

ESTUDIO QSAR INDEPENDIENTE DE LA CONFORMACIÓN MOLECULAR EN ACTIVADORES DEL RECEPTOR X DE PREGNANO

Pablo R. Duchowicz¹, Agustín Albanesi², Germán Cordero² y Eduardo A. Castro¹

¹ Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA, UNLP, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Suc. 4, 1900 La Plata, Argentina Correo: pabloducho@gmail.com

² Cátedra de Química Teórica y Computacional, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calle 115 y 47, 1900 La Plata, Argentina

1. Introducción

El receptor X de pregnano (PXR) es un receptor nuclear que al ser activado es capaz de detectar la presencia de sustancias tóxicas o extrañas en el cuerpo humano, y en respuesta, regula la expresión de las enzimas metabólicas implicadas en la detoxificación y eliminación de estas sustancias. El potencial rol de PXR también se reconoce en la resistencia a drogas de quimioterapia en varios tipos de cáncer, así como su asociación a enfermedades inflamatorias como la enfermedad de Crohn y la colitis ulcerativa, y a enfermedades cardiovasculares [1].

La proteína PXR es activada por un gran número de ligandos endógenos y exógenos de estructura heterogénea, incluyendo esteroides, ácidos biliares, drogas clínicas, suplementos alimenticios, hierbas medicinales y contaminantes ambientales

En base a información biológica experimental disponible de la literatura [2], decidimos desarrollar un estudio QSAR (Relaciones Cuantitativas Estructura-Actividad) [3] con el fin de predecir la activación del receptor PXR. Para este propósito, se emplean 160 ligandos de estructura heterogénea con actividad EC_{50} (concentración que demuestra una eficacia del 50 %). Se utilizan descriptores moleculares para representar las estructuras químicas que no dependan de las conformaciones, de manera que contemplen únicamente la naturaleza constitucional y topológica de las moléculas estudiadas. Además, se consideran descriptores flexibles (dependientes de la actividad experimental estudiada) en los modelos establecidos.

2. Metodología

El programa E-Dragon [4] permite calcular 941 descriptores que no dependen de la conformación molecular. El programa Coral [5] permite obtener 6 descriptores flexibles a través de diferentes definiciones. Ambos programas son de libre acceso. La representación de compuestos se realiza con los programas HyperChem [6] y ACD/ChemSketch [7].

Se divide el conjunto molecular en un conjunto de calibración (cal) con $N_{cal} = 107$ moléculas (67%) y otro de validación (val) con las restantes, mediante el método de partición Análisis de Agrupamiento de K-Medias [8].

Los modelos QSAR se establecen con un algoritmo matemático muy útil basado en la técnica del Análisis de Regresión Lineal Multivariable, llamado "Método del Reemplazo" (RM) [9, 10]. Permite seleccionar los mejores descriptores a partir de miles, de manera que se minimice la raíz cuadrada de la desviación cuadrática media (RMSD).

Cada modelo obtenido se valida por medio de las técnicas de Validación Cruzada 'Dejar-Uno-Afuera' (loo) [11] y Aleatorización-Y (rand) [12] con el fin de estimar su desempeño predictivo, y con el conjunto de validación externo. Todos los algoritmos

de Matlab [13] utilizados se desarrollaron en nuestro grupo de QSAR y están a disposición.

3. Resultados

Se presenta a continuación la mejor relación cuantitativa estructura-actividad encontrada a través del método RM, que involucra cinco descriptores moleculares:

$$-\log_{10} EC_{50} = -0.17(\pm 0.09) \cdot GATS3m - 0.091(\pm 0.06) \cdot H-049 - 0.030(\pm 0.01) \cdot H-052 + 0.047(\pm 0.01) \cdot DCW(EC1) + 0.070(\pm 0.008) \cdot DCW(EC2) + 4.37(\pm 0.1) \quad (1)$$

$$N_{cal} = 107, R_{cal} = 0.91, RMSD_{cal} = 0.30, F = 92$$

$$R_{ij}^{max} = 0.85, o(>3.S) = 1, R_{loo} = 0.90, RMSD_{loo} = 0.32, RMSD^{rand} = 0.60$$

$$N_{val} = 53, R_{val} = 0.70, RMSD_{val} = 0.40$$

R es el coeficiente de correlación lineal, F es el parámetro de Fisher, R_{ij}^{max} es el coeficiente de correlación lineal máximo entre descriptores, $o(>3.S)$ es el número de moléculas con residuo mayor a 3 veces la desviación estándar del modelo (S).

El poder predictivo de este modelo lineal es satisfactorio, según sugieren los parámetros estadísticos de calibración, validación externa y Validación Cruzada ($R \geq 0.7$) y Aleatorización-Y ($RMSD^{rand} > RMSD_{cal}$). La Ec. 1 posee $RMSD_{cal} = 0.30$ y $RMSD_{val} = 0.40$. Estos parámetros se comparan con los reportados en la literatura, $RMSD_{cal} = 0.27$ y $RMSD_{val} = 0.28$ [2], donde se utiliza un método híbrido más sofisticado denominado Máquina de Soporte Vectorial/ensamble de farmacóforo, pero que es fuertemente dependiente de la geometría tridimensional.

Se aprecia que los descriptores de Ec. 1 son [14]: el descriptor de autocorrelación 2D $GATS3m$, donde se usa la masa atómica para discernir entre átomos; $H-049$: el número de hidrógenos unidos a carbonos con diferente número de oxidación formal e hibridación (C3(sp3)/ C2(sp2)/ C3(sp2)/ C3(sp)); $H-052$: el número de hidrógenos unidos a carbonos de tipo C0(sp3), con un átomo electronegativo enlazado al próximo carbono; y los descriptores flexibles $DCW(EC1)$ y $DCW(EC2)$, que consideran en su formulación el número de átomos vecinos a un átomo considerado.

4. Conclusiones

El modelo establecido tiene la tendencia de predecir acertadamente la activación de PXR. La importancia del estudio actual radica en la habilidad de establecer un modelo QSAR que dependa solamente de los aspectos constitucionales y topológicos de las estructuras moleculares. Se evita así considerar el complejo problema conformacional, que significa modelar cada conformación de los ligandos para el posterior cálculo de descriptores dependientes de la geometría. Además, de esta manera se busca no depender de información sobre las conformaciones moleculares experimentales que causan la actividad biológica observada.

Agradecimientos. Se agradece a CONICET (PIP0151).

Referencias

[1] Rathod, V., Jain, S., Nandekar, P., Sangamwar, A. T., Drug Discovery Today, 19, 2014, 63.

- [2] C.- N. Chen, Y.- H. Shih, Y.- L. Ding, M. K. Leong, *Chemical Research in Toxicology*, 24, 2011, 1765.
- [3] C. Hansch, A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, D. C., 1995.
- [4] E-Dragon, Milano Chemometrics and QSAR Research Group, <http://michem.disat.unimib.it/chm/index.htm>
- [5] Coral Sea, <http://www.insilico.eu/coral/>
- [6] Hyperchem 6.03 (Hypercube, Inc.), <http://www.hyper.com>
- [7] ACD/ChemSketch 12.0, www.acdlabs.com
- [8] Kaufman, L., Rousseeuw, P. J. *Introduction to Cluster Analysis*, Wiley, New York, 2005.
- [9] P.R. Duchowicz, E. A. Castro, F. M. Fernández, M. P. González, *Chemical Physics Letters*, 412, 2005, 376.
- [10] P. R. Duchowicz, D. O. Bennardi, D. E. Baselo, E. L. Bonifazi, C. Rios-Luci, J. M. Padrón, G. Burton, R. I. Misico, *European Journal of Medicinal Chemistry*, 77, 2014, 176.
- [11] D. A. Konovalov, L. E. Llewellyn, Y. V. Heyden, D. Coomans, *Journal of Chemical Information and Modeling*, 48, 2008, 2081.
- [12] S. Wold, L. Eriksson, in: H. van de Waterbeemd (Ed.), *Chemometrics Methods in Molecular Design*, VCH, Weinheim, 1995.
- [13] Matlab 7.0, The MathWorks, Inc., <http://www.mathworks.com>
- [14] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley-VCH, Weinheim, 2009.