

Aplicación de aprendizaje por ensamblado para la predicción de biodisponibilidad de fármacos en el sistema nervioso central

Autores: Morales, JF; Ruiz, ME; Talevi, A.

Lugar de trabajo: Cátedra de Química Medicinal, Depto. de CS. Biológicas, Fac. de Cs. Exactas, Universidad Nacional de La Plata. Calle 47 y 115 (B1900AVV). La Plata, Buenos Aires, Argentina (atalevi@biol.unlp.edu.ar).

Cátedra de Control de Calidad de Medicamentos, Depto. de CS. Biológicas, Fac. de Cs. Exactas, Universidad Nacional de La Plata. Calle 47 y 115 (B1900AVV). La Plata, Buenos Aires, Argentina.

Introducción

La barrera hematoencefálica constituye una barrera física y bioquímica que limita la biodisponibilidad de fármacos en el Sistema Nervioso Central (SNC). Por tal motivo es de interés el desarrollo de filtros *in silico* e *in vitro* para la predicción temprana de la biodisponibilidad de candidatos a fármacos: mientras que en el caso de los fármacos cuyo blanco molecular se encuentra en el cerebro interesa que los mismos presenten buena biodisponibilidad en el SNC, los fármacos que actúan periféricamente son frecuentemente optimizados para minimizar el riesgo de efectos adversos en el SNC, restringiendo su llegada al cerebro. Históricamente, la mayoría de los métodos predictivos se enfocaban en la distribución total del compuesto entre el plasma y el cerebro¹. Sin embargo, recientemente se ha propuesto que la razón de las concentraciones libres de plasma a cerebro en estado estacionario (K_p) es más relevante, dado que la acción farmacológica de un fármaco en el SNC depende no de la cantidad total sino de la cantidad de fármaco libre en el mismo^{2,3}. La predicción del parámetro K_p es un buen estimador de la biodisponibilidad *efectiva* de un fármaco a nivel del SNC, por cuanto se considera no sólo la capacidad del compuesto de superar la barrera hematoencefálica, sino también su interacción con proteínas plasmáticas y elementos de los tejidos. Para poder realizar esta estimación desarrollamos una serie de modelos QSAR clasificadores que posteriormente fueron ensamblados selectivamente para mejorar su capacidad predictiva.

Experimental

Se realizó una búsqueda bibliográfica de datos de K_p en mamíferos, obtenidos por las técnicas de Microdiálisis y *Slice*⁴, incluyendo en el *dataset* compilado únicamente compuestos con valores del parámetro menores a 1.4 (se considera que valores mayores indican pasaje activo del compuesto desde la sangre al cerebro). Se obtuvo un *dataset* de

89 fármacos. Se definieron dos categorías de fármacos utilizando un valor de corte de 0.4. Según nuestro criterio de clasificación nuestro *dataset* estaba compuesto por 31 fármacos con buena capacidad de difundir pasivamente a través de la BHE (categoría A) y 58 fármacos con biodisponibilidad central limitada (categoría B). Debido al escaso número de compuestos en la primera categoría, el conjunto de entrenamiento se integró con el total de compuestos de la categoría A y 31 compuestos de la categoría B elegidos al azar. Los 27 fármacos restantes de la categoría B conformaron nuestro conjunto de prueba.

Se calcularon 2013 descriptores moleculares independientes de la conformación con el programa Dragon 6.0 (Talete SRL, 2012); tales descriptores fueron aleatoriamente divididos en 100 subconjuntos de no más de 200. Los modelos se derivaron con el programa STATISTICA 8.0, aplicando el módulo Análisis Lineal Discriminante con una tolerancia de 0.5 y un número de pasos máximo de 6 a fin de minimizar la probabilidad de sobreajuste. Se obtuvieron así 100 modelos clasificadores, que fueron validados mediante técnicas estándar de validación interna y externa, y comparados mediante curvas *Receiver Operating Characteristic* (ROC). Los mejores modelos fueron combinados mediante aprendizaje por ensamblado selectivo (*ensemble learning*) utilizando los operadores Promedio, Mínimo y Máximo como técnicas de combinación.

Resultados y Discusión

El mejor modelo individual presentó un 93% y 85% de buenas clasificaciones en el conjunto de entrenamiento y conjunto de prueba, respectivamente. En la curva ROC correspondiente al conjunto de entrenamiento se obtuvo un área bajo la curva de 0.969. La combinación de los mejores modelos utilizando el operador Mínimo resultó en un rendimiento superior en términos de especificidad, dando lugar a un 87% y 96% de buenas clasificaciones en el conjunto de entrenamiento y conjunto de prueba, respectivamente, y un área bajo la curva de 0.974. Se prevé la validación experimental de la capacidad predictiva de esta combinación de modelos.

Referencias

1. Lanevskij, K., Japertas, P. & Didziapetris, R. Improving the prediction of drug disposition in the brain. *Expert Opin. Drug Metab. Toxicol.* **9**, 473–86 (2013).
2. Martin, I. Prediction of blood-brain barrier penetration: are we missing the point? *Drug Discov. Today* **9**, 161–2 (2004).
3. Summerfield, S. G. & Dong, K. C. In vitro, in vivo and in silico models of drug distribution into the brain. *J. Pharmacokinet. Pharmacodyn.* **40**, 301–14 (2013).
4. Fridén, M., Gupta, A., Antonsson, M., Bredberg, U. & Hammarlund-Udenaes, M. In vitro methods for estimating unbound drug concentrations in the brain interstitial and intracellular fluids. *Drug Metab. Dispos.* **35**, 1711–9 (2007).