

¿Qué sabe Ud. acerca de...QSAR?

What do you know about QSAR?

Jorge Lozano-Aponte, Thomas Scior
Facultad de Ciencias Químicas, BUAP

Dedicado al Prof. Dr. Joachim Seydel, pionero de QSAR y CADD desde hace 25 años¹

Preguntas

1. ¿Qué es la metodología QSAR?
2. ¿Qué tipo de información se requiere para diferentes tipos de estudios QSAR?
3. ¿Qué son los descriptores QSAR?
4. ¿Cuál es el reto de QSAR?
5. ¿Cuánto tiempo se necesita para realizar un estudio de tipo QSAR?
6. ¿Cuáles son las ventajas y desventajas de esta metodología?
7. ¿Cuáles fármacos se han descubierto por diseño computacional con modelos QSAR?

Respuestas

1. En el área del desarrollo de nuevos fármacos de las industrias farmacéuticas, institutos de investigación privados y públicos (*research and development*, R&D) se han desarrollado herramientas computacionales (*in silico drug design*) y se han establecido nuevas disciplinas en el entorno de la Química Farmacéutica (*Medicinal Chemistry*, *Computer Aided Molecular/Drug Design*, CAMD, CADD).^{2,3} Una de ellas es la metodología que relaciona numéricamente estructuras (químicas) con sus actividades biológicas (*Quantitative Structure-Activity Relationships*, **QSAR**)⁴ QSAR reúne un conjunto de técnicas computacionales relacionadas con diseño y visualización espacial virtual de moléculas (también llamado *in silico*), cálculo de propiedades fisicoquímicas moleculares (Descriptores), bioinformática y estadística. Todo esto con el fin de hacer una predicción teórica de la actividad biológica que permita el diseño teórico de posibles futuros nuevos fármacos (aun no existentes), evitando pasar por el proceso a prueba y error de síntesis orgánica. Al ser una ciencia que existe sólo en un entorno virtual – desmaterializado de necesidades de infraestructura (tales como equipo, instrumentos, materiales y personal de laboratorio) – con enfoque en las relaciones estructura (química) – actividad (biológica), el diseño de candidatos a nuevos fármacos es mucho más económico y rápido.
2. Para llevar a cabo un estudio de tipo QSAR se necesitan básicamente tres tipos de información: (A) Estructura molecular de diferentes compuestos que tengan un mismo mecanismo de acción (farmacodinamia), y por lo tanto se consideran como ligandos ante una diana biomolecular común; (B) Datos de actividad biológica de cada uno de los ligandos incluidos en el estudio (por ejemplo: DE₅₀, DL₅₀, Ki, IC₅₀, etc.) y (C) Propiedades fisicoquímicas (descriptores numéricos) de los ligandos calculados por medios computacionales a partir de la estructura molecular generada *in silico*.

En general, es recomendable que los datos de actividad biológica provengan, si es posible, de un solo laboratorio de investigación, esto reduce en cierta medida el error humano en las mediciones experimentales ya que se considera que los bioensayos fueron realizados bajo las mismas condiciones de estandarización o normas y bajo los mismos errores sistemáticos operativos y de manipulación humana. Dado que los bioensayos son más costosos, es común que los experimentos no se repiten de forma debida para asegurar un valor estadísticamente validado y al ser un método estadístico, el estudio QSAR incluye este componente de incertidumbre de la Variable dependiente por lo que la predicción puede

fallar. En contraste, los valores de las variables independientes o descriptores, no tienen dicha naturaleza, ya que cada valor es reproducible e invariable pues descansa en un algoritmo de cálculo definido en el software. Eso no quiere decir que los descriptores no fallan, al contrario, hay que discriminar entre descriptores inadecuados (mala selección en ciertos casos), redundantes (co-varianza o colinealidad) u otros mal programados (mal diseño del algoritmo).⁴

Cuando se cuenta con dicha información colectada de la literatura o de un laboratorio, se utiliza una herramienta estadística llamada regresión lineal múltiple (*Multiple Linear Regression*, MLR) tomando como variable dependiente los valores de actividad biológica de los ligandos y como variables independientes, los descriptores calculados. El resultado es una ecuación lineal del tipo $y = a_1x_1 + a_2x_2 + \dots + a_nx_n$, también llamada ecuación o modelo QSAR.

Existen dos tipos de uso de datos de entrada: (1) superposición de los ligandos de tal forma que idénticas o semejantes subestructuras se alinean (*Structure Based Alignment*, **SBA-QSAR**) o (2) si existe información 3D disponible sobre el blanco biológico, se puede realizar un alineamiento en base a la proteína-blanco (diana biomolecular) (*Protein Based Alignment*, **PBA-QSAR**).^{2,3}

Los resultados en forma de ecuación o modelo QSAR permiten predecir la actividad biológica de compuestos aun no sintetizados que se generan virtualmente en sólo un par de minutos, pero deben compartir características estructurales de los ligando incluidos en el estudio para no salir de la reglas / patrón químico / rango de valores de los descriptores (**tipo prospectivo**).

El otro tipo es **retrospectivo**, es decir, analizar las moléculas ya existentes (las de la síntesis y bioensayos) para entender sus interrelaciones no evidentes entre “estructuras y actividades”.

Como se nota por la extensión de la respuesta, la preparación de los datos de entrada (*input data*) es el paso más crucial ya que el resultado – literalmente – se obtiene de forma automatizada (oprimiendo botones de la PC) y únicamente depende del *input*.

3. Los descriptores son propiedades fisicoquímicas que se calculan a partir de una estructura virtual 3D por métodos computacionales, por lo que son un reflejo cuantitativo o describen numéricamente a cada una de las moléculas. Estos pueden ser de diferentes tipos, y ello depende de la complejidad de información que requiera el cálculo. Así algunos descriptores son bastante “globales” describiendo toda la molécula y por ende son muy simples (como el peso molecular) y otros pueden resultar muy complejos (como la variabilidad topológica, distribución de carga eléctrica, superficie polar, por mencionar algunos). El número de descriptores a considerar está en función de las herramientas computacionales de cálculo con que se cuente y del número de moléculas incluidas en el estudio. El error más frecuente consiste en permitir a la operación matemática de MLR agregar un espacio numérico sobredimensionado para describir cada molécula de forma independiente e individual sin poder establecer una regla con poder predictivo y confiable (ecuación lineal). Eso sucede cuando el número de descriptores rebasa al número de las moléculas (*overfitting*⁴).
4. La metodología QSAR es interdisciplinaria, por lo que recibe información de la Química Orgánica y de la Farmacología. La forma en que QSAR retribuye esta situación y que constituye el objetivo de esta metodología, es mediante el diseño dirigido de ligandos aun no existentes, pero que mediante las ecuaciones generadas han mostrado una alta probabilidad de éxito farmacológico pues como se ha visto, estas ecuaciones permiten una predicción de la actividad biológica. En concreto, QSAR persigue el diseño molecular dirigido de compuestos con potencial farmacológico, permitiendo el ahorro de recursos económicos y humanos.
5. Para un experto en simulaciones moleculares computacionales el tiempo para la realización de un estudio con unos 50 ligandos – sin o con diana – es de un par de semanas, y mucho menor si se compara con el que llevaría la síntesis y bioensayos de nuevos compuestos en meses y hasta años. Esta ventaja permite tomar una serie pequeña de moléculas (20 a 60) y gracias a la rapidez de contar con los resultados, directamente retroalimentar el laboratorio de síntesis en el proceso continuo del proyecto. Así, QSAR predice nuevas estructuras jamás vistas y las propone a los químicos orgánicos para ser llevados a los bioensayos cuyos resultados confirman o contradicen los valores predichos por el modelo QSAR. En un caso óptimo, mediante este ciclo operativo se obtienen mejores candidatos que por el puro proceso de prueba y error (*trial & error*). Esto ahorra tiempo, dinero, recursos y evita fracasos para quienes desarrollan nuevos fármacos (véase Tabla, Respuesta 7).
6. Las ventajas de QSAR son el bajo costo (programas gratuitos, información relativamente accesible, no se utiliza instrumental ni reactivos químico, etc.), el uso de interfaces que facilitan el manejo y diseño, además de que la construcción de las moléculas y el cálculo de descriptores pueden ser sumamente rápidos. Sus desventajas son la familiarización con metodologías computacionales (diferentes sistemas operativos e interfaces gráficas, manejo de bases de datos, desarrollo del software) y en este sentido, la resolución de diferentes problemas de computo (compatibilidad, actualizaciones, registros, formatos de datos) así como el hecho de tener que disponer de datos de actividad biológica de las moléculas que provengan de una misma fuente, el cambio de perspectiva en la forma de trabajo.

7. Por último, en la tabla, se resumen tres fármacos (ahora medicamentos) cuyo descubrimiento se atribuye al uso de la metodología QSAR.

Tabla. Ejemplos de fármacos comerciales desarrollados por QSAR. [Modificado de la tabla en ref. 5]. (Ref. 5, citada en ref. 2 y 3).

Compuesto	Nombre comercial (Compañía farmacéutica)	Uso	Método computacional	Ref.
Norfloxacina	NOROXIN® (Kyorin Pharmaceutical)	Antibacteriano	QSAR	Koga, 1980 (Citado en 5)
Captopril	CAPOTENA® (Squibb)	Antihipertensivo	QSAR	6
Losartan	COZAAR® (DuPont (BMS) Merck)	Antihipertensivo	Modelado molecular y QSAR	Duncia, 1990 (Citado en 5)

Referencias

- Silverman RB. 1995. Medizinische Chemie für Organiker, Biochemiker und Pharmazeutische Chemiker, VCH: Weinheim; Título original: The Organic Chemistry of Drug Design and Drug Action. Comentario: El Dr. Seydel, Research Center Borstel, Alemania, tradujo el libro del autor americano R. B. Silverman, durante su estancia en México y cuya hija hoy es catedrática de la UNAM.
- Scior T, Lozano-Aponte J, Echeverría D. 2009. CAMD Y CADD. Simulaciones moleculares computacionales de fármacos. Parte 1. Informacéutico, 16 (5), 46-50.
- Scior T, Lozano-Aponte J, Echeverría D. 2009. CAMD Y CADD. Simulaciones moleculares computacionales de fármacos. Parte 2. Informacéutico, 16 (6), 32-35.
- Scior T, Medina-Franco JL, Do Q-T, Martínez-Mayorga K, Yunes Rojas JA and Bernard P. 2009. How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review. Current Medicinal Chemistry, 16, 4297-4313. <http://www.benthamscience.com/cmc/openaccessarticles/cmc16-32/0006C%5B1%5D.pdf>
- Medina Franco JL. 2005. Diseño de fármacos asistido por computadora. Inhibidores de la transcriptasa reversa del virus VIH, hipocolesterolémiantes y antiparasitarios. Tesis Doctoral. Facultad de Química, UNAM. México D.F.
- Smith CG, Vane JR. 2003. The discovery of captopril. FASEB J., 17 (8): 788-9. <http://www.fasebj.org/cgi/pmidlookup?view=long&pmid=12724335>