

© 2018 Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Zaragoza.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

TIP Revista Especializada en Ciencias Químico-Biológicas, 21(2): 124-134, 2018.

DOI: 10.22201/fesz.23958723e.2018.2.6

DISEÑO DE FÁRMACOS ASISTIDO POR COMPUTADORA: CUANDO LA INFORMÁTICA, LA QUÍMICA Y EL ARTE SE ENCUENTRAN

Fernando D. Prieto-Martínez¹ y José L. Medina-Franco²

*Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad #3000, C.P. 04510, Ciudad de México, México

E-mails: ¹ferdpm4@hotmail.com, ²medinajl@unam.mx

RESUMEN

La industria farmacéutica está en constante evolución, siendo el desarrollo y descubrimiento de nuevos fármacos el motor de la misma. Antaño, el diseño de fármacos estaba directamente relacionado a los productos naturales, que posteriormente se modificaban por síntesis. Si bien, la estrategia sigue vigente, el desarrollo de nuevos fármacos implica una inversión considerable de dinero y tiempo. Actualmente, el desarrollo de computadoras más potentes y eficientes ha permitido generar metodologías y simulaciones que han optimizado y en cierto grado modificado el panorama farmacéutico. Aquí se presenta un texto introductorio, a los principales métodos y técnicas computacionales auxiliares en el proceso de diseño y modelado molecular. Se hace especial énfasis en métodos quimioinformáticos, sus conceptos, bases y sus aplicaciones.

Palabras Clave: acoplamiento molecular, modelo de farmacóforo, modelo por homología, quimioinformática, similitud molecular.

Computer-aided drug design: when informatics, chemistry and art meets

ABSTRACT

The pharmaceutical industry is in constant evolution being the driving force, discovery and development of new drugs. In the past, drug discovery was basically based on natural products that were later modified by chemical synthesis. Despite the fact such strategy continues to be valuable, the cost and time of current drug discovery is very high. Currently, the advancement in the development of more powerful and efficient computers has enabled to develop methods and simulations that are optimizing at certain point the drug discovery outlook. In this work we introduce major computational methods and techniques that aid the drug discovery process emphasizing chemoinformatics concepts, their basis and applications.

Key Words: molecular docking, pharmacophore model, homology model, chemoinformatics, molecular similarity.

INTRODUCCIÓN

El desarrollo e investigación de nuevos fármacos es un proceso complejo. Históricamente, los productos naturales han sido la fuente clásica para su búsqueda. Cada año se invierten fuertes cantidades de dinero con la meta de evaluar e identificar fármacos potenciales. Actualmente, gracias al desarrollo tecnológico es posible realizar en menor tiempo simulaciones de procesos biológicos, cálculo de propiedades y comparar estructuras químicas de manera sistemática. En un principio estas herramientas formaron parte de la denominada *Química Computacional*. Con el paso del tiempo éstas fueron aplicadas de manera exitosa al descubrimiento y desarrollo de fármacos. Una disciplina relacionada a la Química Computacional es la *Quimioinformática* que hace uso de ecuaciones, modelos y técnicas computacionales, para resolver problemas en química. La Química Computacional y la Quimioinformática forman parte de un campo del conocimiento más general y aplicado a la Química Farmacéutica: el Diseño de Fármacos Asistido por Computadora (DIFAC).

Desafortunadamente no existe suficiente difusión sobre los avances logrados en DIFAC, principalmente por la poca cantidad de revistas especializadas en el tema. Aunado a lo anterior, la literatura en español referida a la “quimioinformática” es limitada. Por tanto, el objetivo de este trabajo es la mención descriptiva de los procesos más importantes en este campo, con un enfoque particular en los métodos de similitud molecular.

CAMPOS DE FUERZA, CIMIENTOS DE LA QUÍMICA COMPUTACIONAL

A principios del siglo XX, la química avanzó significativamente. Un ejemplo notable es el desarrollo de la mecánica cuántica. Es precisamente por la química cuántica que ocurre el encuentro con la informática, ya que las ecuaciones que describen el carácter electrónico de una molécula sencilla serían bastante laboriosas mediante los métodos tradicionales. Tras la implementación de cálculos cuánticos en programas computacionales, investigadores en distintas universidades decidieron realizar modelos moleculares con ayuda del ordenador. En principio esto podría parecer una tarea tediosa, no obstante, es posible simplificar el modelo: si se considera a una molécula como una

serie de esferas unidas por resortes podemos aplicar ecuaciones matemáticas que describen su movimiento y la carga que pueden tener. Después de todo las computadoras ya han demostrado su capacidad de realizar cálculos complejos. A este proceso se le denomina parametrización. Debido a las diferentes consideraciones y combinaciones posibles es que al conjunto descriptivo de ellas se le denomina campo de fuerza (*force field*). Mediante estas descripciones es posible realizar cálculos de mecánica molecular y comenzar a pensar en simulaciones dinámicas de moléculas en diferentes ambientes. No obstante, los campos de fuerza desarrollados hasta ahora no son universales. Esto quiere decir que de acuerdo al cálculo o simulación que sea de interés, será el campo a elegir, pues algunos son específicos para modelar proteínas, glucósidos o ácidos nucleicos. Si bien el análisis detallado de los campos de fuerza escapa del enfoque de este texto los modelos utilizados sirven de base para realizar simulaciones más complejas, por ejemplo, relajación de una proteína, solvatación de la misma, emular el reconocimiento ligando-proteína. Algunos campos de fuerza usados comúnmente son: AMBER, CHARMM, GROMOS, MARTINI, MMFF94 y OPLS. Si el lector desea ahondar en el tema recomendamos la referencia (Ponder & Case, 2003).

De los métodos *in vitro* a los métodos *in silico*

Ya se ha establecido cómo se puede indicar a una computadora que genere modelos moleculares. Ahora, habrá que preguntarse ¿Qué programas usar para ello? Es probable que el lector esté familiarizado con programas para construir y visualizar estructuras químicas como ChemDraw, ChemSketch, Marvin, PyMol. Fuera de estas funciones, estos programas son limitados para hacer DIFAC. Por otra parte, existen bases de datos en internet donde es posible buscar estructuras, actividades biológicas, propiedades fisicoquímicas, etc. (Tabla I).

La Tabla II resume ejemplos de programas especializados que se pueden usar para hacer cálculos teóricos, química computacional y DIFAC. Dependiendo de los objetivos del proyecto se usa generalmente más de un programa. ¿Es posible comenzar a diseñar y descubrir fármacos potenciales con alguno de estos programas? Probablemente, sí ¿Para ello basta con usar el programa y “un par de clics”? Definitivamente no.

Nombre	URL	Descripción
Chemspider	http://chemspider.com/	Repositorio de moléculas pequeñas
ChEMBL	https://www.ebi.ac.uk/chembl/	Repositorio de fármacos y otras moléculas pequeñas
Protein Data Bank	http://www.rcsb.org/	Repositorio de estructuras cristalográficas de proteínas
PubChem	https://pubchem.ncbi.nlm.nih.gov/	Repositorio de moléculas pequeñas
Drugbank	http://drugbank.ca/	Repositorio de fármacos
ZINC	http://zinc.docking.org/	Repositorio de moléculas para cribado virtual

Tabla I. Bases de datos públicas que permiten la búsqueda de estructuras químicas y sus principales propiedades.

Nombre	Capacidades	¿Comercial o académico?	URL
Autodock	Acoplamiento molecular	Académico	www.autodock.scripps.edu/
Chimera	Visualizador, modelado de proteínas y moléculas	Académico	www.cgl.ucsf.edu/chimera/
Datawarrior	Visualizador, cálculo de propiedades y otros descriptores	Académico	www.openmolecules.org/datawarrior/
Gaussian	Cálculo de propiedades cuánticas y cargas semiempíricas	Comercial	www.gaussian.com/
Maestro	Visualizador, modelado de proteínas y dinámica molecular	Comercial/Académico	www.schrodinger.com/maestro
Mayachemtools	Cálculo de propiedades y otros descriptores	Académico	www.mayachemtools.org/
Modeller	Modelado de proteínas y modelos por homología	Comercial/Académico	https://salilab.org/modeller/
MOPAC	Cálculo de propiedades cuánticas y cargas semiempíricas	Académico	www.openmopac.net
NAMD	Dinámica molecular	Académico	www.ks.uiuc.edu/Research/namd/
Openbabel	Conversión de formatos, cálculo de cargas, torsiones y fragmentos moleculares	Académico	www.openbabel.org
Vega ZZ	Cribado virtual	Académico	www.vegazz.net
Vina	Acoplamiento molecular	Académico	www.vina.scripps.edu
VMD	Cribado virtual	Académico	www.ks.uiuc.edu/Research/vmd/

*Se presentan sólo como ejemplos ilustrativos, también se ha dado prioridad a programas de fácil uso y acceso.

Tabla II. Ejemplos de software que permite realizar análisis quimioinformático y/o DIFAC*.

El uso correcto de los programas requiere conocimiento, aunque no extenso ni totalmente especializado. Por ello, a continuación se describirán algunos métodos quimioinformáticos, ejemplos y aplicaciones.

CRIBADO VIRTUAL: CUANDO LA COMPUTADORA FUNGE COMO UN FILTRO ANTES DE PRUEBAS DE LABORATORIO

Al proceso de identificar, diseñar o analizar el potencial farmacológico de una sustancia por métodos computacionales se le conoce como cribado virtual. Un protocolo de cribado virtual incluye cuando menos tres pasos: el curado (o preparación) de las estructuras, el filtrado y pruebas *in silico*. Una vez terminado este ciclo se pueden añadir pruebas, como un modelo descriptivo con los resultados obtenidos o iniciar el protocolo nuevamente para refinar los resultados.

El curado es el paso más lento pero determinante. Consiste en precisar de manera correcta las estructuras químicas a emplear, asignar la geometría, cargas y quiralidad adecuadas de los compuestos de interés. Los compuestos pueden proceder de bases de datos públicas, mismos que deben depurarse, ya que suelen

tener errores. La alternativa es usar compuestos identificados por un grupo de investigación o hacer la propuesta teórica de nuevas estructuras.

El filtrado consiste en la selección de compuestos que mejor cumplan con los criterios de interés. Las pruebas *in silico* son simulaciones que pueden incluir, por ejemplo, “similitud química” (descrita en párrafos posteriores), acoplamiento molecular y dinámica molecular. Hay que resaltar que los resultados del cribado virtual deben tomarse *cum grano salis*; es decir, no se puede aseverar que se ha descubierto o diseñado un fármaco hasta comprobarlo con pruebas biológicas. Sin embargo, aunque las técnicas computacionales no podrán sustituir a las pruebas biológicas, sí son herramientas invaluable para optimizar tiempo y recursos.

Una pregunta lógica es ¿el cribado virtual ha sido exitoso? La Tabla III presenta ejemplos representativos de fármacos identificados con la participación del cribado virtual. En las siguientes secciones se abordarán los métodos más comunes usados en quimioinformática.

Nombre	Farmacéutica que lo desarrolló	Uso
Captopril	Bristol Myers-Squibb	Inhibidor de la ECA (Antihipertensivo)
Oseltamivir	Gilead Sciences	Inhibidor de la neuroaminidasa (Antiviral)
Dorzolamida	Merck	Inhibidor de la AC- α (Glaucoma)
Aliskiren	Novartis	Inhibidor de renina (Antihipertensivo)
Boceprevir	Schering-Plough	Hepatitis C (Antiviral)

Tabla III. Ejemplos de fármacos desarrollados por métodos DIFAC.

ESPACIO QUÍMICO: EL MULTIVERSO DE LA QUÍMICA

El espacio químico se ha comparado con el universo cósmico y se han hecho analogías entre el número de compuestos químicos que puedan existir con el número de estrellas. Sobre el número de compuestos, estudios recientes hablan de valores que van desde 10^{40} hasta 10^{200} (Virshup *et al.*, 2013). Esto lleva a la reflexión de cuánto y qué tanto se conoce este “universo químico”. En la misma analogía con el universo cósmico se ha intentado catalogar y delimitar este espacio, por ejemplo, definir “galaxias químicas” en función de su actividad química, su reactividad, toxicidad, etc.

Para estudiar al espacio químico de N compuestos en forma sistemática (y matemática), se tienen que establecer las características de interés, que servirán como referencia para delimitar al espacio. Las características pueden ser una o más y usualmente se les llama descriptores. De esta manera, se habla en forma genérica de un espacio químico de N compuestos descritos por M características. Estas características pueden ser propiedades fisicoquímicas o propiedades electrónicas, entre otras. Si dos compuestos están cercanos en el espacio químico se infiere que tienen características semejantes y se puede inferir que tendrán un comportamiento químico semejante. Al enunciado anterior se le conoce como “principio de similitud” y aunque intuitivo, es una herramienta útil en la búsqueda de fármacos (Maggiore *et al.*, 2013), ya que la similitud (estructural) en principio puede extenderse a otros rubros como la actividad biológica.

El análisis y visualización del espacio químico se hace cada vez más complicado cuando con cada descriptor la representación física deseada va aumentando en complejidad, por ejemplo, una clasificación con cinco propiedades será entonces un espacio pentadimensional (difícil de imaginar en términos cotidianos). De manera que la representación visual del espacio químico requiere de una simplificación para facilitar la tarea y pasar de las cinco del ejemplo (pentadimensional) a tres dimensiones.

La reducción dimensional puede hacerse por diferentes métodos matemáticos. Uno de ellos se conoce como análisis de componentes principales (Wallisch, 2014). El método consiste en identificar las propiedades de “mayor peso” en

la representación o sea, la distribución de las moléculas en el espacio químico se ve más afectada por las propiedades que se conservan menos entre compuesto y compuesto. Una vez identificadas se pueden usar estas propiedades, para representar el espacio multidimensional que tendrá una distribución análoga al original y poder así identificar compuestos semejantes.

Representación visual del espacio químico: dependencia con la representación molecular

El espacio químico depende directamente de los descriptores usados; por tanto, es un concepto *relativo* a las M características empleadas para definirlo. Para ilustrar este punto la Figura 1 presenta dos representaciones visuales del espacio químico de la misma serie de compuestos. La diferencia entre ambos gráficos es el grupo de descriptores empleados para hacer la representación. En la Figura 1A las características o descriptores corresponden a la “complejidad estructural” de las moléculas; en la Figura 1B se representa la “flexibilidad”. La figura muestra que la distribución y, por tanto, la “similitud” relativa de los compuestos puede cambiar significativamente cuando se modifica la representación molecular. Así, a diferencia del espacio cósmico, el espacio químico es *relativo* a la descripción molecular. Esta discusión nos lleva a la siguiente pregunta: ¿cómo cuantificar la similitud de dos moléculas en forma consistente?

SIMILITUD MOLECULAR, UNA CUESTIÓN DE PERSPECTIVA

Cuando se habla de objetos semejantes se infiere que existe algún parecido, generalmente de naturaleza física. Por ejemplo, una naranja y una mandarina son notablemente semejantes, por ejemplo, respecto al tamaño, color, forma, etc. En cambio, una manzana y una naranja no resultan similares a primera vista. No obstante, es posible encontrar similitudes entre estos dos objetos (frutas) si se consideran otros aspectos, por ejemplo, el contenido de azúcar o vitamina C.

Una forma directa y rápida para comparar estructuras químicas en forma sistemática es mediante el tipo de átomos y cómo están conectados entre sí para establecer la semejanza de los grupos funcionales. Ahora, considere las tres comparaciones de dos moléculas en la Figura 2. La similitud del tipo y conexión de los átomos en el ciclohexano y benceno (Figura 2A) es muy alta. Sin embargo, reconocer características semejantes se va haciendo más difícil al tratar de igualar la identidad y conectividades de

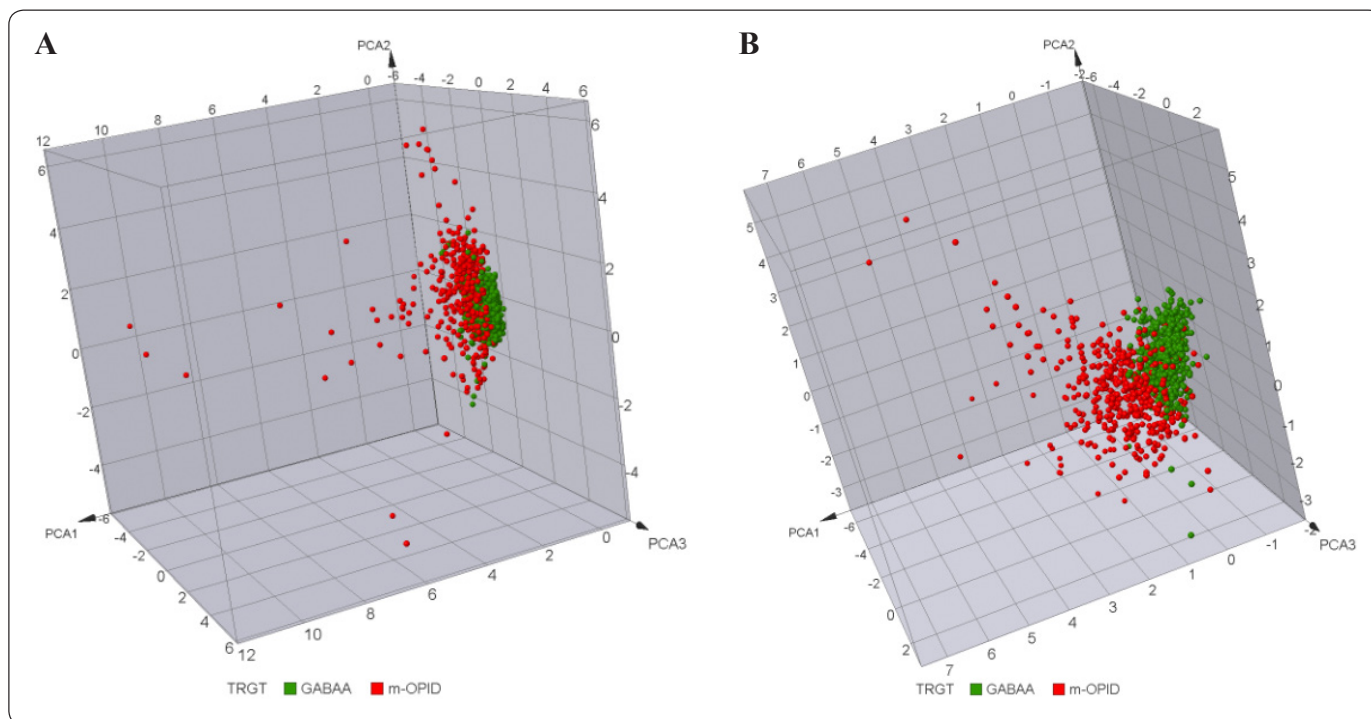


Figura 1. Espacio químico de antagonistas opioides y agonistas del receptor de GABA. A) Usando descriptores de “complejidad molecular”. B) Usando descriptores de flexibilidad molecular. Ver la versión a color en línea.

los átomos en los pares de moléculas de las Figuras 2B (ácido shikímico y L-tirosina) y 2C (zolpidem y alprazolam).

Para medir cuantitativamente la similitud entre dos moléculas en forma consistente, son necesarios dos elementos: A) una representación sistemática y consistente de las estructuras y B) una métrica (o coeficiente) de similitud (Bender *et al.*, 2009).

Una representación sistemática de estructuras químicas: Huellas digitales moleculares

Una forma de codificar de manera única (sistemática) la identidad y conectividad de los átomos es a través de las “huellas digitales moleculares” (*molecular fingerprints*). Se pueden distinguir dos tipos principales: unas basadas en los fragmentos y otras en la topología (Saldívar-González *et al.*, 2017). Los primeros utilizan fragmentos de referencia y se hace un “pase de lista” sobre las estructuras químicas marcando como presente (1) o ausente (0) cada fragmento del listado. Una de las limitaciones de éste método es que, al conocer una huella digital de una molécula, se puede saber si este fragmento está presente o no, pero no se conoce su ubicación en la estructura. Además, la representación depende de los fragmentos que comprenden la lista de referencia. Esto quiere decir que se pierde información sobre la estructura, específicamente sobre su conectividad. Ejemplos de estas huellas digitales son MACCS *keys* que se compone de 166 fragmentos (o más) y Pubchem *fingerprint* con 881 fragmentos.

En las huellas digitales basadas en la topología, se emplea la conectividad propia de las moléculas. El proceso inicia seleccionando un átomo central, para determinar los enlaces y átomos vecinos a una distancia determinada. Estos “caminos”, serán los que se comparen entre una estructura y otra. Por tanto, en este caso se construye una representación “más fiel” de la molécula, con el inconveniente de que la similitud será en general, más baja respecto a la representación por fragmentos pues las estructuras deberán tener conectividad semejante.

Es importante mencionar que las huellas digitales moleculares no son infalibles, ya que existen situaciones donde no serán capaces de distinguir estructuras con suficiente sensibilidad, para arrojar una representación única, esto es muy frecuente en isómeros. Por esta razón, cuando se trabaja con enantiómeros hay que estar bien seguro de que la estereoquímica sea considerada correctamente en pasos posteriores de un cribado.

Cuantificación de la similitud molecular

En la sección anterior se discutió que hay huellas digitales moleculares que representan a las moléculas por medio de fragmentos o subestructuras. Entonces, es posible hacer una relación matemática sencilla para medir su parecido. Esta operación se conoce como Coeficiente de Tanimoto:

$$T(X, Y) = \frac{c}{a + b - c}$$

donde a y b representan al número de fragmentos totales que tienen los compuestos X e Y ; mientras que c representa el número de fragmentos que comparten ambos compuestos. Al ser un cociente, el valor de similitud sólo podrá ir de 0 (compuestos que no comparten ningún fragmento por ejemplo, totalmente diferentes) y 1 (compuestos iguales o que comparten exactamente el mismo número de fragmentos) (Prieto-Martínez *et al.*, 2016). Otro caso sería, retomando la pareja de estructuras en la Figura 2C: usando MACCS *keys* el coeficiente de Tanimoto entre ambos compuestos es de 0.44 mientras que al usar una representación topológica (ECFP4) es de 0.09.

Si bien el coeficiente de Tanimoto es una métrica ampliamente usada, hay otras métricas que permiten evaluar la similitud. Particularmente, como una función de distancia, se pueden hacer

comparaciones matemáticas como la Distancia Euclidiana, Manhattan o Soergel. Otros coeficientes que se pueden emplear son: Similitud coseno o Coeficiente Sorensen-Dice (Bajusz *et al.*, 2015).

En la práctica, ¿cuál es la utilidad de cuantificar la similitud entre dos moléculas? Buscar un uso alternativo para fármacos aprobados usando el principio de similitud, esto se conoce como *reposicionamiento*. Por ejemplo, se sabe que el omeprazol y sus derivados, además de ser útiles en el tratamiento de la gastritis, tienen una actividad antiparasitaria significativa *in vitro* (Pérez-Villanueva *et al.*, 2011).

La similitud molecular también es una herramienta que permite identificar moléculas que se unen a más de un receptor.

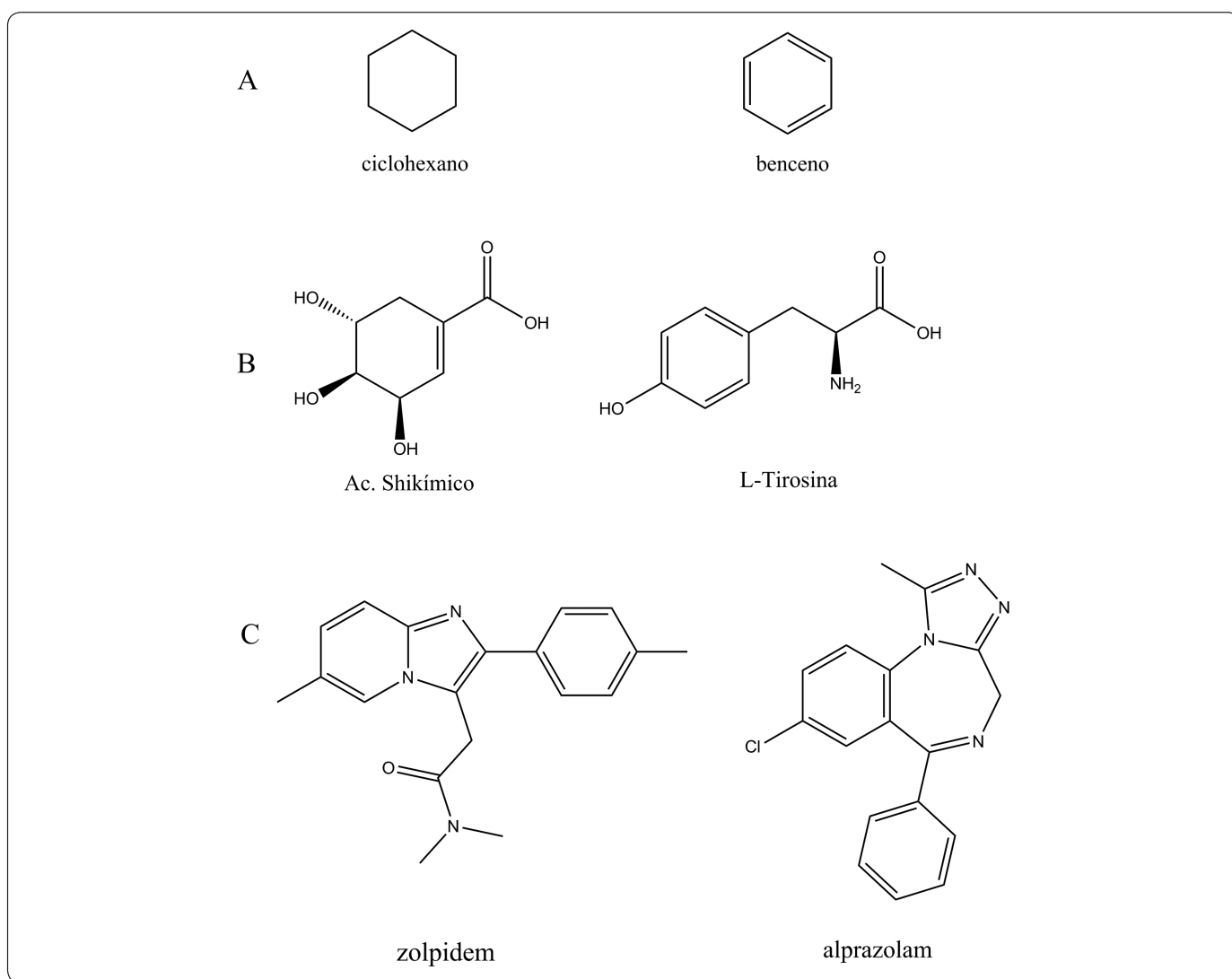


Figura 2. Comparaciones entre pares de moléculas con algún grado de similitud. A) Similitud estructural. B) Similitud bioquímica. C) Similitud farmacológica.

Anteriormente, estas moléculas se descartaban usualmente por sus potenciales efectos secundarios no deseables (Hu & Bajorath, 2010). Sin embargo, mediante el reposicionamiento es posible identificar moléculas que actúan como “llaves maestras” y efectivas en varios padecimientos (Méndez-Lucio *et al.*, 2016). Un ejemplo son los flavonoides; particularmente hesperidina, pues posee actividades antiparasitarias, antiinflamatorias, sedantes, antioxidantes, etc., convirtiéndolos en polifármacos.

La baja similitud estructural también es útil. Considere el caso siguiente: de acuerdo al espacio químico resulta que una serie de compuestos propuestos tiene propiedades semejantes a fármacos antivirales. Tras realizar comparaciones de Tanimoto resultan valores de similitud estructural muy bajos. ¿Esto quiere decir que los compuestos deben descartarse? Por el contrario, ya que por el principio de similitud es posible que exista una actividad semejante, valdría la pena probar éstos compuestos. Si al final resultan activos, se habrá identificado una nueva estructura antiviral (un bioisómero).

En la siguiente sección se aborda el tema de acoplamiento molecular y la simulación de interacciones proteína-ligando.

ACOPLAMIENTO MOLECULAR

Actualmente, es posible (con ciertas limitaciones), hacer aproximaciones del reconocimiento e interacción molecular. Un acoplamiento básico puede realizarse sin mayor problema en una computadora portátil y no tarda más de unos minutos (en algunos casos se hace en segundos). El proceso se puede realizar en tres formas generales: acoplamiento rígido, flexible y covalente. Este último no se tratará a detalle, por ser un caso muy específico que implica otras metodologías auxiliares. Si el lector está interesado en conocer más del tema, recomendamos estas excelentes referencias (De Cesco *et al.*, 2017; Kumalo *et al.*, 2015).

El acoplamiento rígido se basa en el modelo clásico de Emil Fischer: la llave y la cerradura. En esta aproximación se asume que ligando y proteína están rígidos durante su interacción. Otro tipo de acoplamiento se deriva del modelo de ajuste inducido, considerando que la unión del ligando y la proteína se debe a una influencia directa y flexibilidad entre sí (Ferreira *et al.*, 2015). Para modelar las interacciones, se usan campos de fuerza y las ecuaciones de Coulomb, además de las atracciones electrostáticas y fuerzas débiles como la de Lennard-Jones y Van der Waals (Pagadala *et al.*, 2017). Por tanto, detrás de escena, lo que realiza la computadora es una serie de cálculos matemáticos que describen las interacciones entre ambas partes. El proceso global se puede resumir en dos pasos: el programa propone una serie de “soluciones” al problema y posteriormente selecciona a la más óptima como la correcta.

Actualmente, se usa con frecuencia un punto intermedio como aproximación: el acoplamiento semi-flexible donde el ligando

se considera “libre” y desplazándose, mientras que el receptor permanece inmóvil. Esto permite un buen balance entre el tiempo de cálculo y la fiabilidad de los resultados. La búsqueda se realiza mediante un algoritmo que genera diferentes conformaciones del ligando en forma sistemática o estocástica (al azar) y las empaqueta con la proteína. Una vez identificadas las diferentes formas de organización molecular, el resultado es la diversidad artísticamente configurada y apreciada en las figuras de este trabajo, evaluadas por una función matemática que calcula la energía de cada conformación al interactuar con la proteína. Este valor de puntaje es el criterio de selección para la solución más óptima, es decir aquella de menor energía. Este proceso, a pesar de simple y elegante, tiene varios problemas. Para comenzar, la selección del modo de unión de menor energía como la correcta no es realista: en prácticamente ningún caso el ligando se encuentra en su conformación más estable cuando interactúa con una proteína. Otro problema actual es que hasta ahora no se ha desarrollado un algoritmo capaz de considerar en forma “natural” la flexibilidad y su influencia en el reconocimiento. A esto se añade, que en pocos casos se considera la contribución del agua durante la unión proteína-ligando, debido a que aún es complicado modelar en forma rápida y precisa las moléculas del agua en forma explícita (Spyrakakis & Cavasotto, 2015). Sin embargo, existen varias excepciones a la regla de forma que hay que conocer bien el sistema que se va a simular.

Por todo lo expuesto hasta ahora debe validarse un protocolo de acoplamiento. La validación consiste en determinar la capacidad del programa y su algoritmo para identificar el modo de unión de una molécula de referencia que puede ser el sustrato natural o un inhibidor característico de la misma (Verdonk *et al.*, 2007). Generalmente es posible encontrar el complejo de referencia en el Protein Data Bank (Tabla I). Una vez identificados referencia y modos de unión, se realiza el protocolo de acoplamiento, calculados con el programa (ver Figura 3). Idealmente, lo que se espera es que ambos modos de unión se superpongan en su totalidad. En la práctica se utiliza la métrica de la raíz cuadrada de la desviación media (RMSD, por sus siglas en inglés) para hacer la comparación. Lo que este valor nos dice es la distancia promedio entre todos los átomos de cada modo de unión. Esta distancia debe encontrarse por debajo de los 2 Å para considerar una validación exitosa (Kitchen *et al.*, 2004).

Una vez validado el protocolo de acoplamiento, es posible considerar la energía asignada como un criterio de búsqueda para proponer nuevos compuestos activos. No obstante, es importante recordar que la energía de unión no es relevante si no se conservan las interacciones conocidas con la proteína. Por ejemplo, considere el caso de los bromodominios, pequeñas proteínas que participan en la regulación y expresión genética (Ferri *et al.*, 2016). Recientemente se ha observado que los bromodominios inician el reclutamiento de genes y factores asociados a la inflamación, neoplasia y metabolismo.

En la literatura se reporta que una característica crítica de los inhibidores y sustratos es formar dos puentes de hidrógeno con los residuos 97 y 140 (ver Figura 4). La figura presenta un inhibidor característico (llamado JQ1) con las interacciones esperadas con la diana, mientras que las dos moléculas restantes tienen energías de unión comparables, pero no así sus interacciones.

Para superar estas limitaciones es posible “orientar” al programa de acoplamiento. Indicando las características que debe buscar en las moléculas y las interacciones que forman. Esto se conoce como “modelo farmacofórico” y puede obtenerse antes o después de un protocolo de acoplamiento. Otro problema puede surgir cuando la proteína que se desea acoplar no se encuentra en el Protein Data Bank. Esto puede deberse a que es una proteína recién descubierta o simplemente no ha logrado cristalizarse. Ante esto, puede realizarse un modelo tridimensional de la proteína o un llamado “modelo por homología”. Ambos

“modelos” se comentarán brevemente, así como su utilidad en el diseño racional y optimización de fármacos.

MODELOS FARMACOFÓRICOS

¿Qué le confiere actividad biológica a una molécula? En farmacología se dice que la actividad está fundamentada en las relaciones estructura-actividad. En Química Farmacéutica estas relaciones son la base para optimizar las propiedades de las moléculas a partir de un núcleo activo. Como se comentó en la sección de similitud molecular (*vide supra*), estructuras diferentes pueden tener actividades semejantes (bioisómeros). El descubrimiento de núcleos alternativos es muy útil, ya sea para aumentar la selectividad o mejorar el perfil farmacocinético de una sustancia. No obstante, las implicaciones que esto conlleva indican que las relaciones estructura-actividad tienen limitaciones importantes. Además, se han reportado numerosos casos donde las modificaciones químicas sobre un núcleo activo tienen un resultado paradójico:

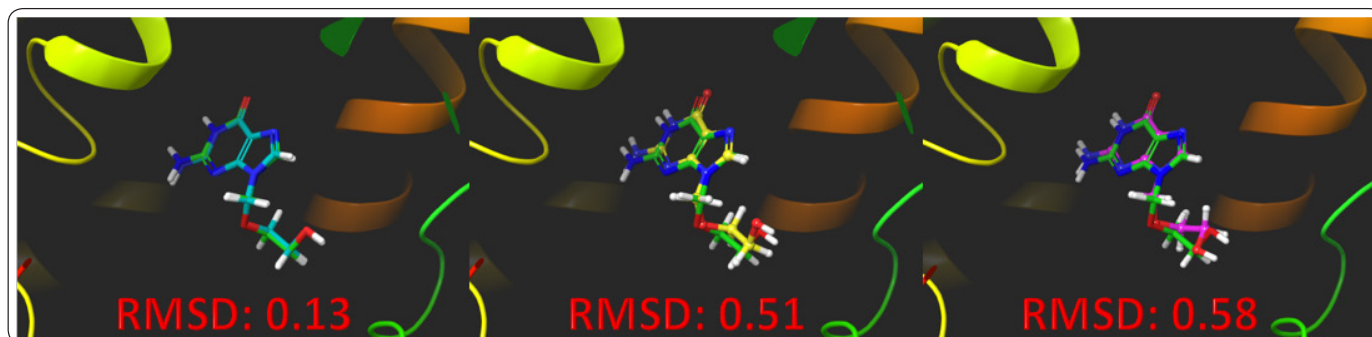


Figura 3. Validación de un acoplamiento molecular hipotético, usando RMSD. Ver la versión a color en línea.

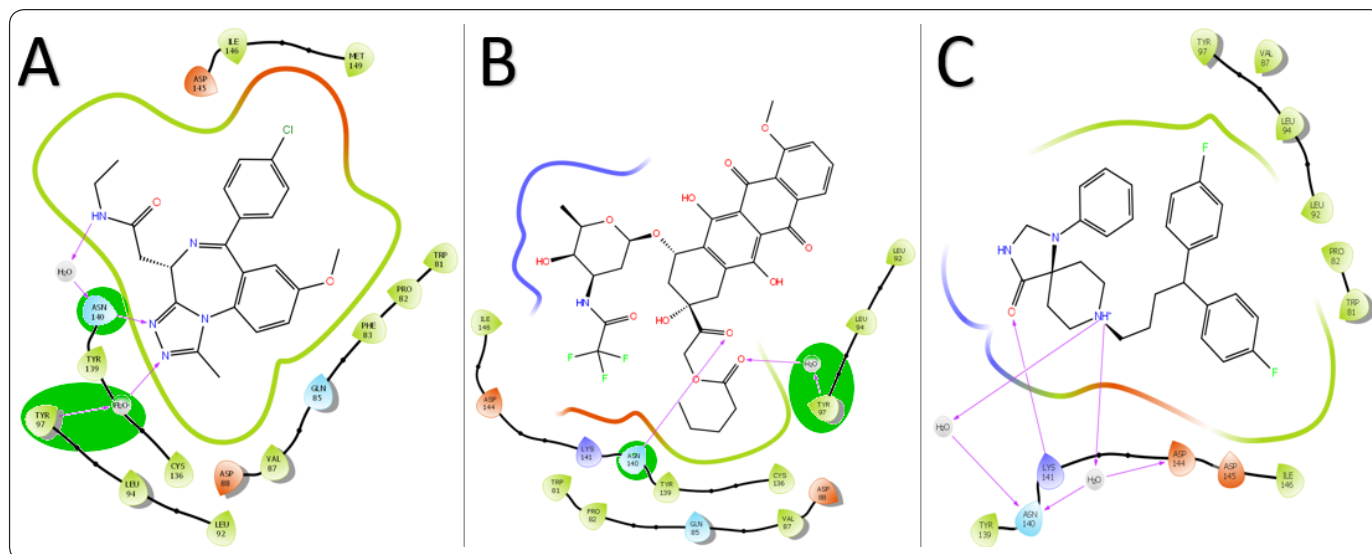


Figura 4. Interacciones moleculares para el BRD4 determinadas por acoplamiento molecular. A) Fármaco de referencia. B) Inhibidor potencial. C) Molécula inactiva. Las interacciones más importantes se resaltan en óvalos grises grandes. Ver la versión a color en línea.

modificaciones que deberían mejorar la actividad, provocan una disminución notable o una pérdida total de la misma. Estos “acantilados de actividad” no han podido explicarse en forma satisfactoria (y pueden estar relacionados con errores experimentales (Medina-Franco, 2013)). Sin embargo, se ha observado que núcleos específicos son más propensos a éste comportamiento. En cualquier caso, todo lo anterior obliga al desarrollo de una mejor representación.

Un modelo farmacofórico, es una representación abstracta de los requerimientos estructurales que “exige” el sitio activo de una proteína/receptor, para reconocer una molécula. Esta descripción incluye: átomos donadores/aceptores de hidrógeno, elementos hidrofóbicos; así como la distancia y ubicación espacial de cada elemento (Koes & Camacho, 2011). La Figura 5 muestra la estructura de Erlotinib, un fármaco antineoplásico. Note los diferentes puntos en la estructura: todos son elementos farmacofóricos. No obstante, de todos ellos, sólo las esferas (donadores de hidrógeno) y los grupos aromáticos son los que confieren su actividad biológica (como antineoplásicos).

Recientemente, la clasificación se ha extendido a elementos de carácter aromático y elementos voluminosos. Cuando se identifican las características cruciales para el reconocimiento

y actividad, se pueden identificar los rasgos “haptofóricos”, es decir, aquellos que ayudan al reconocimiento, pero son prescindibles, ya que no confieren actividad a un compuesto. De esta manera, las características haptofóricas son las que se pueden modificar para optimizar una estructura.

Las ventajas de un modelo farmacofórico son notables. En primera instancia, permiten identificar estructuras diversas, siempre y cuando se ajusten a los requerimientos. Además, pueden usarse como criterios de orientación para refinar los resultados de un acoplamiento molecular. La principal desventaja es que, para ser robusto, el modelo debe tener la mayor cantidad de información posible sobre el receptor e inhibidores conocidos, por lo que deberá ser actualizado constantemente para mantener su aplicabilidad y vigencia (Clayton *et al.*, 2007).

De manera semejante al acoplamiento molecular, un modelo farmacofórico debe validarse para determinar su aplicabilidad y capacidad de predicción. La manera de hacerlo es mediante un análisis de Característica Operativa del Receptor (ROC, por sus siglas en inglés). En general, el análisis consiste en determinar la capacidad del modelo para distinguir entre resultados positivos y falsos positivos

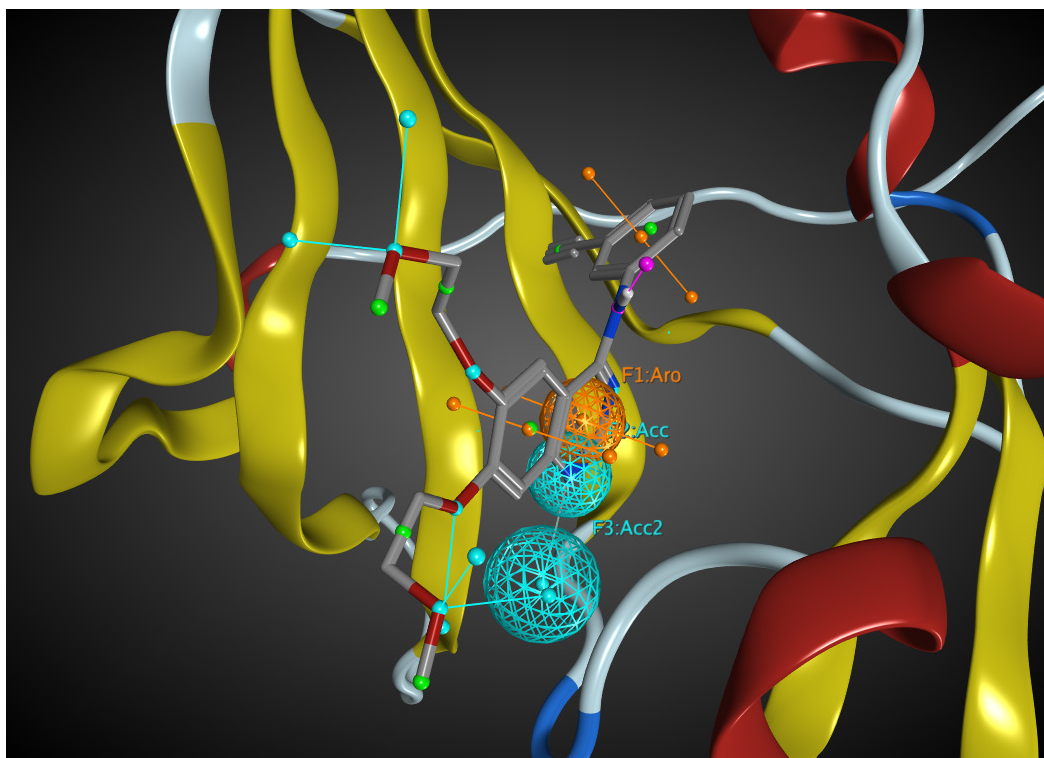


Figura 5. Modelo farmacofórico de Erlotinib, los puntos son elementos de posible interacción; las esferas son las asociadas directamente con la actividad biológica. Ver la versión a color en línea.

(Brown & Davis, 2006) (ver Figura 6). Lo anterior se hace al graficar el número de ambos eventos y ver el comportamiento de la curva. Cuando los resultados son puramente al azar (no existe ningún modelo valioso) se describe una línea recta que corta el plano en diagonal. Un modelo ideal, será aquel que identifica correctamente sólo a los casos positivos formando una “escuadra” en el plano. En consecuencia, un modelo propuesto se encontrará entre la diagonal y la escuadra. Si por alguna razón, la curva se ubica por debajo de la diagonal, quiere decir que el modelo es “engañado” fácilmente por falsos positivos, de hecho, hay mejor oportunidad de encontrar un caso positivo al azar.

MODELOS POR HOMOLOGÍA

Cristalizar una proteína es un proceso muy delicado y laborioso para lograr una buena resolución de la estructura. Esto se debe principalmente a las limitaciones de los métodos y la pericia del cristalógrafo para asignar correctamente los aminoácidos y su densidad electrónica. Se considera como buena resolución, un margen de error entre 1.5 y 3 Å en la asignación.

Además, existen proteínas que son bastante difíciles de cristalizar como las proteínas de membrana. Actualmente, el Protein Data Bank es una base de datos pública que recoge más de 100,000 estructuras cristalográficas. Lamentablemente el porcentaje de estructuras únicas y de resolución razonable es pequeño (Warren *et al.*, 2012). Ante esto cabe la pregunta: ¿cuántas proteínas son “modulables” mediante fármacos? Algunos estimados de la década pasada hablaban de alrededor de 300 (actualmente este

número se ha duplicado (Santos *et al.*, 2016). En ese entonces el Protein Data Bank se componía de 30,000 estructuras, donde cerca del 60% representaba sólo 34 proteínas de interés farmacéutico (Mestres, 2005).

De esta manera, en ocasiones no hay acceso a receptores o proteínas específicas para hacer pruebas de acoplamiento. Una alternativa a esto es crear un modelo por homología de la proteína de interés. El proceso consiste en crear la “estructura de una “proteína virtual” usando como base (o plantilla) estructuras ya conocidas y cristalizadas. Es probable que el resultado sea una “quimera” proteica, que igualmente requiere un análisis para determinar que fue construida correctamente y que sea útil para el acoplamiento molecular. En resumen, es necesario determinar que la secuencia del modelo sea válida, que su geometría sea correcta, que la estructura se mantenga estable y que el sitio activo sea capaz de reconocer al ligando (Hillisch *et al.*, 2004).

Es cierto que la utilidad real de un modelo por homología puede ser limitada, pero es una herramienta útil para apoyar el diseño de fármacos a dianas, cuyas estructuras tridimensionales no son bien conocidas. Los resultados pueden ser muy significativos (Duan *et al.*, 2015; Leffler *et al.*, 2017).

CONCLUSIONES

Los métodos computacionales son muy diversos y versátiles. Actualmente la capacidad de éstos son bastante notables, probando su utilidad práctica, aunque todavía no sea posible contar con la simulación precisa de sistemas muy complejos. Lo importante es tener expectativas realistas aprovechando al máximo los recursos computacionales. Además, es necesario erradicar paradigmas erróneos como el “descubrimiento en un par de clics” y el uso de cálculos o análisis sin intervención racional del usuario. Debe existir un balance adecuado entre el conocimiento químico, el aporte computacional y el valor farmacológico.

De esta forma, se podrán optimizar tiempo y recursos. Ya sea, dando prioridad a la síntesis de ciertos compuestos o derivados, la presunta evidencia de un mecanismo de acción, reposicionando fármacos o simplemente para disminuir la cantidad de pruebas biológicas. Los métodos computacionales representan un nuevo horizonte en la ciencia, abriendo la puerta a nuevas posibilidades.

AGRADECIMIENTOS

Fernando D. Prieto Martínez, agradece al CONACYT la beca doctoral, No. 660465/576637. También se agradece a la Facultad de Química de la Universidad Nacional Autónoma de México el financiamiento recibido por medio del Programa de Apoyo a la Investigación y el Posgrado (PAIP), número 5000-9163. Agradecemos a Tatiana Enríquez Gómez y Ana Martínez Jiménez por la lectura crítica a este manuscrito.

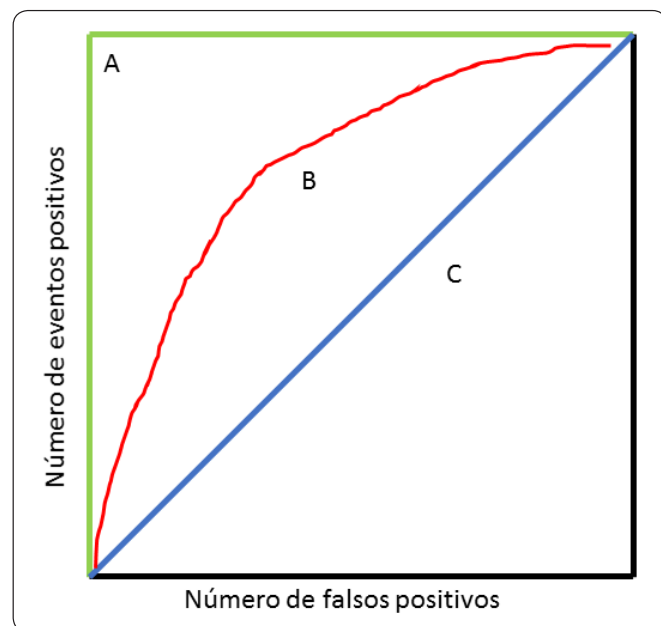


Figura 6. Representación de una curva ROC. A) Representa un modelo ideal. B) Representa un modelo hipotético. C) Representa resultados al azar. Ver la versión a color en línea.

REFERENCIAS

- Bajusz, D., Rácz, A., & Héberger, K. (2015). Why is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminformatics*, **7**, 20. DOI: 10.1186/s13321-015-0069-3
- Bender, A., Jenkins, J. L., Scheiber, J., Sukuru, S. C. K., Glick, M., & Davies, J. W. (2009). How Similar are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *J. Chem. Inf. Model.*, **49**, 108–119. DOI: 10.1021/ci800249s
- Brown, C. D., & Davis, H. T. (2006). Receiver Operating Characteristics Curves and Related Decision Measures: A Tutorial. *Chemometr. Intell. Lab.*, **80**, 24–38. DOI: 10.1016/j.chemolab.2005.05.004
- Clayton, T., Chen, J. L., Ernst, M., Richter, L., Cromer, B. A., Morton, C. J. & Cook, J. M. (2007). An Updated Unified Pharmacophore Model of the Benzodiazepine Binding Site on Gamma-Aminobutyric Acid(A) Receptors: Correlation with Comparative Models. *Curr. Med. Chem.*, **14**, 2755–2775. DOI: 10.2174/092986707782360097
- De Cesco, S., Kurian, J., Dufresne, C., Mittermaier, A., & Moitessier, N. (2017). Covalent Inhibitors Design and Discovery. *Eur. J. Med. Chem.*, **138**, 96–114. DOI: 10.1016/j.ejmech.2017.06.019
- Duan, X., Zhang, M., Zhang, X., Wang, F., & Lei, M. (2015). Molecular Modeling and Docking Study on Dopamine D2-Like and Serotonin 5-HT2A Receptors. *J. Mol. Graph. Model.*, **57**, 143–155. DOI: 10.1016/j.jmgm.2015.01.014
- Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, **20**, 13384–13421. DOI: 10.3390/molecules200713384
- Ferri, E., Petosa, C., & McKenna, C. E. (2016). Bromodomains: Structure, Function and Pharmacology of Inhibition. *Biochem. Pharmacol.* **106**, 1–18. DOI: 10.1016/j.bcp.2015.12.005
- Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of Homology Models in the Drug Discovery Process. *Drug Discov Today*, **9**, 659–669. DOI: 10.1016/S1359-6446(04)03196-4
- Hu, Y., & Bajorath, J. (2010). Molecular Scaffolds with High Propensity to form Multi-Target Activity Cliffs. *J. Chem. Inf. Model.*, **50**, 500–510. DOI: 10.1021/ci100059q
- Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.*, **3**, 935–949. DOI: 10.1038/nrd1549
- Koes, D. R., & Camacho, C. J. (2011). Pharmer: Efficient and Exact Pharmacophore Search. *J. Chem. Inf. Model.*, **51**, 1307–1314. DOI: 10.1021/ci200097m
- Kumalo, H. M., Bhakat, S., & Soliman, M. E. S. (2015). Theory and Applications of Covalent Docking in Drug Discovery: Merits and Pitfalls. *Molecules*, **20**, 1984–2000. DOI: 10.3390/molecules20021984
- Leffler, A. E., Kuryatov, A., Zebroski, H. A., Powell, S. R., Filipenko, P., Hussein, A. K. & Holford, M. (2017). Discovery of Peptide Ligands Through Docking and Virtual Screening at Nicotinic Acetylcholine Receptor Homology Models. *P. Natl. Acad. Sci. USA*, **114**, E8100–E8109. DOI: 10.1073/pnas.1703952114
- Maggiore, G. M., Vogt, M., Stumpfe, D., & Bajorath, J. J. (2013). Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.*, **57**, 3186–3204. DOI: 10.1021/jm401411z
- Medina-Franco, J. L. (2013). Activity Cliffs: Facts or Artifacts? *Chem. Biol. Drug Des.*, **81**, 553–556. DOI: 10.1111/cbdd.12115
- Méndez-Lucio, O., Naveja, J. J., Vite-Caritino, H., Prieto-Martínez, F. D., & Medina-Franco, J. L. (2016). Review. One Drug for Multiple Targets: A Computational Perspective. *J. Mex. Chem. Soc.*, **60**, 168–181.
- Mestres, J. (2005). Representativity of Target Families in the Protein Data Bank: Impact for Family-Directed Structure-Based Drug Discovery. *Drug Discovery Today*, **10**, 1629–1637. DOI: 10.1016/S1359-6446(05)03593-2
- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for Molecular Docking: A Review. *Biophys. Rev.* **9**, 91–102. DOI: 10.1007/s12551-016-0247-1
- Pérez-Villanueva, J., Romo-Mancillas, A., Hernández-Campos, A., Yépez-Mulia, L., Hernández-Luis, F., & Castillo, R. (2011). Antiprotozoal Activity of Proton-pump Inhibitors. *Bioorg. Med. Chem. Lett.*, **21**, 7351–7354. DOI: 10.1016/j.bmcl.2011.10.028
- Ponder, J. W., & Case, D. A. (2003). Force Fields for Protein Simulations. *Adv. Protein Chem.*, **66**, 27–85. DOI: 10.1016/S0065-3233(03)66002-X
- Prieto-Martínez, F. D., Fernández-de Gortari, E., Méndez-Lucio, O., & Medina-Franco, J. L. (2016). A Chemical Space Odyssey of Inhibitors of Histone Deacetylases and Bromodomains. *RSC Adv.*, **6**, 56225–56239. DOI: 10.1039/C6RA07224K
- Saldívar-González, F., Prieto-Martínez, F. D., & Medina-Franco, J. L. (2017). Descubrimiento y Desarrollo de Fármacos: Un Enfoque Computacional. *Educación Química*, **28**, 51–58. DOI: 10.1016/j.eq.2016.06.002
- Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G. & Overington, J. P. (2016). A Comprehensive Map of Molecular Drug Targets. *Nature Reviews Drug Discovery*, **16**, 19–34. DOI: 10.1038/nrd.2016.230
- Spyrakakis, F., & Cavasotto, C. N. (2015). Open Challenges in Structure-Based Virtual Screening: Receptor Modeling, Target Flexibility Consideration and Active Site Water Molecules Description. *Arch. Biochem. Biophys.*, **583**, 105–109. DOI: 10.1016/j.abb.2015.08.002
- Verdonk, M. L., Taylor, R. D., Chessari, G., & Murray, C. W. (2007). Illustration of Current Challenges in Molecular Docking. In *Structure-Based Drug Discovery* (pp. 201–221). Dordrecht: Springer Netherlands. DOI: 10.1007/1-4020-4407-0_8
- Virshup, A. M., Contreras-García, J., Wipf, P., Yang, W., & Beratan, D. N. (2013). Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *J. Am. Chem. Soc.* **135**, 7296–7303. DOI: 10.1021/ja401184g
- Wallisch, P. (2014). Principal Components Analysis. *MATLAB for Neuroscientists*, 305–315. DOI: 10.1016/B978-0-12-383836-0.00017-5
- Warren, G. L., Do, T. D., Kelley, B. P., Nicholls, A., & Warren, S. D. (2012). Essential Considerations for Using Protein-Ligand Structures in Drug Discovery. *Drug Discovery Today*, **17**, 1270–1281. DOI: 10.1016/j.drudis.2012.06.011