



**UNIVERSIDAD
DEL AZUAY**

FACULTAD DE MEDICINA

Trabajo de titulación previo a la obtención de título de médico

**Aplicación de la metodología QSAR para el desarrollo de un modelo
matemático para la predicción de efectos secundarios en
psicofármacos y hormonas**

Autoras:

Doménica Cristina Congacha Ochoa

Lorena Patricia Padrón Alvarado

Director:

Dr. Johann Radax

Asesor metodológico:

Dr. Piercosimo Tripaldi Cappelletti

Cuenca 2021

Resumen:

En este estudio se aplica la metodología QSAR para establecer los parámetros moleculares encargado de provocar efectos secundarios y posteriores fabricar un modelo de predicción toxicológica. Tras llevar a cabo el proceso de selección de moléculas, los efectos adversos fueron clasificados como activos (clase 2) e inactivos (clase 1). Se comprobó la validez externa de este modelo mediante 3 métodos; No paramétrico KNN, No paramétrico N3, y No paramétrico SIMCA. De igual manera se demostró tener una buena sensibilidad y especificidad; sin embargo, en cada modelo hay clases que no son bien clasificadas y por este motivo se considera que el modelo de Consensus tiene mayor utilidad para predecir la aparición de efectos adversos (clase 1, Precisión: 0.92, Sn: 1.00, Sp: 0.73); (clase 2, Precisión: 1.00 Sn: 0.73, Sp: 1.00).

Con la realización de este estudio, se desarrolló un modelo matemático que permite predecir la propiedad molecular de psicofármacos y hormonas para producir efectos secundarios, por otro lado, además este modelo sirve como precedente para investigaciones subsiguientes y lograr así avances en la ingeniería molecular en el contexto de la creación de fármacos.

Palabras Clave:

QSAR, Efectos adversos, Predicción, Ingeniería Molecular, Psicofármacos, Hormonas.

Abstract:

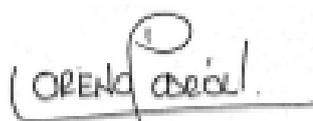
In the present study, the QSAR methodology was used to determine the molecular parameters responsible for causing secondary/adverse effects and to develop a toxicological prediction model. After a molecule selection process, adverse effects were classified as inactive (class 1) and active (class 2). The validity was verified using 3 methods; Nonparametric KNN, Nonparametric N3, and Nonparametric SIMCA. It was demonstrated the good sensitivity and specificity, however, in each of the models, some classes are not well classified, and that is why it is considered that the consensus model is better to predict adverse effects (class 1, Precision: 0.92, Sn: 1.00, Sp: 0.73); (class 2, Accuracy: 1.00 Sn: 0.73, Sp: 1.00).

This study allowed the development of a mathematical model to predict the molecular property of producing secondary/adverse effects of both psychotropic drugs and hormones, furthermore, it serves as a model for subsequent research and to achieve an advance in molecular engineering in the context of drug development.

Keywords:

QSAR, Adverse effects, Prediction, Molecular Engineering, Psychopharmaceuticals, Hormones.

Translated by:



Lorena Padrón



Doménica Congacha



Lista de abreviaciones:

BNN = Binned Nearest Neighbours (vecinos cercanos desechados)

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

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)

SIMCA = Soft Independent Modeling of Class Analogy (Modelo independiente de la analogía de clases)

Introducción:

El incremento de trastornos endocrinológicos y trastornos psiquiátricos, mismos que en la actualidad representan un 30% aproximadamente de la carga mundial de enfermedad no mortal (1), ha acarreado la necesidad de uso de terapias farmacológicas para mantener en equilibrio la calidad de vida de los pacientes e incluso mejorarla. (2) Sin embargo, además del impacto beneficioso que llegan a brindar, se pueden presentar una serie de efectos adversos, que pueden llegar incluso a disminuir el apego terapéutico de los pacientes, lo que ha motivado a la mejora en las técnicas e instrumentos diagnósticos. Dichos efectos están dados por cambios en procesos o funciones fisiológicas que surgen tras la interacción de moléculas farmacológicas con receptores del cuerpo, a este fenómeno se lo denomina mecanismo de acción.

Así mismo el mecanismo se relaciona directamente con la composición molecular del medicamento, teniendo la capacidad para corregir o complicar una condición en particular, causando así diferentes reacciones que pueden ir desde cumplir con su objetivo principal o producir reacciones de hipersensibilidad, anafilaxia hasta desenlaces fatales.(3)

Esto conlleva al estudio de moléculas que generan desenlaces no deseados en los medicamentos con la finalidad de determinar las características moleculares de hormonas y psicofármacos, consideramos necesario emplear la metodología del modelo QSAR ya que además de ser una herramienta de gran impacto por su amplia aplicación en las ciencias químicas, biológicas y farmacológicas, esta herramienta reúne un grupo de técnicas computacionales que se relacionan con la visualización espacial y diseño virtual de moléculas, así como bioinformática, estadística y el cálculo de propiedades fisicoquímicas moleculares (4)

“El método QSAR (quantitive structure – activity relationships) es definido como una ecuación matemática en la cual se correlaciona la actividad biológica con descriptores moleculares o información fisicoquímica de una estructura.” (5) para promover la creación de un diseño teórico para nuevos medicamentos, evitando pasar por el desarrollo de prueba y error de síntesis orgánica. Su beneficio radica en que el esquema de candidatos a nuevos fármacos, es mucho más económico y rápido ya que al ser una metodología virtual, la necesidad de infraestructura, personal de laboratorio y materiales dejan de ser necesarios. (6) (7)

Los descriptores QSAR son propiedades físico químicas que se calculan por métodos computacionales, por lo que reflejan la forma de cada molécula (4), cumpliendo un rol crucial al desarrollar el modelo, ya que estos representan de manera cuantitativa la información química que ha sido codificada, no participan únicamente en la deducción de una correlación matemática entre la información de la estructura química y la respuesta deseada, sino también en la identificación de un proceso bioquímico. (5) (8)

Al identificar estos descriptores moleculares, se hallan también las causas moleculares que producen los efectos adversos tanto de psicofármacos como de hormonas, obteniendo de esta manera una herramienta de predicción útil que permite y facilita la innovación en el área farmacológica.(9) Dichos avances inspiran a industrias farmacéuticas a la creación, modificación y mejora de fármacos.

Este trabajo de investigación desarrolla un modelo matemático por medio del método QSAR relacionando parámetros estructurales de 311 fármacos entre psicotrópicos y hormonas, para pronosticar los efectos adversos de estos en un y facilitar la creación de nuevos y mejores medicamentos en base a los anteriores con similares composiciones moleculares.

Materiales y Métodos:

Obtención de Moléculas:

Los datos obtenidos sobre los efectos secundarios de las moléculas estudiadas respecto a psicofármacos fueron obtenidos de la base de datos SIDER 4 (Side Effect resource), la cual poseía 311 moléculas, cada una fue asociada a su código (10) misma que es la manera en la que la molécula es representada a manera de una secuencia de letras.

Mediante los códigos SMILES se obtuvieron los descriptores moleculares bidimensionales de cada molécula de forma inmediata utilizando el programa DRAGON versión 7. Se obtuvieron 761 descriptores.

Un código SMILES permite describir la estructura y propiedades moleculares de forma automática.

Las reacciones adversas fueron divididas en 2 grupos: clase 1 – inactiva o poco activa con 249 fármacos y clase 2 – activa con 62 medicamentos. El fraccionamiento se llevó a cabo usando como umbral el 80vo percentil de la distribución en los resultados. La base de datos formada por 311 fármacos ha sido distribuida en grupos de calibración formado por 218 medicamentos y predicción que agrupa los 93 fármacos restantes; esto para efectos de validación externa.

Para el desarrollo del modelo, el grupo de calibración ha sido utilizado a través de tres variedades de clasificación:

- 1) Método no paramétrico KNN (11) a través del método de los algoritmos genéticos, este se combina y se ajusta a la selección de descriptores que son significativos (12).

Si hablamos desde la apreciación teórica, este método es sencillo: en función del grupo dominante de los K vecinos, una muestra es encasillada.

- 2) Método no paramétrico N3 (13) este debe ser aplicado a los descriptores que anteriormente fueron elegidos con KNN. En este método se considera mientras se clasifica, el aporte de todos los objetos en el espacio multidimensional, ponderados por una colaboración conforme a la cercanía del objeto a clasificar. El parámetro alfa óptimo fue 1, en esta ocasión.

- 3) Método no paramétrico SIMCA. Un modelo SIMCA consiste en una colección de modelos de Componentes principales cada uno por cada clase. De esta manera SIMCA define, en este trabajo, 2 modelos de clase. Cada nueva muestra es proyectada en cada sub espacio y se calcula su distancia de cada espacio de clase. Como métrica se usan los residuos Q normalizados y los T2 valores de Hotelling al 95% de confianza. Las muestras se asignan a la clase para la cual tienen distancia menor. Una característica de este método es que puede ocurrir que la distancia de dos clases sea igual, en este caso el objeto no viene clasificado. (14)

Para el modelo de clasificación y validación se usan parámetros de calidad que engloban la tasa de aciertos (NER), especificidad (Sp), sensibilidad (Sn) y precisión (Pr). (15)

La precisión “representa la capacidad del modelo de clasificación de no incluir en una clase determinada elementos de otras clases”

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

Donde n'g es el número total de muestras asignadas a la g-esma clase y ngg es el número de elementos de la clase g asignados correctamente a la clase g

La sensibilidad describe la habilidad del modelo de clasificar los elementos de la clase g en la misma clase g

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

Donde n_g es el número total de elementos perteneciente a la clase g .

La especificidad esta hace referencia a la técnica de la clase g para rechazar las muestras de las demás clases.

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

Donde n'_k es el número total de elementos asignados a la k -ésima clase

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

La tasa de no aciertos (N.E.R) se obtiene al calcular el promedio de la sensibilidad en las clases

$$NER = \frac{\sum_{g=1}^G Sn_g}{G}$$

Donde G representa el total de elementos.

Consensus: para estimar la capacidad de integración del proceso de clasificación, se otorgan los subgrupos de los tres métodos de asignación aplicando el criterio de mayoría de voto. (16)

Resultados

Los Algoritmos genéticos acoplados con el método N3 facilitaron la selección de un modelo óptimo de seis descriptores, detallados en la Tabla 1.

Tabla No 1. Descriptores moleculares para el modelo N3 y su descripción

No	Descritor	Bloque	Descripcion
1	B09[C-N]	Pares de atomos	Presencia/ausencia de enlaces C - N a la distancia topologica 9
2	nR09	No. anillos	Number of 9-membered rings
3	SpMax_AEA(b)	Indices espectrales	valores pesados por el orden de enlace de laa matriz de adiacencia aumen
4	B02[N-N]	Pares de atomos	Presencia/ausencia de enlaces N-N a disstancia topologica 2
5	B05[N-N]	Pares de atomos	Presencia/ausencia de enlaces N-N a disstancia topologica 5
6	D/Dtr09	Descriproes de anillo	Distanca del camino sobre el anillo de orden 9

En la tabla No.2 se detallan los parámetros de calidad de los tres modelos, descritos en su clasificación y predicción.

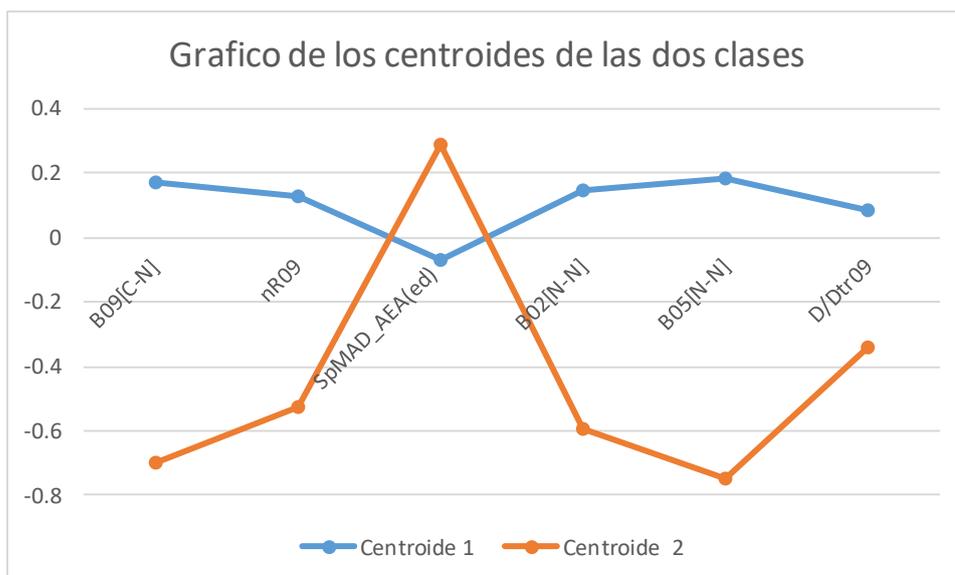
Tabla No 2. Descripción de los parámetros de calidad de los modelos

Modelo	Ner	Clase 1			Clase 2		
		Precision	Sensitivity	Specifity	Precision	Sensitivity	Specifity
KNN clasificacion	0.65	0.85	0.91	0.39	0.53	0.39	0.91
KNN prediccion	0.63	0.86	0.88	0.39	0.44	0.39	0.88
N3 clasificacion	0.69	0.89	0.75	0.64	0.39	0.64	0.75
N3 prediccion	0.70	0.91	0.68	0.72	0.35	0.72	0.68
SIMCA clasificacion	0.72	0.89	0.76	0.68	0.45	0.68	0.76
SIMCA prediccion	0.68	0.91	0.75	0.6	0.3	0.6	0.75
Consensus KNN clasificacion	0.88	0.95	0.95	0.81	0.81	0.81	0.95
Consensus KNN prediccion	0.86	0.92	1.00	0.73	1.00	0.73	1.00

Al analizar la tabla 2 se deduce que el modelo de consensus es óptimo para predecir efectos adversos, mientras que en cada uno de los otros se tiene siempre una clase no bien clasificada.

Se realizó un cálculo de los valores centrales en cada clase, con el fin de asimilar el interés de los descriptores en la predicción de la clase de pertenencia.

Figura No. 1: representación gráfica de los centroides pertenecientes a cada descriptor en cada clase.



Podemos apreciar en el gráfico que el grupo de fármacos de la clase 1 – inactivos o poco activos, tienen valores de los descriptores B09[C-N], nR09, B02[N-N], B05[N-N], y D/Dtr09 más altos que los fármacos de la clase 2 – activa, así como el valor promedio del descriptor SpMAD_AEA(ed) es más alto.

Con esto podemos decir que los resultados pueden interpretarse en esta forma: el grupo de moléculas que forma parte de la clase 1 contienen, en general más enlaces carbono nitrógeno y nitrógeno-nitrógeno respecto a los de la clase 2. Los fármacos de la clase más activa tienen valores de los descriptores B09[C-N], nR09, B02[N-N], B05[N-N], y D/Dtr09 que los de la clase menos activa, entonces tienen menos enlaces C-N y N-N, pero como los valores del descriptor SpMAD_AEA(ed) es más alto, significa que sus moléculas son más complejas en cuanto tienen más dispersión de los auto valores de la matriz de adyacencia pesada por el orden de los átomos vértices.

Discusión

La relevancia de este estudio radica en que para pronosticar la clase de pertenencia de nuevos fármacos bastaría con introducirlos en el test de validación y aplicar la predicción

utilizando como base el training set actual, lo que permite predecir los efectos adversos mediante los descriptores, y de esa manera poder disminuir su impacto en el organismo, ya que hoy en día tanto los psicofármacos como las hormonas son ampliamente utilizados. Es importante señalar que los estudios farmacológicos QSAR existentes no tratan sobre estos fármacos y por este motivo no se puede realizar una comparación. Sin embargo, mediante este estudio pretendemos ampliar la visión de las investigaciones farmacológicas para la creación de nuevos fármacos y el mejoramiento de los ya existentes, mediante moléculas que identifiquen los supuestos efectos adversos de dichos fármacos. Otra de las ventajas que presenta este modelo es la capacidad de disminuir los recursos económicos, humanos y de tiempo en el proceso de investigación y desarrollo de un fármaco.

La sensibilidad y especificidad se obtuvo a través del modelo de consenso, dicho modelo es muy bueno para predecir efectos adversos, además de describir la predicción de cada uno. (Ver tabla No.2)

Nuestro modelo predictivo se obtuvo utilizando los siguientes descriptores B09[C-N], nR09, B02[N-N], B05[N-N], y D/Dtr09, que representan a los fármacos de clase más activa, dicha clase tiene menos enlaces carbono nitrógeno y nitrógeno-nitrógeno, lo que demuestra que tienen menos efectos adversos en comparación con el valor promedio del descriptor SpMAD_AEA(ed) que es más alto y refleja mayores efectos adversos. (Ver figura 1)

Durante la elaboración del presente estudio se destaca como ventaja la mínima dificultades en su ejecución, no se requiere de recursos humanos para pruebas, es un estudio innovador en nuestro país que seguramente despertara el futuro interés de investigaciones farmacológicas.

Sin embargo presenta debilidades en cuanto a la falta de estudios de metodología y análisis QSAR, lo que dificulta realizar comparaciones o búsquedas más exhaustivas, además la base de datos de moléculas a estudiar es limitada, y no se describen todos los efectos adversos en los programas, lo que representa una desventaja en el análisis del presente estudio y en la elaboración de futuros fármacos .

Conclusiones:

En el contexto de nuestro estudio, como sabemos hoy en día la prescripción de fármacos es una de las piezas clave en el manejo de enfermedades , tanto a nivel psiquiátrico como

a nivel endocrinológico, con tal importancia que en términos económicos, la mayoría de patologías psiquiátricas ameritan un tratamiento con dos fármacos, lo que implica un costo mensual promedio $30,67 \pm 24,94$ dólares. (17) Además, la facturación de farmacia en general en el año 2016, según datos del Ministerio de Sanidad, era de más de 900 millones de recetas en el Sistema Nacional de Salud.

Sin embargo, a pesar de la gran importancia de la industria farmacéutica, pocos son los que conocen el riesgo de prescribir fármacos nuevos o poco estudiados. Por todo ello, se recomienda que los pacientes no tengan contacto con este tipo de fármacos, ya que se desconoce de su comportamiento al entrar en contacto con el cuerpo. Sin embargo, durante las fases de investigación clínica, se esperaría poder trabajar directamente con pacientes, pero por cuestiones éticas evidentemente no puede ser así, en todos los casos, lo que conlleva a prácticas con animales, que no siempre van a ser precisas, por ejemplo, manifestaciones clínicas como somnolencia, tinnitus o confusión mental, etc. no son fácilmente reconocidas en los estudios preclínicos con animales. (18)

Es por lo expuesto, que el modelo matemático puede contribuir, al reducir el riesgo de exposición de los pacientes a efectos secundarios que pueden ser potencialmente peligrosos, evitar que se trabaje con animales lo que mejoraría la precisión al definir la presencia de efectos secundarios, además de disminuir costos a nivel de sanitario, y mejorar el desarrollo de medicamentos en el entorno de la ingeniería molecular.

Finalmente, cabe recalcar que se debe tomar en cuenta para desarrollar modelos matemáticos con metodología QSAR, la necesidad de obtener bases de datos más extensas, por lo que recomendamos que posteriormente se realicen nuevos estudios, con nuevos datos que incluyan más efectos secundarios, con el propósito de obtener un concepto más global de las consecuencias de estos. (19)

Al ser ésta una metodología nueva, y poco investigada, en un futuro se espera que, por sus beneficios, se convierta en pieza angular en la estructuración, producción de nuevos fármacos y mejore la eficiencia de los ya existentes.

Agradecimientos

A nuestro director de tesis , a nuestro asesor metodológico, extendemos nuestro agradecimiento por su direccionamiento durante todo el proyecto, agradecemos también a nuestros padres por su apoyo incondicional en cada paso del camino.

BIBLIOGRAFIA

1. Organización Panamericana de la Salud. OPS/OMS | [Internet]. «Depresión: hablemos», dice la OMS, mientras la depresión encabeza la lista de causas de enfermedad. 2017. Disponible en: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=13102:depr

ession-lets-talk-says-who-as-depression-tops-list-of-causes-of-ill-health&Itemid=1926&lang=es

2. Claudio Liberman G. Prevalencia e incidencia de los principales trastornos endocrinos y metabólicos. Rev. Medica Clínica Las Condes. 2013;24(5):735-41.
3. Pichler W. Drug hypersensitivity: Classification and clinical features - UpToDate [Internet]. UpToDate. 2019 [citado 6 de mayo de 2021]. Disponible en: https://www.uptodate.com/contents/drug-hypersensitivity-classification-and-clinical-features?search=Drug%20hypersensitivity:%20Classification%20and%20clinical%20features&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1
4. Galan N, Pedroza A. ESTUDIO HOLOGRAMICO DE LA RELACIÓN CUANTITATIVA ESTRUCTURA - ACTIVIDAD (HQSAR) PARA LA PREDICCIÓN DE ACTIVIDADES BIOLÓGICAS DE ANTAGONISTAS DEL RECEPTOR DE ADENOSINA A2A RELACIONADO CON LA ENFERMEDAD DE PARKINSON. [Colombia]: Universidad de Cartagena; 2009.
5. Lozano-Aponte J, Scior T. ¿Qué sabe Ud. acerca de...QSAR? Rev Mex Cienc Farm. 2012;43(2):82-4.
6. Granados-Tavera K, Tilvez EA, Ahumado-Monterrosa M. Modelado de las Relaciones Cuantitativas Estructura-Actividad (QSAR) de Análogos del Tipifarnib con Actividad Antichagásica. Inf Tecnológica. 2019;30(1):3-14.
7. WERMUTH C, GANELLIN C, LINDBERG P, MITSCHER L. GLOSSARY OF TERMS USED IN MEDICINAL CHEMISTRY. Pure & App/. 1998;70(5):1129-43.
8. Radax J, Torres G, Jara J. Métodos de correlación estructura-actividad aplicados a efectos adversos de fármacos cardiovasculares, Cuenca, 2018. [Cuenca-Ecuador]: Universidad del Azuay; 2019.
9. Cajamarca M. Aplicación de métodos QSAR y quimiométricos para determinar la relación estructura-actividad de pesticidas en mamíferos (rata). [Cuenca-Ecuador]: Universidad Politécnica Salesiana; 2015.
10. Hähnke VD, Kim S, Bolton EE. PubChem chemical structure standardization. J Cheminformatics [Internet]. 2018 [citado 6 de mayo de 2021];10(36). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6086778/>
11. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theory. 1967;13(1):21-7.
12. Leardi R, Lupiáñez González A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemom Intell Lab Syst. 1998;41(2):195-207.
13. Cassotti M, Ballabio D, Todeschini R, Consonni V. A similarity-based QSAR model for predicting acute toxicity towards the fathead minnow (*Pimephales promelas*). SAR QSAR Environ Res. 2015;26(3):217-43.

14. Wold S. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* 1976;8(3):127-39.
15. Consonni V, Ballabio D, Todeschini R. Comments on the Definition of the Q^2 Parameter for QSAR Validation. *J Chem Inf Model.* 2009;49(7):1669-78.
16. Ballabio D, Biganzoli F, Todeschini R, Consonni V. Qualitative consensus of QSAR ready biodegradability predictions. *Toxicol Environ Chem.* 2016;1-24.
17. Camino A, Chávez C, Hernández V, Villarroel C. Análisis económico del costo de los medicamentos usados en patologías psiquiátricas: el tercio de los ingresos de un paciente se consumen en fármacos. *Revista De La Facultad De Ciencias Médicas.* 2017;42(2):91-5.
18. Guerrero GAM. Las fases en el desarrollo de nuevos medicamentos. *Rev Fac Med UNAM.* 2009;52(6):260-4.
19. Tripaldi P, Pérez B. Actividad Antihipertensiva de Péptidos obtenidos por hidrolisis y/o fermentación de alimentos – Modelización QSAR. [Cuenca-Ecuador]: Universidad del Azuay; 2014.