



NUEVOS ENFOQUES PARA LA DETERMINACIÓN
DE PROPIEDADES FARMACOCINÉTICAS DE
MOLÉCULAS TERAPÉUTICAS ACTIVAS Y SU
REPERCUSIÓN EN EL DESCUBRIMIENTO DE
NUEVOS FÁRMACOS

Trabajo Fin de Grado
Grado en Farmacia
Facultad de Farmacia
Universidad de Sevilla

Alumna: Marta Contreras García-Trevijano
Tutora: M^a Violante de la Paz Bañez

RESUMEN

Este Trabajo versa sobre nuevas técnicas y enfoques para el descubrimiento de nuevas moléculas bioactivas y la mejora de las ya existentes. Se debe destacar la farmacocinética (PK) de cada molécula y su proceso de absorción, distribución, metabolización, excreción y toxicidad (ADMET), así como la búsqueda para agilizar el proceso de análisis y desarrollo de nuevos fármacos. De esta manera, la Quimioinformática introdujo una rutina de búsqueda en la que se relaciona la estructura-actividad y estructura-propiedad de cada una de las moléculas bioactivas a estudiar. Como consecuencia del avance informático, surge la Quimiometría que, usando señales analíticas y datos, aporta información. A raíz del progreso en estas disciplinas y su unión con las nuevas tecnologías se abre camino al descubrimiento de mecanismos de acción de determinados compuestos que, junto con la Química Combinatoria, dan lugar a las quimiotecas. Todo esto llevó a lo denominado “*Machine Learning*”. Como objetivos se han planteado, entre otros, la revisión de las técnicas ya existentes, conocer los programas informáticos y metodologías disponibles para la investigación y desarrollo de fármacos, así como la evaluación del potencial científico de las distintas técnicas. Los principales resultados que se han obtenidos en esta revisión bibliográfica parten de la idea de que las propiedades ADMET juega un papel crucial en el descubrimiento de nuevos fármacos y moléculas.

PALABRAS CLAVE

ADMET; Farmacocinética; SVM; *Random Forest*; Predicción; *drug-like*; QSAR; k-NN

ABSTRACT

This work deals with new techniques and approaches for the discovery of new bioactive molecules and the improvement of existing ones. The pharmacokinetics (PK) of each molecule and its process of absorption, distribution, metabolization, excretion and toxicity (ADMET) should be highlighted, as well as the search to speed up the process of analysis and development of new drugs. In this way, Chemoinformatics introduced a search routine in which the structure-activity and structure-property of each of the bioactive molecules to be studied are related. Because of the computer science advance, Chemometrics arises and, using analytical signals and data, provides valuable information. As a result of the progress in these disciplines and their union with new technologies, the way is opened to the discovery of mechanisms of action of certain

compounds that, together with Combinatorial Chemistry, give rise to chemotheques. All this led to what is called "Machine Learning". The objectives have been set, being one of the main ones the review of existing techniques, the exploration the computer programs and methodologies available for research and development of drugs, as well as the evaluation of the scientific potential of the different techniques. The main results obtained in this scientific literature revision are based on the idea that ADMET properties plays a crucial role in the discovery of new drugs and molecules.

KEYWORD

ADMET; *Pharmacokinetics*; SVM; *Random Forest*; *Prediction*; *Drug-like*; QSAR; k-NN.

Índice

1. Introducción.....	6
2. Objetivos de la revisión.....	7
3. Metodología.....	8
3.1. Búsqueda de material.....	8
4. Resultados y discusión.....	8
4.1. Quimioinformática y búsqueda de nuevos fármacos	8
4.2. Control del ADMET con las técnicas y aprendizajes existentes	12
4.3. Técnicas actuales en el estudio de las propiedades ADMET	13
4.4. Herramientas útiles y accesibles a través de la web	18
5. Conclusiones.....	22
6. Bibliografía	23

Índice de Figuras

Figura 1. Tendencia o interés en la búsqueda de “Machine Learning” desde 2004 hasta hoy.	7
Figura 2. Representación de la estrategia de síntesis mediante química combinatoria. .	11
Figura 3. representación de síntesis de mezclas y en paralelo.....	11
Figura 4. Método k-NN para predicción de la toxicidad en hepatoblastoma humano. ...	12
Figura 5. Técnicas y métodos ML para la evaluación del ADMET	14
Figura 6. Metodología para la clasificación y predicción de la farmacocinética y toxicidad.	17
Figura 7. Diseño de recursos accesibles para la creación de perfiles farmacocinéticos. .	19

Índice de siglas y acrónimos

Abreviatura	Significado
AB	<i>Ada Boost</i>
ADMET	<i>Absorption, distribution, metabolism, excretion and toxicity</i>
AECF	<i>Adaptive Ensemble Classification Framework</i>
ANN	<i>Artificial Neural Network</i>
BB	<i>Building Blocks</i>
CYP	<i>Cytochrome P</i>
DFT	<i>Density-Functional Theory</i>
DT	<i>Decision Tree</i>
ERT	<i>Extremely Randomized Trees</i>
ETC	<i>Extra Trees Classifier</i>
FAME	<i>FAst MEtabolizer</i>
GB	<i>Gradient Boosting</i>
GP	<i>Gaussian Process</i>
HIA	<i>Human Intestinal Absorption</i>
HTS	<i>High Throughput Screening</i>
I+D+i	Investigación, desarrollo e innovación
<i>k</i> -NN	<i>k</i> -Nearest Neighbors
LDA	<i>Linear Discriminant Analysis</i>
logBB	<i>Log [brain]/[blood]</i>
LR	<i>Logistic Regression</i>
MARS	<i>Multivariate Adaptive Regression Splines</i>
ML	<i>Machine Learning</i>
MLM	<i>Mouse Liver Microsomal</i>
MLR	<i>Multiple linear regression</i>
NB	<i>Naive Bayes</i>
P-gp	<i>P-glycoprotein</i>
PK	<i>Pharmacokinetics</i>
PLS	<i>Partial Least Square</i>
PNN	<i>Probabilistic Neural Network</i>
PSA	<i>Polar Surface Area</i>
QSAR	<i>Quantitative Structure-Activity Relationship</i>

QSPR	<i>Quantitative Structure-Property Relationship</i>
RF	<i>Random Forest</i>
SARS-CoV-2	<i>Severe Acute Respiratory Syndrome Coronavirus 2</i>
SMO	<i>Sequential minimal optimization</i>
SNC	<i>Sistema Nervioso Central</i>
SoM	<i>Site of Metabolism</i>
SVM	<i>Support Vector Machines</i>

1. Introducción

En los últimos años, y más aun si cabe en el 2020 debido a la aparición del SARS-CoV-2 (González-Paz et al., 2020) nos hemos visto en la necesidad de agilizar el análisis y desarrollo de nuevos fármacos. Esto nos ha llevado a los avances en las técnicas ya existentes. La mayoría están basadas en la optimización de la farmacocinética (PK) y farmacodinámica de las diferentes moléculas activas y sus mecanismos de absorción, distribución, metabolismo, excreción y toxicidad (ADMET), todo ello analizando cada una de las técnicas con más importancia. (Ferreira & Andricopulo, 2019).

Previo al uso de modelos computacionales en la investigación farmacéutica, el proceso de actuación era el cribado fenotípico. Esto es, una estrategia clásica basada en la búsqueda de un organismo que muestre un fenotipo asociado a una enfermedad, dando lugar a la consiguiente investigación y los ensayos con distintos fármacos evaluando la eficacia y toxicidad. Esto llevó al desarrollo de ligandos con máxima selectividad, los cuales actúan sobre un único blanco molecular el cual está asociado a la enfermedad objetivo. Este enfoque se basaba en dos puntos clave:

- Fármacos altamente específicos que no dan lugar a efectos secundarios, produciendo así terapias más seguras.
- Fármacos que actúan sobre un único blanco, aunque esto solo ocurra en algunas enfermedades.

Esta hipótesis se comprobó errónea, ya que los fármacos se unen a más de un objetivo molecular. Esta comprobación fue un hecho importante para llegar a la perspectiva de lo que se conoce como “reposicionamiento de fármacos o búsqueda de segundos usos médicos”. Todo esto desembocó en el diseño y la búsqueda racional de fármacos (Gavernet, 2021).

A raíz de las hipótesis y comprobaciones previas, cada día está más en auge lo que se denomina *Machine Learning* (ML). Esta es una rama de la inteligencia artificial que desarrolla modelos analíticos, por lo que desempeña un papel muy importante en el desarrollo de nuevos modelos de moléculas para estudiar sus diferentes acciones. *Machine Learning* es una disciplina enfocada en responder algunas preguntas, entre la que se encuentra: “¿Cómo puede un sistema computacional mejorar a través de su propia experiencia?” (Jordan & Mitchell, 2015).

Esta disciplina cada vez muestra más importancia en base al buscador *Google Trends*. Desde el año 2016 hasta la actualidad se han visto aumentadas sus búsquedas en este buscador (Google), mostrando su máximo auge en estos últimos años (Figura 1).

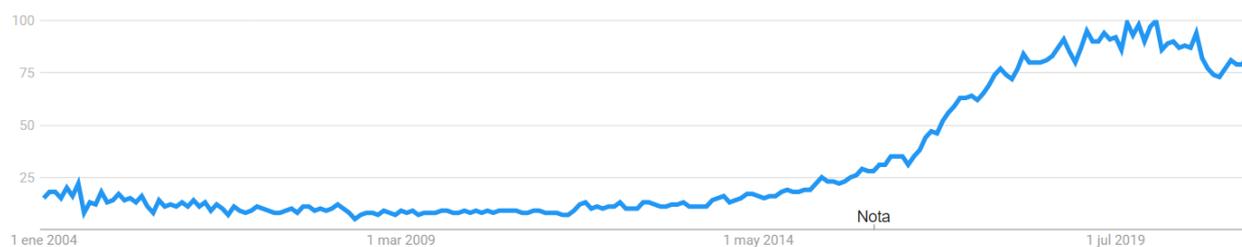


Figura 1. Tendencia o interés en la búsqueda de “Machine Learning” desde 2004 hasta hoy.

Obtenida de Google Trends (Machine Learning - Explorar - Google Trends, s. f.)

Hay que destacar la importancia del conocimiento de procesos ADMET a la hora de trabajar con ML ya que toda la información recogida de las formas de actuar que tiene cada molécula nos ayuda a ir creando una base de datos cada vez mayor.

Por otro lado, la pérdida de eficacia y seguridad juega un papel importante en la investigación, desarrollo e innovación (I+D+i) de fármacos y estos, a su vez, pueden influir en las propiedades farmacocinéticas, aunque bien es cierto que este impacto ha disminuido debido al mayor control existente desde la última década.

2. Objetivos de la revisión

Se exponen, a continuación, los objetivos que se han postulado en este Trabajo:

- Realizar una revisión de la bibliografía actual en referencia a las técnicas y métodos existentes sobre la búsqueda de nuevos fármacos o la evaluación de los ya existentes a través de los medios informáticos.
- Conocer las metodologías para investigar y desarrollar nuevos fármacos de manera eficaz y eficiente.
- Diferenciar ventajas e inconvenientes de las diferentes técnicas.
- Evaluar el potencial científico de las distintas técnicas.
- Comprender la incorporación de las nuevas tecnologías y el cambio de paradigma que estas le han aportado a la investigación de diferentes grupos de moléculas y las actuaciones de estas en distintos niveles.

3. Metodología

3.1. Búsqueda de material

La búsqueda de materiales (documentos bibliográficos) se ha realizado en diferentes bases de datos de forma sistemática. De esta manera, se han usado diversas bases de datos, como son: *PubMed*, *Scopus* y *Google Scholar*.

Así, se han utilizado diferentes palabras claves como descriptores, unidas a los operadores booleanos “AND” y “OR”. Entre las palabras claves utilizadas se encuentran: ADMET, PK, k-NN, SVM, RF y QSAR (tanto siglas como palabras completas, diferenciando igualmente los términos en español e inglés). De esta manera quedarían, por ejemplo, conjunciones como las siguientes: “ADMET AND k-NN” o “ADMET AND SVM”.

Se ha filtrado los artículos resultantes por título, esto es, que estuvieran escritos en español e inglés, obviando otros idiomas. Posteriormente se filtraron por resumen o *abstract* interesantes y relacionado con los diferentes objetivos. Por último, se han utilizado referencias bibliográficas de los artículos previamente leídos.

4. Resultados y discusión

4.1. Quimioinformática y búsqueda de nuevos fármacos

Hoy, entre los métodos más usados están lo que llamamos la quimioinformática. Fue propuesto en 1957 cuando Ray y Kirsch publicaron una rutina de búsqueda de estructuras químicas (Ray & Kirsch, 1957). En 1962, Hansch perfeccionó estas búsquedas en las que se desarrolló de manera más clara la relación estructura-actividad y estructura-propiedad (Hansch et al., 1962; Hansch & Fujita, 1964) Se crearon programas informáticos como QSAR (del inglés *Quantitative Structure-Activity Relationship*). Teniendo como referencia el libro *Introducción a la Química Medicinal* (Gavernet, 2021), el método QSAR se basa en la Ecuación:

$$A = f(d_1, d_2, d_3, \dots, d_x)$$

Ecuación 1

Donde “A” es la actividad biológica de un compuesto químico, se considera una función matemática con ciertas características estructurales cuantificadas mediante descriptores moleculares (d_1, d_2). La relación “f” puede ser obtenida por diferentes métodos estadísticos, métodos evolutivos, etc. Una vez obtenida esta relación, la actividad A de un compuesto nuevo o no testeado frente al blanco molecular seleccionado puede ser predicha a partir de su estructura molecular mediante el cálculo del valor que asumen los

descriptores moleculares (d1, d2, etc.) (incluidos en el modelo QSAR) para ese compuesto.

Siguiendo a Gavernet (Gavernet, 2021), existen diferentes formas de clasificar los modelos QSAR:

Teniendo en cuenta la naturaleza de la variable dependiente, podemos clasificarlos:

- Modelos QSAR cuantitativos: son aquellos que “darán como resultado una variable dependiente continua y predecirán el valor numérico de la actividad modelada”.
- Modelos QSAR cualitativos: son aquellos que “darán como resultado una respuesta cuyo valor estará asociado a una categoría previamente establecida”.

En relación con la dimensionalidad de los descriptores moleculares usados en los modelos, estos se podrán agrupar en:

- Modelos QSAR de baja dimensionalidad: son aquellos que “utilizan descriptores moleculares independientes de la conformación. Estos tienen la ventaja de que no requieren la optimización conformacional de las moléculas”.
- Modelos QSAR de alta dimensionalidad: son aquellos que “utilizan descriptores dependientes de la conformación de las moléculas”. Aquí encontraríamos los modelos QSAR-3D y 4D.

Teniendo en cuenta la manera en que se seleccionan las variables independientes y la forma en que se correlacionan con la actividad de interés, los modelos pueden ser:

- Modelos QSAR lineales: supone una relación lineal de la variable dependiente e independiente. Estas pueden ser cuantitativas (regresión lineal) o cualitativas (análisis lineal discriminatorio).
- Modelos QSAR no lineales: “los descriptores se asignan principalmente a un espacio relacional no lineal y ayudan a superar algunas de las limitaciones de los métodos lineales”.

Ya en la década de los 80 apareció la Quimiometría, que se define como una disciplina que tiene como finalidad transformar señales y datos analíticos en información. Utiliza métodos de origen matemático y estadístico, aunque también tiene gran influencia la química analítica. Uno de los pioneros en este campo fue Hinderling (Hinderling et al., 1984).

Fue en 1990 cuando Lipinski, pionero en la correlación de unas variables fisicoquímicas con propiedades PK y contando con un gran número de datos disponible, forjó la famosa Regla de los Cinco (Lipinski et al., 2012).

Ya en el 2000 surgieron nuevas pautas relevantes en relación con la química médica donde hay que destacar el uso de los farmacóforos (“conjunto de características necesarias para asegurar las interacciones supramoleculares óptimas con un objetivo biológico específico y para disparar o bloquear su respuesta biológica” (Wermuth et al., 1998)) para clasificar cada molécula como fármaco o no fármaco (Ghose et al., 1999; Muegge et al., 2001).

En este periodo surge lo que se denomina la era de “los grandes datos” (Ferreira & Andricopulo, 2019) donde se puede destacar la HTS (*High Throughput Screening*) (Gavernet, 2021), metodología capaz de ejecutar una gran cantidad de ensayos de manera simultánea valiéndose de la robótica y las tecnologías avanzadas. En relación con el descubrimiento de nuevos fármacos, esta técnica nos da la posibilidad de evaluar una amplia variedad de prototipos activos. Una de las desventajas que presenta es el gran coste económico que supone, pero a pesar de esto su gran ventaja es la velocidad y automatización. El campo de actuación donde podría tener mas relevancia su uso se encuentra en el descubrimiento de mecanismos farmacológicos de determinados compuestos o fármacos.

También es de gran importancia la Química Combinatoria (Figura 2) (Gavernet, 2021), que se define como “un conjunto de procedimientos que hace posible la generación eficiente, simultánea y rápida de una gran cantidad de estructuras”. A todas estas sustancias se les conoce como biblioteca de compuestos o quimiotecas. La química combinatoria nos resulta útil como fuentes de compuestos líderes y como herramienta para la preparación de varios miembros de una misma familia estructural para la optimización de fármacos.

La forma de preparar estas quimiotecas es tomar un conjunto de compuestos diferentes, pero que estén químicamente relacionados, para que, reaccionando con otro grupo de estructuras, generen bajo las mismas condiciones de reacción, todos los productos posibles que surgen de la combinación de ambos grupos del BB (*building blocks* BB, o bloques de construcción).

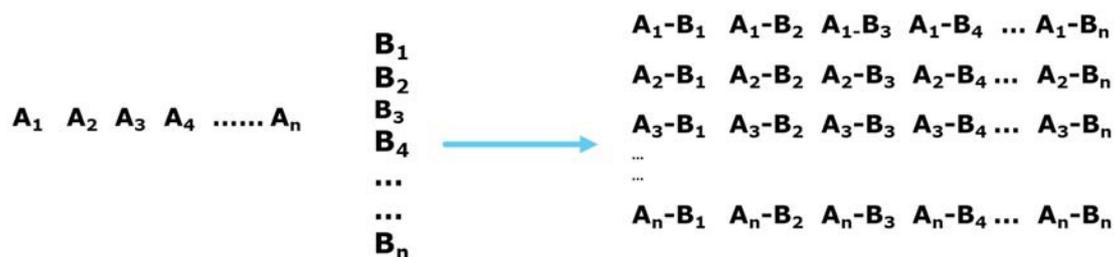


Figura 2. Representación de la estrategia de síntesis mediante química combinatoria.

Obtenida de (Gavernet, 2021)

En la síntesis combinatoria las quimiotecas pueden obtenerse de dos maneras: mediante síntesis de mezclas o síntesis en paralelo (Figura 3). Mediante el primer procedimiento se obtienen distintos productos dentro del mismo medio de reacción y en el caso, de la segunda síntesis cada compuesto se prepara de forma separada en un recipiente individual.

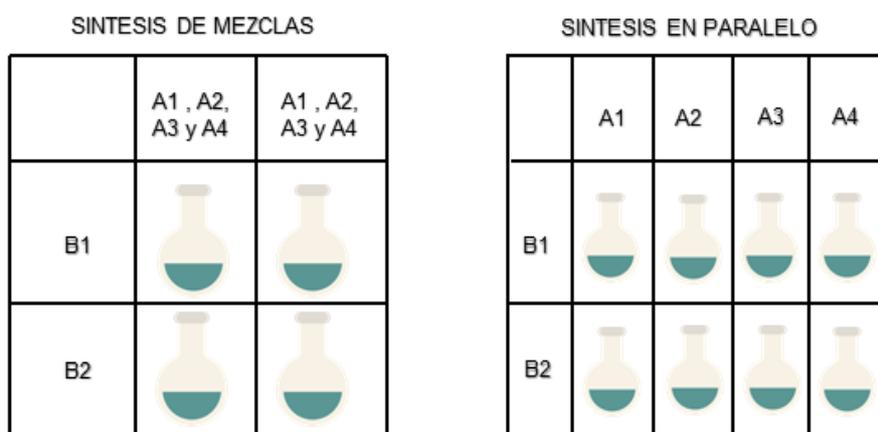


Figura 3. representación de síntesis de mezclas y en paralelo.

Obtenida de (Gavernet, 2021)

Con todo esto se puede decir que se ha pasado de las técnicas denominadas “convencionales” a los métodos más avanzados del “aprendizaje automático” o *Machine Learning* (ML).

4.2. Control del ADMET con las técnicas y aprendizajes existentes

ML se utiliza en el descubrimiento de fármacos y en la predicción de sus PK. Hay varias técnicas que se van a destacar: “*k* vecino más cercano” (*k*-Nearest Neighbors algorithm, *k*-NN), “máquinas de vectores de soporte” (*Support Vector Machines*, *SVM*), “bosque aleatorio” (*Random Forest*, *RF*) y “redes neuronales artificiales” (*Artificial Neural Network*, *ANN*) (Tao et al., 2015).

Se explica a continuación en qué consiste cada una de ellas:

- El *k* vecino mas cercano (*k*-NN) es un método no lineal para el reconocimiento de patrones (Figura 4).

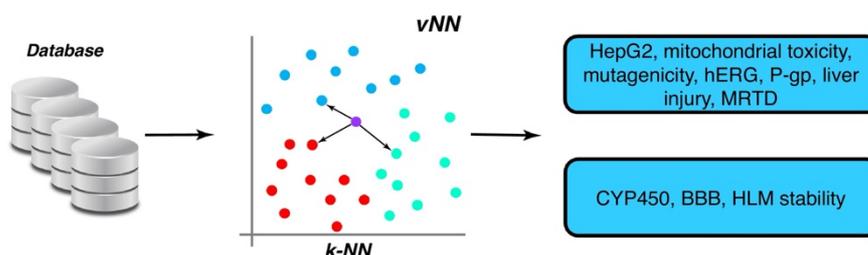


Figura 4. Método *k*-NN para predicción de la toxicidad en hepatoblastoma humano.

Obtenida de (Ferreira & Andricopulo, 2019)

- SVM son métodos relacionados con la clasificación y regresión. Los algoritmos de SVM buscan la posición de dos puntos en el espacio, separando las clases a dos espacios lo más amplios posibles. La separación existente entre estos puntos se llama “vector soporte” (Heikamp & Bajorath, 2014).

Hay que destacar un estudio que aplica SVM al modelado de la PK, donde Doniger y colaboradores investigaron el paso de la barrera hematoencefálica por un conjunto de 179 compuestos activos a nivel del Sistema Nervioso Central (SNC) y 145 compuestos inactivos. En este ensayo se utilizaron 30 compuestos de prueba y el modelo dio un rendimiento del 81,5% (Doniger et al., 2002).

- RF es un método de conjunto que, al igual que SVM, se puede usar para tareas de clasificación y regresión. Este método selecciona automáticamente descriptores moleculares, influyendo en el tiempo de cálculo disminuyéndolo drásticamente y permitiendo una exploración de conjuntos de datos mucho más grandes (Breiman, 2001; Cano et al., 2017).
- ANN es uno de los métodos más importante en ML ya que tiene una gran importancia en el descubrimiento de fármacos tanto en su análisis de regresión como en su clasificación. Lo que se pretende conseguir con ANN es imitar la

estructura de los circuitos neuronales biológicos para que se puedan procesar datos de manera similar a como lo hace el cerebro humano. El funcionamiento consiste en la conexión de muchas unidades de procesamiento que trabajan en paralelo. Estas unidades de procesamiento o neuronas se clasifican en capas: la capa de entrada que recibe los datos, capas ocultas que procesan los datos por medio de una función de activación y una capa de salida que genera la respuesta al problema en cuestión. Las ANN profundas, en comparación con las redes superficiales convencionales, tienen un mayor número de capas ocultas lo que les permite manejar mejor grandes conjuntos de datos (Zhang et al., 2017).

Los pioneros en el uso de ANN en la predicción de propiedades PK fueron Gobburu y Shelver (Gobburu & Shelver, 1995) . Utilizaron unos antagonistas de los receptores β -adrenérgicos para predecir algunas propiedades, como pueden ser la unión a proteínas plasmáticas, el volumen de distribución, el aclaramiento renal y no renal y el tiempo medio de residencia.

A pesar de los avances descritos y desarrollados, estos métodos no permiten profundizar en el conocimiento de tantas moléculas activas y se hizo necesaria la investigación y desarrollo de otras técnicas para el estudio de las propiedades ADMET.

4.3. Técnicas actuales en el estudio de las propiedades ADMET

Los sistemas de clasificadores múltiples se basan en la combinación de modelos individuales y se han aplicado para paliar los problemas de alta dimensionalidad y los conjuntos de datos desequilibrados (Haixiang et al., 2017). Esto a veces nos conduce a una serie de datos de las propiedades ADMET sesgados. Por ello se introdujo por parte de Yang y colaboradores un enfoque novedosos denominado marco de clasificación de conjunto adaptativo (AECF, del inglés *Adaptive Ensemble Classification Framework*) (Figura 5) (Yang et al., 2018).

Este método abarca varios pasos como son la generación de modelos individuales, la integración de modelos y la optimización del conjunto para tratar simultáneamente los problemas que comentamos antes de la dimensionalidad y el desequilibrio. Este método se usó para construir modelos de permeabilidad de Caco-2 (modelo de barrera epitelial intestinal), absorción intestinal humana (HIA, del inglés *Human Intestinal Absorption*), biodisponibilidad oral y unión de glucoproteína P (P-gp).

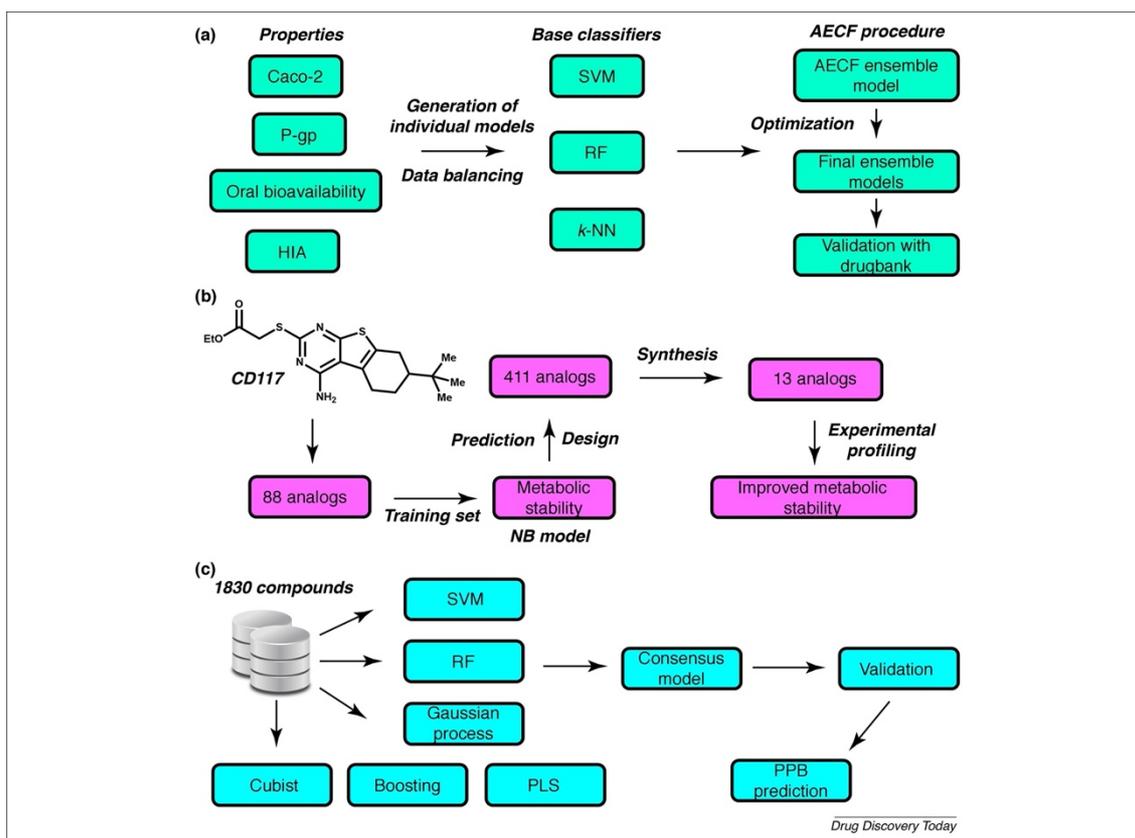


Figura 5. Técnicas y métodos ML para la evaluación del ADMET

(a) Marco de clasificación de conjuntos adaptativos (AECF) para predecir la permeabilidad de Caco-2, la absorción intestinal humana (HIA) y la unión a glucoproteína P. La maquinaria utilizada es SVM, RF y k-NN. (b) Análisis de Bayes para estimar la estabilidad metabólica de una serie de análogos del agente antituberculoso CD117. (c) El QSPR basado en PLS, RF, SVM, proceso gaussiano (GP) y Boosting para predecir la unión a las proteínas plasmáticas (PPB, del inglés plasma protein binding). Obtenida de (Ferreira & Andricopulo, 2019).

Entre todas las técnicas existentes de ML, la rutina AECF seleccionó como los mejores clasificadores de datos: SVM, RF y k-NN.

En el caso de Stratton y colaboradores, estos investigadores desarrollaron un modelo Baves (*Naive Bayes*, NB) para predecir la estabilidad microsomal del hígado del ratón (*Mouse Liver Microsomal*, MLM) (Stratton et al., 2017). Se usaron en total unos 88 análogos del agente antituberculoso metabólicamente inestable CD117 como compuesto de referencia, a partir de los cuales se predijeron 411 análogos. A continuación, se sintetizaron las 13 moléculas que se predijo que serían metabólicamente estables y se utilizaron en ensayos para determinar la estabilidad de MLM. Todas las moléculas sintetizadas superaron al compuesto original en términos de estabilidad metabólica. Con estos resultados quedó demostrada la solidez del modelo NB para predecir la estabilidad MLM de nuevos agentes antituberculosos.

La capacidad de cruzar la barrera hematoencefálica, mediada por el coeficiente de partición sangre-cerebro (logBB), se investigó para 287 compuestos en un estudio de Zhu y colaboradores (Zhu et al., 2018). Se construyeron modelos *Quantitative Structure-Property Relationship* (QSPR) basados en SVM, *Multiple Linear Regression* (MLR), ranuras de regresión adaptativa multivariantes (*Multivariate Adaptive Regression Splines*, MARS) y RF. Se evaluó la certeza estadística y el poder predictivo de estos modelos y se realizó un análisis de su dominio de aplicación. El algoritmo de RF (coeficiente de correlación predictiva) superó a los otros métodos ML e identificó el área de superficie polar (PSA) y el coeficiente de partición octanol-agua (logP) como los principales parámetros determinantes en la penetración del fármaco a través de la barrera hematoencefálica.

Wang y colaboradores investigaron la unión a proteínas plasmáticas con un conjunto de datos de 1830 compuestos estructuralmente distintos (Wang et al., 2017). Se construyeron 6 modelos predictivos utilizando los *Partial Least Square* (PLS), RF, SVM, *Cubist*, Proceso Gaussiano (GP) y *Boosting* (Figura 5c). Individualmente, el análisis PLS mostró los peores resultados, en comparación con los resultados de RF, que dieron lugar al mejor rendimiento. Seguidamente se construyó un modelo donde se combinaban los cinco mejores modelos. La lipofilia, la presencia de anillos aromáticos y la distribución de carga parcial se identificaron como los parámetros claves impulsores de la unión a proteínas plasmáticas. El estudio demostró un rendimiento superior de las técnicas más novedosas como la RF, con respecto a algunas muy utilizadas, como PLS, en la evaluación del perfil de unión a proteínas plasmáticas.

Kumar y colaboradores ampliaron los métodos utilizados para la predicción de HIA (R. Kumar et al., 2017) y emplearon la red neuronal probabilística (*Probabilistic Neural Network*, PNN) y el análisis discriminador lineal (*Linear Discriminant Analysis*, LDA). En este caso se utilizó un conjunto de datos de 1242 fármacos y compuestos *drug-like* para construir los modelos. Se obtuvo una predicción de un 90,38% para el conjunto de 745 compuestos, mientras que con 497 compuesto se obtuvo un 91,54% de predicción. Se observó que el poder predictivo disminuía con esta secuencia: SVM, ANN, PNN, *k*-NN, LDA y PLS. Una vez más SVM superó a otros métodos en cuanto a su poder predictivo y el método que obtuvo el rendimiento más pobre fue LDA (R. Kumar et al., 2017).

Finkelmann y colaboradores (Finkelmann et al., 2017) desarrollaron un método para la identificación de sitios de metabolismo (*Site of Metabolism*, SoM) basado en el entorno electrónico y estérico del átomo y la ubicación relativa en la molécula. Se usaron un conjunto de datos obtenidos a partir de 678 sustratos del citocromo P (CYP) (figura 6). Los cálculos que se llevaron a cabo estaban basados en la mecánica cuántica de la teoría del funcional de la densidad (*Density-Functional Theory*, DFT) para calcular las cargas parciales. Los entornos estéricos y las ubicaciones de los átomos se atribuyeron mediante descriptores de correlación, descriptores de funciones de distribución radial y la distancia topológica. Los modelos se pusieron a prueba utilizando diferentes técnicas, como árboles extremadamente aleatorios (*Extremely Randomized Trees*, ERT), SVM y RF, siendo esta última la que proporcionó los mejores resultados en términos del coeficiente de correlación de Matthews. Además, se utilizaron unos 25 compuestos con diferentes perfiles metabólicos como conjunto de validación externa. Este estudio representa una combinación de mecánica cuántica y medidas topológicas; por otra parte, están basadas en la distancia aplicada a la predicción de SoM

Se actualizó un conjunto de modelos de RF para la predicción de SoM llamado metabolizador rápido (FAME, del inglés *Fast Metabolizer*) (Kirchmair et al., 2013). En la nueva versión (FAME 2), Šícho y colaboradores añadieron un método de conjunto basado en ERT para investigar la influencia de los descriptores moleculares topológicos y cuánticos en la precisión de la predicción (Geurts et al., 2006; Šícho et al., 2017). Además, se desarrollaron e integraron modelos para predecir la regioselectividad de CYP3A4, CYP2D6 y CYP2C9 en FAME 2. El paquete incluye una salida gráfica que señala las posibles SoM e indica la probabilidad de que un átomo participe en reacciones metabólicas (Figura 6b). El software está disponible de forma gratuita bajo petición.

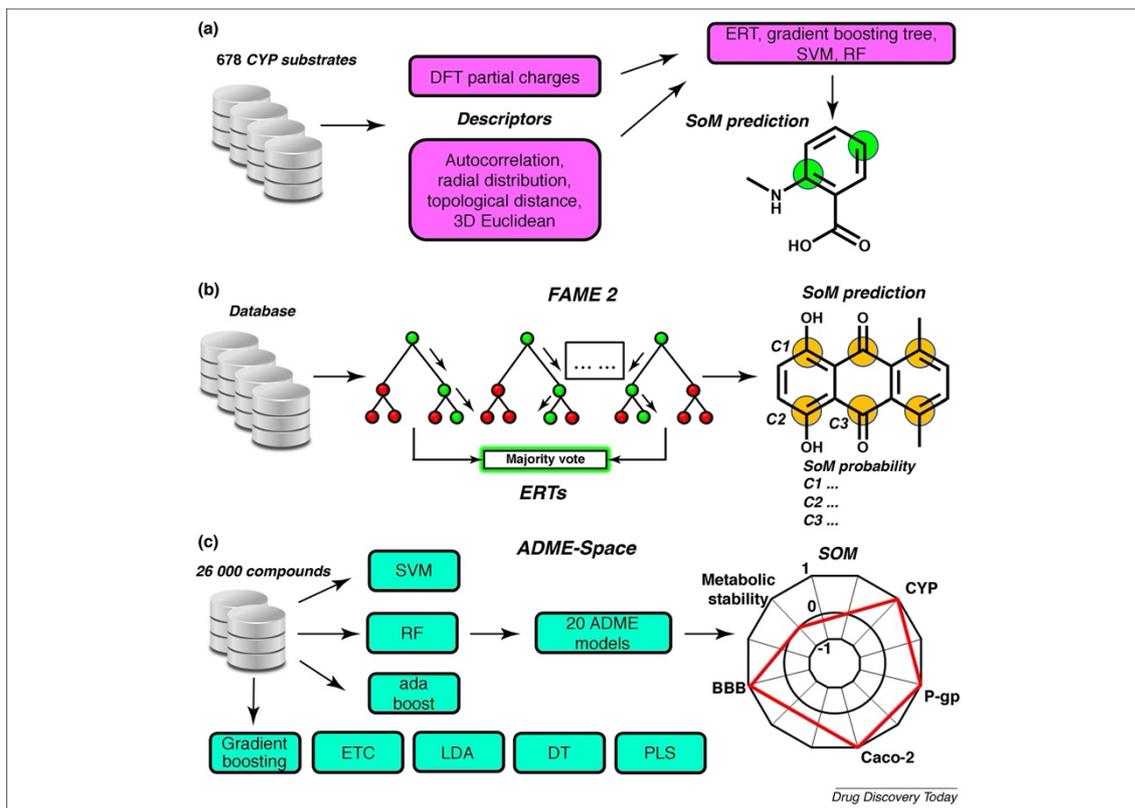


Figura 6. Metodología para la clasificación y predicción de la farmacocinética y toxicidad.

(a) El uso de ERT, árbol de aumento de gradiente, SVM y RF para identificar de sitios de metabolismo (SoM). La DFT: cargas parciales derivadas, métricas topológicas y basadas en la distancia se utilizaron como descriptores para desarrollar los modelos. (b) La incorporación de ERT en FAME 2 para la predicción de SoM. (c) ADME-Space: un conjunto de 20 modelos cuantitativos de QSPR para la predicción de estabilidad metabólica, la permeabilidad de Caco-2, la permeabilidad de la barrera hematoencefálica, la unión de la glucoproteína P y la inhibición de la CYP450. Los modelos se basan en SVM, RF, Ada Boost (AB), LDA, GB, DT, PLS y ETC. Obtenida de (Ferreira & Andricopulo, 2019).

Bocci y colaboradores desarrollaron *ADMET-Space*, una herramienta que incluye 20 modelos QSPR para la predicción de propiedades ADMET, incluida la estabilidad metabólica, la permeabilidad de Caco-2, la permeabilidad de BB, la unión de glucoproteínas P y la inhibición de CYP450 (Bocci et al., 2017). Los modelos se construyeron en base a los datos de 26.000 moléculas y se asentaron en los siguientes algoritmos: SVM, RF, *Ada Boost* (AB), LDA, potenciación del gradiente (*Gradient Boosting*, GB), árboles de decisión (*Decision Tree*, DT), PLS y clasificador de árboles extra (*Extra Trees Classifier*, ETC) (Figura 6c). En general, RF y SVM dieron los mejores resultados. El análisis de mapas se basa en un ANN que convierte los datos de una matriz n-dimensional en un mapa 2D. La traducción de múltiples modelos en una sola huella digital 2D permite monitorizar simultáneamente las diferentes propiedades y la detección de posibles inconvenientes de ADMET. Además, es una forma sencilla de evaluar cómo de lejos están los compuestos de nuevo diseño del espacio PK óptimo (Paper et al., 1990).

4.4. Herramientas útiles y accesibles a través de la web

Dong y colaboradores anunciaron un recurso basado en la web de libre acceso para la PK y la predicción de toxicidad llamado *ADMETlab* (<http://admet.scbdd.com>) (Dong et al., 2018). Los modelos de regresión se construyeron con RF, SVM, RP y PLS, y los modelos de clasificación se generaron utilizando RF, SVM, NB y DT (Figura 7a). En general, RF produjo los mejores resultados para los análisis de clasificación y regresión. *ADMETlab* permite al usuario realizar análisis de similitud con un fármaco según cinco reglas: Lipinski, Ghose, Veber, Varma y Oprea (Ghose et al., 1999; Lipinski et al., 2012; Tian et al., 2018; Varma et al., 2010; Veber et al., 2002). Se encuentran disponibles 31 modelos predictivos de propiedades fisicoquímicas (tres modelos), absorción (seis), distribución (tres), metabolismo (diez), eliminación (dos) y toxicidad (siete). Dada la cantidad de datos recopilados y la cantidad de modelos QSPR, *ADMETlab* es una de las plataformas más completas que se ha puesto a disposición recientemente para la predicción de ADMET.

Tian y colaboradores han desarrollado una herramienta que está disponible gratuitamente para predecir los reactivos CYP450 (Tian et al., 2018). *CypReact* (https://bitbucket.org/Leon_Ti/cypreact). Esta herramienta pronostica si es probable que las moléculas de consultadas sean metabolizadas por cualquiera de las isoformas CYP450: 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1 y 3A4 (Figura 7b). El algoritmo se entrenó con un conjunto de datos obtenidos a partir del metabolismo de 1632 reactivos. Se utilizaron SVM, regresión logística (LR), DT, RF y un método de conjunto para construir los modelos de clasificación para distinguir entre reactivos y no reactivos CYP450. El mejor modelo para cada isoforma produjo puntuaciones AUROC que variaron del 83% al 92%. El modelo de RF fue el mejor para cada isoforma 1A2, 2A6, 2B6, 2C8, 2C19, 2E1 y 3A4, mientras que el método de conjunto fue el mejor para las isoformas 2C9 y 2D6.

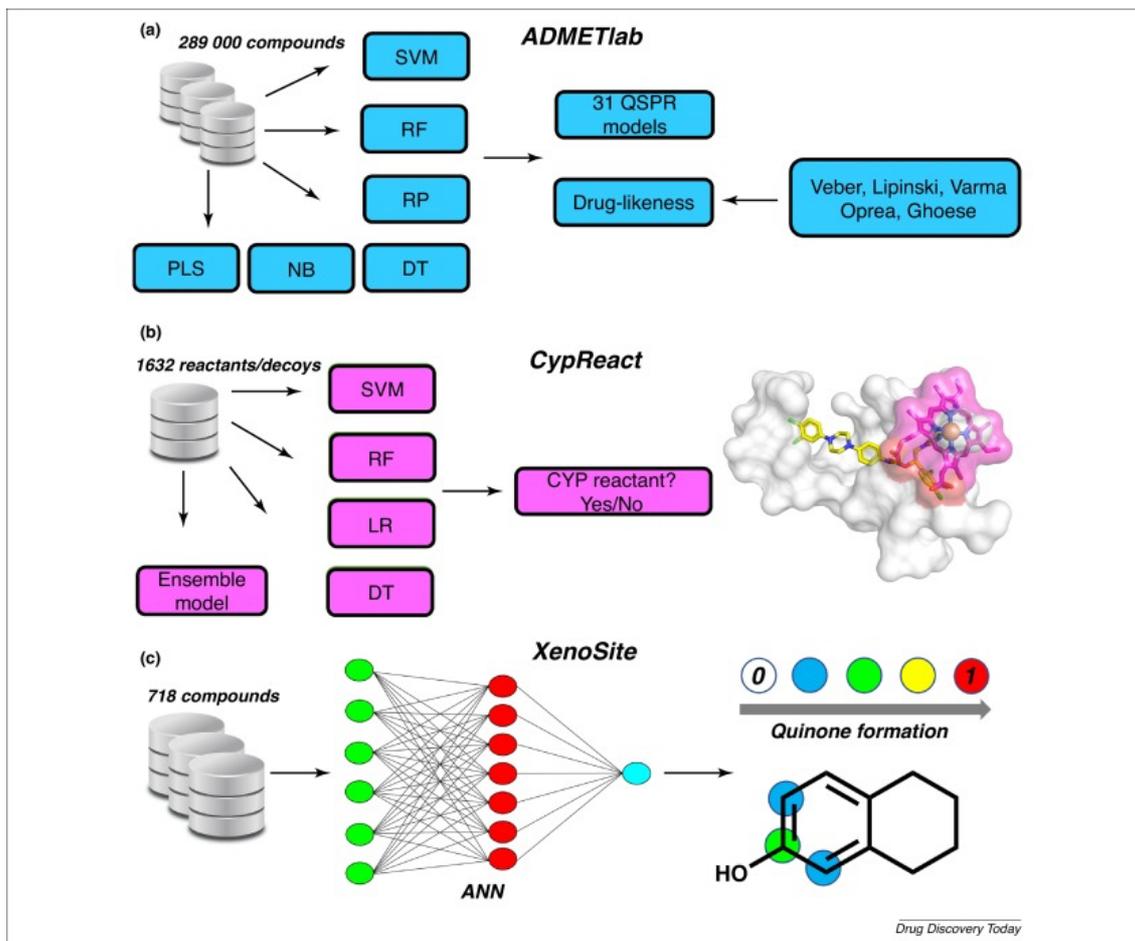


Figura 7. Diseño de recursos accesibles para la creación de perfiles farmacocinéticos

ADMETlab: servidor web para la predicción de parámetros farmacocinéticos y de toxicidad. Los modelos se construyeron utilizando particiones recursivas (RP), DT, RF, SVM, PLS y NB. (b) La herramienta CypReact para predecir los reactivos CYP450. Se utilizó un conjunto de 1632 reactivos para construir los modelos de clasificación de conjuntos, regresión logística (LR), DT, RF y SVM. (c) XenoSite: una red neuronal artificial (ANN) para predecir los metabolitos de quinona producidos por oxidación metabólica a través de las enzimas del CYP450. Obtenido de (Ferreira & Andricopulo, 2019)

Los metabolitos de quinona producidos por las enzimas CYP450 y peroxidasa son aceptores de Michael altamente reactivos y son responsables de una variedad de efectos adversos (Bolton & Dunlap, 2017). Usando *deep learning* ANN con 718 compuestos como grupo de aprendizaje, Hughes y Swamidass, fueron los primeros en desarrollar un modelo ML para predecir la formación de quinonas por oxidación metabólica (Figura 7c) (Hughes & Swamidass, 2017). El modelo logró distinguir entre moléculas formadoras de quinonas de las no formadoras de quinonas. Además, el ANN identificó pares de átomos que probablemente originen quinonas. Esta funcionalidad incluye una salida gráfica codificada por colores que resaltan y se atribuye una puntuación a los átomos que generan quinonas.

El llamado modelo *XenoSite* está disponible de manera gratuita (<http://swami.wustl.edu/xenosite/p/quinone>) y es muy práctico para identificar moléculas susceptibles de producir metabolitos tóxicos (Figura 7c). Más adelante se creó una versión mejorada del servidor web *FAF-Drugs* (Lagorce et al., 2017; Miteva et al., 2006). La nueva versión, *FAF-Drugs4*, incluye varias funciones centradas en PK y se puede acceder de manera libre (<http://fafdrugs4.mti.univ-paris-diderot.fr/>) (Lagorce et al., 2017). *FAF-Drugs4* puede filtrar grandes colecciones de compuestos para perfilar ADMET, marcar grupos tóxicos, predecir propiedades fisicoquímicas y proporcionar alertas estructurales. La nueva versión también incluye una estimación cuantitativa de las similitudes con los medicamentos que se comparó con un conjunto de datos de 711 medicamentos orales.

Podlewska y Kafel desarrollaron de *Met-StabOn*, un servidor en línea disponible de manera gratuita para la predicción de la estabilidad metabólica (Podlewska & Kafel, 2018). Se utilizaron conjuntos de compuestos como base de aprendizaje para comprobar la estabilidad metabólica en microsomas: del plasma de hígado humano 5.234 compuestos, de rata 2.829 compuestos y de ratón 1.136 compuestos. El modelo de regresión se generó utilizando la optimización secuencial mínima (SMO, del inglés *Sequential Minimal Optimization*), que es una adaptación de SVM (Shevade et al., 2000), y se generaron cinco sistemas de clasificación utilizando SMO, NB, k-NN y DT. Los modelos SMO, k-NN y RF superaron a los basados en DT y NB en cuanto a términos de su poder predictivo para el conjunto de compuestos usado como base de aprendizaje. Cuando se utilizaron compuesto de prueba externos, SMO, NB y RF superaron a las otras técnicas de ML. Además de predecir el aclaramiento plasmático, la vida media de un compuesto y la estabilidad metabólica, para cada uno de los compuestos consultados, esta plataforma proporciona una lista de estructuras similares a el conjunto de compuestos usados como entrenamiento.

SwissADME se introdujo como una herramienta para poder modelar el proceso ADMET (<http://www.swissadme.ch/index.php>) (Daina et al., 2017). *SwissADME* permite la predicción de diez propiedades fisicoquímicas, además de comprender seis modelos de lipofilia y tres de solubilidad en agua. También están disponibles nueve modelos de PK: absorción gastrointestinal, permeabilidad de la barrera hematoencefálica, unión a glucoproteína P, permeabilidad cutánea e inhibición de la CYP450 para cinco de las isoformas existentes. La similitud de los distintos compuestos se puede evaluar aplicando diferentes filtros: Lipinski, Ghose, Veber, Egar y Muegge. Esta funcionalidad destaca por

darnos información acerca de los compuestos que no cumplen con estas reglas y de cuyas propiedades tampoco están recogidas en estas. Nos dan información en forma de alertas químicas médicas (Brenk et al., 2008). Los modelos *SwissADME* se crearon utilizando SVM en relación con grandes conjuntos de datos, características moleculares y fisicoquímicas como descriptores. El resultado de *SwissADME* incluye el gráfico llamado “*Boiled-Egg*”, que permite evaluar cómo es la absorción gastrointestinal y el paso a la barrera hematoencefálica varían con el logP y el área de superficie polar topológica (Daina & Zoete, 2016). También proporciona información acerca de la biodisponibilidad. Estos modelos ofrecen la opción de consultar toda la información de manera rápida con simplemente introducir la estructura de un compuesto y esto queda guardado en formato de tabla. En el informe final recoge un resumen fácilmente visible del modelo adjuntado. El servidor web vNN (<https://vnnadmet.bhsai.org/>) es un recurso de libre acceso que incluye 15 modelos para la predicción de puntos finales del ADMET (Schyman et al., 2017). Como característica a destacar, el sitio permite al usuario personalizar los modelos incorporados, pero agregando sus propios datos experimentales. También es posible generar análisis de regresión y clasificación completamente nuevos basándose únicamente en los datos del usuario. Los modelos vNN están basados en el algoritmo *k*-NN (Liu et al., 2012). La plataforma proporciona predicciones sobre propiedades relevantes sobre la seguridad del uso del compuesto como son citotoxicidad, toxicidad mitocondrial, mutagénesis, cardiotoxicidad, daño hepático y dosis terapéutica máxima recomendada.

Y por último cabe nombrar la herramienta online *Hit Dexter 2.0* (<http://hitdexter2.zbh.uni-hamburg.de>) donde los usuarios pueden acceder a los modelos y reglas de ML incorporados para identificar a las subestructuras que no son adecuadas para el descubrimiento de fármacos.

Actualmente, la pandemia mundial producida por el SARS-CoV-2 hizo necesaria la aceleración en el estudio de conjuntos de compuestos en relación a los virus productores de enfermedades respiratorias (González-Paz et al., 2020), llegando incluso a desarrollar programas específicos para detectar posibles actividades inhibitoras producidas por este virus (V. Kumar & Roy, 2020).

5. Conclusiones

Se exponen, a continuación, las conclusiones a las que se ha llegado en este Trabajo a partir de la revisión bibliográfica realizada en relación con los objetivos planteados.

Los artículos revisados en este Trabajo proponen la existencia de técnicas valiosas que permiten desarrollar nuevos fármacos o grupos de fármacos previendo cuáles van a ser su campo de actuación, su toxicidad y sus efectos adversos con una gran fiabilidad y eficacia.

Las propiedades ADMET juegan un papel crucial en el descubrimiento de nuevos fármacos ya que en ellas radica en gran parte la efectividad de los tratamientos con dichas moléculas bioactivas. De esta manera las técnicas de ML siguen siendo una parte importante en el campo de la investigación de nuevos fármacos y por ello que siguen estando presente en las nuevas técnicas y programas informáticos.

Por último, es de destacar la importancia de los programas informáticos con base de datos de acceso universal en la búsqueda de nuevos prototipos y fármacos específicos para distintas patologías y que requieran poseer propiedades farmacocinéticas específicas. Su uso agiliza la búsqueda de nuevos fármacos, optimiza el tiempo y los recursos económicos disponibles y, por tanto, aportar mayor rapidez en todo el proceso de investigación. El desarrollo de programas específicos para algunas enfermedades o patologías está en auge y supondrá un hito muy significativo en salud pública, en el tratamiento de enfermedades emergentes y otras patologías que, aun siendo conocidas, presentan una elevada prevalencia en la sociedad y para las que todavía no se dispone de un tratamiento adecuado.

6. Bibliografia

- Bocci, G., Carosati, E., Vayer, P., Arrault, A., Lozano, S., & Cruciani, G. (2017). ADME-Space: A new tool for medicinal chemists to explore ADME properties. *Scientific Reports*, 7(1), 25-27. <https://doi.org/10.1038/s41598-017-06692-0>
- Bolton, J. L., & Dunlap, T. (2017). Formation and biological targets of quinones: Cytotoxic versus cytoprotective effects. *Chemical Research in Toxicology*, 30(1), 13-37. <https://doi.org/10.1021/acs.chemrestox.6b00256>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>
- Brenk, R., Schipani, A., James, D., Krasowski, A., Gilbert, I. H., Frearson, J., & Wyatt, P. G. (2008). Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, 3(3), 435-444. <https://doi.org/10.1002/cmdc.200700139>
- Cano, G., Garcia-Rodriguez, J., Garcia-Garcia, A., Perez-Sanchez, H., Benediktsson, J. A., Thapa, A., & Barr, A. (2017). Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Systems with Applications*, 72, 151-159. <https://doi.org/10.1016/j.eswa.2016.12.008>
- Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7(March), 1-13. <https://doi.org/10.1038/srep42717>
- Daina, A., & Zoete, V. (2016). A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem*, 1117-1121. <https://doi.org/10.1002/cmdc.201600182>
- Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., Lu, A. P., & Cao, D. S. (2018). Admetlab: A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *Journal of Cheminformatics*, 10(1), 1-11. <https://doi.org/10.1186/s13321-018-0283-x>
- Doniger, S., Hofmann, T., & Yeh, J. (2002). Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms. *Journal of Computational Biology*, 9(6), 849-864. <https://doi.org/10.1089/10665270260518317>

- Ferreira, L. L. G., & Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discovery Today*, 24(5), 1157-1165. <https://doi.org/10.1016/j.drudis.2019.03.015>
- Finkelmann, A. R., Göller, A. H., & Schneider, G. (2017). Site of Metabolism Prediction Based on ab initio Derived Atom Representations. *ChemMedChem*, 12(8), 606-612. <https://doi.org/10.1002/cmdc.201700097>
- Gavernet, L. (2021). Introducción a la Química Medicinal. En Editorial de la UNLP (Ed.), *Universidad Nacional de la Plata*. <https://libros.unlp.edu.ar/index.php/unlp/catalog/view/1627/1606/5240-1>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *Journal of Combinatorial Chemistry*, 1(1), 55-68. <https://doi.org/10.1021/cc9800071>
- Gobburu, J. V. S., & Shelver, W. H. (1995). Quantitative Structure–Pharmacokinetic Relationships (QSPR) of Beta Blockers Derived Using Neural Networks. *Journal of Pharmaceutical Sciences*, 84(7), 862-865. <https://doi.org/10.1002/jps.2600840715>
- González-Paz, L., Paz, J. L., Vera-Villalobos, J., & Alvarado, Y. J. (2020). Compuestos Fitoquímicos Dirigidos al Bloqueo de la Polimerasa Viral del SARS-CoV-2 Causante del COVID-19: un Análisis Comparativo de Funciones de Puntuación para Acoplamiento con Interés Biomédico. *Revista Politécnica*, 46(1), 7-20. <https://doi.org/10.33333/rp.vol46n1.01>
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, 220-239. <https://doi.org/10.1016/j.eswa.2016.12.035>
- Hansch, C., & Fujita, T. (1964). ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, 86(8), 1616-1626. <https://doi.org/10.1021/ja01062a035>
- 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>
- Heikamp, K., & Bajorath, J. (2014). Support vector machines for drug discovery. *Expert Opinion on Drug Discovery*, 9(1), 93-104. <https://doi.org/10.1517/17460441.2014.866943>
- Hinderling, P. H., Schmidlin, O., & Seydel, J. K. (1984). Quantitative relationships between structure and pharmacokinetics of beta-adrenoceptor blocking agents in man. *Journal of Pharmacokinetics and Biopharmaceutics*, 12(3), 263-287. <https://doi.org/10.1007/BF01061721>
- Hughes, T. B., & Swamidass, S. J. (2017). Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism. *Chemical Research in Toxicology*, 30(2), 642-656. <https://doi.org/10.1021/acs.chemrestox.6b00385>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kirchmair, J., Williamson, M. J., Afzal, A. M., Tyzack, J. D., Choy, A. P. K., Howlett, A., Rydberg, P., & Glen, R. C. (2013). FASt METabolizer (FAME): A rapid and accurate predictor of sites of metabolism in multiple species by endogenous enzymes. *Journal of Chemical Information and Modeling*, 53(11), 2896-2907. <https://doi.org/10.1021/ci400503s>
- Kumar, R., Sharma, A., Siddiqui, M. H., & Tiwari, R. K. (2017). Prediction of Drug-Plasma Protein Binding Using Artificial Intelligence Based Algorithms. *Combinatorial Chemistry & High Throughput Screening*, 21(1), 57-64. <https://doi.org/10.2174/1386207321666171218121557>
- Kumar, V., & Roy, K. (2020). Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. *SAR and QSAR in Environmental Research*, 31(7), 511-526. <https://doi.org/10.1080/1062936X.2020.1776388>
- Lagorce, D., Bouzlama, L., Becot, J., Miteva, M. A., & Villoutreix, B. O. (2017). FAF-Drugs4: Free ADME-tox filtering computations for chemical biology and early stages drug discovery. *Bioinformatics*, 33(22), 3658-3660. <https://doi.org/10.1093/bioinformatics/btx491>

- Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 64(SUPPL.), 4-17. <https://doi.org/10.1016/j.addr.2012.09.019>
- Liu, R., Tawa, G., & Wallqvist, A. (2012). Locally weighted learning methods for predicting dose-dependent toxicity with application to the human maximum recommended daily dose. *Chemical Research in Toxicology*, 25(10), 2216-2226. <https://doi.org/10.1021/tx300279f>
- Machine Learning - Explorar - Google Trends*. (s. f.). Recuperado 20 de junio de 2021, de <https://trends.google.es/trends/explore?date=all&q=Machine Learning>
- Miteva, M. A., Violas, S., Montes, M., Gomez, D., Tuffery, P., & Villoutreix, B. O. (2006). FAF-Drugs: Free ADME/tox filtering of compound collections. *Nucleic Acids Research*, 34(WEB. SERV. ISS.), 738-744. <https://doi.org/10.1093/nar/gkl065>
- Muegge, I., Heald, S. L., & Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, 44(12), 1841-1846. <https://doi.org/10.1021/jm015507e>
- Paper, I., Introduction, I., Models, N., & Maps, S. (1990). Kohen self organization map. En *Proceedings of the IEEE* (Vol. 78, Número 9, pp. 1464-1480).
- Podlewska, S., & Kafel, R. (2018). Metstabon-online platform for metabolic stability predictions. *International Journal of Molecular Sciences*, 19(4). <https://doi.org/10.3390/ijms19041040>
- Ray, L. C., & Kirsch, R. A. (1957). Finding chemical records by digital computers. *Science*, 126(3278), 814-819. <https://doi.org/10.1126/science.126.3278.814>
- Schyman, P., Liu, R., Desai, V., & Wallqvist, A. (2017). vNN web server for ADMET predictions. *Frontiers in Pharmacology*, 8(DEC), 1-14. <https://doi.org/10.3389/fphar.2017.00889>
- Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, 11(5), 1188-1193. <https://doi.org/10.1109/72.870050>
- Šicho, M., De Bruyn Kops, C., Stork, C., Svozil, D., & Kirchmair, J. (2017). FAME 2:

- Simple and Effective Machine Learning Model of Cytochrome P450 Regioselectivity. *Journal of Chemical Information and Modeling*, 57(8), 1832-1846. <https://doi.org/10.1021/acs.jcim.7b00250>
- Stratton, T. P., Perryman, A. L., Vilchèze, C., Russo, R., Li, S. G., Patel, J. S., Singleton, E., Ekins, S., Connell, N., Jacobs, W. R., & Freundlich, J. S. (2017). Addressing the Metabolic Stability of Antituberculars through Machine Learning. *ACS Medicinal Chemistry Letters*, 8(10), 1099-1104. <https://doi.org/10.1021/acsmedchemlett.7b00299>
- Tao, L., Zhang, P., Qin, C., Chen, S. Y., Zhang, C., Chen, Z., Zhu, F., Yang, S. Y., Wei, Y. Q., & Chen, Y. Z. (2015). Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Advanced Drug Delivery Reviews*, 86, 83-100. <https://doi.org/10.1016/j.addr.2015.03.014>
- Tian, S., Djoumbou-Feunang, Y., Greiner, R., & Wishart, D. S. (2018). CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes. *Journal of Chemical Information and Modeling*, 58(6), 1282-1291. <https://doi.org/10.1021/acs.jcim.8b00035>
- Varma, M. V. S., Obach, R. S., Rotter, C., Miller, H. R., Chang, G., Steyn, S. J., El-Kattan, A., & Troutman, M. D. (2010). Physicochemical space for optimum oral bioavailability: Contribution of human intestinal absorption and first-pass elimination. *Journal of Medicinal Chemistry*, 53(3), 1098-1108. <https://doi.org/10.1021/jm901371v>
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 45(12), 2615-2623. <https://doi.org/10.1021/jm020017n>
- Wang, N. N., Deng, Z. K., Huang, C., Dong, J., Zhu, M. F., Yao, Z. J., Chen, A. F., Lu, A. P., Mi, Q., & Cao, D. S. (2017). ADME properties evaluation in drug discovery: Prediction of plasma protein binding using NSGA-II combining PLS and consensus modeling. *Chemometrics and Intelligent Laboratory Systems*, 170, 84-95. <https://doi.org/10.1016/j.chemolab.2017.09.005>
- Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. a. (1998). Glossary of Terms Used in Medicinal Chemistry. *Pure & Appl. Chem.*, 70(5), 1129-1143.

- Yang, M., Chen, J., Xu, L., Shi, X., Zhou, X., Xi, Z., An, R., & Wang, X. (2018). A novel adaptive ensemble classification framework for ADME prediction. *RSC Advances*, 8(21), 11661-11683. <https://doi.org/10.1039/c8ra01206g>
- Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today*, 22(11), 1680-1685. <https://doi.org/10.1016/j.drudis.2017.08.010>
- Zhu, L., Zhao, J., Zhang, Y., Zhou, W., Yin, L., Wang, Y., Fan, Y., Chen, Y., & Liu, H. (2018). ADME properties evaluation in drug discovery: in silico prediction of blood–brain partitioning. *Molecular Diversity*, 22(4), 979-990. <https://doi.org/10.1007/s11030-018-9866-8>