

AVANCES EN ESTUDIOS QSAR/QSPR CON APLICACIONES AGRONÓMICAS

José F. Aranda, Pablo R. Duchowicz y Eduardo A. Castro

Instituto de Investigaciones Físicoquímica Teóricas y Aplicadas (INIFTA), Dpto. de Química, Facultad de Ciencias Exactas, UNLP, Calle 64 Diag. 113, CP (1900), La Plata, Buenos Aires, Argentina.

jfaranda10@gmail.com

RESUMEN: El interés por las Relaciones Cuantitativas Estructura/Actividad-Estructura/Propiedad se ha originado en las últimas décadas, para predecir el impacto de un pesticida antes de ser liberado en el ambiente, mediante modelos matemáticos. Se establece un modelo QSPR para la predicción del coeficiente de sorción en suelo (K_{oc}) de 643 compuestos heterogéneos. Se obtiene el mejor modelo encontrado en base a un descriptor flexible. El poder predictivo es satisfactorio en la calibración, validación interna y externa.

Palabras clave (tres): Teoría QSPR-QSAR, Propiedades Agronómicas, Técnica de MLRA

El principal objetivo de la ciencia de los plaguicidas, es ser capaz de predecir el impacto de un pesticida antes de ser liberado en el ambiente [1]. Mediante la Teoría QSAR-QSPR se pueden realizar tales predicciones a partir de modelos matemáticos cuantitativos para describir la asociación entre la estructura molecular de un compuesto y sus propiedades fisicoquímicas y biológicas [2,3].

En Agroquímica, el coeficiente de sorción en suelo de un plaguicida (funguicida, herbicida o insecticida) tiene gran interés porque mide la retención de estos en la matriz del suelo.

Se establece un modelo QSPR para la predicción de coeficientes de sorción en suelo (K_{oc}) de 643 compuestos orgánicos heterogéneos, cuya información experimental es extraída de la literatura [4]. El programa Coral [5] permite obtener los descriptores flexibles a través de diferentes definiciones. 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 $N_{cal}=93$ moléculas y otro de validación (val) con las restantes $N_{val}=550$ al igual que en el trabajo previo [4]. La función que vincula los descriptores moleculares con la propiedad estudiada puede establecerse mediante un algoritmo matemático muy útil basado en la técnica del Análisis de Regresión Lineal Multivariable, llamado "Método del Reemplazo" (RM) [8], permite seleccionar los mejores descriptores a partir de miles, de manera que se minimice la desviación estándar del conjunto de calibración. Cada modelo obtenido se valida por medio de las técnicas de Validación Cruzada "Dejar-Uno-Afuera" (loo) [9] y Aleatorización-Y (rand) [10] con el fin de estimar su desempeño predictivo, y con el conjunto de validación externo. Todos los algoritmos de Matlab [11] utilizados se desarrollaron en nuestro grupo de QSAR y están a disposición. El mejor modelo QSPR encontrado en base a un descriptor flexible es:

$$-\log_{10} K_{oc} = 0.64 + 0.21DCW \quad (1)$$

$$N_{cal} = 93, d = 1, R_{cal}^2 = 0.90, S_{cal} = 0.45, F = 594$$

$$o(3S) = 0, R_{loo}^2 = 0.86, S_{loo} = 0.46, S^{rand} = 1.13$$

$$N_{val} = 550, R_{val}^2 = 0.71, S_{val} = 0.69$$

F es el parámetro de Fisher y $o(3S)$ es el número de moléculas con residuo mayor a 3 veces la desviación estándar del modelo (S). El poder predictivo es satisfactorio en la calibración, validación interna y externa, Validación Cruzada ($R^2 > 0.5$) y Aleatorización-Y ($S^{rand} > S_{cal}$). La calidad de la Ec. 1 se compara a la encontrada en la literatura [4].

Referencias

- [1] R.D. Wauchope, T. Buttler, A. Hornsby, P. Augustijn-Beckers, J. Burt, "The SCS/ARS/CES pesticide properties database for environmental decision-making", En *Reviews of environmental contamination and toxicology*, Springer, (1992), pp. 1-155.
- [2] D.W. Salt, N. Yildiz, D.J. Livingstone, C.J. Tinsley, *The use of artificial neural networks in QSAR, Pestic. Sci* 36 (1992) 161-170.
- [3] C. Hansch, *Quantitative approach to biochemical structure-activity relationships, Acc. Chem. Res.* 2 (1969) 232-239.
- [4] P. Gramatica, E. Giani, E. Papa, *Statistical external validation and consensus modeling: A QSPR case study for K_{oc} prediction, J. Mol. Graph. Model.* 25 (2007) 755-766.
- [5] Coral Sea, <http://www.insilico.eu/coral/>.
- [6] Hyperchem, 6.03, (Hypercube, Inc.), <http://www.hyper.com>.
- [7] ACD/ChemSketch 12, www.acdlabs.com.
- [8] P. R. Duchowicz, E. A. Castro, F.M. Fernández., *Alternative Algorithm for the Search of an Optimal Set of Descriptors in QSAR-QSPR Studies*, Commun. Math. Comput. Chem. (MATCH) 55 (2006) 179-192.
- [9] D.A. Konovalov, L.E. Llewellyn, Y. Vander Heyden, D. Coomans, Robust cross-validation of linear regression QSAR models, *Chem. Inf. Model.* 48 (2008) 2081-2094.
- [10] S. Wold, L. Eriksson, S. Clementi, "Statistical validation of QSAR results", *Chemometric methods in molecular design*, (1995), pp. 309-338.
- [11] Matlab 7.0, The MathWorks, Inc., <http://www.mathworks.com>.