos Ames/QSAR, basados en datos del ensayo de Ames, se han consolidado como herramientas clave para predecir la mutagenicidad de impurezas presentes en productos farmacéuticos, pesticidas y metabolitos (Furuhama y col., 2023).

La confiabilidad de estos modelos depende en gran medida de la calidad de los datos químicos y biológicos utilizados, así como de criterios adicionales de validación que aseguren la solidez de las predicciones (Chung y col., 2023). En toxicología regulatoria, la predicción de la mutagenicidad y la carcinogenicidad es fundamental, y los modelos QSAR han demostrado un valor significativo al identificar fragmentos estructurales asociados al daño genético. En este contexto, los sistemas de predicción basados en la prueba de Ames han mostrado una notable precisión, especialmente cuando

se sustentan en bases de datos amplias y en conocimiento experto (Honma y col., 2019). Además, el uso de baterías de modelos, que integran múltiples QSAR, mejora la robustez y la confiabilidad de las predicciones en comparación con los modelos individuales.

La aplicabilidad de QSAR se extiende más allá de la mutagenicidad y abarca diversos escenarios toxicológicos. La Agencia de Protección Ambiental de los Estados Unidos (EPA) ha sido pionera en la adopción de estos enfoques, desarrollando modelos QSAR para una amplia gama de sustancias químicas y utilizándolos en procesos regulatorios (Muto y col., 2024). Esta integración contribuye a acelerar la evaluación de riesgos y a reducir el uso de animales de experimentación, en consonancia con los principios de las 3Rs. Paralelamente, la investigación continúa perfeccionando estos modelos mediante descriptores moleculares más sofisticados que capturan con mayor precisión la relación entre la estructura química y la actividad biológica (Furuhama y col., 2023). No obstante, la utilidad de un modelo QSAR depende de su aplicación dentro de un dominio de aplicabilidad claramente definido, que delimita el espacio químico en el que las predicciones son confiables. La transparencia en la construcción, validación y definición de las limitaciones del modelo es esencial para su correcta interpretación y aceptación en la evaluación de la seguridad química (Bossa y col., 2021).

ENFOQUE RELACIONADO A LA TOXICIDAD REPRODUCTIVA Y DEL DESARROLLO

La capacidad de una sustancia para interferir con la función reproductiva

o para dañar a un embrión o feto en desarrollo es una de las preocupaciones más serias en toxicología. Los modelos QSAR están emergiendo como herramientas prometedoras para la identificación temprana de estos peligros, lo que permite una intervención proactiva para proteger la salud reproductiva. Por lo tanto, predecir la toxicidad reproductiva y de desarrollo es un desafío complejo, ya que implica una multitud de procesos biológicos. Sin embargo, los modelos QSAR pueden abordar esta complejidad mediante el uso de descriptores moleculares que capturan las propiedades fisicoquímicas y estructurales que se sabe que están asociadas con estos efectos adversos (von Hellfeld y col., 2023).

La integración de datos de diferentes fuentes, como la biblioteca Tox21 Data Challenge 2014 y Toxi 21 10K, que incluyen ensayos *in vitro* y datos *in silico* (computacionales, como algoritmos como XGBoost), es clave para construir modelos QSAR robustos que permiten predecir cómo estos químicos interfieren. Este enfoque, conocido como "Toxicología del Siglo XXI", busca reducir la dependencia de los ensayos con animales y avanzar hacia una evaluación de la seguridad más eficiente y basada en los mecanismos de toxicidad en cuanto a la identificación de químicos que podrían potencialmente dirigirse a distintas vías u objetivos biológicos (*Endpoints*) y interactuar de modo agonista o antagonista en los distintos receptores nucleares, receptores endocrinos, proteínas, citoquinas en la vías de respuestas al estrés dentro del cuerpo humano y conducir a la toxicidad (Banerjee y col., 2016; Kurosaki y col., 2020).

ANÁLISIS TOXICOCINÉTICOS MEDIANTE QSAR

Los modelos *in silico*, como las relaciones cuantitativas estructura actividad (QSAR), han adquirido un papel central en la toxicocinética por su capacidad para acelerar la evaluación de nuevos compuestos, reducir costos y mejorar la seguridad de los medicamentos y las sustancias químicas. Su utilidad radica en la predicción confiable de propiedades ADME, lo que permite identificar perfiles toxicológicos potenciales desde etapas tempranas del desarrollo (Siramshetty y col., 2021).

En la fase de absorción, los QSAR permiten estimar la biodisponibilidad al identificar propiedades que favorecen el paso a través de las membranas biológicas. Para la distribución, predicen la unión a proteínas plasmáticas y la posible acumulación en órganos específicos, factores determinantes de la eficacia y la toxicidad del compuesto. En la metabolización, ayudan a anticipar rutas metabólicas y la formación de metabolitos activos o tóxicos, mientras que en la eliminación permiten estimar la tasa de excreción y el riesgo de acumulación (Rovida y col., 2020).

La metodología QSAR integra algoritmos matemáticos que correlacionan descriptores moleculares —físico-químicos, topológicos y electrónicos— con actividades biológicas o con parámetros farmacocinéticos. Su aplicación en estudios ADME permite predecir el comportamiento de moléculas candidatas antes de su síntesis, optimizando recursos y reduciendo la necesidad de ensayos experimentales.

Entre las técnicas más empleadas destacan el 2D QSAR, basado en descriptores bidimensionales y métodos estadísticos; el 3D QSAR, que incorpora estereoquímica mediante enfoques como CoMFA y CoMSIA; el *docking* molecular, que simula las interacciones ligando-receptor; y los modelos predictivos de ADME, que evalúan la absorción intestinal, el metabolismo hepático (CYP450), la excreción y la toxicidad (Abdullahi y col., 2022).

La confiabilidad de estos modelos exige una validación rigurosa mediante métricas como $R^2 > 0,6$, $Q^2 > 0,5$, un RMSE bajo y el cumplimiento de los principios de la Organización para la Cooperación y el Desarrollo Económicos (OCDE): endpoint definido, algoritmos reproducibles, dominio de aplicabilidad delimitado y validación externa. Herramientas como SwissADME permiten verificar las reglas de filtrado farmacocinético (Lipinski, Ghose, Veber, Egan y Muegge), asegurando que los compuestos candidatos posean propiedades óptimas para la biodisponibilidad oral (Abdullahi y col., 2022).

En paralelo, se han desarrollado modelos globales de QSPR basados en aprendizaje automático para predecir propiedades ADME mediante algoritmos de aprendizaje multitarea. Estos modelos integran múltiples ensayos —permeabilidad (Papp, MDCK, PAMPA, Caco 2), *clearance* metabólico (CLint), unión a proteínas plasmáticas (PPB), lipofilicidad ($\log P$, $\log D$) e inhibición de CYP450— y emplean arquitecturas como MPNN (*Message Passing Neural Networks*) y DNN (*Deep Neural Networks*) para capturar patrones complejos. En efecto, las arquitecturas MPNN y DNN son modelos de aprendizaje profundo capaces de aprender patrones complejos: las DNN mediante representaciones jerárquicas de

datos estructurados o continuos, y las MPNN mediante el intercambio de información entre los nodos de un grafo, lo que permite capturar relaciones topológicas, como las presentes en estructuras moleculares. Su validación temporal y su aplicación a degradadores dirigidos a proteínas (TPDs) demuestran su utilidad incluso en contextos con datos limitados (Peteani y col., 2024).

En conjunto, la integración estratégica de QSAR, QSPR, aprendizaje automático y validación prospectiva constituye un marco robusto para la predicción de propiedades ADME y la toma de decisiones tempranas en el desarrollo farmacéutico, lo que aumenta la probabilidad de éxito traslacional y minimiza los riesgos en las etapas preclínicas.

USO DE QSAR EN ECOTOXICOLOGÍA

En el marco de la evaluación y predicción de la seguridad química, múltiples organismos internacionales —como la EPA, la OECD, la Autoridad Europea de Seguridad Alimentaria (EFSA), la Agencia Europea de Sustancias y Mezclas (ECHA) y la Administración de Alimentos y Medicamentos (FDA) — han impulsado directrices y herramientas orientadas a reducir el impacto ambiental y a proteger los ecosistemas acuáticos y terrestres. Entre estas herramientas destacan bases de datos como OpenFoodTox y ChemPortal, así como plataformas *in silico* como OECD Toolbox y AMBIT 2.0, que integran información toxicológica y permiten aplicar enfoques basados en el umbral de preocupación toxicológica (TTC) a compuestos relacionados con los alimentos. Estas iniciativas proporcionan un marco esencial para la ecotoxicología, facilitando el estudio de los efectos de las sustancias

químicas en organismos y ecosistemas completos (Dorne y col., 2021).

En este contexto, los modelos QSAR han adquirido un papel transformador en la predicción de la toxicidad ambiental. Su capacidad para estimar la toxicidad de sustancias en organismos como algas, crustáceos, peces y otras especies acuáticas permite anticipar riesgos, establecer límites de vertido y promover el diseño de compuestos más seguros (Chung y col., 2023). La lipofilia, un descriptor clave, es particularmente relevante en ecotoxicología debido a su relación con la bioacumulación y la persistencia en tejidos biológicos, factores determinantes del potencial tóxico de un compuesto (Bunmahotama y col., 2022).

La investigación en ecotoxicología continúa expandiéndose, lo que impulsa el desarrollo de modelos QSAR más precisos y mecanísticos. Se están explorando enfoques que incorporan mecanismos de acción específicos, como los asociados a los insecticidas, lo que mejora la capacidad predictiva y la relevancia biológica de los modelos. Asimismo, se están desarrollando modelos para predecir la toxicidad de mezclas químicas, un avance crucial dado que los organismos suelen estar expuestos simultáneamente a múltiples contaminantes en condiciones reales (Xu y col., 2023).

El uso de QSAR en ecotoxicología no solo fortalece la evaluación de riesgos, sino que también contribuye al diseño de sustancias químicas más seguras desde su concepción, en línea con los principios de la química verde. Al anticipar la toxicidad antes de la síntesis, es posible evitar el desarrollo de compuestos con efectos adversos sobre el medio ambiente, lo que

promueve prácticas más sostenibles y responsables (Bueso y col., 2024).

TOXICIDAD DE NANOPARTÍCULAS Y BIOMATERIALES

El rápido avance de la nanotecnología ha impulsado el desarrollo de nanomateriales, nanopartículas y biomateriales con propiedades fisicoquímicas únicas — tamaño, forma, área superficial, reactividad— que ofrecen aplicaciones prometedoras en medicina, energía y electrónica. Sin embargo, estas mismas características plantean desafíos significativos para evaluar su toxicidad y garantizar su seguridad ambiental y sanitaria. En este contexto, los modelos QSAR adaptados a nanomateriales, conocidos como QSAR nano o QNAR, han emergido como herramientas esenciales en la informática de nanopartículas, apoyados por algoritmos de inteligencia artificial que evolucionan rápidamente para abordar esta complejidad (Muratov y col., 2020; Bossa y col., 2021).

A diferencia de los compuestos químicos tradicionales, la toxicidad de las nanopartículas depende no solo de su composición, sino también de propiedades físicas como el tamaño, la forma, la cristalinidad y las características superficiales. Estas propiedades influyen en su capacidad de penetrar en las células, de interactuar con biomoléculas y de causar daño, lo que exige la incorporación de descriptores específicos en los modelos QNAR (Bossa y col., 2021).

El campo se encuentra en constante evolución, con esfuerzos orientados a desarrollar modelos más precisos y mecanísticos. Un ejemplo reciente es el modelo nano qRASTR, diseñado para

predecir la toxicidad de nanopartículas de óxidos metálicos empleadas en aplicaciones médicas y ambientales. Este enfoque considera fenómenos cuánticos que modifican la reactividad superficial de partículas extremadamente pequeñas, lo cual puede influir en su capacidad para inducir estrés oxidativo y otros efectos tóxicos (Kar y col., 2024).

La comprensión de la nanotopografía y de las propiedades superficiales de los nanomateriales es fundamental para interpretar su comportamiento biológico. Estas propiedades determinan su biodistribución, degradación, biocompatibilidad y potencial tóxico. Por ello, el estudio de nanopartículas requiere integrar la química cuántica, la física de superficies, los ensayos experimentales, las simulaciones computacionales y la catálisis. Este enfoque multidisciplinario es indispensable para avanzar hacia el diseño seguro de nanomateriales en una era en la que la nanotecnología tiene un impacto creciente en múltiples sectores (Gulumian y col., 2021).

QSAR EN REACH

Los QSAR se utilizan en REACH (*Registration, Evaluation, Authorisation and Restriction of Chemicals*) (Registro, Evaluación, Autorización y Restricción de Sustancias Químicas) para predecir diversas propiedades peligrosas de los compuestos, como la toxicidad aguda, la carcinogenicidad, la mutagenicidad y los efectos sobre la reproducción. Esto permite a las empresas realizar evaluaciones de riesgo sin necesidad de ensayos en animales, lo cual se alinea con las políticas de protección animal y reduce costos y tiempos (Huang y col., 2021). Además,

el uso de QSAR en REACH requiere que los modelos sean científicamente válidos, transparentes y bien documentados, lo que garantiza que las predicciones sean confiables y aceptadas por las autoridades regulatorias. Así, los QSAR constituyen una herramienta clave para cumplir con los requisitos regulatorios, facilitar el registro de sustancias químicas y promover la gestión de riesgos en la industria química de manera ética y eficiente (Huang y col., 2021).

QSAR EN LA FDA

La *Food and Drug Administration* (FDA) ha reconocido el valor de los modelos QSAR como herramientas aceptables para evaluar la seguridad de los compuestos en ciertos contextos, especialmente en las industrias farmacéutica y alimentaria, siempre que se cumplan requisitos estrictos de validez y transparencia. Esto está respaldado por regulaciones y guías oficiales, como la *Guidance for Industry: Use of Quantitative Data in Drug Development* y las directrices de *Toxicology in the 21st Century* (Tox21), que promueven el uso de métodos salubres y basados en el conocimiento científico (Wang y col., 2024).

Para que un modelo QSAR sea aceptado, debe cumplir con los criterios de las OECD *Principles of Good Laboratory Practice* (GLP) y también debe ser validado según los criterios establecidos en el QMRF (*Quantitative Structure-Activity Relationship Model Reporting Format*). La FDA requiere que los datos predictivos estén bien justificados y acompañados de documentación que explique la solidez del modelo, incluyendo las limitaciones y la calidad de los datos de entrada y de predicción (Wang y col., 2024).

QSAR EN EPA

El principal marco legal que respalda el uso de modelos QSAR en la evaluación de sustancias químicas es la *Toxic Substances Control Act* (TSCA), que faculta a la EPA (*Environmental Protection Agency*, Agencia de Protección Ambiental de los Estados Unidos) para evaluar la toxicidad de compuestos sin requerir ensayos de laboratorio en todos los casos. En este contexto, la agencia promueve el empleo de metodologías computacionales—incluidos los QSAR— conforme a sus *Guidelines for Chemical Testing*, entre ellas la *Guideline for Data Quality and Predictive Models*, que establece que los modelos predictivos pueden complementar o sustituir estudios experimentales cuando son científicamente válidos, transparentes y reproducibles (Browne y col., 2022).

El enfoque Read-Across (RAx) permite extrapolar datos toxicológicos de sustancias análogas a compuestos con vacíos de información, lo que reduce la necesidad de ensayos *in vivo*. Su aplicabilidad es objeto de investigación por organismos como la OCDE, mientras que marcos regulatorios como EU REACH ya contemplan su uso, junto con acuerdos de cooperación con la EPA, el gobierno de Canadá y otras jurisdicciones. Su aceptación regulatoria depende de optimizar elementos críticos, como la definición de la similitud, la interpretación de la toxicidad y las diferencias en los parámetros de ADME. En esta línea, el proyecto EU ToxRisk desarrolla guías para fortalecer las Nuevas Metodologías de Abordaje (NAMs), estandarizar criterios y validar el procedimiento (Rovida y col., 2020). Asimismo, la EPA ha adoptado los Principios de la OCDE para la validación de modelos QSAR, que exigen transparencia en

la descripción del modelo, la justificación del dominio de aplicabilidad y la validación frente a datos experimentales independientes. La agencia también mantiene listados de modelos aceptados, como QSAR Toolbox, utilizados para predecir toxicidad, persistencia, bioacumulación y otros efectos adversos relevantes (Browne y col., 2022).

Desafíos y limitaciones del QSAR en toxicología de calidad y disponibilidad de datos

A pesar de sus avances, el QSAR se enfrenta a obstáculos que definen las futuras áreas de investigación y requieren una aplicación responsable. La calidad y disponibilidad de datos persisten como la principal limitación, ya que la escasez de datos toxicológicos de alta calidad y consistencia, especialmente para nuevos compuestos o efectos crónicos, puede conducir a modelos erróneos o no predictivos (Shkil y col., 2024).

A pesar del gran avance y descubrimientos y mejoras en los modelos, la complejidad de los mecanismos de toxicidad representa otro desafío significativo, la importancia en la selección rigurosa de los conjuntos de datos y descriptores utilizados para la formación y validación de modelos es lo que permite garantizar el desarrollo de modelos QSAR, dado que muchos efectos tóxicos resultan de interacciones moleculares y múltiples vías biológicas que no se ajustan a una simple relación estructura-actividad. En este contexto, la predicción de la toxicidad de las mezclas sigue siendo particularmente compleja. El Dominio de Aplicabilidad (AD) es el mayor obstáculo para la extrapolación, lo que limita la confianza en las predicciones para compuestos fuera de su rango. Existen

métodos que determinan el dominio de aplicabilidad, su validación se encuentra en estudio y las recomendaciones son difíciles de encontrar (Héberger, 2023).

La interpretación de modelos complejos (“caja negra”), como las redes neuronales avanzadas, puede resultar limitada, lo que dificulta la comprensión de las relaciones causales subyacentes. Actualmente, la investigación se enfoca en métodos de interpretabilidad, sobre todo en *Shapley Additive Explanations* (SHAP), para “abrir” estas cajas negras, lo que requiere una fusión multimodelo para mejorar la precisión y la capacidad predictiva de la actividad biológica (Dong y col., 2025). La variabilidad interespecífica e interindividual también afecta la generalización de los modelos debido a las diferencias biológicas y genéticas que pueden introducir incertidumbre en el estudio. Además, la falta de descriptores adecuados para ciertos fenómenos complejos o propiedades emergentes sigue siendo un obstáculo (von Hellfeld y col., 2023). Finalmente, el desarrollo de modelos robustos y predictivos para la toxicidad de mezclas es un área incipiente y desafiante, dada la complejidad de las interacciones sinérgicas, antagónicas o aditivas (Zhan y col., 2023).

Perspectivas Futuras y Direcciones

El futuro del QSAR en toxicología es prometedor, impulsado por la innovación tecnológica y la colaboración multidisciplinar, con el fin de superar las limitaciones actuales y ampliar su impacto. La creación y consolidación de las distintas bases de datos toxicológicas, como OpenFoodTox, que se encuentra estructurada según plantillas armonizadas de la OCDE que facilitan el intercambio y la evaluación de

riesgos químicos, son cruciales para el desarrollo de nuevos modelos de relación estructura-actividad cuantitativa (QSAR) y su integración en herramientas como la OECD QSAR Toolbox. Su aplicación actual abarca desde la evaluación de huellas ambientales hasta la definición de umbrales de preocupación toxicológica (TTC), pero su futuro, con OpenFoodTox 2.0, promete una mayor integración de datos fisicoquímicos, de exposición y de toxicocinética. Esta evolución es fundamental para el futuro de la toxicología predictiva, ya que permitirá consolidar diversas líneas de evidencia (*in vivo*, *in vitro* e *in silico*) mediante enfoques de ponderación de la evidencia, lo que apoyará las metodologías de nuevo enfoque (NAM) que buscan reducir significativamente el uso de ensayos con animales en la evaluación de la seguridad química (Dorne y col., 2021).

En el ámbito de la evaluación de la seguridad y del análisis QSAR, la ontología emerge como una tecnología fundamental para superar la fragmentación de los datos toxicológicos y la falta de un entendimiento mecanicista profundo. Su importancia radica en su capacidad para proporcionar un marco conceptual explícito que sistematiza el conocimiento toxicológico a través de múltiples escalas de granularidad, desde el nivel de organelo hasta el de órgano, lo que permite definir con precisión las dianas toxicológicas y las relaciones causales de los efectos adversos. A futuro, el enfoque ontológico es indispensable para el desarrollo de la toxicología computacional, ya que no solo estandariza datos dispares, como los valores de NOAEL, para mejorar la calidad y la compatibilidad de los análisis, sino que también facilita la extrapolación a humanos al cerrar la brecha entre los datos clínicos y los de investigación básica.

Además, potencia la inteligencia artificial y el aprendizaje automático al aportar contexto biológico y mayor explicabilidad a los biomarcadores predictivos, convirtiéndose en un pilar para la integración de conocimiento multidisciplinario y la toma de decisiones críticas sobre la seguridad en el desarrollo de fármacos (Yamagata y col., 2019).

QSAR basados en mecanismos: modelos predictivos que incorporan información sobre mecanismos de acción

La integración de mecanismos de acción en modelos QSAR ha permitido avanzar hacia predicciones de toxicidad más precisas y biológicamente relevantes. Este enfoque se basa en incluir información detallada sobre cómo un compuesto interactúa con blancos biológicos específicos, lo que facilita no solo la estimación del efecto adverso, sino también la comprensión del proceso molecular que conduce a esa toxicidad (Chen y col., 2022). Una estrategia común es el uso de descriptores derivados de estudios de *docking* molecular, que predicen la afinidad y las interacciones específicas entre un compuesto y su receptor o enzima diana. Estos descriptores reflejan detalles como la energía de unión, las interacciones hidrofóbicas y las fuerzas electrostáticas, lo que permite al modelo identificar las características estructurales críticas para activar un mecanismo toxicológico específico (Uesawa y col., 2020).

Desarrollo de nuevos Descriptores

Entre estos, destacan los descriptores basados en propiedades electrónicas y de interacción molecular. Los orbitales HOMO y LUMO son fundamentales, pues indican la capacidad de un compuesto

para donar o aceptar electrones, aspectos clave en las reacciones de activación o de detoxificación en sistemas biológicos. La energía del orbital HOMO se relaciona con la reactividad nucleofílica, mientras que la del LUMO predice la afinidad por especies electrofílicas, aspectos críticos en los mecanismos de sensibilización y de daño molecular (Clark y col., 2020).

Asimismo, los índices de reactividad específica, como los índices de Fukui, permiten identificar regiones de alta reactividad nucleofílica o electrofílica dentro de la molécula, directamente vinculadas a sitios activos en los mecanismos de toxicidad. Complementariamente, los potenciales electrostáticos generados a partir de mapas de interacción electrostática proporcionan información sobre cómo la molécula se comporta en ambientes biológicos, lo cual influye en su afinidad por blancos específicos (Wold y col., 2021).

Por otro lado, los descriptores conformacionales y de flexibilidad molecular cobran cada vez más importancia. Los campos electrostáticos y de energía de interacción capturan la igualación conformacional y la adaptabilidad de la molécula a diferentes receptores, aspectos fundamentales en los mecanismos de unión y activación. Otros descriptores incluyen medidas de flexibilidad molecular, como el número de rotaciones libres o de conformaciones estables, que influyen en la capacidad de un compuesto para interactuar con blancos dinámicos y activar vías de toxicidad (Amrein y col., 2022).

Avances en redes neuronales para QSAR

Uno de los enfoques más destacados es el uso de redes neuronales convolucionales

que operan sobre imágenes y mapas de campos electrostáticos, lo que permite al modelo aprender patrones directamente a partir de representaciones visuales. Por su parte, los modelos basados en gráficos (*Graph Neural Networks, GNN*) aprovechan la estructura intrínseca de las moléculas, considerando los átomos como nodos y los enlaces como aristas, para capturar relaciones topológicas y electrónicas de manera integrada y eficiente (Balabin y col., 2018).

Además, técnicas como los transformers, aplicadas a modelos moleculares como MolBERT, han demostrado una capacidad sobresaliente para aprender representaciones contextualizadas y profundas que pueden relacionarse directamente con mecanismos de acción biológica. Estas arquitecturas permiten incorporar información bioquímica y farmacológica adicional, lo que enriquece los modelos predictivos y facilita su interpretación (Balabin y col., 2018).

La integración de IA y *Deep Learning* en QSAR también ha permitido superar limitaciones tradicionales, como la escasez de datos y la dependencia de descriptores predefinidos, mediante el aprendizaje automático en conjuntos de datos extensos y diversos. Esto ha abierto caminos para la predicción más rápida y confiable de perfiles de toxicidad, en particular para compuestos nuevos o poco estudiados (Abbod y col., 2024).

Plataformas computacionales y bases de datos

La evolución de las tecnologías de almacenamiento de datos ha sido

fundamental para optimizar la aplicación y el desarrollo de modelos QSAR en toxicología y en química computacional. Las bases de datos estructuradas y de fácil acceso facilitan la recopilación, gestión y análisis de grandes volúmenes de información molecular y experimental, constituyendo un pilar esencial para la construcción y validación de modelos predictivos robustos y confiables (Cozac y col., 2025).

Una de las principales ventajas de contar con bases de datos bien diseñadas radica en la disponibilidad de datos de alta calidad y con una granularidad adecuada. La integración de conjuntos de datos provenientes de diversas fuentes—como publicaciones científicas, agencias regulatorias y laboratorios privados—permite incrementar la diversidad y el tamaño de las muestras, lo que, a su vez, mejora la precisión y la generalizabilidad de los modelos QSAR. Además, el acceso automatizado y estandarizado a estos datos reduce errores y acorta los tiempos de preparación y depuración de información, que suelen ser procesos laboriosos y propensos a errores subjetivos (Cozac y col., 2025).

Colaboración Multidisciplinaria

La resolución de problemas complejos en química, toxicología y bioinformática requiere una colaboración estrecha y coordinada entre estas disciplinas. La química aporta conocimiento sobre el diseño, la síntesis y la caracterización de moléculas, proporcionando las bases estructurales necesarias para comprender sus posibles efectos. La toxicología, por su parte, contribuye a la comprensión de los mecanismos de acción, los efectos

biológicos y los perfiles de toxicidad, lo que permite que los modelos predictivos sean relevantes y contextualizados en la evaluación de riesgos (Meyer y col., 2021). La bioinformática complementa esta integración mediante el análisis y gestión de datos moleculares y biológicos, así como el desarrollo y aplicación de algoritmos avanzados, como *machine learning* y *Deep Learning*, que permiten construir modelos QSAR cada vez más precisos y explicables. Además, la bioinformática facilita la interpretación mecanística y la identificación de biomarcadores, enriqueciendo los modelos predictivos con información biológica y molecular (Meyer y col., 2021).

Estandarización y armonización de datos y protocolos

La estandarización y armonización de protocolos para la elaboración y validación de modelos QSAR son esenciales para garantizar la confiabilidad, la reproducibilidad y la aceptación internacional de estas herramientas predictivas. La variabilidad en los procedimientos de generación de datos, cálculo de descriptores, selección de algoritmos y validación puede afectar significativamente la calidad y la aplicabilidad de los modelos (Gini, 2022). Al establecer estándares internacionales, como los de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), se asegura que los modelos QSAR sean construidos y evaluados bajo criterios uniformes y transparentes, lo que facilita su uso en procesos regulatorios y en la toma de decisiones en salud pública y en el medio ambiente. La OCDE, por ejemplo, ha desarrollado principios para la validación y aceptación de QSAR que incluyen requisitos

de robustez, precisión, aplicabilidad y claridad (Gini, 2022).

Conflictos de interés

Los autores declaran que no existen conflictos de interés alguno relacionados con esta publicación. No se recibió financiamiento alguno de ninguna industria, institución, organización, organismo ni ente.

Referencias Bibliográficas

- Abbod M, Mohammad A. 2024. Combined interaction of fungicides binary mixtures: experimental study and machine learning-driven QSAR modeling. *Sci Rep* 14(1):12700.
- Abdullahi M, Uzairu A, Shallangwa GA, Mamza PA, Ibrahim MT. 2022. Computational modelling studies of some 1,3-thiazine derivatives as anti-influenza inhibitors targeting H1N1 neuraminidase via 2D-QSAR, 3D-QSAR, molecular *docking*, and ADMET predictions. *Beni-Suef Univ J Basic Appl Sci* 11(1):104.
- Amrein S, Matter M. 2022. Conformational flexibility descriptors in molecular modeling: state of the art. *J Mol Graph Model* 116:108209.
- Andrade-Ochoa S, Sánchez-Aldana D, Rodríguez-Valdez LM, Nevárez-Moorillón GV. 2023. *In vitro* and QSAR evaluation of antifungal activity of terpenoid constituents of essential oils against *Alternaria alternata* and *Fusarium oxysporum*. *Biomedica* 43(Suppl 1):156–169.
- Balabin IA, Judson RS. 2018. Exploring non-linear distance metrics in the structure-activity space: QSAR models for human estrogen receptor. *J Cheminform* 10(1):47.
- Belfield SJ, Cronin MTD, Enoch SJ, Firman JW. 2023. Guidance for good practice in the application of machine learning in development of toxicological QSARs. *PLoS One* 18(5):e0282924.
- Beltrán-Pérez C, Serrano AA, Solís-Rosas G, Martínez-Jiménez A, Orozco-Cruz R, Espinoza-Vázquez A, Miralrio A. 2022. A general use QSAR-ARX model to predict corrosion inhibition efficiency of drugs using quantum mechanical descriptors and experimental comparison for lidocaine. *Int J Mol Sci* 23(9):5086.
- Banerjee P, Siramshetty VB, Drwal MN, Preissner R. 2016. Computational methods for the prediction

- of *in vitro* effects of new chemical structures. *J Cheminform* 8:51.
- Bishop PL, Mansouri K, Eckel WP, Lowit MB, Allen D, Blankinship A, Lowit AB, Harwood DE, Johnson T, Kleinstreuer NC. 2024. Evaluation of *in silico* model predictions for mammalian acute oral toxicity and regulatory application in pesticide hazard and risk assessment. *Regul Toxicol Pharmacol* 149:105614.
- Bossa C, Andreoli C, Bakker M, Barone F, De Angelis I, Jeliakova N, Nymark P, Battistelli CL. 2021. FAIRification of nanosafety data to improve applicability of QSAR approaches: a case study on *in vitro* Comet assay genotoxicity data. *Comput Toxicol* 20:100190.
- Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS, Allen DG. 2022. New approach methodologies under TSCA: Opportunities and challenges. *Regulatory Toxicology and Pharmacology* 131:105164.
- Bueso-Bordils JI, Antón-Fos GM, Martín-Algarra R, Alemán-López PA. 2024. Vista general de métodos de toxicología computacional aplicados al descubrimiento de drogas y productos químicos verdes. *Diario de Xenobióticos* 14(4):1901–1918.
- Bunmahotama W, Vijver MG, Peijnenburg W. 2022. Development of a quasi-QSAR model for prediction of immobilization response of *Daphnia magna* exposed to metal-based nanomaterials. *Environ Toxicol Chem* 41(6):1439–1450.
- Chen Y, Dong Y, Li L, Jiao J, Liu S, Zou X. 2022. Toxicity rank order as a new approach for toxicity prediction by QSAR models. *Int J Environ Res Public Health* 20(1):701.
- Chung E, Russo DP, Ciallella HL, Wang YT, Wu M, Aleksunes LM, Zhu H. 2023. Data-driven QSAR modeling for human carcinogenicity by chronic oral exposure. *Environ Sci Technol* 57(16):6573–6588.
- Clark T, Hénon E. 2020. Electrostatic potential surface descriptors for prediction of biological activity. *Eur J Med Chem* 188:112056.
- Cozac R, Hasic H, Choong JJ, Richard V, Beheshti L, Froehlich C, Koyama T, Matsumoto S, Kojima R, Iwata H, Hasegawa A, Otsuka T, Okuno Y. 2025. kMOL: an open-source machine and federated learning library for drug discovery. *J Cheminform* 17(1):22.
- Cronin MTD, Basiri H, Chrysochou G, Enoch SJ, Firman JW, Spinu N, Madden JC. 2025. The predictability of QSARs for toxicity: recommendations for improving model performance. *Comput Toxicol* 33:100338.
- Da Xue Xue Bao Yi Xue Ban. 2017. Quantitative structure-activity relationship model for prediction of cardiotoxicity of chemical components in traditional Chinese medicines. *J Peking Univ Health Sci* 49(3):551–556.
- Dablander M, Hanser T, Lambiotte R, Morris GM. 2023. Exploring QSAR models for activity-cliff prediction. *J Cheminform* 15(1):47.
- Danishuddin, Khan AU. 2016. Descriptor selection methods in QSAR analysis: paradigm for drug design. *Drug Discov Today* 21(8):1291–1302.
- De P, Kumar V, Kar S, Roy K, Leszczynski J. 2022a. Repurposing FDA approved drugs as possible anti-SARS-CoV-2 medications using ligand-based computational approaches: sum of ranking difference-based model selection. *Struct Chem* 33(5):1741–1753.
- De P, Kar S, Ambure P, Roy K. 2022b. Prediction reliability of QSAR models: an overview of various validation tools. *Arch Toxicol* 96(5):1279–1295.
- de Souza AS, Ferreira LLG, de Oliveira AS, Andricopulo AD. 2019. QSAR for structurally diverse chemotypes with anti-*Trypanosoma cruzi* activity. *Int J Mol Sci* 20(11):2801.
- Dong Z, Chen H, Yang Y, Hao H. 2025. Optimization model of anti-breast cancer candidate drugs based on machine learning. *Front Genet* 16:1523015.
- Dorne JLCM, Richardson J, Livaniou A, Carneseccchi E, Ceriani L, Baldin R, Kovarich S, Pavan M, Saouter E, Biganzoli F, Pasinato L, Zare Jeddi M, Robinson TP, Kass GEN, Liem AKD, Toropov AA, Toropova AP, Yang C, Tarkhov A, Georgiadis N, Kienzler A, Paini A, Rorije E, Worth A, Barroso J, Whelan M, Benfenati E, Battistelli CL, Bossa C, Benigni R, Mekenyan O, Petkov P, Tcheremenskaia O, DeMeo C, Cross KP, Myatt GJ. 2021. EFSA's OpenFoodTox: an open-source toxicological database on chemicals in food and feed. *Environ Int* 146:106293.
- Fan T, Sun G, Zhao L, Cui X, Zhong R. 2018. QSAR and classification study on prediction of acute oral toxicity of N-nitroso compounds. *Int J Mol Sci* 19(10):3015.
- Fattouche M, Belaidi S, Abchir O, Al-Shaar W, Younes K, Al-Mogren MM, Chtita S, Soualmia F, Hochlaf M. 2024. ANN-QSAR, molecular *docking*, ADMET predictions, and molecular dynamics studies of isothiazole derivatives as inhibitors of HCV polymerase NS5B. *Pharmaceuticals* 17(12):1712.
- Forest V, Hochepeid JF, Pourchez J. 2019. Importance of choosing relevant biological *Endpoints* to

- predict nanoparticle toxicity with computational approaches. *Chem Res Toxicol* 32(7):1320–1326.
- Furuhama A, Kitazawa A, Yao J, Matos Dos Santos CE, Rathman J, Yang C, Chakravarti S, Myatt GJ, Cross KP, Benfenati E, Raitano G, Mekenyan O, Petkov P, Bossa C, Benigni R, Battistelli CL, Giuliani A, Tcheremenskaia O, DeMeo C, Honma M 2023. Evaluation of QSAR models for predicting mutagenicity: outcome of the Second Ames/QSAR international challenge project. *SAR QSAR Environ Res* 34(12):983–1001.
- Gadaleta D, Serrano-Candelas E, Ortega-Vallbona R, Colombo E, García de Lomana M, Biava G, Roncaglioni A, Benfenati E, Worth A, Paini A, Kienzler A, Rorije E, Barroso J, Whelan M. 2024. Benchmarking computational tools for predicting toxicokinetic and physicochemical properties of chemicals. *J Cheminform* 16(1):145.
- Gandhi A, Masand V, Zaki ME, Al-Hussain SA, Ghorbal AB, Chapolikar A. 2021. QSAR evaluation of MDA-MB-231 antiproliferative leads. *Molecules* 26(16):4795.
- Garthwaite C. 2025. Economic markets and pharmaceutical innovation. *J Econ Perspect* 39(2):3–26.
- Gini G. 2022. QSAR methods. *Methods Mol Biol* 2425:1–26.
- Greco S, Bossa C, Battistelli CL, Giuliani A. 2025. Statistical exploration of QSAR models in cancer risk assessment: pesticides case study. *Toxics* 13(4):299.
- Gulumian M, Andraos C, Afantitis A, Puzyn T, Coville NJ. 2021. Importance of surface topography in biological activity and catalysis of nanomaterials. *Int J Mol Sci* 22(15):8347.
- Hansch C, Fujita T. 1964. ρ σ π Analysis: correlation of biological activity and chemical structure. *J Am Chem Soc* 86:1616–1626.
- Hatmal MM, Abuyaman O, Taha M. 2021. *Docking*-generated ligand poses for machine learning classification: repurposing covalent inhibitors for TMPRSS2. *Comput Struct Biotechnol J* 19:4790–4824.
- Héberger K. 2023. Selection of optimal validation methods for QSAR and applicability domain. *SAR QSAR Environ Res* 34(5):415–434.
- Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T, Saiakhov R, Chakravarti S, Myatt GJ, Cross KP, Benfenati E, Raitano G, Mekenyan O, Petkov P, Bossa C, Benigni R, Battistelli CL, Giuliani A, Tcheremenskaia O, DeMeo C, Dorne JLCM, Johnson C, Kruhlak NL, Stavitskaya L, Teasdale A, Vock E. 2019. Improvement of QSAR tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR project. *Mutagenesis* 34(1):3–16.
- Huang T, Sun G, Zhao L, Zhang N, Zhong R, Peng Y. 2021. QSAR studies on toxic effects of nitroaromatic compounds: systematic review. *Int J Mol Sci* 22(16):8557.
- Janicka M, Śliwińska A. 2022. Quantitative retention–activity relationships for predicting pharmaceutical and toxic properties of pesticides. *Molecules* 27(11):3599.
- Kar S, Yang S. 2024. Third-generation periodic table descriptors for nano-qRASTR modeling of zebrafish toxicity. *Beilstein J Nanotechnol* 15:1142–1152.
- Khan MZI, Ren JN, Cao C, Ye HY, Wang H, Guo YM, Yang JR, Chen JZ. 2024. Comprehensive hepatotoxicity prediction using ensemble machine learning and deep learning. *Front Pharmacol* 15:1441587.
- Kim T, You BH, Han S, Shin HC, Chung KC, Park H. 2021. Quantum artificial neural network approach to derive predictive 3D-QSAR for BBB passage. *Int J Mol Sci* 22(20):10995.
- Király P, Kiss R, Kovács D, Ballaj A, Tóth G. 2022. Relevance of goodness-of-fit, robustness, and prediction categories of OECD QSAR validation principles. *Mol Inform* 41(11):e2200072.
- Kurosaki K, Wu R, Uesawa Y. 2020. Toxicity prediction tool for agonist/antagonist activities in molecular initiating events. *Int J Mol Sci* 21(21):7853.
- Kuz'min VE, Muratov EN, Artemenko AG, Gorb L, Qasim M, Leszczynski J. 2008. Effect of nitroaromatics composition on *in vivo* toxicity: non-additive 1D QSAR analysis. *Chemosphere* 72(9):1373–1380.
- Li J, Wang C, Yue L, Chen F, Cao X, Wang Z. 2022. Nano-QSAR modeling for predicting cytotoxicity of metallic and metal oxide nanoparticles: review. *Ecotoxicol Environ Saf* 243:113955.
- Liu JY, Peeples J, Sayes CM. 2024. Evaluation of machine learning QSAR models for classification of lung surfactant inhibitors. *Environ Health* 2(12):912–917.
- Lin Z, Chou WC. 2022. Machine learning and artificial intelligence in toxicological sciences. *Toxicol Sci* 189(1):7–19.
- Madden JC, Enoch SJ, Paini A, Cronin MTD. 2020. Review of *in silico* tools as alternatives to animal testing. *Altern Lab Anim* 48(4):146–172.
- Makhouri FR, Ghasemi JB. 2018. *In silico* studies in drug research against neurodegenerative diseases. *Curr Neuropharmacol* 16(6):664–725.

- Matveieva M, Polishchuk P. 2021. Benchmarks for interpretation of QSAR models. *J Cheminform* 13(1):41.
- Meyer M, Friedrich P. 2021. Bioinformatics approaches for QSAR modeling and chemical toxicity prediction. *Bioinformatics* 37(15):2343-2350.
- Mei Y, Wu K. 2022. Multi-objective optimization in anti-breast cancer candidate drug studies. *Sci Rep* 12(1):19347.
- Meyer H. 1899. On the theory of alcohol narcosis: which property of anesthetics gives narcotic activity. *Arch Exp Pathol Pharmacol* 42:109-118.
- Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, Tropsha A. 2020. QSAR without borders. *Chem Soc Rev* 49(11):3525-3564.
- Muto S, Furuhashi A, Yamamoto M, Otagiri Y, Koyama N, Hitaoka S, Koyama T, Matsumoto S, Kojima R, Iwata H, Hasegawa A, Otsuka T, Okuno Y, Mishima M. 2024. Local QSAR for stability of nitrenium ions to reduce false positives in mutagenicity prediction. *Genes Environ* 46(1):1-10.
- Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, Anger LT, Barber C, Beilke L, Bercu J, Booth ED, Bower D, Brigo A, Cammerer Z, Cross KP, Custer L, Dobo K, Ford KA, Fortin MC, Gad-McDonald SE, Gervais V, Gerets HHJ, Glover KP, Glowienke S, Gomez S, Gompel JV, Harvey J, Hasselgren C, Honma M, Johnson C, Kenyon MO, Kruhlak NL, Leavitt P, Lindberg J, Miller S, Muster W, Nicolette J, O'Donovan M, Parenty A, Powley MW, Quigley DP, Reddy MV, Roberts DW, Schilter B, Stavitskaya L, Teasdale A, Vock E, White AT, Wichard J, Witt KL, Yang C, Zwickl C.. 2018. *In silico* toxicology protocols. *Regul Toxicol Pharmacol* 96:1-17.
- Nha Tran TT, Thuan Tran TD, Thuy Bui TT. 2023. Machine-learning integration in 3D-QSAR CoMSIA models for lipid antioxidant peptides. *RSC Adv* 13(48):33707-33720.
- OECD. 2013. Guidance document on developing and assessing adverse outcome pathways. Series on Testing and Assessment No. 184.
- Peteani G, Huynh MTD, Gerebtzoff G, Rodríguez-Pérez R. 2024. Machine learning models for predicting properties of targeted protein degraders. *Nat Commun* 15:5764.
- Playe B, Stoven V. 2020. Evaluation of deep and shallow learning methods in chemogenomics for prediction of drug specificity. *J Cheminform* 12(1):11.
- Qin L, Zhang X, Chen Y, Mo L, Zeng H, Liang Y. 2017. Predictive QSAR models for toxicity of disinfection byproducts. *Molecules* 22(10):1671.
- Raies AB, Bajic VB. 2016. *In silico* toxicology: computational methods for prediction of chemical toxicity. *Wiley Interdiscip Rev Comput Mol Sci* 6(2):147-172.
- Ren R, Gao L, Li G, Wang S, Zhao Y, Wang H, Liu J. 2023. 2D and 3D-QSAR study and *docking* of vascular endothelial growth factor receptor 3 inhibitors for potential treatment of retinoblastoma. *Front Pharmacol* 14:1177282.
- Richet C. 1893. On the relationship between the toxicity and the physical properties of substances. *Compt Rendus Seances Soc Biol* 9:775-776.
- Rovida C, Barton-Maclaren T, Benfenati E, Caloni F, Chandrasekera PC, Chesné C, Cronin MTD, De Knecht J, Dietrich DR, Escher SE, Fitzpatrick S, Flannery B, Herzler M, Hougaard Bennekou S, Hubesch B, Kamp H, Kisitu J, Kleinstreuer N, Kovarich S, Leist M, Luechtefeld T, Maxwell G, Munn S, Pallocca G, Pamies D, Price A, Prieto P, Rovida E, Schmutz M, Schwarz M, Smirnova L, Toma C, Vinken M, Whelan M, Hartung T. 2020. Internationalization of read-across as a validated new approach method for regulatory toxicology. *ALTEX* 37(4):579-606.
- Scheufen Tieghi R, Moreira-Filho JT, Martin HJ, Wellnitz J, Otoch MC, Rath M, Kleinstreuer N. 2024. Machine learning model and web portal for predicting human skin sensitization effects of chemical agents. *Toxics* 12(11):803.
- Selvestrel G, Lavado GJ, Toropova AP, Toropov AA, Gadaleta D, Marzo M, Baderna D, Benfenati E. 2022. Monte Carlo models for sub-chronic repeated-dose toxicity: systemic and organ-specific toxicity. *Int J Mol Sci* 23(12):6615.
- Shahlaei M. 2013. Descriptor selection methods in QSAR studies: review. *Chem Rev* 113(10):8093-8103.
- Shayanfar S, Shayanfar A. 2022. Comparison of methods for validity evaluation of QSAR models. *BMC Chem* 16(1):63.
- Shi P, Ray S, Zhu Q, Kon MA. 2011. Top-scoring pairs for feature selection in machine learning and applications to cancer outcome prediction. *BMC Bioinformatics* 12:375.
- Shiri F, Pirhadi S, Ghasemi JB. 2016. Alignment-independent 3D-QSAR, quantum calculations and molecular *docking* of Mer-specific tyrosine kinase inhibitors as anticancer drugs. *Saudi Pharm J* 24(2):197-212.
- Shkil DO, Muhamedzhanova AA, Petrov PI, Skorb EV, Aliev TR, Steshin YS, Tumanov AV, Kisilinskiy AS, Fedorov MV. 2024. Expanding predictive capabilities in toxicology: hackathon-enhanced

- data and model segmentation. *Molecules* 29(8):1826.
- Siramshetty V, Williams J, Nguyễn ĐT, Neyra J, Southall N, Mathé E, Xu X, Shah P. 2021. Validating ADME QSAR models using marketed drugs. *SLAS Discov* 26(10):1326-1336.
- Toropov AA, Di Nicola MR, Toropova AP, Roncaglioni A, Carnesecchi E, Kramer NI, Dorne JLC. 2022. Regression-based QSAR model to predict acute toxicity of aromatic chemicals in tadpoles of *Rana japonica*. *Sci Total Environ* 830:154795.
- Tuenter E, Segers K, Kang KB, Viaene J, Sung SH, Cos P, Pieters L. 2017. Antiplasmodial activity, cytotoxicity and SAR of cyclopeptide alkaloids. *Molecules* 22(2):224.
- Uesawa Y. 2020. AI-based QSAR modeling for prediction of active compounds in MIE/AOP. *Yakugaku Zasshi* 140(4):499-505.
- Verma J, Khedkar V, Coutinho C. 2010. 3D-QSAR in drug design: review. *Curr Top Med Chem* 10(1):95-115.
- Von Hellfeld R, Gade C, Vargesson N, Hastings A. 2023. Considerations for future QSAR modeling of heavy metals: a mercury case study. *Toxicol* 499:153661.
- Wang W, Kim MT, Sedykh A, Zhu H. 2015. Enhanced blood-brain barrier permeability models integrating external bioassay data in QSAR modeling. *Pharm Res* 32(9):3055-3065.
- Wang Y, Jia S, Wang F, Jiang R, Yin X, Wang S, Jin R, Guo H, Tang Y, Wang Y. 2024. 3D-QSAR, scaffold hopping, virtual screening and molecular dynamics simulations of pyridin-2-one as mIDH1 inhibitors. *Int J Mol Sci* 25(13):7434.
- Wei Y, Li W, Du T, Hong Z, Lin J. 2019. Targeting HIV/HCV coinfection using machine-learning-based multiple QSARs. *Int J Mol Sci* 20(14):3572.
- Wold S, Andersson CA. 2021. Molecular reactivity indices in QSAR modeling: overview. *Front Chem* 9:662558.
- Wunsch FM, Wünsch B, Bernal FA, Schmidt TJ. 2021. QSAR of natural-product-inspired aminoalkyl-substituted 1-benzopyrans as antiplasmodial agents. *Molecules* 26(17):5249.
- Xu JY, Wang K, Men SH, Yang Y, Zhou Q, Yan ZG. 2023. QSAR-QSIR-based prediction of bioconcentration factor using machine learning. *Environ Int* 177:108003.
- Yamagata Y, Yamada H, Horii I. 2019. Computational toxicology in drug safety assessment under ontological intellection. *J Toxicol Sci* 44(11):721-735.
- Zhang F, Wang Z, Peijnenburg WJGM, Vijver MG. 2023. Machine learning-driven QSAR models for predicting mixture toxicity of nanoparticles. *Environ Int* 177:108025.