



**UNIVERSIDAD
DEL AZUAY**

FACULTAD DE MEDICINA

Trabajo de titulación previo a la obtención de título de Médico

**Métodos de Correlación Estructura-Actividad aplicados a
efectos adversos de fármacos con aplicación en el sistema
digestivo, Cuenca 2019**

Autores

**Jorge Enrique Torres León
Milton Andrés Vintimilla Torres**

Director

Dr. Johann Radax

Asesor Metodológico

Dr. Piercosimo Tripaldi Cappelletti

Cuenca, 10 de Junio de 2021

Resumen

El método QSAR relaciona un efecto biológico y una estructura molecular, se utiliza en el presente estudio como un predictor en base a un modelo matemático para fármacos de aplicación en el sistema digestivo. La base de datos fue extraída de SIDER 4.1 Side Effects Resource, usando como fuente para los cálculos la frecuencia de los efectos secundarios los ponderamos como leves (1), intermedios (5) más graves (10). Se trabajó con 91 moléculas activas. Calculamos descriptores moleculares con DRAGON. Se obtuvo 2 clases de respuestas: Clase 1: pocos efectos secundarios, clase 2: grandes efectos secundarios. Para establecer un umbral óptimo se utilizó N.E.R. y aplicamos el método toolbox de MATLAB, con lo que se realizó dos conjuntos de training set y test set de validación externa, con lo cual se calcula sensibilidad, especificidad y precisión. Posteriormente se aplicó un algoritmo SIMPLEX obteniendo el punto de partición 4 con los respectivos valores de umbral: 0.0035, N.E.R: 0.632, Sp: 0.572 Sn: 0.692 Pr: 0.6 Número de descriptores: 3. Con la selección de los 3 descriptores obtenemos el modelo predictor para la clase 1: Sp: 0.6969697, Sn: 0.70967742, Pr: 0.6875, Clase 2: Sp: 0.70967742, Sn: 0.6969697, Pr: 0.71875. Para validación externa en la clase 2: Sp: 0.572, Sn: 0.692, Pr: 0.6, N.E.R: 0.632. Se proporciona con el modelo una herramienta de predicción útil para definir en fase de planificación pre síntesis de un fármaco.

Palabras clave (MESH):

Efectos adversos, Modelos matemáticos, Relación cuantitativa estructura-actividad, Sistema digestivo, Desarrollo de fármacos.

Abstract

The QSAR method relates a biological effect and a molecular structure, it was used in the present study as a predictor based on a mathematical model for drugs for application in the digestive system. The database was extracted from SIDER 4.1 Side Effects Resource, using the frequency of side effects as a source for calculations, we weighted them as mild (1), intermediate (5) and more severe (10). We worked with 91 active molecules. We calculated molecular descriptors with DRAGON. There were 2 kinds of responses: Class 1: few side effects, Class 2: large side effects. To establish an optimal threshold we use N.E.R. and we applied the MATLAB toolbox method, with which two sets of training set and external validation test set were performed, with which sensitivity, specificity and precision were calculated. Subsequently, a SIMPLEX algorithm was applied obtaining partition point 4 with the respective threshold values: 0.0035, N.E.R: 0.632, Sp: 0.572 Sn: 0.692 Pr: 0.6 Number of descriptors: 3. With the selection of the 3 descriptors we obtain the model predictor for class 1: Sp: 0.6969697, Sn: 0.70967742, Pr: 0.6875, Class 2: Sp: 0.70967742, Sn: 0.6969697, Pr: 0.71875. For external validation in class 2: Sp: 0.572, Sn: 0.692, Pr: 0.6, N.E.R: 0.632. A useful prediction tool was provided with the model to define a drug in the pre-synthesis planning phase.

Keywords (MESH):

Adverse Effects, Mathematical model, Quantitative Structure-Activity Relationship, Digestive System, Drug Development.

Translated by
Milton Vintimilla Torres



Lista de abreviaturas:

- GA-toolbox = Genetic Algorithm Toolbox (Algoritmo genético toolbox)
- K.N.N = K Nearest Neighbours (K vecinos más cercanos)
- N.E.R = No error rate (Proporción de clasificados correctos)
- MedDRA = Medical Dictionary for Regulatory Activities (Diccionario médico para actividades regulatorias)
- N3 = N - Nearest Neighbours (N vecinos más cercanos)
- Pr = precisión
- Q.S.A.R = 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)
- Sn = Sensibilidad
- SOC = System Organ Classes (clases de órganos y sistemas)
- Sp = Especificidad

Introducción

El método QSAR se origina de hipótesis donde se relaciona un efecto biológico y la estructura molecular. Este método permitió un desarrollo tecnológico para construir modelos a través de descriptores moleculares. Llegó a ser utilizado en el desarrollo farmacéutico y consiguió importancia para las interacciones entre los fármacos y sus receptores, permitiendo posteriormente utilizarlos en otros ámbitos. En la actualidad se presenta 2 grandes propósitos, el primero enfocado en la creación de nuevos fármacos con la predicción de su posible efecto, y en segunda instancia, predice las propiedades con un interés relevante para los seres biológicos. Pero por razones de variabilidad en las propiedades fisicoquímica en la naturaleza, puede afectar en la reproductividad de las investigaciones de los mismos (1). Existen 2 tipos de resultados, el tipo prospectivo, en el cual se generan nuevas moléculas o compuestos, pero compartiendo un patrón químico; el otro tipo es retrospectivo, se basa en analizar las moléculas ya existentes para descifrar sus interacciones o actividades (2).

Es importante la implementación de estos modelos, al permitir conocer los efectos toxicológicos ante la creación de un fármaco, y evitar los estudios clínicos que ponen a prueba a seres vivos con un elevado costo económico.

Las relación entre estructura y actividad ha sido ampliamente utilizada en diferentes fases de desarrollo de un fármaco, como en estudios sobre afinidad en ligando a su sitio de unión, constantes de inhibición y de velocidad, propiedades lipofílicas, o métodos actuales como explorar las moléculas en su estructura tridimensional. Todo esto permite poseer una preselección de los compuestos más prometedores (3).

La utilidad es amplia tanto en el ámbito médico como en la producción agroindustrial. Constituye un predictor de efectos en la toxicidad de pesticidas, por ejemplo, analizando varios factores de propiedades físico-químicas, por lo que en la actualidad existe un amplio uso en varias áreas que evidencia el potencial del método. (4)

Las enfermedades digestivas representan un número importante de patologías, que ocurren con frecuencia elevada en la población general. Por ejemplo, según la OMS, la enfermedad diarreica aguda es la segunda causa de muerte en niños menores de 5 años. El *Helicobáctter pylori* presenta una prevalencia de 80% en países en vías de desarrollo, relacionado también con el cáncer gástrico (5). Actualmente los tratamientos antidiarreicos siguen siendo cuestión de controversia por su baja evidencia de efectividad, sin embargo, la población sigue utilizando estos fármacos como terapia de manera errónea y sin supervisión médica (6).

En el mundo entero la prevalencia de la enfermedad por reflujo gastroesofágico representa un 52%, es una causa frecuente de visita a la consulta y de automedicación (7). Por ejemplo, en España se conoce que son los medicamentos más prescritos en los últimos años, con un aumento del 200% en 8 años en su comercialización (8).

El consumo progresivo e incluso el mal uso de inhibidores de la bomba de protones provocan reacciones adversas documentadas, entre las cuales se encuentra el aumento del riesgo de eventos cardiovasculares y de síndrome coronario agudo, nefritis intersticiales agudas, hipomagnesemia, deficiencia de vitamina b12, etc. (9)

Además, el consumo de antieméticos se ha vuelto frecuente, por ejemplo, la metoclopramida suele provocar entre sus efectos adversos: cefalea, mareos, somnolencia, depresión, reacciones de hipersensibilidad. (10)

Con los datos presentados sabemos que los fármacos que actúan sobre el sistema digestivo se utilizan con frecuencia elevada, en ocasiones automedicada, por lo que es importante conocer los efectos adversos que este tipo de moléculas representan.

Materiales y métodos

La información de las moléculas fue extraída de la base de datos **SIDER 4.1: Side Effect Resource**, como fuente para los cálculos sirvieron las frecuencias de los efectos secundarios como figuran en la misma. Introdujimos un factor de

ponderación para resaltar la peligrosidad del efecto secundario; a los efectos más graves se atribuyó el peso de 10, a los intermedios el peso de 5, a los leves el peso de 1. En la tabla número 1 se presenta una muestra de la matriz de los pesos asignados a los efectos secundarios. Donde se utilizó alrededor de 700 efectos secundarios para la matriz base en la terminología MedDRA.

Tabla N° 1. Matriz de los pesos de los efectos secundarios

#	SOC	Términos preferidos	Ponderación
58	Infections and infestations	Abscess periodontal	5.00
510	Infections and infestations	Application site folliculitis	1.00
699	Infections and infestations	Bacterial infection	5.00
1158	Infections and infestations	Candida infection	5.00
1494	Infections and infestations	Clostridium difficile colitis	10.00
1554	Infections and infestations	Common cold	1.00
1850	Infections and infestations	Dentoalveolar abscess	5.00
2072	Infections and infestations	Dysentery	5.00
2559	Infections and infestations	Flu symptoms	5.00
2627	Infections and infestations	Fungal infection	5.00
2688	Infections and infestations	Gastrointestinal candidiasis	5.00
3082	Infections and infestations	Herpes simplex	1.00
3087	Infections and infestations	Herpes zoster	5.00
3416	Infections and infestations	Infection	1.00
3437	Infections and infestations	Infestation NOS	1.00
3443	Infections and infestations	Influenza	5.00
3445	Infections and infestations	Influenza like illness	1.00
3446	Infections and infestations	Influenza viral infections	5.00
3447	Infections and infestations	Influenza-like symptoms	1.00
3897	Infections and infestations	Lower respiratory tract infection	5.00

#	SOC	Términos preferidos	Ponderación
4354	Infections and infestations	Nasopharyngitis	1.00
4623	Infections and infestations	Onychomycosis	1.00
4644	Infections and infestations	Oral candidiasis	1.00
4668	Infections and infestations	Orchitis	5.00
4726	Infections and infestations	Otitis	5.00
4729	Infections and infestations	Otitis media	5.00
4970	Infections and infestations	Periodontitis	1.00
5041	Infections and infestations	Pharyngitis	1.00
5150	Infections and infestations	Pneumonia	10.00
5468	Infections and infestations	Pulpitis dental	5.00
5499	Infections and infestations	Pyelonephritis	5.00
5691	Infections and infestations	Respiratory tract infection	1.00
5758	Infections and infestations	Rhinitis	1.00
5888	Infections and infestations	Sepsis	10.00
5981	Infections and infestations	Sinusitis	1.00
6009	Infections and infestations	Skin infection	1.00
6487	Infections and infestations	Tooth abscess	5.00
6494	Infections and infestations	Tooth infection	5.00
6664	Infections and infestations	Upper respiratory tract infection	1.00
6706	Infections and infestations	Urinary tract infection	1.00
6784	Infections and infestations	Vaginal moniliasis	1.00
6786	Infections and infestations	Vaginal mycosis	1.00
6797	Infections and infestations	Vaginitis bacterial	1.00
6904	Infections and infestations	Viral diarrhoea	5.00
6905	Infections and infestations	Viral infection	1.00
6910	Infections and infestations	Viral upper respiratory tract infection	1.00

#	SOC	Términos preferidos	Ponderación
6976	Infections and infestations	Vulvovaginal mycotic infection	5.00
7052	Infections and infestations	Yeast infection	1.00

Se han eliminado del conjunto original de fármacos aquellos para los cuales los nombres no permitían determinar inequívocamente la molécula activa (por ejemplo, CALCIO). En total se trabajó con 91 moléculas.

Para cada una se han obtenido sus códigos S.M.I.L.E.S a partir de Pub-Chem y se han calculado, a partir de estos códigos 939 descriptores moleculares utilizando el paquete informático DRAGON. (11)

Estos descriptores son el resultado de la eliminación de los descriptores constantes o casi-constantes y de lo correlacionado entre sí con más del 95%.

Creación de las clases de peligrosidad

Las respuestas obtenidas para cada fármaco forman un conjunto continuo de valores, se ha decidido separar los fármacos y sus relativas respuestas en dos clases de peligrosidad: clase 1 (fármacos con relativamente pocos efectos secundarios), clase 2 (fármacos con grandes efectos secundarios).

El problema que se presenta es aquel de establecer el umbral óptimo de separación entre las respuestas de las dos clases. Como criterio de óptimo los autores han utilizado el promedio del No Error Rate (N.E.R) en validación cruzada de la clasificación, el promedio del NER en predicción sobre un grupo de fármacos no utilizado para entrenar el algoritmo, la especificidad, la sensibilidad, la precisión, de la predicción con respecto a la clase más peligrosa (clase 2); en caso de casi igualdad de puntuación se optó por utilizar la partición que produjo un modelo con menos descriptores (razor de Occam). Como algoritmo de clasificación se ha utilizado el método N3 de Todeschini et al (12).

La N3 es una variación del procedimiento de los K vecinos más cercanos (K.N.N) (13), para realizar esto se emplean los K vecinos próximos al objeto a analizar proporcionando información de los mismos para su clasificación.

Entonces, el método N3 considera los demás objetos en comparación al método K.N.N.

Se ubica la información de los objetos en forma descendente y en función del intervalo del n-1 ésimo objeto desde el objeto a clasificarse y se pondera en forma exponencialmente negativa. Un exponente alfa se incluye en el peso con el cual la ponderación del objeto es medida. Un valor óptimo de alfa permite disminuir el error de clasificación.

Se utiliza la herramienta toolbox de MATLAB para aplicar este método, el cual está descrito por los autores del mismo y disponible en: <http://michem.disat.unimib.it/chm/download/n3info.htm> (14).

Conjunto de entrenamiento (training set) y de evaluación externa (test set)

Se divide la base de datos en 2 agrupaciones a raíz de la base de datos: El test set para validación (40%) y para la fabricación del modelo el training set (60%). En ambos conjuntos la proporción ha sido similar.

Evaluación de la calidad los modelos

Para evaluar la calidad del training set como del test set se ha implementado: la razón entre casos clasificados correctamente y casos totales por medio de un N.E.R de validación cruzada.

Para el training set se obtuvo la sensibilidad (S_n) y la especificidad (S_p) y la precisión de la clasificación de los fármacos en ambas clases. (15).

Para validar el modelo se aplica al test set externo, además se obtiene el N.E.R, especificidad, sensibilidad y precisión.

Precisión

Se obtiene como la razón entre casos clasificados correctamente de la clase g y la totalidad de casos asignados por el modelo a la clase g . Demostrando la capacidad del modelo de clasificación de no incorporar casos de otras clases en la clase considerada.

$$Pr_g = \frac{n_{gg}}{n'_g}$$

Total de casos asignados a la clase g : n'_g

Casos de la clase g clasificados correctamente: n_{gg} .

Sensibilidad

Es la facultad del modelo de consignar los elementos de una clase a la propia clase de manera correcta.

$$Sn_g = \frac{n_{gg}}{n_g}$$

Total de casos de la clase g : n_g

En el cálculo de la sensibilidad no se considera los casos de la clase g no incluidos dentro de la propia clase.

Especificidad

Es la capacidad de la clase g de exceptuar los casos de todas las demás clases y se la determina como:

$$Sp_g = \frac{\sum_{k=1}^G (n'_k - n_{gk})}{n - n_g} \quad \text{for } k \neq g$$

Total de casos asignados a la clase k : n'_k

$$n'_k = \sum_{g=1}^G n_{gk}$$

El cálculo de la especificidad no considera los casos no asignados.

Umbral de clase óptimo

El umbral entre las dos clases se ha optimizado de manera progresiva desde valores arbitrarios y se modifica con el algoritmo SIMPLEX de acuerdo con Shavers et al (16) hasta alcanzar el modelo que tenga el mayor valor de la puntuación previamente establecida.

El SIMPLEX inicial y sus respectivos N.E.R. y demás parámetros de calidad en predicción se presentan en la tabla N° 2.

Tabla N° 2.

Umbrales, NER externo, demás parámetros de calidad en predicción y numero de descriptores del modelo.

Puntos	Umbral	N.E.R predicció n	Sp. clase 2 en predicción	Sn. clase 2 en predicción	Pr. clase 2 en predicción	N°. descriptores
1	0	0.505	0.857	0.154	0.5	7
2	0.001	0.544	0.857	0.231	0.857	6
3	0.002	0.659	0.857	0.462	0.75	39
4	0.0035	0.632	0.572	0.692	0.6	3

La partición del punto 3 da un buen resultado como N.E.R en predicción, pero la sensibilidad (0.462) es insuficiente para utilizar el modelo.

La partición del punto 4 da un N.E.R en predicción ligeramente más bajo que el anterior, pero la especificidad, predicción y sensibilidad son mejores y en general más equilibradas, además utiliza sólo tres descriptores.

En este trabajo se ha optado por utilizar como umbral de clase el valor de 0.0035.

Resultados

91 moléculas y 939 descriptores conforman la base de datos, existen descriptores que al ser redundantes se deben eliminar y por lo tanto conservar aquellos que favorezca la construcción de un modelo predictivo.

Para realizar la selección se utiliza un algoritmo genético en la versión de Leardi (17) implementado en un utilitario GA-toolbox de MATLAB.

Los descriptores, seleccionados para el training set han sido: nArNH₂, B02[N-S], Infective-80; el modelo de tres descriptores tiene un N.E.R en clasificación de 0.70, el valor de alfa es de 1.75.

La matriz de especificidad, sensibilidad y precisión del modelo se informa en la tabla N° 3.

Tabla N° 3. Matriz de especificidad, sensibilidad y precisión del modelo de clasificación obtenido.

Clase	Especificidad	Sensibilidad	Precisión
1	0.6969697	0.70967742	0.6875
2	0.70967742	0.6969697	0.71875

Validación Externa

Para obtener parámetros de validación para péptidos que no fueron incluidos en la fabricación del modelo, se aplicó el modelo a los fármacos del training set externo.

Tabla N°4. Matriz de precisión, sensibilidad y especificidad del modelo obtenido en predicción externa

Sp. clase 2 en predicción	Sn. clase 2 en predicción	Pr. clase 2 en predicción	N.E.R predicción externa
0.572	0.692	0.6	0.632

Los resultados obtenidos no dan valores experimentales sino información sobre los efectos secundarios reportados, el resultado en predicción es interesante y proporciona un primer instrumento para definir en fase de planificación pre síntesis de un fármaco su eventual clase de peligrosidad frente a sus efectos secundarios.

Discusión

En el estudio realizado en la Universidad del Azuay por Torres y Jara (18), el modelo de predicción para efectos adversos de fármacos cardiovasculares usó una metodología similar a la aplicada en este estudio, pero con variación en los descriptores utilizados y su número: CATS2D_05_AL, SpMin7_Bh (m), B10 [N-O], B04 [C-O], F10 [C-F], en total 5 descriptores, mientras que los descriptores en el trabajo actual fueron 3: nArNH₂, B02 [N-S], Infective-80.

Existió diferencia entre el uso de un N.E.R. de clasificación y el alfa que previamente dio un valor de 0.8 y 0.75 respectivamente, mientras que ahora se obtuvo un valor de 0.70 y 1.75.

Además las clases utilizadas fueron dirigidas hacia el efecto cardiaco sea leve o grave, a diferencia del presente estudio que plantea los efectos secundarios como relativamente pocos y grandes efectos secundarios, los valores luego de la validación externa reportaron sensibilidad y especificidad y N.E.R de predicción externa de 91%-83% y 0.46 respectivamente, esto fue el resultado de la aplicación de un algoritmo que permitió incrementar los valores, el actual método obtuvo directamente un valor luego de la validación externa de sensibilidad, especificidad y N.E.R. de predicción externa 59%-69% y 0.632, obteniendo en estas investigaciones un modelo predictor de efectos cardiacos severos de fármacos que serán pre sintetizados, así como la peligrosidad de los efectos secundarios.

Al ser un grupo de fármacos de mayor variabilidad dentro de las distintas categorías se logra un modelo con un mayor potencial predictivo, tal es el caso de

la escasa diversidad de familias farmacológicas que pertenece al conjunto fármacos cardiovasculares, a diferencia de los utilizados en el sistema digestivo, por lo que se puede afirmar lo que previamente discuten los autores acerca de un nuevo estudio con mayor inclusión de diversos medicamentos.

La capacidad de ejecutar un estudio con este objetivo es sencillo, permite evitar la participación de seres humanos y limitando los problemas éticos y legales, permitiendo abrir paso a nuevos estudios y validar de manera óptima la predicción.

La base de datos utilizada es incompleta, por lo que uno de los problemas al momento de realizar este tipo de estudios es conocer la totalidad y la frecuencia exacta de los efectos adversos, sin embargo, contiene información consistente sobre los mismos.

Conclusiones

El método QSAR brinda una herramienta de predicción innovadora para el proceso de planificación pre-síntesis de un fármaco de una manera más simplificada, debido a que el conocimiento de los posibles efectos adversos limitaría el número de moléculas a evaluar.

La aplicación de un predictor ofrece información que puede ser útil para efectos beneficiosos, como previamente se ha utilizado, y efectos adversos, como en el presente estudio.

Este tipo de modelos predictivos permite – de manera sencilla y práctica – llevar la seguridad farmacológica a la primera fase del desarrollo de un nuevo medicamento con el efecto de una selección temprana de las moléculas idóneas lo que, por un lado, puede ahorrar a las empresas farmacéuticas millones de dólares de inversión y, por otro lado, debería materializarse en un menor costo de los fármacos para los pacientes. De esta forma, el humilde modelo matemático contribuye de manera substancial a la calidad de vida de todos y logra liberar

recursos... tal vez para estudios en el ámbito desatendido de los fármacos huérfanos – en un mundo ideal.

Agradecimientos

Agradecemos al director y al asesor metodológico por la entrega de su tiempo y conocimientos a los autores del presente estudio.

Agradecemos a nuestros padres por el esfuerzo y apoyo en este camino.

Bibliografía

1. Gini G. QSAR: What Else? *Methods Mol Biol.* 2018;1800:79-105.
2. Gini G. QSAR methods. *Methods Mol Biol.* 2016; 1425:1-20.
3. Verma J, Khedkar V, Coutinho E. 3D-QSAR in Drug Design - A Review. *Current Topics in Medicinal Chemistry.* 2010; 10(1): p. 95–115.
4. Ghosh S, Ojha PK, Carnesecchi E, Lombardo A, Roy K, Benfenati E. Exploring QSAR modeling of toxicity of chemicals on earthworm. *Ecotoxicology and Environmental Safety.* 2020; 190: p. 110067.
5. Bortoli, N. Tolone, S. Frazzoni, M. et al. Gastroesophageal reflux disease, functional dyspepsia and irritable bowel syndrome: common overlapping gastrointestinal disorders. *Ann Gastroenterol.* 2018 Nov-Dec; 31(6): 639–648.
6. Bruzzese E, Giannattasio A, Guarino A. Antibiotic treatment of acute gastroenteritis in children. *Faculty rev.* 2018; 193.
7. Carmona R, Solana S. La prevalencia de los diferentes grupos que integran la enfermedad por reflujo con endoscopia negativa. *Rev. Gastroenterología.* 2005; 70(1): p. 7-13.
8. AEMPS. Utilización de medicamentos antiulcerosos en España durante el periodo 2000-2012. [Internet]; 2014. Citado 22 de Marzo de 2021. Disponible en:
<https://www.aemps.gob.es/medicamentosUsoHumano/observatorio/docs/antiulcerosos.pdf>.
9. Cardona-Ospina J, Medina-Morales D, Rodríguez-Morales A, Machado-Alba J. Efectos adversos a largo plazo de los inhibidores de la bomba de protones: Perspectiva desde la medicina basada en la evidencia. *Rev Col*

- Gastroenterologia. 2016; 31(4): p. 403-408.
- 10 Alfonso Orta I, Jiménez López G, Chao Cardeso A, Ávila Pérez J. La metoclopramida y sus reacciones adversas sobre el sistema nervioso central. Rev Cubana Med Gen Integr. 2011; 27(2): p. 197-206.
- 11 Kode Chemo Informatics. [Internet]; 2014. [Citado 13 de Abril de 2021]. Disponible en: <https://chm.kode-solutions.net/index.php>.
- 12 Todeschini R, Ballabio D, Cassotti M, Consonni V. N3 and BNN: Two New Similarity Based Classification Methods in Comparison with Other Classifiers. Journal of Chemical Information and Modeling. 2015; 55(11): p. 2365–2374.
- 13 Deng Z, Zhu X, Cheng D, Zong M, Zhang S. Efficient k NN classification algorithm for big data. Neurocomputing. 2016; 195: p. 143–148.
- 14 Matlab toolboxes [Internet]. Milano Chemometrics and QSAR Research Group. [Citado 22 de mayo de 2021]. Disponible en: <https://michem.unimib.it/download/matlab-toolboxes/>
- 15 Mansouri K, Sahigara F, Ballabio D, Mauri A, Consonni V, Todeschini R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. Molecules. 2012; 17(5): p. 4791-810.
- 16 Shavers CL, Parsons ML, Deming SN. Simplex optimization of chemical systems. Journal of Chemical Education. 1979; 56(5): p. 307.
- 17 Leardi R, Lupiáñez González A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemometrics and Intelligent Laboratory Systems. 1998; 41(2): p. 195–207.

18. Torres G, Jara J. Métodos de correlación estructura-actividad aplicados a efectos adversos de fármacos cardiovasculares [Internet][Tesis de pregrado] [Cuenca, EC] UDA: 2018. [Citado el 22 de marzo de 2021] Disponible en: <http://dspace.uazuay.edu.ec/bitstream/datos/9453/1/15089.pdf>