

Sección: Química Medicinal

COMPARACION SISTEMATICA DEL DESEMPEÑO DE DIFERENTES APROXIMACIONES DE CRIBADO VIRTUAL 2D Y 3D BASADAS EN EL LIGANDO PARA EL DESCUBRIMIENTO DE FARMACOS ANTICONVULSIVOS

Mauricio E. Di Ianni, Melisa E. Gantner, María E. Ruiz, Alan Talevi, Luis E. Bruno-Blanch.

Química Medicinal, Departamento de Ciencias Biológica, Facultad de Ciencias Exactas, Universidad Nacional de La Plata. Calle 47 y 115, La Plata (B1900AJI), Buenos Aires, Argentina. atalevi@biol.unlp.edu.ar.

INTRODUCCION. El cribado virtual es una técnica ampliamente utilizada en la búsqueda de nuevos fármacos así como también, más recientemente, en el reposicionamiento de compuestos activos [Drug Discovery Today, **2011**, 16, 298-310]. Consiste en explorar el extenso universo químico a través de algoritmos o modelos computacionales [Advanced Methods and Applications in Chemoinformatics, Castro, E; Haghi, A., Eds. Editorial IGI Global **2011**; pp 229-245]. Esencialmente, las técnicas de cribado virtual pueden clasificarse en métodos basados en el ligando o métodos basados en la estructura o blanco molecular [Current Medicinal Chemistry, 20, Issue 23, **2013**, 2839-2860]. Se han desarrollado una inmensa cantidad de algoritmos para describir los sistemas moleculares, entre los que podemos mencionar, huellas dactilares moleculares (*Molecular Fingerprints*) 2D, Fingerprints farmacofóricas, modelos QSAR 2D y 3D, funciones de puntaje basadas en Docking molecular, entre muchos otros [J. Biotechnol. Comput. Biol. Bionanotechnol. **2011**, 92, 249-264]. Estos métodos han surgido, en parte, por la necesidad de reducir los costos asociados al cribado de alto rendimiento (HTS) explorando el creciente espacio químico de compuestos tipo-fármaco de manera eficiente. Las diferentes técnicas poseen diversos grados de rendimiento en términos de porcentaje de enriquecimiento de activos y reconocimiento de andamiajes novedosos [J. Chem. Inf. Model. **2012**, 52, 1103–1113]. Se ha reportado que el uso de algunas aproximaciones 2D presenta generalmente mejores rendimientos en el cribado virtual en términos de enriquecimiento de activos y velocidad [J. Chem. Inf. Model. **2010**, 50, 2079–2093]. En contraste, las herramientas más complejas (aproximaciones basadas en el receptor y aproximaciones 3D basadas en el ligando) han mostrado en general mejores resultados con respecto a la detección de nuevos andamiajes [J Chem Inf Model **2012**, 52 (5), 1103-13]. Debe destacarse que el desempeño de cada metodología depende fuertemente del blanco molecular de interés. En función de estos resultados se han discutido las ventajas de combinar distintas aproximaciones [Curr. Comput. Aided Drug Des. **2009**, 5, 23–37]. En el presente trabajo, comparamos el rendimiento (en términos de enriquecimiento en activos) de diferentes herramientas computacionales basadas en el ligando para la búsqueda de nuevos agentes antiepilépticos a través de cribado virtual.

MATERIALES Y METODOS. Se generó una base de datos de 1000 estructuras, dentro de la cual se sembraron 50 moléculas con actividad reportada en el ensayo MES-test para determinación de actividad anticonvulsiva; las 950 estructuras restantes, de las cuales se desconoce su actividad frente al ensayo MES-test, fueron extraídas aleatoriamente de la base de datos de compuestos de acceso público ZINC [Chem. Inf. Model. **2005**, 45, 177–182] y categorizadas como inactivas. La conformación de mínima energía de cada

una de las moléculas fue hallada mediante un análisis conformacional estocástico utilizando el módulo *Conformational Search* del programa Hyperchem 8.0.3, utilizando el campo de fuerza MM+ con un gradiente de convergencia de 0.05 kcal/mol y refinando posteriormente el conformero de menor energía por el método PM3 (gradiente 0.01 kcal/mol).

Las herramientas de cribado virtual utilizadas fueron en su mayoría algoritmos de similitud basados en distintos sistemas de representación molecular, a saber, Fingerprints clásicas (FP2, FP3, FP4) calculadas con OpenBabel [*J Cheminform* **2011**, 3, 33], Extended Conectivity Fingerprints (ECFP) de JChem 6.1.0 (Chemaxon 2013), Functional Class Fingerprints (FCFP) de JChem, Fingerprints farmacofóricas 2D con JChem y PowerMV [*J Chem Inf Model* **2005**, 45 (2), 515-22], similitud basada en pares de átomos, pares de átomos de Carhart y pares de fragmentos con PowerMV (National Institute of Statistical Sciences 2005) y los scores Shape y Match de Jchem [*J Med Chem* **2006**, 49 (5), 1536-48]. Adicionalmente se utilizaron algoritmos *in-house*, entre los cuales se encuentra una función discriminante basada en descriptores topológicos 2D capaz de identificar fármacos activos al ensayo MES-test [*J. Comput. Aided Mol. Des.* **2007**, 21, 527-538], un algoritmo de similitud basado en pares de átomos de Carhart [*J. Enzyme Inhib. Med. Chem.* **2007**, 22(3), 253-265], y distintos esquemas de ordenamiento basados en la superposición de cada una de las estructuras minimizadas de la base de datos con un farmacóforo Sybyl [*Bioorg Med Chem* **2004**, 12:3857-3869]. Para las herramientas que implican una medida de similitud, se utilizó el coeficiente de Tanimoto y como molécula de referencia al antiepiléptico carbamazepina. Para evaluar la capacidad de recuperación de activos de cada uno de los métodos se utilizaron las curvas ROC (Receiver Operating Characteristic), una herramienta sencilla capaz de balancear los errores tipo I y II. El área bajo la curva ROC (AUC) fue nuestra medida de comparación, siendo que una clasificación perfecta o ideal daría un AUC igual a 1 mientras que una clasificación al azar es representada como una línea recta de pendiente 1 con un AUC igual a 0,5 [*Pattern Recogn. Lett.* **2006**, 27, 861-874.]

Adicionalmente, se estudió el efecto de realizar búsquedas por similitud con más de una referencia (utilizando como moléculas de referencia adicionales a Fenitoina, Fenobarbital, Diazepam y Felbamato) y finalmente el desempeño de distintos algoritmos de fusión de datos [*Journal of Cheminformatics* **2011**, 3:29] [*J. Chem. Inf. Model.* **2013**, 53, 1531-1542].

RESULTADOS Y CONCLUSIONES.

Nuestros resultados indican que la capacidad clasificatoria de ninguno de los modelos individuales es destacada, siendo el desempeño de los algoritmos de fusión significativamente superior a cualquiera de estos. Dentro de los algoritmos individuales, las huellas digitales circulares mostraron consistentemente mejores resultados. Se observó además que el porcentaje de activos recuperados por búsqueda por similitud, guarda una relación directa con el número de referencias utilizadas, en concordancia con publicaciones anteriores [*Journal of Cheminformatics* **2011**, 3:29]. Este estudio expone una vez más, las ventajas de aplicar una combinación de técnicas frente al uso de modelos individuales durante el filtrado de activos de grandes bases de datos.