



**FACULTAD DE MEDICINA**

**Trabajo de titulación previo a la obtención de título de médico**

**Uso de descriptores moleculares en la identificación de efectos gastrointestinales por fármacos antiinflamatorios, analgésicos y anestésicos, Cuenca 2018**

**Autoras:**

María Gabriela Palomeque Zeas

Lorena Micaela Pesántez Rodríguez

**Director:**

Dr. Johann Radax

**Asesor metodológico:**

Dr. Piercosimo Tripaldi Cappelletti

**Cuenca, 15 de Octubre de 2018**

## **Resumen**

En el presente estudio correlacional se empleó la metodología QSAR para determinar los parámetros moleculares responsables de causar efectos adversos gastrointestinales y elaborar un modelo de predicción toxicológica. Luego de un proceso de selección de moléculas y de ponderación de los efectos secundarios gastrointestinales, se usaron algoritmos genéticos para calcular descriptores moleculares relevantes. A continuación, se procedió a la clasificación mediante el método N3 («N-nearest neighbor»). El modelo matemático elaborado demostró una sensibilidad y especificidad buenas, además se comprobó su validez externa. El valor predictivo del modelo alcanza 0,67, la sensibilidad y la especificidad son del 87% y 77%, respectivamente. Este estudio piloto constituye el primero de su género que logró desarrollar un modelo matemático para predecir la propiedad molecular de producir efectos adversos gastrointestinales, y podría servir de fórmula para futuras investigaciones a lo largo de las mismas líneas y para la optimización de la ingeniería molecular en la elaboración de fármacos.

### **Palabras clave:**

QSAR, efectos adversos, modelo matemático, predicción, ingeniería molecular

Use of molecular descriptors in the identification of gastrointestinal effects by anti-inflammatory, analgesic and anesthetic drugs. Cuenca 2018.

**Abstract:**

In this correlational study, the QSAR methodology was used to determine the molecular parameters responsible for adverse gastrointestinal effects and develop a toxicological prediction model. Genetic algorithms were used to calculate relevant molecular descriptors after a process of molecule selection and weighting of gastrointestinal side effects. Next, classification was carried out using the N3 method ("N-nearest neighbor"). The developed mathematical model showed good sensitivity, specificity and its external validity was proved. The predictive value of the model reached 0.67, the sensitivity was 87% and the specificity was 77%. This pilot study was the first of its kind that managed to develop a mathematical model to predict the molecular property of producing gastrointestinal adverse effects. It could serve as a formula for future research along the same line and for the optimization of molecular engineering in the elaboration of drugs.

**Keywords:**

QSAR, adverse effects, mathematical model, prediction, molecular engineering.



  
Translated by  
Ing. Paul Arpi

**Lista de abreviaciones:**

BNN = Binned Nearest Neighbours (vecinos más cercanos desechados)

C.A.S. = Chemical Abstracts Service (servicio de resúmenes químicos)

FDA = Food and Drug Administration (Administración de alimentos y fármacos)

FAERS = FDA Adverse Event Reporting System (Sistema de información de eventos adversos de la FDA)

ICH = International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (Consejo Internacional para la armonización de requerimientos técnicos para fármacos de uso humano)

kNN = k Nearest Neighbours (k vecinos más cercanos)

MedDRA = Medical Dictionary for Regulatory Activities (diccionario médico para actividades regulatorias)

N.E.R = No error rate (porcentaje de los clasificados correctos)

N3 = N-Nearest Neighbours (N vecinos más cercanos)

OLS = Ordinary least squares (Mínimos cuadrados ordinarios)

QSAR = Quantitative structure–activity relationship (relación cuantitativa estructura-actividad)

S.M.I.L.E.S. = Simplified Molecular Input Line Entry Specification (especificación de introducción lineal molecular simplificada)

## Introducción

La noción de una conexión entre la estructura química y la acción fisiológica o farmacológica de una sustancia ya lleva más de 150 años (Brown & Fraser, 1868). En el transcurso del siglo que siguió se publicaron estudios que comprobaron una correlación inversa entre la solubilidad en agua y la citotoxicidad de sustancias («*plus ils sont solubles, moins ils sont toxiques*») (Richet, 1893), que la actividad narcótica de ciertos compuestos orgánicos se relacionaba directamente con su constante de partición entre agua y aceite de oliva (Meyer, 1899; Overton, 1901), la correlación de la actividad narcótica con la saturación relativa de compuestos volátiles en el medio de administración (Ferguson, 1939), y muchos más.

Ha transcurrido más de una cincuentena de años desde que Corwin Hansch publicara el método de modelaje de la relación cuantitativa estructura-actividad QSAR (Hansch, Maloney, Fujita, & Muir, 1962; Hansch et al., 1963). Desde entonces, esta metodología ha conquistado amplios campos de investigación química y farmacéutica, hasta el punto de que el profesor Dominic Tildesley, presidente de la británica Real Sociedad de Química, predijera que «en 15 años ningún químico realizará experimentos de laboratorio sin previamente construir un modelo computarizado» (Steven, 2013).

En el medio siglo de uso y desarrollo, el modelaje QSAR diversificó grandemente su potencial en metodología y aplicación (Cherkasov et al., 2014) como, por ejemplo, su incursión en la química de los alimentos (Rojas et al., 2016; Rojas et al., 2017), como protagonizado por integrantes de nuestra Universidad.

Este estudio a nivel molecular aspira aportar otro camino más, novedoso y original, mediante el desarrollo de un modelo QSAR que permite identificar las causas moleculares de los efectos adversos gastrointestinales producidos por los AINE y, de esta forma, proveer un instrumento de predicción útil para el desarrollo farmacológico.

## Materiales y Métodos

### Lista de moléculas

Extrajimos los fármacos para investigar – antiinflamatorios y/o analgésicos -- de la base de datos SIDER 4.1: Side Effect Resource (Kuhn, Letunic, Jensen, & Bork, 2015b). A continuación procedimos con la obtención, para cada uno de ellos, del C.A.S. y del S.M.I.L.E.S., usando la base de datos de PubChem (PubChem BioAssay Database) (National Center for Biotechnology Information, 2018). La mayoría de programas de ingeniería molecular pueden importar estas cadenas y reconstruir de ellas los modelos bi o tridimensionales.

Tabla N°1: Codificación C.A.S. y S.M.I.L.E.S

Fármaco	C.A.S.	S.M.I.L.E.S.
Acetaminophen	103-90-2	<chem>CC(=O)NC1=CC=C(C=C1)O</chem>
Alfentanil	69049-06-5	<chem>CCC(=O)N(C1=CC=CC=C1)C2(CCN(CC2)CCN3C(=O)N(N=N3)CC)COC</chem>
Almotriptan	154323-57-6	<chem>CN(C)CCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3</chem>

Los datos de los efectos adversos también provienen de la base de datos en línea SIDER 4.1: Side Effect Resource (Kuhn, Letunic, Jensen, & Bork, 2015a). Los términos de los efectos adversos fueron recopilados y transformados en el formato MedDRA (ICH, 2015) para garantizar su estandarización, usando el programa MedDRA Desktop Browser Version Beta 4.0.0.97 (2015). Como valores expresando los efectos adversos se usaron los porcentajes de los usuarios que experimentan el efecto respectivo. Aparte se realizó una ponderación de los valores por un factor 5 en efectos secundarios graves y un factor 10 en efectos secundarios muy preocupantes según el criterio de los autores, pues no existe ejemplo de guía en la literatura, siendo este estudio el primero de su naturaleza. Un ejemplo de un efecto adverso con ponderación de 5 es: «Blood in stool» (sangre en heces), un ejemplo de ponderación 10 es: «Pancreatitis aggravated» (pancreatitis, agravada). La lista completa de las ponderaciones se encuentra en el anexo a este estudio. En el caso de tratarse de un rango de porcentajes, se usó el valor más alto. Excluimos los efectos adversos que aparecieron en el posmercadeo y donde no constan valores numéricos. De esta forma se logró acumular una base de datos primaria de 118 moléculas, de las cuales 68 presentan efectos adversos reportados (base de datos secundaria).

### Cálculo de descriptores

Se calcularon los descriptores moleculares mediante el software DRAGON 7.0.

En la fase inicial del desarrollo del modelo matemático, intentamos construir un modelo de regresión OLS utilizando los algoritmos genéticos para la selección de las variables significativas (es decir, los descriptores significativos, considerando que Dragon 7.0 puede calcular cerca de 6000 descriptores moleculares por molécula), pero se han obtenido modelos inadecuados y no válidos en predicción.

Por esta razón se ha decidido construir un modelo de clasificación N3 dividiendo las moléculas en dos clases: una de nula actividad (es decir, sin efectos adversos gastrointestinales) y otra con actividad adversa reportada. Optimizamos el umbral entre las dos clases de forma secuencial, partiendo de valores arbitrarios y modificándolos

con el algoritmo S.I.M.P.L.E.X (Shavers, Parsons, & Deming, 1979) hasta alcanzar el modelo con la mayor N.E.R. en predicción externa.

### Clasificación mediante N3

El método N3 fue derivado del método k-Nearest Neighbour (KNN) (Altman, 1992). Se basa en el principio del método KNN, a saber, se considera la información de ponderación local para realizar la clasificación de cada objeto. En contraste con KNN, donde se busca el mejor valor K, en el método N3 se toma en cuenta todos los objetos  $n - 1$  para la clasificación del objeto  $i$ -ésimo, pero los vecinos  $n - 1$  se ordenan desde el más similar hasta el menos similar y se ponderan de acuerdo a su posición de rango position  $r$ , cuyo papel es modificado por un por un exponente alfa, al que hay que optimizar. El mejor valor de alfa es aquel que minimiza el error de clasificación. (Todeschini, Ballabio, Cassotti, & Consonni, 2015)

Para la aplicación de este método, se ha utilizado un toolbox de MATLAB el toolbox N3-BNN para MATLAB, una colección de módulos para calcular los métodos de clasificación N3 y BNN, con la implementación del método tradicional de clasificación kNN para su comparación (Todeschini, 2018).

### Umbral de clase

Para la determinación del umbral de clase, procedimos a dividir la base de datos en dos conjuntos: el conjunto de entrenamiento (*training set*) para la construcción del modelo (60% de las moléculas) y el conjunto de ensayo (*test set*) para su validación (40% de las moléculas). La proporción de las clases en ambos conjuntos ha sido la misma.

El SIMPLEX inicial y sus respectivos N.E.R. en predicción se presentan en la Tabla N°2.

Tabla N°2: Umbrales y N.E.R externo iniciales

Puntos	Umbral	N3_pred. Ext
1	10	0,463
2	12,66	0,56
3	10,6	0,5
4	15,32	0,63
5	17,37	0,5

En la determinación del umbral óptimo se partió del cálculo del N.E.R de dos particiones arbitrarias.

En un segundo paso, calculamos el centroide entre los dos mejores N.E.R:

$$\text{Ecuación 1: } x_{i,c} = \frac{(x_{1,1} + x_{1,3})}{2}$$

El valor del centroide que nos interesa ha sido de 8,97. Con este valor se determinó el nuevo punto experimental (punto 2) cuyo umbral es:

$$\text{Ecuación 2: } x_{1,n} = x_{i,c} + (x_{i,c} - x_{1,\text{descartado}})$$

El valor del nuevo umbral (punto 2) fue de 12,66. Sobre la base de este valor se redistribuyeron los datos en dos clases y se calculó el modelo N3 acoplado al algoritmo genético obteniendo un N.E.R en predicción de 0,56 que representa un leve mejoramiento con respecto a los resultados anteriores.

A continuación, se calculó el nuevo centroide entre los puntos 1 y 2 con la ecuación 1, obteniendo el valor de 10,30. La ecuación 2 arrojó el nuevo umbral (punto 3) que resultó ser 10,60.

Continuamos recalculando el nuevo modelo siguiendo el procedimiento descrito, aplicado al punto 3, obteniendo un N.E.R en predicción de 0,5

Se determinó el centroide entre los puntos 2 y 3 y luego se calculó el umbral de 15,32 (punto 4) y el N.E.R en predicción para el modelo N3 con la nueva partición, obteniendo el valor de 0,63

Finalmente se ha calculado el nuevo umbral (punto 5) de 17,37 el N.E.R del modelo con esta partición ha sido de 0,5 inferior al de 0,63 por lo que se consideró terminado el SIMPLEX teniendo como umbral optimo entre las dos clases el valor de 15,32.

#### Selección de los descriptores

Teniendo una base de datos constituida por 68 moléculas y 660 descriptores, resultó necesario seleccionar los descriptores para eliminar los redundantes y retener solamente los que permiten construir un modelo predictivo.

La selección se realizó con algoritmo genético en la versión de Leardi (Leardi & Lupiáñez González, 1998; Leardi, 2002) implementado en un utilitario ga-toolbox de MATLAB para obtener el mínimo número de descriptores útiles en la construcción del modelo (normalmente entre el 20 y el 25% de las moléculas)

#### Dominio de Aplicación

Los modelos QSAR son de tipo local, es decir, tienen validez solo en el espacio definido por los valores de los descriptores del training set. Afuera de este espacio los resultados son extrapolados y la predicción no es segura. En otras palabras, los modelos QSAR son reduccionistas. Encuentran límites en su aplicabilidad, pues, por el tipo de las estructuras químicas consideradas (del training set), sus propiedades físicas y químicas, sus mecanismos de acción y sus efectos adversos (como en nuestro caso), pueden producir predicciones confiables solo cuando las moléculas investigadas por el modelo matemático elaborado presenten similitudes estructurales con aquellas usadas para elaborar el modelo. (Rojas Villa, 2017)

Para optimizar la evaluación de la calidad de un modelo es esencial saber si el modelo es robusto. Además, tiene que lograr predecir la actividad de moléculas externas dentro de cierto margen de confianza. La bondad del ajuste y la robustez de un modelo se denominan validación interna; su capacidad predictiva se denomina validación externa. (Rojas Villa, 2017)

Con el fin de determinar el dominio de aplicación, se ha utilizado un método basado sobre matrices de distancias topológicas de las molécula externas (del test set u otra), calculando las distancias de un punto definido en el espacio de los descriptores del training set. Se compara la distancia con una predefinida distancia umbral. El valor de este se fija para maximizar la separación de las regiones más densamente pobladas del espacio de los datos originales (Sahigara et al., 2012). El cálculo correspondiente se realizó utilizando la metodología KNN con 5 vecinos para definir las clases y distancia de Mahalanobis (Mahalanobis, 1936) para acercarse más a las condiciones de cálculo de N3. Las moléculas de test set que no pertenecen al dominio se eliminan del mismo.

#### Cálculo del modelo

Una vez seleccionadas las variables significativas y definido el dominio, se continuó con el cálculo del modelo QSAR aplicando el algoritmo N3 sobre el training set óptimo seleccionado por la optimización tipo S.I.M.P.L.E.X.

#### Evaluación de los Modelos

Los parámetros de calidad utilizados tanto para el training set, como para el test set han sido el N.E.R. con validación cruzada, es decir, la razón entre los casos clasificados correctamente y los casos totales.

Adicionalmente se calculó para el training set la sensibilidad, a saber, la «habilidad de separación en clases», y la especificidad, la «capacidad de producir clases puras».

Tabla N°3: Sensibilidad y Especificidad del modelo de clasificación obtenido

	<b>Sn</b>	<b>Sp</b>
<b>Class1</b> <b>(no act)</b>	0.778	0.879
<b>Class2</b> <b>(act)</b>	0.879	0.778

## Resultados

Tabla N°4: Descriptores moleculares, su descripción y tipo

Descriptor molecular	Descripción	Original inglés	Tipo
MATS5i	Autocorrelación moran de intervalo de 5, ponderado según el potencial de ionización	Moran lag 5 weighted by ionisation potential	Autocorrelación 2D
F06 [O-O]	Frecuencia de O-O en una distancia topológica de 6 enlaces	Frequency of O-O at topological distance 6	Frecuencia 2D de pares de átomos a una distancia de 6 enlaces
NOHt	Número de alcoholes terciarios	Number of tertiary alcohols	Recuento de grupos funcionales

C-011	Fragmento centrado en el átomo C-011 CR3X	Atom-centred fragment C-011 CR3X	Fragmentos centrados en átomo
CATS2D_02_LL	Grado de interacción entre dos grupos lipofílicos puestos a una distancia de dos enlaces.	CATS2D Lipophilic-Lipophilic at lag 02	Descriptores CATS (Chemically Advanced Template Search) 2D

Los descriptores, seleccionados para el training set fueron: MATS5i, F06 [O-O], nOht, C-011, CATS2D\_02\_LL; el modelo de cinco descriptores tiene un N.E.R con validación cruzada de 0,77, el valor alfa es de 1,5

Los descriptores seleccionados se relacionan con la estructura de las moléculas.

Validación

Validación externa

Se ha aplicado el modelo a las moléculas del training set externo con el fin de obtener los parámetros de validación sobre moléculas no utilizados de ninguna manera en la construcción del modelo.

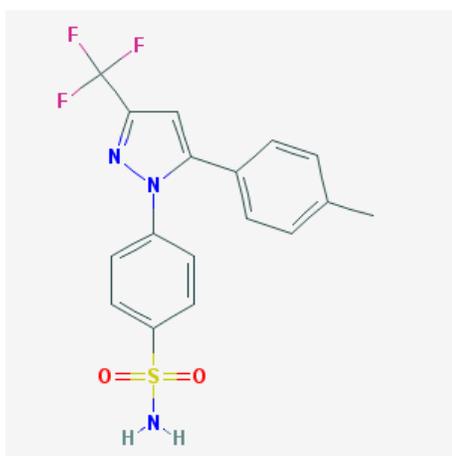
El N.E.R. en predicción externa ha sido de 0,67, mientras que los valores de sensibilidad y especificidad obtenidos se presentan en la Tabla N°5.

Tabla N°5: Sensibilidad y Especificidad del modelo obtenido en predicción externa

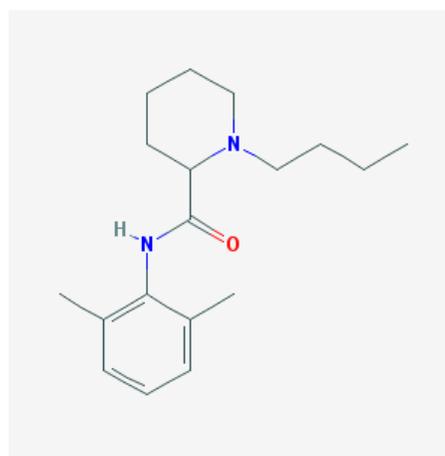
	Sn	Sp
Class1	0.333	100
Class2	100	0.333

Para comprender como la estructura molecular influye en los efectos secundarios sobre el sistema gastrointestinal se pueden comparar dos moléculas: la molécula de Celecoxib, que es clasificada de baja actividad (A), y la de Bupivacaína, que es clasificada de alta actividad (B).

Figura N°1: Fórmulas estructurales de las molécula de Celecoxib (A) y Bupivacaína (B)



A



B

Los valores de MATS5i para la molécula A es de -0,0033, mientras que para la molécula B es de -0,064, lo que indica que en las moléculas de la clase con menos

efectos secundarios hay menos grupos con los mismos potenciales de ionización a la distancia de cinco enlaces.

Los valores del segundo descriptor F06 [O-O] son idénticos en ambos casos.

Ocurre lo mismo ocurre para el tercer descriptor nOH-t y con el cuarto C-011.

El quinto descriptor, CATS2D\_02, presenta valores diferentes para la molécula A con respecto a la molécula B: para la A es de 8, mientras que para la B es de 11.

## **Discusión**

Para la elaboración del modelo predictivo se obtuvieron 5 descriptores moleculares, sin embargo, el descriptor principal, que nos brinda más información acerca de los efectos adversos, es MATS5i, lo que indica que las moléculas de alto potencial de producir efectos adversos gastrointestinales presentan más grupos con los mismos potenciales de ionización a la distancia de cinco enlaces.

La validación externa, la sensibilidad y la especificidad resultaron buenas en el estudio.

Siendo este estudio el primero en su clase, no hemos podido recopilar información externa a fin de compararlo. Sin embargo, de acuerdo a los valores estadísticos de significancia opinamos que este modelo tiene el potencial de ser un prototipo de modelo predictivo para futuros estudios en el campo de la toxicidad farmacéutica.

Las fortalezas del presente estudio son:

- Es nuevo en su clase
- Es fácil de realizar
- No se necesitan sujetos de prueba
- No se incumple con los principios de la bioética

Las debilidades son:

- No existen estudios similares comparativos
- El número total de moléculas estudiadas fue pequeño

- La base de datos se compone de datos reportados que pueden llevar cierto sesgo de información
- La ponderación se basa sobre el criterio de los autores, pues no existen trabajos previos que hubiesen podido servir de guía

Al ser este un estudio piloto podrá servir como referente para futuras investigaciones similares sobre otros fármacos, o sobre los mismos con una base de datos más amplia y, quizá, más confiable.

Esta metodología quizá encuentre aplicación para guiar la ingeniería molecular en la síntesis de moléculas de reducido potencial tóxico.

## Conclusiones

«Pocas personas conocen que los nuevos fármacos de prescripción médica tienen una probabilidad de 1 en 5 de causar serias reacciones adversas después de que estos han sido aprobados. Esta es la razón por la que los médicos recomiendan no tomar medicamentos nuevos durante al menos cinco años, salvo si los pacientes ya han probado primero opciones mejor establecidas o tienen la necesidad de tomarlos.» Así se lee en el blog del Edmond J. Safra Center for Ethics de la Universidad de Harvard (Light, 2014).

La FDA de los Estados Unidos publicó el alarmante número de muertes y eventos severos a causa de los fármacos de prescripción médica, reproducido en la Tabla N°6.

Tabla N°6: Resultados informados por el sistema FAERS (FDA, 2015)

Año	Muertes	Serios*	---	Año	Muertes	Serios*
2006	37.309	264.227	---	2011	98.469	572.992
2007	36.689	272.324	---	2012	117.202	656.613

2008	49.699	318.536	---	2013	116.388	707.593
2009	63.830	373.471	---	2014	123.927	807.270
2010	82.704	471.243	---	2015 (Q1)	44.693	253.017

---

\* Incluye muerte, hospitalización, discapacidad que amenaza la vida, anomalías congénitas y/u otros resultados graves.

Es fácilmente comprensible que el estudio y la vigilancia de los efectos secundarios llevan mucha importancia en la salud pública, la medicina y la farmacéutica.

Es allí donde nuestro modelo matemático puede contribuir su parte: Reducir el riesgo de producir daño desde el inicio, desde el desarrollo de los fármacos en la ingeniería molecular. Cabe destacar, que este enfoque de la metodología QSAR nunca antes ha sido publicado.

Para cerrar el círculo, podemos volver a las palabras de Dominic Tildesley de la británica Real Sociedad de Química: «...en 15 años ningún químico realizará experimentos de laboratorio sin previamente construir un modelo computarizado» (Steven, 2013).

Sí, y es justamente allí donde encaja nuestro programa, anticipando futuras tendencias y, ojalá, poniendo su granito de arena.

### **Agradecimiento**

Agradecemos a nuestras respectivas familias por el apoyo incondicional; a nuestro tutor y asesor: sin ellos este proyecto no hubiera sido posible.

## Bibliografía

- Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 46(3), 175–185. <https://doi.org/10.1080/00031305.1992.10475879>
- Brown, A. C., & Fraser, T. R. (1868). On the Connection between Chemical Constitution and Physiological Action. Part. I.—On the Physiological Action of the Salts of the Ammonium Bases, derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *Trans. R. Soc. Edinburgh*, 25, 151.
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., ... Tropsha, A. (2014). QSAR Modeling: Where have you been? Where are you going to? *J Med Chem.*, 57(12), 4977–5010. <https://doi.org/10.1021/jm4004285.QSAR>
- FDA. (2015). FAERS Reporting by Patient Outcomes by Year. Recuperado el 25 de septiembre de 2018, a partir de <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070461.htm>
- Ferguson, J. (1939). The use of chemical potentials as indices of toxicity. *Proc. R. Soc. London (Ser. B)*, 127, 387–404.
- Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature*, 194(4824), 178–180. <https://doi.org/10.1038/194178b0>
- Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., & Streich, M. (1963). The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *Journal of the American Chemical Society*, 85(18), 2817–2814. <https://doi.org/10.1021/ja00901a033>
- ICH. (2015). MedDRA - Basics - How to Use. Recuperado el 20 de junio de 2016, a partir de <https://www.meddra.org/basics>
- Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2015a). Sider 4.1: Browse the side effects by name. Recuperado el 20 de junio de 2016, a partir de <http://sideeffects.embl.de/se/>
- Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2015b). SIDER 4.1 Browse drugs by ATC hierarchy. Recuperado el 13 de junio de 2017, a partir de <http://sideeffects.embl.de/drugs/>
- Leardi, R. (2002). Chemometrics: From Classical to Genetic Algorithms. *Grasas y Aceites*, 53(1), 115–127. Recuperado a partir de <http://grasasyaceites.revistas.csic.es/index.php/grasasyaceites/article/viewFile/294/>

- Leardi, R., & Lupiáñez González, A. (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometrics and Intelligent Laboratory Systems*, 41, 195–207. [https://doi.org/10.1016/s0169-7439\(98\)00051-3](https://doi.org/10.1016/s0169-7439(98)00051-3)
- Light, D. W. (2014). New Prescription Drugs: A Major Health Risk With Few Offsetting Advantages. Recuperado el 25 de septiembre de 2018, a partir de <https://ethics.harvard.edu/blog/new-prescription-drugs-major-health-risk-few-offsetting-advantages>
- Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49–55. Recuperado a partir de [http://insa.nic.in/writereaddata/UpLoadedFiles/PINSA/Vol02\\_1936\\_1\\_Art05.pdf](http://insa.nic.in/writereaddata/UpLoadedFiles/PINSA/Vol02_1936_1_Art05.pdf)
- Meyer, H. H. (1899). Zur Theorie der Alkoholnarkose. Erste Mittheilung. Welche Eigenschaft der Anästhetica bedingt ihre narkotische Wirkung? *Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmacol.*, 42, 109–118.
- National Center for Biotechnology Information. (2018). PubChem Compound. Recuperado el 4 de septiembre de 2017, a partir de <https://www.ncbi.nlm.nih.gov/pccompound>
- Overton, C. E. (1901). *Studien über die Narkose: zugleich ein Beitrag zur allgemeinen Pharmakologie*. Jena: Gustav Fischer.
- Richet, C. (1893). Sur le rapport entre la toxicité et les propriétés physiques des corps. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales*, 9, 775–776.
- Rojas, C., Ballabio, D., Consonni, V., Tripaldi, P., Mauri, A., & Todeschini, R. (2016). Quantitative structure – activity relationships to predict sweet and non - sweet tastes. *Theoretical Chemistry Accounts*, 135(66), 1–13. <https://doi.org/10.1007/s00214-016-1812-1>
- Rojas, C., Todeschini, R., Ballabio, D., Mauri, A., Consonni, V., Tripaldi, P., & Grisoni, F. (2017). A QSTR-Based Expert System to Predict Sweetness of Molecules. *Frontiers in Chemistry*, 5(July), 1–12. <https://doi.org/10.3389/fchem.2017.00053>
- Rojas Villa, C. X. (2017). *Estudio y aplicación de técnicas matemáticas de la teoría QSAR–QSPR en química analítica y química de los alimentos*. Univeersidad Nacional de La Plata.
- Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., & Todeschini, R. (2012). Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, 17, 4791–4810. <https://doi.org/10.3390/molecules17054791>
- Shavers, C. L., Parsons, M. L., & Deming, S. N. (1979). Simplex Optimization of

Chemical Systems. *Journal of Chemical Education*, 56(5), 307–309. [https://doi.org/Shavers, C. L., Parsons, M. L., & Deming, S. N. \(1979\). Simplex optimization of chemical systems. Journal of Chemical Education, 56\(5\), 307. doi:10.1021/ed056p307](https://doi.org/Shavers, C. L., Parsons, M. L., & Deming, S. N. (1979). Simplex optimization of chemical systems. Journal of Chemical Education, 56(5), 307. doi:10.1021/ed056p307)

Steven, V. (2013). Press release: Next RSC president predicts that in 15 years no chemist will do bench experiments without computer-modelling them first. Recuperado a partir de <http://www.rsc.org/AboutUs/News/PressReleases/2013/Dominic-Tildesley-Royal-Society-of-Chemistry-President-Elect.asp>

Todeschini, R. (2018). N3 - Information. Recuperado a partir de <http://michem.disat.unimib.it/chm/download/n3info.htm>

Todeschini, R., Ballabio, D., Cassotti, M., & Consonni, V. (2015). N3 and BNN: Two new similarity based classification methods in comparison with other classifiers. *J. Chem. Inf. Model.*, 55(11), 2365–2374. <https://doi.org/10.1021/acs.jcim.5b00326>