

PREDICCIONES QSPR EN COEFICIENTES DE SORCION EN SUELO

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

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: ifaranda10@gmail.com

1. Introducción

El estudio del aire, agua y suelo presenta una gran complejidad pues abarca las interacciones entre estos sistemas, la biósfera y en particular, el accionar del hombre. Si bien los problemas ambientales son transdisciplinarios, la propuesta de un marco de trabajo y de herramientas disciplinares basados en la Química resultan necesarios para encarar su estudio. En los últimos años se ha prestado especial atención a la predicción de las propiedades de compuestos químicos existentes y nuevos, que puedan presentar una influencia benéfica o adversa al ser liberados en el medio ambiente y la salud humana. En Agroquímica, el coeficiente de sorción en suelo de un plaguicida (funguicida, herbicida o insecticida) tiene gran interés porque mide su persistencia, riesgo y distribución en el ecosistema.

Las distintas formulaciones de la Teoría de las Relaciones Cuantitativas Estructura-Propiedad (QSPR) [1] constituyen una herramienta matemática valiosa que, si son correctamente empleadas, permiten predecir las propiedades de las sustancias en una relación hipotética entre la estructura molecular y la propiedad. Las técnicas QSPR permiten el diseño y optimización de nuevos compuestos: buscar moléculas aún no sintetizadas que puedan tener buena actividad y baja toxicidad para el medio ambiente. Es decir, diseñar productos químicos más seguros, sin contaminar, uno de los objetivos de la Química Verde o Sustentable.

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 [2]. 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 propiedad experimental estudiada).

2. Metodología

El programa E-Dragon [3] permite calcular 941 descriptores que no dependen de la conformación molecular. El programa Coral [4] permite obtener 7 descriptores flexibles a través de diferentes definiciones. Ambos programas son de libre acceso. La representación de compuestos se realiza con los programas HyperChem [5] y ACD/ChemSketch [6].

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 [2].

Los modelos QSPR se establecen con un algoritmo matemático muy útil basado en la técnica del Análisis de Regresión Lineal Multivariable, llamado "Método del Reemplazo" (RM) [7, 8]. 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) [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.

3. Resultados

El mejor modelo QSPR encontrado es el siguiente:

$$-\log_{10} K_{oc} = -0.087(\pm 0.03) \cdot MAXDP + 0.155(\pm 0.05) \cdot ALOGPS_{\log P} + 0.108(\pm 0.01) \cdot DCW(EC2) + 0.620(\pm 0.1) \quad (1)$$

$$N_{cal} = 93, R_{cal} = 0.95, RMSD_{cal} = 0.38, F = 275$$

$$R_{ij}^{max} = 0.76, o(>3.S) = 0, R_{loo} = 0.95, RMSD_{loo} = 0.40, RMSD^{rand} = 0.50$$

$$N_{val} = 550, R_{val} = 0.88, RMSD_{val} = 0.57$$

RMSD es la raíz cuadrada de la desviación cuadrática media del modelo, *F* es el parámetro de Fisher y *o(>3.S)* 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 ($R \geq 0.7$) y Aleatorización-Y ($RMSD^{rand} > RMSD_{cal}$). La calidad de la Ec. 1 se compara a la encontrada en la literatura [2].

4. Conclusiones

Se logra establecer un modelo QSPR que correlaciona el coeficiente de sorción en suelo, y que logra compararse adecuadamente a los publicados en la literatura. La dependencia establecida involucra únicamente a descriptores de aspectos constitucionales y topológicos de los compuestos químicos.

Agradecimientos. Se agradece a CONICET (PIP0151).

Referencias

- [1] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, D. C., 1995.
- [2] P. Gramatica, E. Giani, E. Papa, Journal of Molecular Graphics and Modelling, 25, 2007, 755.
- [3] E-Dragon, Milano Chemometrics and QSAR Research Group, <http://michem.disat.unimib.it/chm/index.htm>
- [4] Coral Sea, <http://www.insilico.eu/coral/>
- [5] Hyperchem 6.03 (Hypercube, Inc.), <http://www.hyper.com>
- [6] ACD/ChemSketch 12.0, www.acdlabs.com
- [7] P.R. Duchowicz, E. A. Castro, F. M. Fernández, M. P. González, Chemical Physics Letters, 412, 2005, 376.
- [8] 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.
- [9] D. A. Konovalov, L. E. Llewellyn, Y. V. Heyden, D. Coomans, Journal of Chemical Information and Modeling, 48, 2008, 2081.

- [10]** S. Wold, L. Eriksson, in: H. van de Waterbeemd (Ed.), Chemometrics Methods in Molecular Design, VCH, Weinheim, 1995.
- [11]** Matlab 7.0, The MathWorks, Inc., <http://www.mathworks.com>