

---

# *Aplicación de la topología molecular para la predicción de biodegradación anaerobia de clorofenoles*

Teresa Delgado-Pérez, Melireth Holmquist, Laura Francés,  
Jorge Gálvez, Ramón García-Domenech\*  
Departamento de Química Física, Facultad de Farmacia, Universitat de València,  
Avda. V.A. Estellés, s/n, 46100- Burjassot, Valencia, España

---

*Recibido: 13 de diciembre de 2012; revisado: 27 de febrero de 2013; aceptado: 14 de marzo de 2013*

*Application of molecular topology for the prediction of anaerobic biodegradation of chlorophenols*

*Aplicació de la topologia molecular per a la predicció de biodegradació anaeròbia de clorofenols*

## RESUMEN

Se ha aplicado la topología molecular y el análisis de regresión multilineal en la búsqueda de un modelo de relación cuantitativa estructura-actividad (QSAR) capaz de predecir la capacidad de biodegradación anaerobia de un grupo de clorofenoles. Tomando el valor de la constante de velocidad de desaparición del derivado clorofenólico,  $\log k$ , como variable dependiente y los descriptores topológicos como variables independientes, se obtuvo una ecuación con tres variables y coeficientes de correlación y de predicción,  $R^2=0.9291$  y  $Q^2=0.870$ . El modelo fue validado a través de una validación interna y un test de aleatoriedad. Los resultados obtenidos muestran una alta capacidad de predicción para la biodegradación de los compuestos analizados.

**Palabras clave:** Biodegradación, Clorofenoles, Análisis QSAR, Análisis de regresión multilineal, Topología Molecular.

## SUMMARY

Molecular topology and multilinear regression analysis have been used to find a QSAR model capable of predicting the anaerobic biodegradation ability of a group of chlorophenols. Taking the value of the rate constant of chlorophenolic derivatives' degradation,  $\log k$ , as the dependent variable and topological descriptors as independent variables, a three-variable equation was obtained having a correlation coefficient ( $R^2$ ) of 0.9291 and cross-validated correlation coefficient  $Q^2 = 0.870$ . The model was tested through a validation and a test of randomness. The results show high predictability for biodegradation of the compounds analyzed.

**Key words:** Biodegradation, Chlorophenols, QSAR Analysis, multilinear regression analysis, Molecular Topology

## RESUM

S'ha aplicat la topologia molecular y la anàlisi de la regressió multilineal en la trobada d'un model de relació cuantitativa estructura-activitat (QSAR) capaç de predir la capacitat de biodegradació anaeròbia d'un grup de clorofenols. Agafant el valor de la constant de velocitat de desaparició del derivat clorofenòlic,  $\log k$ , com a variable dependent i els descriptores topològics com a variables independents, es va obtenir una equació amb tres variables i coeficients de correlació i de predicció,  $R^2=0.9291$  y  $Q^2=0.870$ . El model va ser validat a través d'una validació interna i un test de aleatorietat. Els resultats obtinguts mostren una alta capacitat de predicció per a la biodegradació dels compostos analitzats.

**Mots clau:** Biodegradació, Clorofenols, Anàlisi QSAR, Anàlisi de regression multilineal, Topologia Molecular.

---

\* Correspondencia autor: ramon.garcia@uv.es;  
Tel. +34 963544291

## 1. INTRODUCCIÓN

Los compuestos fenólicos, en particular los clorados, CPs, son compuestos orgánicos tóxicos, recalcitrantes y potencialmente carcinogénicos y/o mutagénicos [1]. Casi todos los CPs se absorben fácilmente en el organismo al ser ingeridos o a través de la piel. Los monoclorofenoles no permanecen por mucho tiempo en el cuerpo, son transformados en productos menos perjudiciales y la mayoría se eliminan por la orina en un período no superior a 24 horas. Los otros CPs (diclorofenoles, triclorofenoles y tetraclorofenoles), que también pueden eliminarse por la orina en la forma de otras sustancias químicas menos perjudiciales, pueden permanecer en el organismo durante varios días [2].

Estos compuestos orgánicos tienen una enorme importancia económica por su uso como biocidas y su empleo como precursores en la manufactura de productos farmacéuticos, pesticidas, productos químicos, etc. Es por ello que pueden hallarse en las aguas residuales de industrias tales como: petroquímica, papelería, química, farmacéutica, siderúrgica y curtidora. Además, en el pasado, éstos han sido la causa de enormes impactos ambientales –contaminando suelos y acuíferos– debido a su uso generalizado como preservantes en la industria de la madera. Actualmente, su uso para este fin está prohibido en casi todo el mundo, sin embargo, los problemas de contaminación persisten en muchos países madereros debido a su muy elevada vida media en el medio ambiente. La mayoría de los clorofenoles liberados en el medio ambiente va al agua, y en muy pocas cantidades al aire. Los compuestos que tienen más probabilidades de acumularse en el aire son los mono- y los diclorofenoles, ya que son los más volátiles. Una vez en el aire, la luz solar ayuda a deshacer estos compuestos y la lluvia los elimina del aire. Los clorofenoles se adhieren al suelo y a sus sedimentos en el fondo de los lagos, los ríos y los arroyos. Sin embargo, algunos niveles bajos de clorofenoles en el agua, el suelo o los sedimentos se descomponen por la acción de los microorganismos y se eliminan del medio ambiente en cuestión de días o semanas.

Son importantes los procesos de biotransformación y biodegradación anaeróbica y en la determinación del destino de los contaminantes en el subsuelo y los sedimentos. La biodegradación anaeróbica de los CPs de inicia con el proceso de descloración reductiva, donde el cloruro enlazado es eliminado con la adición simultánea de electrones a la molécula [3].

La descloración reductiva es muy importante desde el punto de vista medioambiental, ya que genera productos menos clorados y menos tóxicos que son degradados más fácilmente por microorganismos aeróbicos y anaeróbicos [4].

Desde la década de los 80, se han realizado predicciones de biodegradabilidad usando distintos datos como parámetros en los métodos de predicción, incluyendo modelos, tendencias y sistemas de predicción de biodegradación. Uno de los métodos más recientes empleados en este campo es el QSAR (sigla en inglés de relaciones cuantitativas estructura-actividad). El modelo QSAR es, en este caso, una ecuación lineal que relaciona las variaciones en la actividad biológica con las variaciones en los valores de las propiedades calculadas (o medidas) para una serie de moléculas. Para que el método sea eficiente, el grupo de prueba debe ser diverso estructuralmente. En

muchos grupos, los compuestos sintetizados son similares a la estructura líder [5]. En definitiva, el análisis QSAR establece la relación entre la estructura química y las propiedades experimentales de las moléculas. La caracterización de la estructura molecular se puede realizar usando diferentes formalismos como la química cuántica [6], las propiedades fisicoquímicas [3], la mecánica molecular [6,7] y, más recientemente, la topología molecular [8,9].

En este sentido, la topología molecular se ha convertido en una herramienta útil en la búsqueda de modelos QSAR. A partir de ella se puede caracterizar de una forma rápida y sencilla aspectos estructurales de una molécula relacionados con el tamaño, forma y ramificación a través de los correspondientes descriptores topológicos [10,11].

La aplicación de la topología molecular en la búsqueda de modelos de predicción QSAR se ha utilizado recientemente con éxito en el campo de la química sostenible [12-14]. Por la versatilidad de la topología molecular y el uso de técnicas QSAR, ha sido posible la realización del presente estudio, en donde se analiza la velocidad de la degradación de los CPs, a través de la constante de velocidad,  $\log K$ , siendo ésta considerada como un descriptor de la tendencia del proceso de la descloración reductiva.

El objetivo del presente trabajo es, por tanto, construir un modelo topológico-matemático basado en el análisis de regresión multilínea para predecir la biodegradación anaerobia de un grupo de clorofenoles.

## 2. MATERIALES Y MÉTODOS

### 2.1 Compuestos analizados

Un conjunto de dieciséis clorofenoles ha sido analizado en este estudio, cuya nomenclatura química [15] y descriptores seleccionados por el modelo QSAR han sido recopilados en la Tabla I. Los valores de  $\log k$  experimentales para cada uno de los CPs analizados fueron obtenidos por Takeuchi et al. en procesos con cinéticas de primer orden en un laboratorio anaerobio con lodos metanogénicos que no habían sido expuestos a clorofenoles [16].

**Tabla I** - *Compuestos analizados, valores experimentales de biodegradación,  $\log k$ , y descriptores seleccionados por el modelo QSAR.*

Compuesto	Nomenclatura	$\log k_{\text{exp}}^{\dagger}$	${}^3C_p^v$	$J_s^v$	${}^3C_c/\beta C_c^v$
1	Pentaclorofenol (PCP)	-0.793	3.710	0.00064	0.928
2	2,3,4,5-tetraclorofenol (2,3,4,5-TeCP)	-1.137	3.029	0.00071	0.962
3	2,3,4,6-tetraclorofenol (2,3,4,6-TeCP)	-1.097	2.802	0.00071	0.942
4	2,3,5,6-tetraclorofenol (2,3,5,6-TeCP)	-1.319	2.816	0.00000	0.942
5	2,3,4-triclorofenol (2,3,4-TCP)	-1.092	2.446	0.00079	0.999
6	2,3,5-triclorofenol (2,3,5-TCP)	-1.377	2.092	0.00000	0.985
7	2,3,6-triclorofenol (2,3,6-TCP)	-1.523	2.235	0.00000	0.972
8	2,4,5-triclorofenol (2,4,5-TCP)	-1.252	2.124	0.00079	0.986
9	2,4,6-triclorofenol (2,4,6-TCP)	-1.456	1.850	0.00079	0.961
10	2,3-diclorofenol (2,3-DCP)	-1.602	1.829	0.00000	1.064
11	2,4-diclorofenol (2,4-DCP)	-1.292	1.497	0.00089	1.037

12	2,5-diclorofenol (2,5-DCP)	-1.745	1.514	0.00000	1.037
13	2,6-diclorofenol (2,6-DCP)	-1.620	1.604	0.00000	1.025
14	3,5-diclorofenol (3,5-DCP)	-1.678	1.324	0.00000	1.051
15	3-clorofenol (3-CP)	-1.553	1.065	0.00000	1.211
16	2-clorofenol (2-CP)	-1.456	1.202	0.00000	1.211

\* valores de  $\text{Log } k_{\text{exp}}$  obtenidos de la referencia [16].

## 2.2 Descriptores topológicos

La caracterización de las estructuras de cada uno de los clorofenoles analizados se realizó calculando un conjunto de índices topológicos. Para ello, cada molécula se asimila a un grafo en donde cada vértice representa a un átomo y cada arista un enlace.

A partir del grafo, se construye la matriz topológica cuyos elementos adquieren valores nulos o no en función de la existencia (o no) de conexiones entre dichos elementos en el grafo. La manipulación matemática de la matriz topológica conduce a los índices topológicos que caracterizan a ese grafo de una manera rápida y sencilla. En este estudio se han usado diferentes tipos de descriptores: constitucionales (número de ramificaciones R, nº de vértices del grafo N, longitud del grafo L, pares de ramificaciones PRi, etc) [17], índices topológicos de carga [18] e índices de conectividad de Kier y Hall [19, 20]. Cada compuesto ha sido caracterizado por un conjunto de 62 descriptores. Todos los descriptores han sido obtenidos con el programa Desmol1 [21].

## 2.3 Análisis de regresión multilínea

El objetivo general del análisis de regresión multilínea, ARML, es encontrar la relación entre dos o más variables independientes y una variable dependiente mediante el ajuste de una ecuación lineal a partir de los datos observados. Para la predicción de la velocidad de biodegradación de los clorofenoles el ARML se usó como método estadístico. Con este fin, se usó el software Statistica versión 8.0. [22]. La selección de las variables se ha llevado a cabo por medio del algoritmo Furnival-Wilson de modo que se emplea el conjunto de variables con el mínimo valor del Cp de Mallows [23].

La calidad de la predicción y la robustez del modelo deben ser verificadas usando diferentes tipos de criterios de validación. Normalmente, se adoptan cuatro estrategias [24]: (a) validación cruzada tipo "leave-one-out", LOO, (b) validación interna dividiendo el conjunto de datos en compuestos de entrenamiento y de test; (c) validación externa real por aplicación del modelo a datos externos; y (d) randomización de los datos o Y-scrambling. En este trabajo se han utilizado tres de estos métodos (a, b, y d) como criterios de validación.

## 3. RESULTADOS Y DISCUSIÓN

Existen pocos estudios relacionados con la búsqueda de modelos QSAR para la predicción de la biodegradación de clorofenoles. En 2006 Youzhi Dai et. al., usando descriptores principalmente mecanocuánticos, obtuvieron una ecuación lineal con 15 variables [15]. Al tratarse de un grupo de clorofenoles relativamente pequeño, N=16, hay bastante probabilidad de que el modelo seleccionado se

encuentre sobreajustado y difícilmente aplicable a otros compuestos no utilizados por el grupo de entrenamiento. En este trabajo usamos un reducido grupo de descriptores topológicos, de fácil y rápido cálculo, para la búsqueda de un modelo QSAR de predicción de la biodegradación anaerobia de un grupo de clorofenoles.

Se seleccionó la mejor función de tres variables. La Tabla II recoge los descriptores y coeficientes que los acompañan junto al resto de parámetros estadísticos.

**Tabla II - Parámetros estadísticos de la ecuación seleccionada con tres variables en el análisis de regresión multilínea realizado.**

Descriptor	Coficiente	EE	p
Intercepto	-3,74	0,48	0,00001
${}^3\chi_p^v$	0,35	0,05	0,00001
$J_5^v$	325,68	59,87	0,00015
${}^3\chi_c^e/{}^3\chi_c^v$	1,51	0,40	0,00254

$N=16$   $R^2 = 0,9221$   $Q^2 = 0,8698$   $EEE = 0,080$   $F(3,12)=47,3$   
 $p < 0,000001$

Si pasamos a evaluar los resultados obtenidos, se puede observar que los índices seleccionados son todos estadísticamente significativos por encima del 99% ( $p < 0,002$ ). Esto nos indica que la información topológica y estructural es suficiente para cuantificar la biodegradabilidad de los CPs. En esta ecuación aparecen tres descriptores topológicos distintos: cociente de índices de conectividad de tipo 3-cluster ( ${}^3\chi_c^e/{}^3\chi_c^v$ ), que evalúa la ramificación en posición para  ${}^3\chi_p^v$  que tiene en cuenta el ensamblaje topológico de la molécula y finalmente índices de carga,  $J_5^v$ , que evalúa la transferencia de carga intramolecular entre átomos situados a 5 enlaces de distancia.

La ecuación en su conjunto es altamente significativa, con un valor de  $p < 0,000001$ , es capaz de explicar más del 92% de la varianza de la propiedad correlacionada ( $R^2 = 0,9221$ ) y con un error estándar de estimación por debajo del 10% del rango en el que se mueve la propiedad ( $EES = 0,080$ ).

Los resultados de predicción obtenidos para cada compuesto se muestran en la Tabla III y en la Figura 1.

**Tabla III - Resultados de predicción obtenidos para cada compuesto a partir del modelo QSAR seleccionado.**

Compuesto	$\text{Log } k_{\text{exp}}$	$\text{Log } k_{\text{calc}}$	$\text{Log } k_{\text{calc(ve)}}$
1	-0,793	-0,841	-0,884
2	-1,137	-1,005	-0,969
3	-1,097	-1,115	-1,118
4	-1,319	-1,340	-1,349
5	-1,092	-1,127	-1,134
6	-1,377	-1,528	-1,559
7	-1,523	-1,497	-1,491
8	-1,252	-1,260	-1,261
9	-1,456	-1,392	-1,364
10	-1,602	-1,500	-1,486
11	-1,292	-1,367	-1,409
12	-1,745	-1,650	-1,631
13	-1,620	-1,637	-1,639
14	-1,6778	-1,696	-1,699
15	-1,553	-1,543	-1,535
16	-1,456	-1,496	-1,528

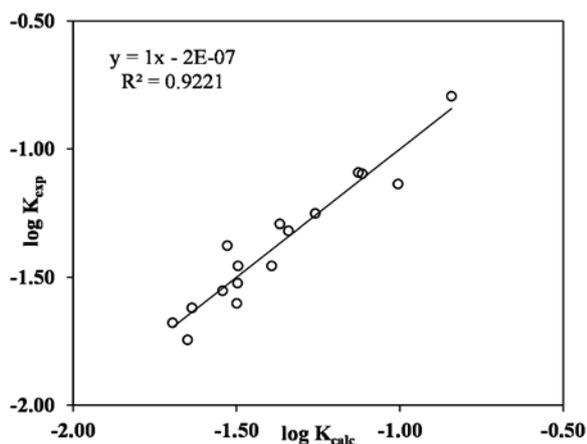


Figura 1.- Representación de  $\log k_{exp}$  frente  $\log k_{calc}$  a partir del modelo QSAR seleccionado.

En la Figura 2 se puede observar como todos los compuestos se acomodan al modelo propuesto, no existiendo ningún compuesto con un residual ( $\log K_{exp} - \log K_{calc}$ ) superior a dos veces el error estándar de estimación, EEE.

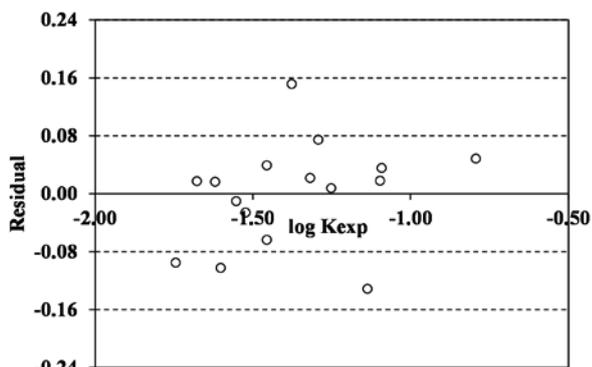


Figura 2.- Representación gráfica de los residuales obtenidos para cada compuesto analizado.

El siguiente paso fue la validación del modelo matemático topológico seleccionado. En primer lugar se llevó a cabo la validación cruzada tipo "leave-one-out". Para ello se reserva un compuesto del grupo de entrenamiento, se busca la mejor ecuación usando los tres descriptores seleccionados y con ella se predice el valor del compuesto reservado. El proceso se repite tantas veces como compuestos forman parte del grupo de trabajo. La Tabla III, columna 4, muestra los resultados obtenidos. El coeficiente de predicción resultante,  $Q^2 = 0.8699$  es similar al valor obtenido con anterioridad para la serie completa,  $R^2 = 0.9221$ , lo cual refuerza la calidad de la ecuación seleccionada.

En segundo lugar se hizo una validación interna, donde se dividió el grupo de entrenamiento en 4 subgrupos (I-IV), cada uno de los cuales contenía el 20% de los compuestos analizados. El primer subgrupo está formado por los compuestos 1, 2, 3 y 4; el segundo subgrupo por los compuestos 5, 6, 7 y 8; el tercer subgrupo contiene a los compuestos 9, 10, 11 y 12 mientras que los compuestos 13, 14, 15 y 16 forman parte del subgrupo IV.

Tres de los cuatro subgrupos (I, II y III), (I, II y IV), (II, III y IV), (I, III y IV) se usaron como grupo de entrenamiento, permaneciendo el cuarto subgrupo (IV), (III), (I), (II) como grupo test. Los resultados obtenidos se encuentran recopilados

en la Tabla IV, donde podemos ver los valores de  $R^2$  y  $R^2_{pred}$  para cada uno de los grupos de entrenamiento y test.

Tabla IV.- Resultados de la validación interna del modelo QSAR seleccionado.

Grupo entrenamiento	Grupo test	$R^2$ (entrenamiento)	$R^2_{pred}$ (test)
I, II, III, IV	-	0,922	-
II, III, IV	I	0,864	0,863
I, III, IV	II	0,945	0,843
I, II, IV	III	0,946	0,862
I, II, III	IV	0,906	0,662
Promedio	-	0,915	0,808

Como se puede observar, para el grupo de entrenamiento los valores de la varianza  $R^2$  del grupo completo y el promedio de las cuatro series son muy similares (0,922 y 0,915, respectivamente) siendo un poco más pequeña para el grupo test (0,808). Estos resultados indican que el modelo es predictivo y estable aun cuando se elimine el 25% de los compuestos del grupo.

El último test de validación del modelo fue el de aleatoriedad. Este test es adecuado realizarlo cuando se trabaja con una propiedad para la cual el intervalo numérico en el que fluctúa es pequeño (en nuestro caso nos movemos entre -0,793 y -1,745). El test consiste en intercambiar de forma aleatoria el valor de la propiedad entre los compuestos del grupo y correlacionar con los descriptores seleccionados en el modelo. El proceso se repite varias veces. La Figura 3 muestra los resultados obtenidos al graficar los valores de  $R^2$  frente a  $Q^2$  para cada regresión realizada (en nuestro caso 10). Como podemos observar, las regresiones obtenidas son muy malas, excepto la que corresponde con los valores reales de la propiedad para cada compuesto (punto señalado en negro en la Figura 3). El test realizado demuestra que el modelo es estable y no aleatorio.

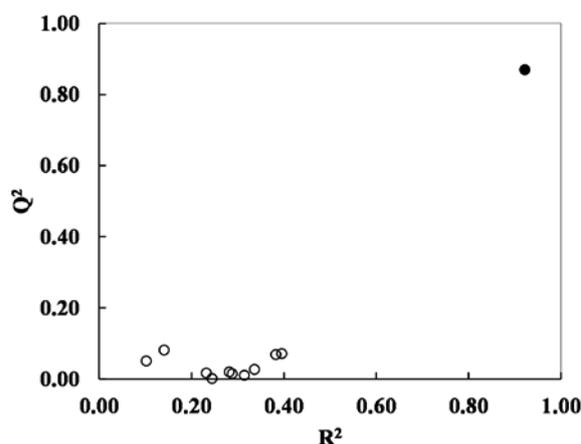


Figura 3.- Representación gráfica del coeficiente de predicción  $Q^2$  frente coeficiente correlación  $R^2$ , obtenidos en el test de aleatoriedad.

Youzhi Dai et al. [15] realizaron un estudio de predicción para el mismo grupo de compuestos, clorofenoles, y propiedad regresionada, velocidad de biodegradación anaerobia. Para ello utilizaron diferentes descriptores estructurales tales como calor de formación, energía electrónica, polarizabilidad molecular, momento dipolar,  $E_{homo}$ ,  $E_{lumo}$ , cargas atómicas etc, todos ellos obtenidos con ayuda del método MP3 usando MOPAC. El modelo de predicción

seleccionado por Youzhi Dai et al. mostraba una varianza  $R^2=0.7827$ , sensiblemente inferior al modelo propuesto en este trabajo,  $R^2=0.9221$ , a pesar de utilizar descriptores estructurales más específicos y complejos en relación con los topológicos. La suma de los errores al cuadrado obtenida ( $SSE=0.2165$ ) era también superior a la obtenida por nuestro modelo ( $SSE=0.0767$ ).

Podemos concluir indicando que la topología molecular es una herramienta útil para construir un modelo QSAR con capacidad predictiva para la velocidad de biodegradación de clorofenoles de manera satisfactoria. Los descriptores moleculares usados en este estudio fueron obtenidos a partir del grafo de las respectivas moléculas y la manipulación matemática de la matriz topológica resultante, de manera sencilla y rápida. El modelo matemático-topológico seleccionado conteniendo tres descriptores, es robusto y estable y, por tanto, con adecuada capacidad predictiva, si bien el dominio de aplicabilidad del modelo se ve restringido al ámbito estructural específico del grupo de clorofenoles.

## AGRADECIMIENTOS

Esta investigación ha sido financiada por el Master Interuniversitario de Química Sostenible (Universitat de Valencia, Curso 2011-2012).

## BIBLIOGRAFÍA

1. <http://www.bvsde.paho.org/bvsaidis/aresidua/peru/argtar033.pdf>.
2. [http://www.atsdr.cdc.gov/es/phs/es\\_phs107.html](http://www.atsdr.cdc.gov/es/phs/es_phs107.html)
3. Dai, Y.; Yang, D. and Zhu, F. The QSPR (quantitative structure-property relationship) study about the anaerobic biodegradation of chlorophenols. *Chemosphere*, 65, 2427-2433, 2006.
4. Tiedje, J.M.; Boyd, S.A. and Fathepure, B.Z. Anaerobic degradation of chlorinated aromatic hydrocarbons. *Developments in Industrial Microbiology*, 27, 117-127, 1987.
5. García-Domenech, R.; García-Mujica, P.; Gil, U.; Casanova, C.; Mireilli Beltrán, J. and Gálvez, J. Búsqueda de modelos QSAR para la actividad repelente de sesquiterpenos naturales frente al mosquito de la Fiebre Amarilla, *Aedes aegypti*. *Afinidad LXVII*, 547, 187-192, 2010.
6. Weinstein, H.; Osman, R. and Green, J. P. The molecular basis of structure-activity relationships: quantum chemical recognition mechanisms in drug-receptor interactions. *J. Am. Chem. Soc.*, 112,161-187, 1979.
7. Johnson, M. A. and Maggiora, G. M. *Concepts and applications of molecular similarity*. Wiley, New York, 1990.
8. Basak, S. C.; Grunwald, G. D. and Niemi, G. J. From chemical topology to three-dimensional geometry, 73, 1997.
9. Garcia-Domenech, R.; Galvez, J.; De Julian-Ortiz, J.V. and Pogliani, L. Some New Trends in Chemical Graph Theory. *Chemical Reviews*, 108, 1127-1169, 2008.
10. Kier, L.B. and Hall, L.H. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, London, 1976.
11. Todeschini, R. and Consonni, V. *Molecular Descriptors for Chemoinformatics*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Second Edition, 2009.
12. Galvez, J.; Gálvez-Llompart, M. and García-Domenech, R. Application of molecular topology for the prediction of the reaction yields under solvent-free conditions. *Green Chem.*, 12, 1056-1061, 2010.
13. Pla-Franco, J.; Gálvez-Llompart, M.; Gálvez, J. and García-Domenech, R. Application of molecular topology for the prediction of reaction yields and anti-inflammatory activity of heterocyclic amidine derivatives. *Int. J. Mol. Sci.*, 12, 1281-1292, 2011.
14. Gálvez, J.; Parreño, M. and Pla, J. Application of molecular topology to the prediction of water quality indices of alkylphenol pollutants. *J. Inter. J. of Chemoinformatics and Chem. Eng.*, 1, 1-11, 2011.
15. Dai, Y.; Yang, D. and Zhu, F. The QSPR (quantitative structure-property relationship) study about the anaerobic biodegradation of chlorophenols. *Chemosphere*, 65, 2427-2433, 2006.
16. Takeuchi, R.; Suwa, Y.; Yamagishi, T. and Yonezawa, Y. Anaerobic transformation of chlorophenols in methanogenic sludge unexposed to chlorophenols. *Chemosphere*, 41, 1457-1462, 2000.
17. Galvez, J.; Garcia-Domenech, R.; De Julián-Ortiz, J. and Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.*, 35, 272-284, 1995.
18. Gálvez, J.; García-Domenech, R.; Salabert, M.T. and R.Soler. Charge indexes: new topological descriptors. *J.Chem. Inf. Comput. Sci.*, 34, 520-525, 1994.
19. Kier, L.B.; Murray, W.J.; Randic, M. and Hall, L.H. Molecular connectivity V: connectivity series concept applied to density. *J. Pharm. Sci.*, 65, 1226-1230, 1976.
20. Kier, L. B. and Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, Letchworth, UK, 1986.
21. Desmol1 software. Molecular Connectivity and Drug Design Research Unit. Department of Physical Chemistry. Faculty of Pharmacy. University of Valencia, 2000.
22. Statistica 8.0,2008.Statsoft Inc <http://www.statsoft.com>
23. Hocking, R.R. Criteria for selection of a subset regression: Which one should be used?. *Technometrics*, 14, 967-970, 1972.
24. Roy P.P.; Leonard, J.T. and Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics Intell. Lab. Syst.*, 90, 31-42, 2008.