



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN

Propuesta de una práctica de laboratorio para la asignatura de
Farmacología general.
Determinación del índice terapéutico y margen de seguridad a
través de un modelo computacional.

T E S I S

Que para obtener el título de
QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA:

JOSUÉ BRIAN ACUÑA GARDUÑO

Asesora:
M. EN F.C. CECILIA HERNÁNDEZ BARBA.

Cuatitlán Izcalli, Estado de México

2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
UNIDAD DE ADMINISTRACIÓN ESCOLAR
DEPARTAMENTO DE EXÁMENES PROFESIONALES**

ASUNTO: VOTO APROBATORIO



DEPARTAMENTO DE
EXÁMENES PROFESIONALES

**M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE**

**ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Exámenes Profesionales
de la FES Cuautitlán.**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de Tesis**

**Propuesta de una práctica de laboratorio para la asignatura de Farmacología general.
Determinación del índice terapéutico y margen de seguridad a través de un modelo computacional.**

Que presenta el pasante: **Josué Brian Acuña Garduño**
Con número de cuenta: **301003577** para obtener el Título de la carrera: **Química Farmacéutico Biológica**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO**.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 28 de Febrero de 2018.

PROFESORES QUE INTEGRAN EL JURADO

| | NOMBRE | FIRMA |
|----------------------|--|-------|
| PRESIDENTE | M.F.C. Ma. Eugenia R. Posada Galarza | |
| VOCAL | M. en C. Lidia Rangel Trujano | |
| SECRETARIO | M.F.C. Cecilia Hernández Barba | |
| 1er. SUPLENTE | M.F.C. Beatriz de Jesús Maya Monroy | |
| 2do. SUPLENTE | Q.F.B. Jonathan Raymundo García Martínez | |

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional (art. 127).

LMCF/mmmg*cga*

–Lo que pasa es que se creen sabios – dice de golpe –. Se creen sabios porque han juntado un montón de libros y se los han comido. Me da risa, porque en realidad son buenos muchachos y viven convencidos de que lo que estudian y lo que hacen son cosas muy difíciles y profundas. En el circo es igual, Bruno, y entre nosotros es igual. La gente se figura que algunas cosas son el colmo de la dificultad, y por eso aplauden a los trapeceistas, o a mí. Yo no sé qué se imaginan, que uno se está haciendo pedazos para tocar bien, o que el trapeceista se rompe los tendones cada vez que da un salto. En realidad, las cosas verdaderamente difíciles son otras tan distintas, todo lo que la gente cree poder hacer a cada momento. Mirar, por ejemplo, o comprender a un perro o a un gato. Esas son las dificultades, las grandes dificultades. Anoche se me ocurrió mirarme en este espejito, y te aseguro que era tan terriblemente difícil que case me tiro de la cama. Imagínate que te estás viendo a ti mismo; eso tan sólo basta para quedarse frío durante media hora.

Julio Cortázar.

Dedicatoria:

Dedico este trabajo a todos los que me han apoyado durante el proceso y la culminación de esta etapa académica de mi vida.

A la Maestra Cecilia, por el seguimiento y el apoyo a este proyecto.

A mi alma máter. Por darme todas las herramientas con las que sigo descubriendo al mundo.

ÍNDICE

| | |
|--|----|
| Índice de tablas..... | I |
| Índice de Ilustraciones..... | II |
| 1. Introducción | 7 |
| 2. Marco teórico | 9 |
| 2.1 Introducción a las relaciones estructura química-actividad biológica..... | 9 |
| 2.1. Descriptores moleculares..... | 11 |
| 2.2. Similitud estructural | 14 |
| 2.3. Métodos estadísticos en QSAR. | 18 |
| 2.4. Modelo de predicción y validación interna..... | 22 |
| 2.5. Validación externa..... | 26 |
| 2.6. Farmacometría: Generalidades | 32 |
| 3. Objetivos | 49 |
| 3.1 Objetivo general | 49 |
| 3.1. Objetivos Particulares..... | 49 |
| 4. Metodología | 49 |
| 4.1. Consideraciones previas | 50 |
| 4.2. Elección del grupo químico..... | 50 |
| 4.3. Preparación de los datos..... | 51 |
| 4.4. Aseguramiento de la similitud estructural..... | 53 |
| 5. Resultados | 56 |
| 5.1. Material y equipo | 56 |
| 5.2. Conceptos previos requeridos para la ejecución de la práctica | 56 |
| 5.3. Preparación previa del equipo de cómputo..... | 56 |
| 5.4. Modelado de estructuras químicas..... | 58 |
| 5.5. Cálculo de descriptores..... | 61 |
| 5.6. Búsqueda de los mejores descriptores para la generación del modelo: | 67 |
| 5.7. Generación del modelo de predicción: | 77 |
| 5.8. Predicción de la DL50 del Diazepam..... | 81 |
| 5.9. Cálculo del índice terapéutico y margen de seguridad..... | 85 |

| | |
|-----------------------|----|
| 6. Discusión..... | 86 |
| 7. Conclusiones | 88 |

Índice de tablas

| | |
|--|--------|
| Tabla 1. Clasificación de los descriptores de acuerdo a su dimensionalidad..... | 13 |
| Tabla 2. Lista de algunos programas para el cálculo de descriptores..... | 13, 14 |
| Tabla 3. Índices de uso común para determinar similitud estructural..... | 15,16 |
| Tabla 4. Compuestos análogos con sus coeficientes de similitud y valores experimentales de toxicidad oral-aguda..... | 17,18 |
| Tabla 5. Compuestos con descriptores 2D y su respuesta biológica..... | 20 |
| Tabla 6. Ejemplos de una validación externa con 6 compuestos..... | 29 |
| Tabla 7 Resultados de los doce análogos seleccionados..... | 53 |

Índice de Ilustraciones

| | |
|---|----|
| Ilustración 1. Estructura química del Diazepam..... | 12 |
| Ilustración 2. Algunos modelos de clasificación y regresión utilizados en QSAR..... | 22 |
| Ilustración 3. Gráfica de valores predichos contra experimentales para compuestos de la validación externa de un modelo QSAR..... | 30 |
| Ilustración 4. Ensayo de unión de radioligando (<i>binding</i>) como ejemplo de una curva concentración-respuesta gradual..... | 36 |
| Ilustración 5. Las curvas de dosis-respuesta gradual permiten la comparación de la actividad de dos fármacos | 37 |
| Ilustración 6. Curva de distribución normal. La probabilidad acumulada (flechas rojas) crece conforme aumenta la dosis..... | 42 |
| Ilustración 7. Curva de probabilidad acumulada generada a partir de la distribución normal de la ilustración 6..... | 43 |
| Ilustración 8. Curva dosis-respuesta cuantil..... | 44 |

| | |
|--|----|
| Ilustración 9. La función <i>probits</i> convierte la probabilidad acumulada a una escala arbitraria..... | 47 |
| Ilustración 10. En una escala de <i>probits</i> , las curvas dosis-respuesta tienden a ser lineales facilitando las interpolaciones..... | 48 |
| Ilustración 11. Representación gráfica de la similitud estructural de las 12 Benzodiazepinas comparadas con el Diazepam..... | 55 |

1. Introducción

El estudio de las propiedades farmacológicas o toxicológicas de los compuestos químicos implica generalmente el uso de modelos basados en la experimentación animal, que a su vez provocan el planteamiento de cuestionamientos bioéticos donde la tendencia ideológica principal es evitar en la medida de lo posible el sufrimiento animal en los ámbitos de la docencia y la investigación.

En los últimos años, los avances de la química computacional y teórica, y el diseño molecular asistido por computadora han permitido que la química y otras disciplinas asociadas a la salud, dejen de ser ciencias completamente experimentales, y que algunos cuestionamientos puedan llevarse al ámbito de la informática para resolverse a través de modelos teóricos y estadísticos que permitan reducir tiempo y costos en los procesos inherentes al descubrimiento de las propiedades de sustancias con interés para el ser humano. De este modo, la química computacional funciona en conjunto con la experimentación para predecir propiedades, las cuales pudieran no existir un modelo experimental real, ya que su implementación resultaría peligrosa, o implicaría el sufrimiento de especies animales.

En el campo de la farmacología, los métodos computacionales pueden ser de gran utilidad para responder cuestionamientos como: ¿Qué posibilidad existe de que un fármaco se una con un determinado receptor? ¿Qué tan efectivo o tóxico puede ser un fármaco? ¿Cómo puede modificarse la molécula de un fármaco para mejorar su efectividad?

En el ámbito docente, la asignatura de farmacología general impartida en la Facultad de Estudios Superiores Cuautitlán, incluye una sesión práctica donde se determinan los efectos del pentobarbital mediante su administración a roedores. La experiencia obtenida persigue que el alumno se familiarice con conceptos elementales como son la curvas dosis-respuesta,

y la obtención de parámetros como el margen de seguridad y el índice terapéutico; estos dos últimos, conceptos básicos para comprender la relación seguridad-eficacia de los fármacos.

En este contexto, y con el fin de aproximar a los alumnos a los conceptos elementales sobre las relaciones de estructura química-actividad biológica, el presente trabajo tiene como objetivo principal; proponer la metodología de una sesión práctica para predecir el índice terapéutico y el margen de seguridad de un fármaco, desarrollando un modelo computacional intuitivo, sencillo y de fácil abordaje a través de software de fácil asequibilidad.

El alumno conocerá conceptos relativos a descriptores moleculares, correlación con la respuesta biológica, predictibilidad y validación, y además podrá asociarlos con conceptos vistos en la asignatura. Asimismo, despertar su interés en metodologías que prescindan del uso de animales.

Del mismo modo, aprenderá cómo los modelos computacionales y experimentales se complementan para responder cuestionamientos sobre el comportamiento de la materia, y así ampliar su panorama profesional, mostrando una opción más dentro de la industria o la investigación científica.

2. Marco teórico

2.1 Introducción a las relaciones estructura química-actividad biológica.

La palabra QSAR, es el acrónimo de *Quantitative structure activity relationship*. Es decir, la relación cuantitativa que existe entre la estructura química de los compuestos y su actividad biológica. Esta definición que es explícita en sí misma, recoge la presunción de que los grupos funcionales y su acomodo espacial, son responsables de que las sustancias químicas tengan algún efecto en los organismos vivos. Este efecto biológico o actividad, puede ser cualquier proceso asociado a mecanismos farmacocinéticos o farmacodinámicos dentro de seres vivos, que se expresen a su vez en un efecto terapéutico o toxicológico. Cuando hablamos en términos inherentes a las técnicas de QSAR, esta actividad biológica recibe el nombre de *endpoints*. Algunos ejemplos de *endpoints* son: la toxicidad oral aguda, la sensibilización dermal, la mutagenicidad y teratogenicidad o la bioacumulación.

La idea de que la estructura química de las sustancias está relacionada con sus efectos biológicos no es nueva. A pesar de que los métodos QSAR son relativamente nuevos dentro del campo de la toxicología, por más de cien años, se han documentado en la literatura diversos intentos de cuantificar las relaciones entre la estructura química y su potencial tóxico. Como ejemplo de lo anterior, se puede citar un viejo artículo de 1863, publicado por la universidad de Estrasburgo en Francia, titulado “Action de l’alcool amylique sur l’organisme”, en el cual Cros y su equipo notaron que existía una relación entre la toxicidad de alcoholes alifáticos primarios y su solubilidad en agua. Esta relación demostró el axioma central del modelo de estructura-toxicidad: la toxicidad de las sustancias se rige por sus propiedades, que a su vez están determinadas por su estructura química. Por lo tanto, hay interrelaciones entre estructura, propiedades y toxicidad (Crum-Brown & Fraser, 1868-69).

Hace más de un siglo, Crum-Brown y Fraser expresaron la idea de que la acción fisiológica de una sustancia en un cierto sistema biológico (Φ) era una función (f) de su constitución química

$$\Phi = f C$$

Así, una alteración en la constitución química, ΔC , se reflejaría en una alteración en la actividad biológica $\Delta\Phi$.

Al estar convencidos de que la estructura de las sustancias químicas es un factor preponderante en su actividad, es preciso evaluar en qué proporción se produce el efecto, y qué cambios en la estructura impactarán en su anulación o potencialización. Un ejemplo clásico de una pequeña variación en la estructura química de un compuesto, puede impactar en un cambio drástico en su actividad biológica, es precisamente la Talidomida. La Talidomida se sintetizó por primera vez en el año de 1954 por los laboratorios Chemie Grünenthal, en Alemania, con el nombre molecular de alfa-Phtalyglutamic-acid-imida. Conocida por sus efectos sedantes e hipnóticos, y utilizada en el tratamiento del vómito asociado al primer trimestre del embarazo, a mediados de los 50 era la tercera droga más vendida, sin toxicidad entonces reconocida, ya que fue probada en ratones a dosis 10,000 mg/kg sin observar efectos colaterales fatales (Powell, 1996; George & Elias, 1999).

A finales de los 50 y principios de los 60 nacieron más de 12,000 niños con graves deformaciones congénitas, de madres que habían ingerido el fármaco, caracterizadas por desarrollos defectuosos en brazos, piernas o bien alteraciones masivas internas (George & Elias, 1999).

Después de las investigaciones, se descubrió que existían dos formas estructurales distintas del fármaco, dos enantiómeros R y S que producían efectos distintos. Mientras la forma R producía el efecto sedante, el isómero S producía efectos teratogénicos. La misma molécula, con el mismo número de átomos, pero diferente orientación espacial tenía un efecto negativo.

Cuatro años después de su creación la talidomida fue retirada del mercado. A pesar de ello la fábrica alemana Chemie Grünenthal continuó comercializándola exclusivamente para su uso en pacientes con lepra. La administración de alimentos y drogas (FDA) en EUA aprobó finalmente la comercialización del fármaco el año de 1998, y la Corporación de Celgena en Nueva Jersey recibió en 1995 la aprobación de la patente para la fabricación de la Talidomida (Leslie, 1998).

2.1.Descriptores moleculares.

El objetivo de las técnicas QSAR es encontrar qué propiedades estructurales están relacionadas con una determinada respuesta biológica. Estas propiedades estructurales pueden resumirse en números llamados descriptores. Por ejemplo, la siguiente estructura química pertenece al Diazepam, un psicofármaco perteneciente al grupo de las benzodiazepinas, de la cual se pueden inferir algunos descriptores a simple vista, por ejemplo, el número de átomos de oxígeno y nitrógeno presentes:

Número de átomos de nitrógeno: $nN=2$

Número de átomos de oxígeno: $nO=1$

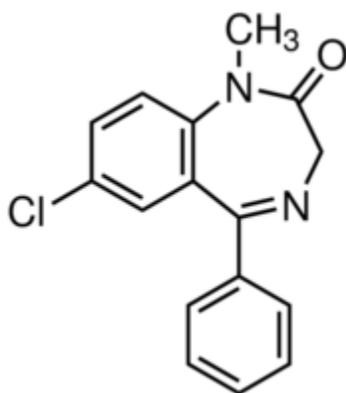


Ilustración 1 Estructura química del Diazepam.

Los parámetros nN y nO han sido inferidos mediante la observación a simple vista de la estructura química, y en virtud de que estas observaciones son traducidas a números, pueden clasificarse como descriptores.

Existen un gran número de mediciones que pueden hacerse a una estructura química, y por muy complejas que sean, si son susceptibles de expresarse numéricamente, son descriptores.

Una definición más formal sobre lo que es un descriptor ha sido propuesta por Todeschini and Consonni:

“El descriptor molecular es el resultado final de un procedimiento lógico y matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado” (Todeschini 2009).

A diferencia de los descriptores que hemos extraído de la estructura química del Diazepam (nO y nN), existen un gran número de descriptores que se obtienen a través de transformaciones matemáticas complejas, o a partir de representaciones o modelos tridimensionales de las sustancias químicas. En los laboratorios de quimioinformática de todo el mundo, se trabaja constantemente por encontrar nuevos descriptores que tengan alguna

correlación o expliquen alguna actividad. Esto ha generado la necesidad de proponer formas de clasificar a los descriptores.

Una de las clasificaciones más usadas, es aquella que agrupa a los descriptores de acuerdo a su dimensionalidad (ver Tabla 1).

Tabla 1 Clasificación de los descriptores de acuerdo a su dimensionalidad

| Dimensión | Parámetros |
|-----------|---|
| 0D | Índices constitucionales, propiedades moleculares, conteo de átomos y enlaces |
| 1D | Conteo de fragmentos, huellas digitales |
| 2D | Parámetros topológicos, parámetros estructurales, parámetros fisicoquímicos incluyendo descriptores termodinámicos. |
| 3d | Parámetros electrónicos, parámetros espaciales, parámetros del análisis de la forma molecular, parámetros del análisis del campo molecular y parámetros del análisis de la superficie del receptor. |

De acuerdo a esta clasificación, los descriptores nO y nN son descriptores 0D.

Actualmente, existen numerosas herramientas libres y privativas para calcular descriptores. Algunas poseen interfaces gráficas, mientras que otras se operan mediante código o poseen un editor gráfico para modelar las estructuras en dos o tres dimensiones. A continuación, se muestra una lista de programas de uso frecuente. Algunos de ellos como CODESSA PRO, MOE u OCHEM, contienen también herramientas para desarrollar modelos con los descriptores obtenidos (ver Tabla 2).

Tabla 2 Lista de algunos programas de uso común para el cálculo de descriptores.

| Software/online platform | Weblink |
|--------------------------|---|
| Cerius2 | http://accelrys.com/ |
| CODESSA PRO | http://www.codessa-pro.com/index.htm |
| Discovery studio | http://accelrys.com/ |
| DRAGON | http://www.taletе.mi.it/products/dragon_description.htm |
| E-Dragon at VCCLAB | http://www.vcclab.org/lab/edragon/ |

| | |
|----------------------|---|
| GRID | http://www.moldiscovery.com/soft_grid.php |
| JME Molecular Editor | http://www.molinspiration.com/jme/index.html |
| Linux4Chemistry | http://www.linux4chemistry.info/ |
| MOE | http://www.chemcomp.com/software.htm |
| MOLCONN-Z | http://www.edusoft-lc.com/molconn/ |
| MOLE db | http://michem.disat.unimib.it/mole_db/ |
| MOLGEN-QSPR | http://www.molgen.de/?src=documents/molgenqspr.html |
| OCHEM | https://ochem.eu/home/show.do |
| OpenBabel | http://openbabel.org/ |
| PaDEL-Descriptor | http://padel.nus.edu.sg/software/padeldescriptor/ |
| PCLIENT | http://www.vcclab.org/lab/pclient/ |
| QSARModel | http://www.molcode.com/ |
| QuaSAR | http://www.chemcomp.com/feature/qsar.htm |
| SYBYL-X | http://tripos.com/index.php?family=modules,SimplePage&page=SYBYL-X |
| TsarTM | http://www.accelrys.com/products/tsar/tsar.html |
| Unscrambler X | http://www.camo.com/rt/Products/Unscrambler/unscrambler.html |
| V-Life MDS | http://www.vlifesciences.com/products/VLifeMDS/Product_VLifeMDS.php |

2.2. Similitud estructural

Generalmente un modelo QSAR comienza recolectando información experimental sobre un grupo de compuestos químicos a los que se les conoce como “data set”. Este grupo de compuestos deben idealmente compartir características estructurales entre sí. Por esa razón, en algunos textos son llamados “análogos”. Teniendo en consideración que sustancias químicas con estructuras similares, producirán efectos biológicos similares, los análogos poseen frecuentemente los mismos grupos funcionales y en algunos casos donde el desarrollo QSAR es más especializado, los análogos se agrupan de acuerdo a criterios más robustos. A este proceso se le conoce como “categorización”. La Organización para la cooperación y desarrollo económico (OCDE) ha publicado la guía 194 para la categorización de sustancias químicas cuyo título original es: Series on Testing & Assessment “GUIDANCE ON GROUPING OF CHEMICALS, SECOND EDITION”

Como ya se mencionó, los análogos estructurales comparten un cierto grado de similitud entre ellos. Esta similitud estructural no puede ser medida de forma subjetiva, y debe tener parámetros a través de algún índice que exprese de forma numérica ese parecido.

La definición de similitud estructural química depende de la característica molecular en análisis, tales como grupos funcionales o subestructuras comunes. En general, se asume ampliamente que las características y el comportamiento de las sustancias están parcialmente condicionados por su estructura, y se han definido numerosas medidas cuantitativas de similitud molecular (Rouvray & D.H., 1995).

La definición de similitud para las moléculas consiste en mapear el espacio químico, es decir, una representación de la molécula en términos de descriptores relevantes en el espacio unidimensional con números reales. La definición también depende de la representación de las moléculas bajo consideración en espacio descriptor. Las moléculas pueden representarse usando una gama de diferentes representaciones.

Algunos índices de uso frecuente para medir similitud estructural se recogen en la siguiente tabla:

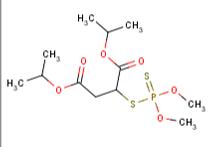
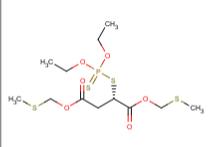
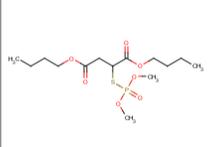
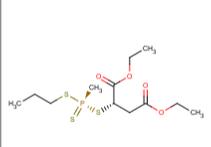
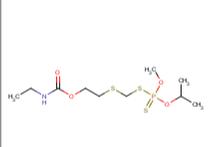
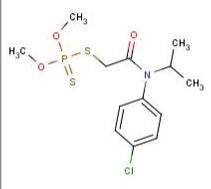
Tabla 3 Índices de uso común para determinar similitud estructural.

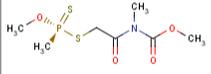
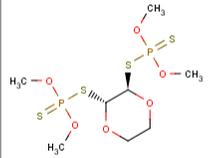
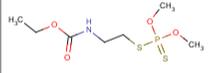
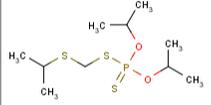
| Medida | Rango | Fórmula |
|--------|---------|--------------------------------|
| Cosine | 0.0,1.0 | $\frac{c}{\sqrt{(a+c)*(b+c)}}$ |
| Dice | 0.0,1.0 | $\frac{2.0*c}{(a+c)+(b+c)}$ |
| Euclid | 0.0,1.0 | $\sqrt{\frac{c+d}{a+b+c+d}}$ |

| | | |
|-----------------|----------|--|
| Forbes | 0.0,∞ | $\frac{c*(a+b+c+d)}{(a+c)*(b+c)}$ |
| Hamman | -1.0,1.0 | $\frac{(c+d)-(a+b)}{a+b+c+d}$ |
| Jaccard | 0.0,1.0 | $\frac{c}{a+b+c}$ |
| Kulczynski | 0.0,1.0 | $0.5*\left(\frac{c}{a+c} + \frac{c}{b+c}\right)$ |
| Manhattan | 1.0,0.0 | $\frac{(a+b)}{(a+b+c+d)}$ |
| Matching | 0.0,1.0 | $\frac{c+d}{a+b+c+d}$ |
| Pearson | -1.0,1.0 | $\frac{(c*d)-(a*b)}{\sqrt{(a+c)*(b+c)*(a+d)*(b+d)}}$ |
| Rogers-Tanimoto | 0.0,1.0 | $\frac{c+d}{(a+b)+(a+b+c+d)}$ |
| Russell-Rao | 0.0,1.0 | $\frac{c}{a+b+c+d}$ |
| Simpson | 0.0,1.0 | $\frac{c}{\min((a+c),(b+c))}$ |
| Tanimoto | 0.0,1.0 | $\frac{c}{a+b+c}$ |
| Yule | -1.0,1.0 | $\frac{(c*d)-(a*b)}{(c*d)+(a*b)}$ |

Ejemplo de un data set de análogos estructurales con su coeficiente de similitud y el valor experimental de DL₅₀ oral en rata transformado a forma logarítmica. (TEST SOFTWARE, 2015)

Tabla 4 Compuestos análogos con sus coeficientes de similitud y valores experimentales de toxicidad oral aguda.

| CAS | Estructura | Coefficiente de similitud | Valor experimental DL50 -Log10(mol/kg) |
|------------|---|---------------------------|--|
| 3700-90-1 |  | 0.96 | 1.95 |
| 32744-43-7 |  | 0.85 | 5.06 |
| 3700-98-9 |  | 0.83 | 2.86 |
| 17581-48-5 |  | 0.76 | 2.71 |
| 3853-74-5 |  | 0.74 | 4.61 |
| 64249-01-0 |  | 0.71 | 2.89 |

| | | | |
|------------|---|------|------|
| 29173-31-7 |  | 0.71 | 3.68 |
| 63717-18-0 |  | 0.70 | 3.06 |
| 5840-95-9 |  | 0.69 | 3.40 |
| 2667-52-9 |  | 0.68 | 4.78 |

2.3.Métodos estadísticos en QSAR.

Cuando se obtienen descriptores de un grupo de compuestos análogos, es preciso averiguar si alguno o algunos, poseen alguna correlación con la respuesta biológica. Por ejemplo, el descriptor LogP, también conocido como coeficiente de distribución, o coeficiente de partición octanol/agua, es además de un descriptor, un parámetro fisicoquímico asociado a los procesos ADME (absorción distribución, metabolismo y excreción) de los fármacos. No sería raro encontrar una relación entre la toxicidad oral aguda en rata y log P. Sin embargo, esta explicación mecanística no es siempre fácil, y debe encontrarse a través de mecanismos matemático-estadísticos. ¿Es fácil concluir con el simple razonamiento si los descriptores nO o nN tienen alguna relación con la toxicidad en ratas? Los algoritmos usados en QSAR, aun cuando estos impliquen un alto grado de complejidad, persiguen los mismos propósitos:

Descubrir cuál o cuáles descriptores tienen correlación alguna con la respuesta biológica estudiada, y la forma y proporción con la que ocurre esta relación.

Como mencionamos anteriormente, los descriptores calculados son al final de cuentas mediciones que se han hecho a un grupo de compuestos químicos, representados en dos o tres dimensiones, y a los cuales se le ha aplicado un tratamiento o transformación específico. En el desarrollo de un modelo QSAR, los descriptores calculados son por lo general vertidos en una matriz o base de datos con el fin de aplicar un algoritmo que descubra cuáles descriptores tienen una relación con la respuesta biológica estudiada, y por ende, cuáles descriptores son útiles para generar un modelo.

Es necesario aclarar, que, aunque no existe concretamente un número adecuado de compuestos para generar un modelo, se sabe con certeza que modelos integrados por datasets pequeños, tienen estadísticos de buena predictibilidad para set interno y externo, pero fallan al predecir con exactitud las actividades biológicas de compuestos externos (Golbraikh & Trosha, 2002). Un interesante estudio, revela que la predictibilidad de un modelo QSAR aumenta cuando se incrementa el número de compuestos que se usan para desarrollarlo (Partha et al , 2008)

La tabla siguiente muestra algunos descriptores 2D calculados con M.O.E 2010.10. La primera columna indica el número CAS de cada compuesto, mientras que la última, la DL50 oral en forma logarítmica, la cual se ha obtenido de la biblioteca nacional de los estados Unidos, en su base de datos TOXNET.

Tabla 5. 10 compuestos con descriptores 2D y su respuesta biológica.

| CAS | VDistEq | VDistMa | vdw_area | vdw_vol | vsa_acc | vsa_hyd | vsa_other | vsa_pol | weinerPath | weinerPol | zagreb | DL50oral |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|-----------|--------|----------|
| 17581-48-5 | 3.1868742 | 8.3703127 | 374.86313 | 426.05524 | 27.133842 | 216.85741 | 114.5006 | 27.133842 | 879 | 24 | 86 | 2.71 |
| 2667-52-9 | 2.9288006 | 7.7054315 | 329.63724 | 373.8913 | 9.3319435 | 237.07358 | 66.860428 | 9.3319435 | 481 | 17 | 70 | 4.78 |
| 29173-31-7 | 2.9587636 | 7.4977479 | 292.16196 | 299.0322 | 31.799814 | 153.92984 | 96.467155 | 31.799814 | 408 | 19 | 66 | 3.68 |
| 32744-43-7 | 3.289257 | 8.78724 | 427.40829 | 468.81296 | 36.465786 | 273.28629 | 101.28493 | 36.465786 | 1276 | 29 | 98 | 5.06 |
| 3700-90-1 | 3.0962324 | 8.5261192 | 379.98282 | 424.04718 | 36.465786 | 225.86079 | 101.28493 | 36.465786 | 966 | 27 | 94 | 1.95 |
| 3700-98-9 | 3.4103076 | 8.7706366 | 400.2905 | 459.13376 | 52.194923 | 257.47778 | 71.399315 | 52.194923 | 1332 | 29 | 98 | 2.86 |
| 3853-74-5 | 3.5255096 | 8.1450882 | 367.17224 | 402.01862 | 22.898865 | 232.00607 | 88.563698 | 28.58144 | 920 | 21 | 80 | 4.61 |
| 5840-95-9 | 3.162246 | 7.4650564 | 291.76324 | 306.34821 | 22.898865 | 160.86784 | 88.563698 | 28.58144 | 454 | 17 | 62 | 3.4 |
| 63717-18-0 | 3.0335944 | 8.3756628 | 377.8244 | 403.39728 | 23.6714 | 207.61983 | 133.72086 | 23.6714 | 810 | 30 | 96 | 3.06 |
| 64249-01-0 | 3.2325237 | 8.4973059 | 374.88602 | 428.23239 | 22.898865 | 258.65308 | 79.809959 | 22.898865 | 974 | 31 | 100 | 2.89 |

Continuando con este ejemplo, el método estadístico que se aplique, deberá encontrar el descriptor o descriptores que mejor se correlacionen con la respuesta biológica (en este caso DL50 oral).

Los métodos estadísticos son un componente esencial del QSAR. Ayudan a construir modelos, a estimar las habilidades predictivas de un modelo y a encontrar relaciones y correlaciones entre variables y actividades. Los métodos de análisis de datos se utilizan para combinar datos en diferentes formas y agrupar observaciones en jerarquías. Los métodos de regresión se utilizan para construir un modelo en forma de una ecuación que da una o más variables dependientes (normalmente actividad) en términos de variables independientes ("descriptores"). El modelo puede usarse para predecir actividades para nuevas moléculas, tal vez priorizando o seleccionando un gran grupo de moléculas cuyas actividades no se conocen.

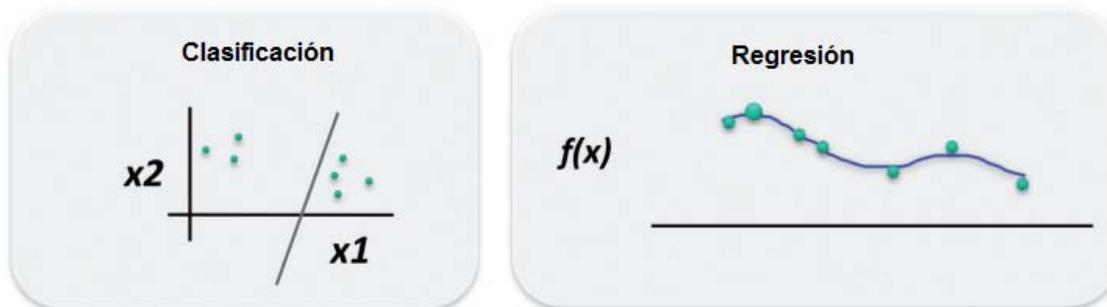
La capacidad de un modelo para proporcionar información sobre el sistema es tan importante como su capacidad de predicción. Posiblemente más valioso que ser capaz de predecir una actividad o propiedad es saber qué aumenta cuando un determinado descriptor aumenta.

Por último, los métodos de validación son necesarios para establecer la predictividad de un modelo en datos no vistos y para ayudar a determinar la complejidad de una ecuación que su cantidad de datos justifica (Devillers, 1996).

En las últimas décadas, además de los miles de descriptores químicos que se han puesto a disposición, también se han desarrollado muchos algoritmos de modelado avanzados y potentes. Los modelos más antiguos QSAR fueron ecuaciones lineales con algunos parámetros. A continuación, se introdujeron otras herramientas, como la red neuronal artificial, la lógica difusa y los algoritmos de minería de datos, lo que hace posibles modelos no lineales y la generación automática de soluciones matemáticas (Devillers, 1996).

Podemos distinguir dos tipos de algoritmos: los de las regresión y clasificación. Los métodos de regresión obtienen un valor continuo. Los clasificadores encuentran la categoría, ej, para clase de toxicidad. La ilustración 10 enumera algunos clasificadores comunes y modelos de regresión utilizados en QSAR (Devillers, 1996).

Ilustración 2 Algunos modelos de clasificación y regresión, utilizados en QSAR.



- * Análisis discriminante
- * CART
- * KNN
- * Lógica difusa
- * Bayesiano
- * Mapa autoorganizado
- * Máquinas de vectores de soporte

- * Análisis multivariado
- * Mínimos cuadrados parciales
- * Redes neurales
- * Otros algoritmos (PCA, algoritmos genéticos).

2.4. Modelo de predicción y validación interna.

Frecuentemente, un algoritmo de clasificación o regresión arrojará información matemático-estadística que permita encontrar qué descriptores se correlacionan o tienen mayor influencia sobre la respuesta biológica y así construir un modelo de predicción que tendrá la forma de una ecuación:

$$\text{Actividad biológica} = \text{descriptor 1} * \text{constante 1} + \text{descriptor 2} * \text{constante 2} + \text{descriptor n} * \text{constante n}$$

Por ejemplo, un artículo publicado (Asadollahi et al, 2011) reporta el desarrollo de modelos QSAR para compuestos antagonistas de los receptores de quimiocina CXCR2, los cuales

surgieron a partir de un algoritmo genético. La ecuación de predicción reportada es la siguiente:

$$pIC50 = -8.92 - 5.41MATS5v - 1.34GATS8p + 31.53MATS2m - 3.54BEHp2$$

Donde la variable dependiente de la ecuación anterior, es el -log de la concentración media inhibitoria máxima, la cual, es una medida de la efectividad de un fármaco al cuantificar la inhibición de una función biológica o bioquímica específica, MATS5v, GATS8p, MATS2m t BEHp2 son los descriptores que componen el modelo, y los valores numéricos son constantes.

Hasta este punto, podría pensarse que al obtener una ecuación como la anterior, esta ya es útil para predecir la actividad de nuevos compuestos. Sin embargo, los modelos obtenidos a partir de descriptores y datos de actividades biológicas experimentales, requieren de rigurosas evaluaciones estadísticas que aseguren que la ecuación puede predecir nuevas entidades químicas.

Uno de esos procedimientos dirigidos a comprobar la confiabilidad de los modelos, se conoce como validación interna. Esta consiste elementalmente en realizar cambios, transformaciones y perturbaciones al modelo obtenido, con el fin de evaluar si este no está sobreentrenado, es decir, que los valores experimentales de la actividad biológica y los descriptores usados para construirlo, no siguen una tendencia que impida a la ecuación predecir nuevos compuestos.

A continuación, se menciona de manera breve algunos métodos de validación interna.

Validación cruzada, Leave One Out (Dejar uno fuera)

Para determinar la validación cruzada LOO, el conjunto de entrenamiento se modifica principalmente eliminando un compuesto del conjunto. A continuación, se reconstruye el modelo QSAR basado en las moléculas restantes del conjunto de entrenamiento utilizando la combinación de descriptor seleccionada originalmente y se calcula la actividad del compuesto suprimido con base en la ecuación QSAR resultante. Este ciclo se repite hasta que todas las moléculas del conjunto de entrenamiento se han suprimido una vez y los datos de actividad previstos obtenidos para todos los compuestos del conjunto de entrenamiento se usan para el cálculo de varios parámetros de validación internos. Por último, se evalúa la predictividad del modelo utilizando la suma residual predicha de cuadrados (PRESS) y se valida R^2 (Q^2) para el modelo mientras que el valor de la desviación estándar del error de predicción (SDEP) se calcula a partir de PRESS

$$\text{PRESS} = \sum (Y_{\text{obs}} - Y_{\text{pred}})^2 \quad \dots(1)$$

$$\text{SDEP} = \sqrt{\frac{\text{PRESS}}{n}}$$

$$Q^2 = 1 - \frac{\sum (Y_{\text{obs}(\text{train})} - Y_{\text{pred}(\text{train})})^2}{\sum (Y_{\text{obs}(\text{train})} - \bar{Y}_{\text{training}})^2} = 1 - \frac{\text{PRESS}}{\sum (Y_{\text{obs}(\text{train})} - \bar{Y}_{\text{training}})^2} \quad \dots(3)$$

En las ecuaciones (1)-(3), Y_{obs} e Y_{pred} corresponden a los valores de actividad observados y LOO previstos, n se refiere al número de observaciones, $Y_{\text{obs}}(\text{train})$ es la actividad observada, $Y_{\text{pred}}(\text{train})$ es la actividad predicha de las moléculas del conjunto de entrenamiento basadas en la técnica LOO. El valor umbral de Q^2 es 0.5 (Roy et al, 2015).

Validación cruzada, Leave many out (Dejar varios fuera)

El principio básico de la técnica de LMO o técnica de dejar algunos fuera leave-some-out (LSO) es que una parte definida del conjunto de entrenamiento se mantiene y se elimina en cada ciclo. Para cada ciclo, el modelo se construye con las moléculas restantes (y utilizando los descriptores seleccionados originalmente) y luego se predice la actividad de los compuestos eliminados utilizando el modelo desarrollado. Después de que se hayan completado todos los ciclos, se utilizan los valores de actividad predichos de los compuestos para el cálculo de Q^2 (Roy et al, 2015).

Bootstrapping

La premisa básica de este método es que el conjunto de datos es representativo de la población de la cual fue modelado. Debido a que sólo hay un conjunto de datos, bootstrapping simula lo que sucedería si la población fuera remuestreada mediante un remuestreo aleatorio del conjunto de datos (Efron y Tibshirani, 1993; Wehrens et al., 2000). En una validación típica de arranque, los grupos K n -dimensionales se generan mediante una selección repetida aleatoriamente de n objetos del conjunto de datos original. Cada grupo de datos es entonces siempre n -dimensional; algunos de estos objetos pueden en el mismo grupo más de una vez, mientras que otros nunca podrían estar dentro. El modelo obtenido en los primeros objetos seleccionados se utiliza para predecir los valores para la muestra excluida. A partir de cada muestra de bootstrap se calcula el parámetro estadístico de interés (aquí Q^2). Esto produce un conjunto de estimaciones utilizadas para obtener una metaestimación (por ejemplo, podríamos generar 100 muestras de bootstrap y calcular Q^2 para cada muestra). Como en el

caso de la validación de leave many out, un alto promedio de Q^2 en la validación de bootstrap es una demostración de la robustez del modelo y la predictividad interna (Gramatica, 2004)

Ajuste por mínimos cuadrados.

El método más común para validar el modelo de forma interna es el ajuste por mínimos cuadrados. Este método de validación es similar a la regresión lineal y se determina R^2 (coeficiente de correlación cuadrado) para la comparación entre las actividades predichas y experimentales. Un método mejorado para determinar R^2 es el ajuste robusto de la línea recta, donde los puntos de datos están alejados de los puntos de datos centrales. Una alternativa a este método es la eliminación de valores atípicos (compuestos del conjunto de entrenamiento) del conjunto de datos en un intento de optimizar el modelo QSAR y sólo es válido si se siguen reglas estadísticas estrictas. Cuando la diferencia entre el valor de R^2 y R^2 adj es menor que 0.3 indica que el número de descriptores implicados en el modelo QSAR es aceptable. El número de descriptores no es aceptable si la diferencia es superior a 0.3.

$$R^2 = \left[\frac{N \sum XY - (\sum X) (\sum Y)}{\sqrt{([N \sum X^2 - (\sum X)^2] [N \sum Y^2 - (\sum Y)^2])}} \right]^2$$

2.5. Validación externa.

La finalidad esencial del desarrollo de los modelos QSAR es la predicción de la actividad biológica de compuestos químicos que aún no han sido sometidos a la experimentación *in vivo*. Lo anterior tiene utilidad en las fases tempranas de diseño y desarrollo de nuevas moléculas, como herramienta premonitoria para la toma de decisiones futuras. Por ejemplo, los estudios QSAR pueden ser valiosos para decidir si vale la pena la síntesis de un compuesto candidato y de este modo, optimizar el proceso y minimizar el gasto innecesario de recursos económicos y físicos.

Los estudios QSAR son usados en áreas como la farmacología, la toxicología, las industrias cosmética, agroquímica y genética, entre otras, donde se precisa tener una idea sólida sobre los posibles efectos de las sustancias químicas en los sistemas biológicos. Como ya se vio antes, la validación de un modelo de predicción consiste elementalmente en asegurar que, en efecto, sea capaz de predecir la actividad biológica de nuevos compuestos que se encuentran dentro de un universo que fue delimitado durante su desarrollo. Esta limitación del modelo es comúnmente llamada -Dominio de aplicabilidad-, y consiste en fijar una región del espacio químico en la cual el modelo tiene poder predictivo. Por ejemplo, si se deseara desarrollar un modelo con capacidad de predecir la toxicidad de fenoles, sería incorrecto pensar que este pudiera funcionar para predecir la toxicidad de aminas terciarias. Es decir, el dominio de aplicabilidad son todas aquellas características del modelo que definen qué compuestos pueden ser predichos y cuáles no.

Una definición más formal del dominio de aplicabilidad, lo describe como una región teórica en el espacio químico conformado tanto por los descriptores del modelo como por la respuesta modelada. El dominio de aplicabilidad juega un papel crucial para estimar la incertidumbre en la predicción de un compuesto particular sobre la base de cómo es similar a los compuestos empleados para construir el modelo QSAR. Por lo tanto, la predicción de una respuesta modelada utilizando QSAR es aplicable sólo si el compuesto que se predijo cae dentro del DA del modelo, ya que es imposible predecir todo el universo de los compuestos utilizando un solo modelo QSAR (Gramatica, 2007).

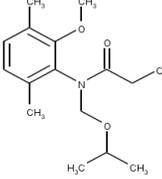
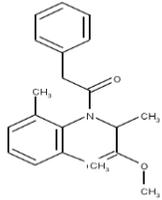
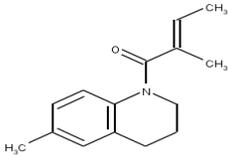
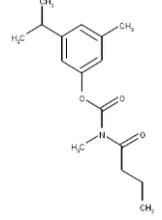
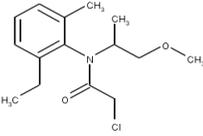
Cuando se tiene la certeza de que un compuesto se encuentra dentro del dominio de aplicabilidad de un modelo, la respuesta biológica puede predecirse comúnmente con dos propósitos: obtener posible información que no ha sido aún probada en modelos experimentales *in vivo*, o comprobar si el modelo es capaz de predecir la actividad biológica de la cual ya se tiene información experimental. En este último caso, el resultado es una respuesta biológica calculada (o predicha) por el modelo, que se compara con la experimental siempre dentro de un intervalo de confianza. La validación externa de un modelo de predicción QSAR o QSPR, consiste precisamente, en determinar si este puede predecir compuestos distintos a los que fueron usados para su construcción. Este grupo de compuestos se conoce como set de prueba, set de validación o set externo.

Los datos obtenidos de una validación externa, son datos experimentales contra calculados, que, al ser graficados tienen el aspecto típico de una recta de regresión.

Los siguientes datos de validación externa provienen del desarrollo de un modelo QSAR para la predicción de la toxicidad de impurezas derivadas del proceso de fabricación de un plaguicida. Este tipo de datos se reportan típicamente en estudios con fines regulatorios en diversas partes del mundo, y en este particular ejemplo, están vertidos en una tabla que contiene cinco compuestos con los números CAS, los valores logarítmicos de la dosis letal media experimental, la dosis letal media calculada con la ecuación de predicción, el valor de los descriptores calculados, y la estructura química (Acuña, 2014) ver Tabla 6.

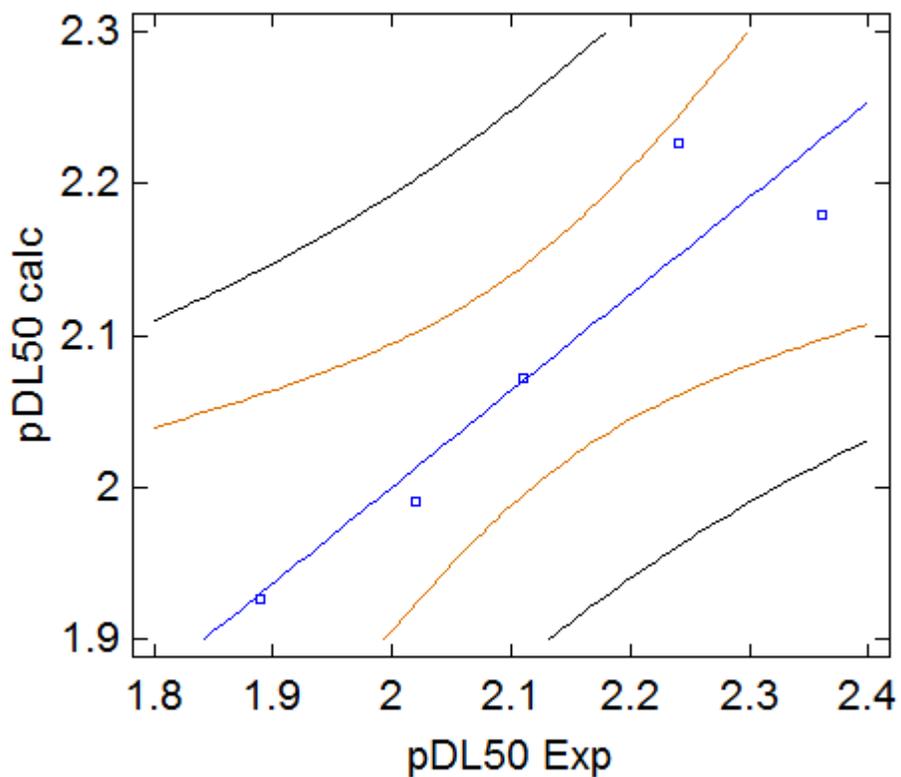
$$\mathbf{pDL_{50} = 2.53107 - 0.154568 * MATS8m + 0.682734 * MATS7p - 0.0150956 * MPC08}$$

Tabla 6. Ejemplo de una validación externa con 6 compuestos.

| Compuesto | DL _{50Exp} -Log (mol/Kg) | MATS8m | MATS7 p | MPC0 8 | DL _{50Calc} -Log (mol/Kg) | Estructura |
|------------|---|--------|------------|-----------|--|---|
| 78194-09-9 | 2.24 | 0.132 | 0.181 | 27 | 2.226 |  |
| 71626-11-4 | 1.89 | -0.057 | -0.059 | 38 | 1.926 |  |
| 81719-55-3 | 2.02 | 0.119 | -0.056 | 32 | 1.991 |  |
| 34264-24-9 | 2.36 | -0.052 | -0.174 | 16 | 2.179 |  |
| 51218-45-2 | 2.11 | 0.112 | 0.039 | 31 | 2.072 |  |

Como se mencionó anteriormente, es común que de los datos experimentales y los calculados sean susceptibles de graficarse y de procesarse estadísticamente para la obtención de parámetros que permiten evaluar la calidad del modelo.

Ilustración 3. Gráfica de valores predichos contra experimentales para compuestos de la validación externa de un modelo QSAR. (Acuña, 2014)



A continuación, se mencionan algunos estadísticos importantes inherentes a la validación externa:

- R^2 predictiva (R^2_{pred} o $Q^2_{(F1)}$)

La R^2_{pred} refleja el grado de correlación entre la actividad predicha y observada del conjunto de prueba.

$$R^2_{pred} = 1 - \frac{\sum (Y_{obs}(prueba) - Y_{pred}(prueba))^2}{\sum (Y_{obs}(prueba) - \bar{Y}_{entrenamiento})^2}$$

Aquí, la Y_{obs} (prueba) e Y_{pred} (prueba) son los datos de la actividad predicha y observada para los compuestos de prueba, mientras $\bar{Y}_{entrenamiento}$ indica la media de la actividad observada de las moléculas del conjunto de prueba. De aquí, que los modelos con valores de R^2 superiores al valor estipulado de 0.5 sean considerados como buenos predictores¹⁸.

- Criterio de Golbraikh y Tropsha

Golbraikh y Tropsha propusieron un conjunto de parámetros para determinar la predictibilidad externa de un modelo QSAR. De acuerdo a Golbraikh y Tropsha, los modelos son considerados satisfactorios, si satisfacen todas las condiciones siguientes¹⁸:

- a) $Q^2_{entrenamiento} > 0.5$
- b) $R^2_{de\ prueba} > 0.6$
- c) $\frac{r^2 - r_0^2}{r^2} < 0.1$ y $0.85 \leq K \leq 1.15$

$$\frac{r^2 - r_0'^2}{r^2} < 0.1 \text{ y } 0.85 \leq k' \leq 1.15.$$

d) $|r_0^2 - r_0| < 0.3$

- Estadístico $r_{m(\text{prueba})}^2$ para validación externa.

Con el fin de verificar la propincuidad entre los datos observados y predichos, el parámetro $r_{m(\text{prueba})}^2$, similar a $r_{m(\text{LOO})}^2$ usado en validación interna, ha sido desarrollado por Roy et al. El valor de este estadístico es calculado usando los coeficientes de correlación cuadrada entre las actividades predicha y observada de los compuestos del conjunto de prueba. Para una predicción aceptable, el valor de $\Delta r_m^2(\text{prueba})$ debería ser preferiblemente menor a 0.2, dado que el valor de $r_{m(\text{prueba})}^2$ es más de 0.5.

Es interesante que Roy y colaboradores establecieran que esta herramienta puede ser extendida a todo el set de datos empleando la actividad predicha LOO (leave one out) para el conjunto de entrenamiento y la actividad predicha para los compuestos del conjunto de prueba. Estos parámetros han sido referidos como $\Delta r_m^2(\text{total})$ el cual refleja la habilidad predictiva del modelo para el total del conjunto de datos (Roy et al, 2015)

2.6.Farmacometría: Generalidades

La farmacometría es la valoración cuantitativa de la actividad biológica de los fármacos. Es la parte de la farmacología encargada de establecer la relación precisa entre dosis y actividad biológica. Esta relación se expresa gráficamente en la curva dosis-efecto, que consideraremos más adelante. Es necesario aclarar que todas las curvas que se presentan en este capítulo son modelos matemáticos, es decir, son expresiones matemáticas con las cuales se intentan

explicar los fenómenos biológicos. Estas son sólo aproximaciones a los fenómenos que se quieren conocer.

Generalidades de los Bioensayos

Cuando un fármaco se administra a un paciente es siempre con el objetivo de producir un efecto terapéutico. Sin embargo, no debemos olvidar que un solo fármaco puede producir diferentes efectos biológicos. Por ejemplo, muchos fármacos pueden presentar uno o más efectos adversos. Adicionalmente, la magnitud de los efectos producidos por un fármaco puede ser variable, dependiendo de diversos factores entre los que se encuentran factores farmacocinéticos, farmacodinámicos e intrínsecos del paciente. Por esto, es importante conocer, tanto los efectos biológicos que un fármaco puede producir, como los mecanismos por los que dichas respuestas se generan y las dosis que se requieren para que se presenten. Los efectos producidos por los fármacos se estudian experimentalmente mediante bioensayos. Un bioensayo consiste en cuantificar una respuesta biológica producida como consecuencia de la exposición a un fármaco. Entonces, en estos estudios la respuesta biológica que el fármaco produce es la variable dependiente (y), mientras que la exposición al fármaco es la variable independiente (x).

Los bioensayos nos sirven como modelos del efecto que un fármaco tendría en la vida real. Por lo tanto, la realización de bioensayos tiene las siguientes aplicaciones:

- 1) Evaluar la actividad de nuevos fármacos.
- 2) Conocer los mediadores endógenos (moléculas de señalización) involucrados en la generación del efecto farmacológico.
- 3) Identificar los efectos secundarios de los fármacos y su interacción con otros compuestos.

4) Conocer la concentración de fármaco necesaria para alcanzar determinado efecto, especialmente los terapéuticos y tóxicos.

Anteriormente, los bioensayos se utilizaban también para la concentración de una sustancia activa, como un fármaco o una hormona, en una muestra biológica. Esto se hacía mediante la comparación de la respuesta producida por la muestra, contra las respuestas producidas por una serie de estándares. Actualmente, los bioensayos no se emplean más con este objetivo pues la cuantificación de fármacos por métodos es menos costosa e igualmente sensible.

Bioensayos con respuestas graduales: curvas concentración-respuesta gradual

Como ya se mencionó, en un bioensayo se mide la respuesta que se genera tras la exposición a un fármaco. Las respuestas graduales se llaman así porque pueden adoptar diferentes valores en la cuantificación. Formalmente, la medición de una respuesta gradual correspondiente a una variable continua, por lo que puede tomar cualquier valor dentro de rango determinado. Desde los primeros experimentos en los que se cuantificaron respuestas farmacológicas graduales, se observó que se generaban respuestas mayores cuando se administraba más fármaco. Por esta razón, en estos bioensayos se administra no sólo una dosis de fármaco, sino varias. El término "dosis" se refiere a la masa de fármaco administrada. Estrictamente, este término debiera usarse sólo en experimentos con animales enteros a los que se les administra por una vía particular determinada masa de fármaco.

En esos casos, la concentración exacta del fármaco en diferentes partes del organismo es desconocida, pero sí se conoce la administrada. En contraste, en bioensayos *in vitro*, cuando utilizamos como muestras biológicas moléculas, células, tejidos u órganos, se aplican como variable independiente, diferentes concentraciones conocidas de fármaco. Aunque por razones históricas las curvas generadas con los resultados de experimentos que miden

respuestas graduales se siguen llamando "curvas dosis respuesta", el término más adecuado es "curvas concentración respuesta".

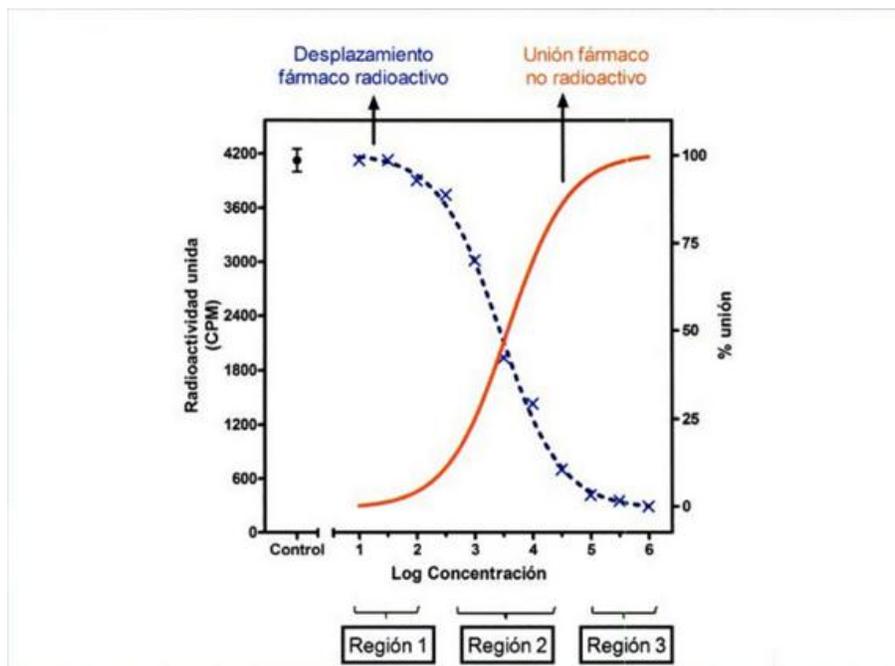
En los bioensayos, la concentración de fármaco debe variar en proporción geométrica. Esto quiere decir que las concentraciones empleadas deben estar igualmente espaciadas en una escala logarítmica. Por ejemplo, pueden utilizarse concentraciones de fármaco de 10, 33, 100, 330, 1000, 3300 y 10000 nM. Cuando se saca el logaritmo a estas concentraciones, dichos valores estarían igualmente espaciados por 0.5 unidades entre uno y otro: 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, y 4.0. Para elegir las concentraciones a utilizar es importante conocer el rango de concentraciones en el que trabajará y el número total de concentraciones que se pueden analizar en el bioensayo. Típicamente se utilizan al menos seis concentraciones diferentes.

Tomemos como ejemplo de una curva concentración-respuesta gradual la medición de la capacidad de un fármaco de unirse a una biomolécula. En estos ensayos de unión o *binding*, las muestras biológicas empleadas son moléculas de importancia biológica como receptores. Para cuantificar la capacidad de unión, se recurre a estrategias que generen una respuesta cuantificable por un método analítico o un equipo especializado. En los ensayos de *binding*, se utilizan fármacos (u hormonas o toxinas) marcados radioactivamente. El procedimiento consiste en tener diferente, cada una con un número equivalente de receptores. A todas las muestras, cada una con un número equivalente de receptores. A todas las muestras se les agrega un exceso de fármaco radioactivo. Si en ese punto, se cuantificara la radioactividad, que es función de la unión del fármaco marcado a los receptores, tendríamos una respuesta máxima, pues todos los receptores estarían ocupados.

Posteriormente a cada muestra se le adiciona una concentración diferente de fármaco que no está marcado radioactivamente. De acuerdo con lo discutido anteriormente, las

concentraciones deben seguir una proporción geométrica, por ejemplo variando diez veces una de la otra. En ese caso, los logaritmos de cada concentración estarían espaciados uno del otro por una unidad. Tras adicionar el fármaco no radiactivo, debe cuantificarse la radioactividad. La disminución en la señal radioactiva es proporcional a la cantidad de fármaco radioactivo que es desplazado de los receptores. Entonces, la unión del fármaco no radiactivo es directamente proporcional a la disminución en la señal ver Ilustración 4.

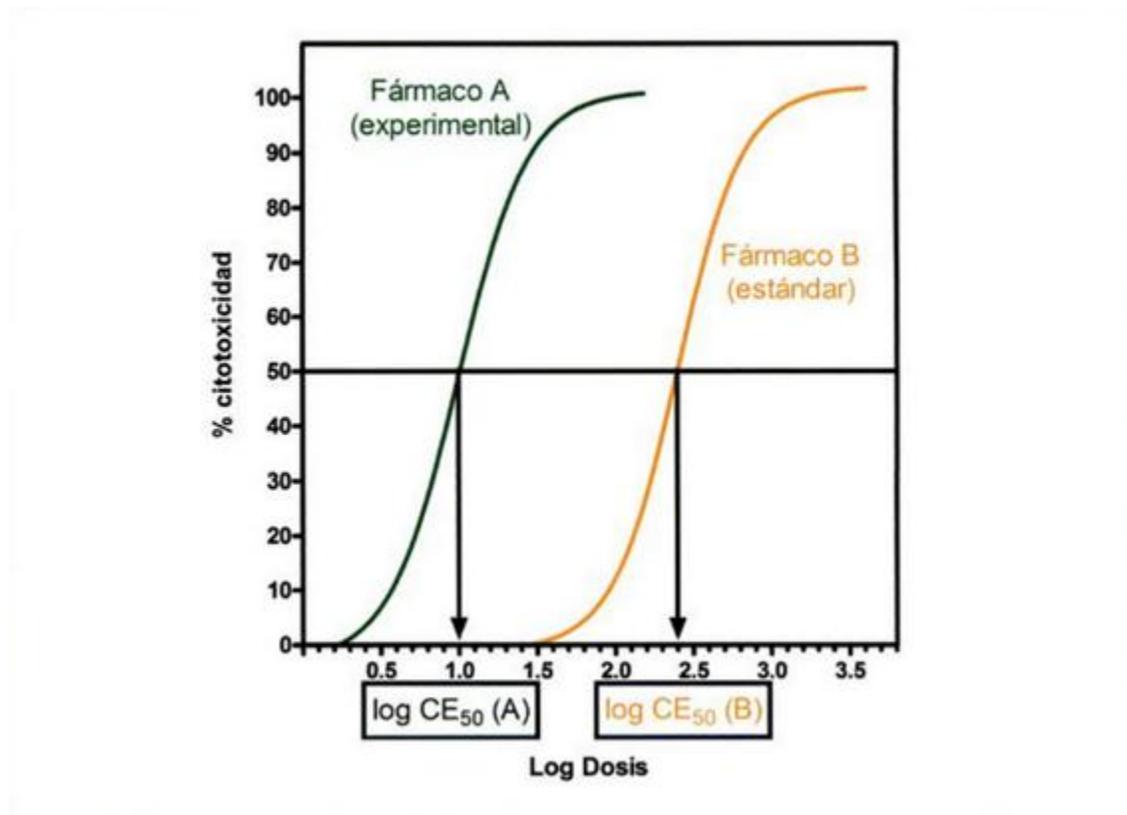
Ilustración 4 Ensayo de unión de radioligando (binding) como ejemplo de una curva concentración-respuesta gradual.



En la Ilustración 4 se pueden observar que existen concentraciones a las cuales aún no hay desplazamiento del fármaco radioactivo (región 1). Esto se debe a que, aunque exista fármaco no radiactivo en el sistema, su afinidad no es suficiente para que comience a unirse a los receptores. De manera equivalente, en la parte derecha de la curva (región 3), la unión alcanzara un máximo aun cuando se utilicen concentraciones mayores de fármaco.

Esto se debe a que el número de receptores en la muestra (al igual que en los tejidos) es limitado y no se pueden ocupar más receptores de los que existen. Sin embargo, también existe una región de la curva (Ilustración 5) en la que la respuesta es directamente a la concentración de fármaco. Es importante notar que esta proporcionalidad se presenta sólo en un rango limitado de concentraciones.

Ilustración 5. Las curvas dosis-respuesta gradual permiten la comparación de la actividad de dos fármacos.



Interpretación de la curva concentración-respuesta

Las características descritas para la curva de unión son comunes a todas las curvas concentración-respuesta. De esta manera, este tipo de curvas están definidas por cuatro parámetros:

- 1) La respuesta en ausencia de fármaco o respuesta basal.
- 2) La respuesta máxima.
- 3) La pendiente de la región en la que la respuesta es directa.
- 4) La concentración efectiva media CE_{50}

Se considera respuesta basal a la producida por el control negativo e idealmente debe ser cero. La respuesta basal depende del sistema en el que se está trabajando, de la respuesta que se mida y del efecto que el vehículo pudiera tener en esa respuesta. De manera similar, la respuesta máxima también depende del sistema y tipo de respuesta, pero depende de manera muy importante de la actividad intrínseca del fármaco para generar esa respuesta.

Dado que la respuesta basal y máxima es función de varios factores, deben determinarse experimentalmente. Por esto, todos los diseños experimentales de curvas de concentración-respuesta deben incluir, además de los controles correspondientes, concentraciones que no produzcan respuesta (concentración subefectivas) y concentraciones que generen la respuesta máxima (concentración subefectivas) y concentraciones que generen la respuesta máxima (concentraciones supreefectivas)

Para facilitar la interpretación de curvas concentración-respuesta gradual, puede ajustarse la respuesta basal a 0 y la respuesta máxima a 100%.

Esta manipulación de datos se conoce como normalización. En una curva normalizada el rango en el que la respuesta varía linealmente con respecto a la concentración de fármaco va aproximadamente de 10 a 90% de respuesta (ver Ilustración 4). A la pendiente de esta región de la gráfica se le conoce como pendiente de Hill. Experimentalmente se ha determinado que muchas respuestas (especialmente en ensayos de “binding”) requieren cerca

de 100 veces (2 unidades logarítmicas) más fármacos para producir 90% de respuesta que para producir 10%. Las curvas que presentan esta característica tienen pendiente de Hill de 1.0. En cambio, cuando la concentración que produce 90% de respuesta es 1000 veces (4 unidades logarítmicas) mayor que la que produce 10%, la pendiente de esta región de la gráfica será dos veces menor, o sea que tendrá pendiente de Hill, tanto menores, como mayores a 1.0. El valor de la pendiente depende de diversos factores fisiológicos y metodológicos, por lo que debe determinarse experimentalmente.

La concentración efectiva media (CE_{50}) es simplemente la concentración de fármaco que provoca una respuesta que está a la mitad de la basal y la respuesta máxima. En una curva normalizada, es muy fácil calcular la CE_{50} , pues corresponde a la concentración que produce 50% de respuesta. Es importante mencionar que la CE_{50} no necesariamente debe coincidir con una de las concentraciones experimentales, sino que comúnmente se obtiene por interpolación en la gráfica. En la ilustración 5 se muestran las curvas concentración-respuesta normalizada del efecto citotóxico de dos fármacos distintos. Para cada fármaco la CE_{50} es diferente, siendo mayor para el fármaco cuya curva se encuentra desplazada a la derecha (fármaco B).

Una vez que se conocen las concentraciones efectivas medias, están sirven para hacer comparaciones entre dos fármacos. Por ejemplo, supongamos que el fármaco B es un principio activo conocido que ya tiene aplicación en la clínica y el fármaco A corresponde a un compuesto experimental.

En estos casos, el fármaco A sirve como estándar de comparación. De la gráfica se observa que para el fármaco A, el log de CE_{50} es 2.4; por lo tanto su CE_{50} es de 251 nM. Para el fármaco B la CE_{50} es 10 nM (antilog de 1). La comparación entre concentraciones de dos

fármacos requeridas para producir una respuesta de la misma magnitud es una medición de la potencia farmacológica. En nuestro ejemplo, podemos concluir que el fármaco B es 25 veces más potente que el fármaco A, pues se requieren concentraciones de B 25 veces menores para producir la misma respuesta.

Cuando se estudian respuestas terapéuticas es esperanzador encontrar fármacos más potentes; sin embargo, cuando se evalúan respuestas tóxicas, un incremento en la potencia significa que el fármaco sería menos seguro.

Bioensayos con respuestas cuantales: curvas dosis-respuesta cuantales

Aunque existen similitudes entre los bioensayos con respuestas graduales y aquellos en los que se miden respuestas cuantales, existen varias diferencias importantes. La primera diferencia es que para este tipo de ensayos sólo se utilizan animales íntegros.

Al utilizarse animales, la variable independiente es la dosis de fármaco y no la concentración (ver arriba). Sin embargo, de manera similar a lo que pasa en las curvas con respuestas graduales, debe existir una proporción geométrica entre las dosis empleadas. La segunda diferencia es obvia: el tipo de respuesta que se mide como variable dependiente. Las respuestas cuantales corresponden a variables discretas que solo pueden adoptar dos valores. Es decir, la respuesta es de todo o nada. Por ejemplo, la inducción o no de anestesia (evaluación de la aparición de un efecto terapéutico) o la muerte o supervivencia de un animal (evaluación de la aparición de un efecto tóxico). La tercera diferencia importante es que por la naturaleza de la respuesta no se puede estudiar su magnitud; en cambio, se analiza la frecuencia con la que ese efecto aparece en la población. Para esto, el diseño experimental

de una curva dosis-respuesta cuantil considera que debe haber un grupo de animales por cada dosis a estudiar.

Tras la administración de la correspondiente dosis de fármaco a cada uno de los animales del grupo, se cuantifica la proporción de individuos que presentaron el efecto. Por ejemplo, si en un grupo de 10 animales que recibieron una dosis X, 8 presentan el efecto estudiado, la proporción para esa dosis será de 0.8 u 80%.

Análisis de la curva dosis- respuesta cuantil

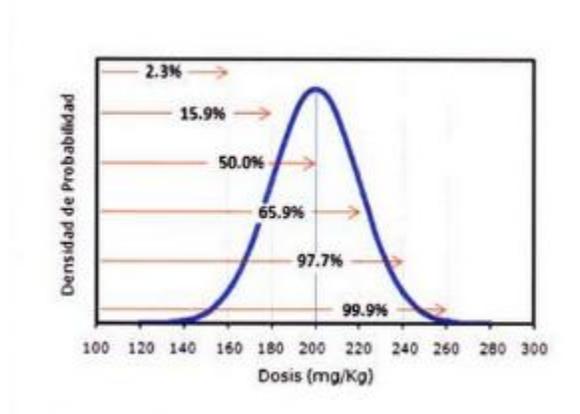
Los animales empleados en estos bioensayos deben pertenecer a la misma población. Es decir, si hablamos de animales de experimentación, estos deber ser de la misma especie, cepa, sexo, edad, y aproximadamente del mismo peso. Sin embargo, aun cuando estas condiciones se cumplan, no todos los individuos reaccionan igual ante un mismo estímulo. Esto se debe a la variabilidad biológica que existe entre individuos. Por ejemplo, si se estudia la dosis de una toxina necesaria para mantener una población de ratones, encontraríamos que la mayoría de ellos requiere dosis similares que se encuentran en un rango pequeño. Sin embargo, algunos requerirían recibir dosis menores o mayores para producir la muerte. Como muchos fenómenos biológicos, la presencia de efecto farmacológico sigue una distribución normal.

La distribución normal (también llamada curva de Gauss), es una distribución continua de probabilidad. Esto quiere decir que la ecuación de la curva describe la probabilidad de que un evento pase. En ensayos farmacológicos, la curva de distribución normal nos sirve para predecir la probabilidad de que un individuo presente el efecto estudiado. La curva de gauss, que tiene una forma de campana, puede ser descrita en función de dos parámetros: el promedio y la desviación estándar.

En promedio queda en el centro de la distribución, por lo que se considera una medida de tendencia central; en cambio, la desviación estándar es una medida de dispersión, pues cuantifica qué tanto se alejan el resto de los datos del promedio. Tanto el promedio, como la desviación estándar tienen unidades de la variable independiente, es decir unidades de masa de fármaco (dosis).

La distribución normal graficada en la ilustración 6 es una distribución hipotética de la frecuencia con la cual diferentes dosis de toxina matan a los animales. En esta distribución, el promedio es 200 mg/kg y la desviación estándar es 20 mg/Kg. Empleando el modelo de distribución normal, podemos calcular el porcentaje de la población que moriría al recibir una dosis igual al promedio menos dos desviaciones (160 mg/Kg) es de 2.3%.

Ilustración 6 Curva de distribución normal. La probabilidad acumulada (flechas rojas) crece conforme aumenta la dosis.

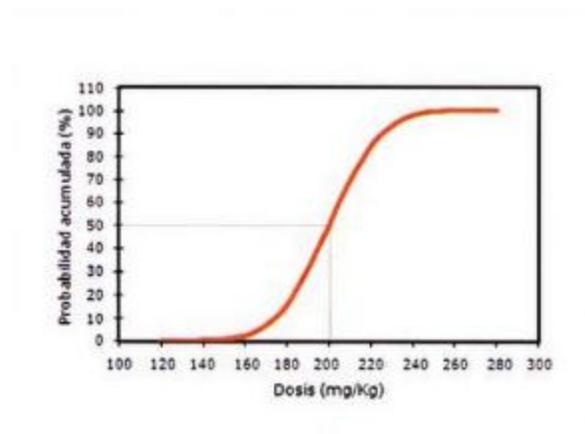


En cambio, si se administrara una dosis de 200 mg/Kg (igual al promedio) 50 % de la población moriría. Finalmente, si la dosis fuera de 260 mg/Kg (promedio más tres desviaciones estándar) 99.9% de la población presentaría un efecto tóxico.

Esto quiere decir que, si tomo un individuo cualquiera perteneciente a esa población, tendría una probabilidad de morir de 2.3% si se administro una dosis de 160 mg/Kg, una probabilidad de 50% si la dosis es de 200 mg/Kg, y se puede tener la certeza (99.9%) que va a morir si la dosis administrada es de 260 mg/Kg.

La función de probabilidad acumulada es la integral de la ecuación que describe a la distribución normal. Por lo tanto, la probabilidad acumulada es también función de la dosis. En la ilustración 7 se observa que la curva de probabilidad acumulada tiene forma de “S”. Es importante hacer notar que el límite inferior es 0% y que la probabilidad de que se presente el efecto es mayor entre mayor sea la dosis, hasta alcanzar 100%. El 50% de probabilidad coincide con el promedio de la dosis.

Ilustración 7 Curva de probabilidad acumulada generada a partir de la distribución normal de la figura 6.



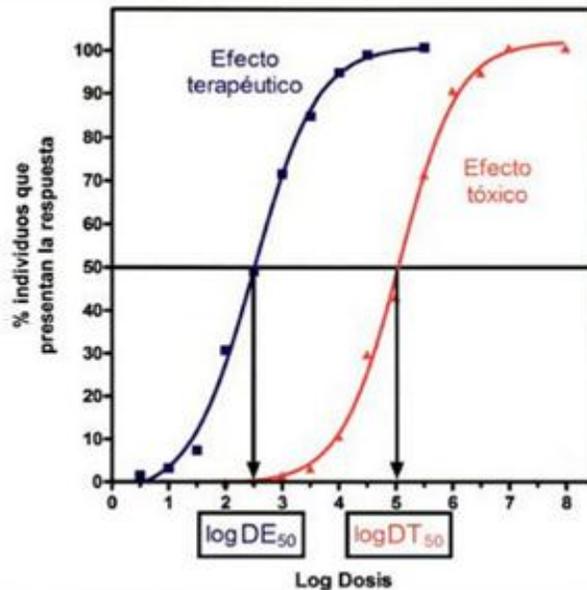
Como ya se mencionó, las curvas dosis respuesta cuantil analizan la frecuencia con la que un efecto se presenta. La distribución de los efectos de los efectos farmacológicos en una población sigue el modelo de distribución normal, por lo tanto, la frecuencia con la que un efecto se presenta tiene también forma de “S”, igual que la función de probabilidad

acumulada. Consecuentemente, en la curva dosis-respuesta cuantales existen dosis que no producen el efecto en ninguno de los individuos (dosis subefectivas).

También existe un rango en la curva en el que la proporción de la población que presenta la respuesta varía proporcionalmente a la dosis. Una vez que se ha alcanzado la dosis que produce el efecto en 100% del grupo correspondiente, la proporción ya no se incrementará aun cuando se administren dosis mayores. Se obtienen gráficas de este tipo independientemente de cuál sea el efecto que se está estudiando.

De hecho, es común estudiar más de un efecto para el mismo fármaco. Por ejemplo, pueden estudiarse un efecto terapéutico y uno tóxico; si ese fuera el caso, los resultados obtenidos serían como los ejemplificados en la Ilustración 8.

Ilustración 8 Curva dosis-respuesta cuantales.



Al igual que con las respuestas graduales podemos conocer, a partir de las curvas de la figura 5, los logaritmos de las dosis que se requieren para producir el efecto estudiado en 50% de la población. Esto se hace mediante interpolación en la parte lineal de las curvas. Para el efecto

terapéutico la dosis efectiva 50 (DE_{50}), sería igual al antilogaritmo de 2.5, o sea $\frac{316 \mu g}{kg}$. En cambio la dosis tóxica 50 (DE_{50}), sería de $100\,000 \frac{\mu g}{kg}$ (antilogaritmo de 5).

Idealmente, todos los fármacos deberían seguir un comportamiento como éste, en el que la dosis necesaria para generar un efecto terapéutico es menor a la que genera un efecto tóxico. Es decir, la curva correspondiente al efecto tóxico debe estar desplazada a la derecha. En nuestro ejemplo, requerirían dosis 300 veces mayores a la DE_{50} para que 50% de los individuos presentaran el efecto adverso. (ver Ilustración 5). Al cociente DT_{50}/DE_{50} se le conoce como índice terapéutico.

Si bien la DE_{50} y la DT_{50} son útiles para describir el comportamiento de un fármaco y para hacer comparaciones de potencia terapéutica o tóxica entre fármacos distintos, su aplicación clínica es limitada. La primera limitante es, por supuesto, que los animales de experimentación no necesariamente responden como los humanos.

Esto se debe a variaciones interespecie, pero también tiene causas experimentales pues en los modelos animales las patologías no cursan exactamente igual que en los humanos.

Por estas razones, los bioensayos realizados en animales no pueden ser fácilmente extrapolables a la clínica. La segunda limitante de estos bioensayos está íntimamente relacionada con el tipo de respuesta que se evalúan.

Cuando se tiene respuesta de todo o nada, se estudia la aparición de dicha respuesta en un tiempo límite tras una administración única de fármaco. Generalmente estos tiempos son cortos (de minutos y horas), por lo que en realidad sólo se analiza efectos agudos. Para un efecto terapéutico esta condición es adecuada, pues se espera que dicho efecto se presente

poco tiempo después de la administración. Sin embargo, los efectos tóxicos pueden requerir tiempos largos o exposiciones repetidas al fármaco para aparecer.

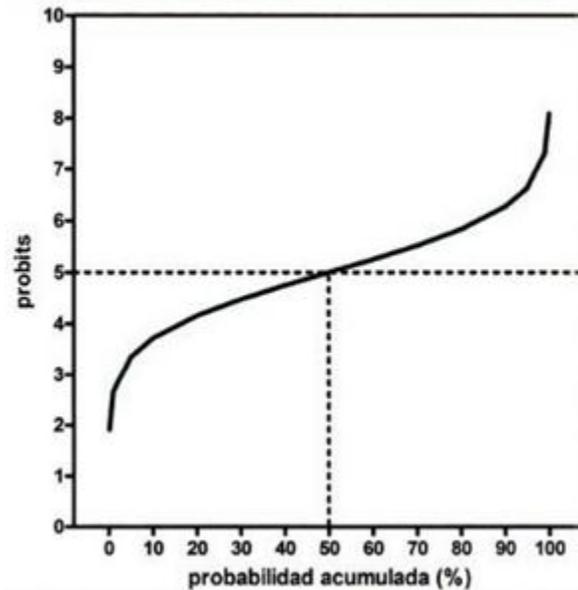
Estas desventajas son más importantes cuando el efecto tóxico es extremo. Por ejemplo, las curvas de letalidad no analizan exposiciones repetidas ni efectos crónicos ni permiten ver si los efectos adversos son reversibles. Sin embargo, la DL_{50} (dosis letal 50 o dosis que mata a 50% de la población) aún tiene aplicaciones importantes, pues sirve para comparar fácilmente la toxicidad de dos sustancias. Conceptualmente es importante que se conozca que la dosis letal 0 es la dosis máxima que no mata a ningún animal, y que la dosis letal 100 es la dosis mínima que mata a todos los animales de experimentación.

Puesto que la DE_{50} obtenida en animales no puede extrapolarse a la clínica, supongamos que el bioensayo de evaluación de efecto terapéutico se realiza con humanos en un ensayo clínico controlado.

En las regiones donde la curva no es lineal, las interpolaciones conllevan errores grandes. Por esto, para facilitar la interpretación de curvas dosis-respuestas cuantales se ha desarrollado un modelo alternativo.

Ese modelo utiliza la función inversa a la función de probabilidad acumulada, con el objetivo de transformar el por ciento de respuesta a una unidad de probabilidad que se conoce como *probits*. La escala de los probits es arbitraria: 0 probits corresponden a una probabilidad de 0.01%, 5 *probits* corresponden a una probabilidad de 99.9%. La ilustración 9 muestra la relación que existe entre los probits y la probabilidad acumulada.

Ilustración 9. La función "probit" convierte la probabilidad acumulada, a una escala arbitraria.



Utilizando la función probit, las curvas con forma de “S”, pueden transformarse a rectas. Por lo tanto, el empleo de probits permite conocer con mayor exactitud las dosis que producen en porcentajes bajos de la población o en porcentajes cercanos a 100. Aunque el modelo de probits sigue siendo válido, está siendo desplazado por programas de cómputo que permiten realizar con relativa facilidad regresiones no lineales y las correspondientes interpolaciones.

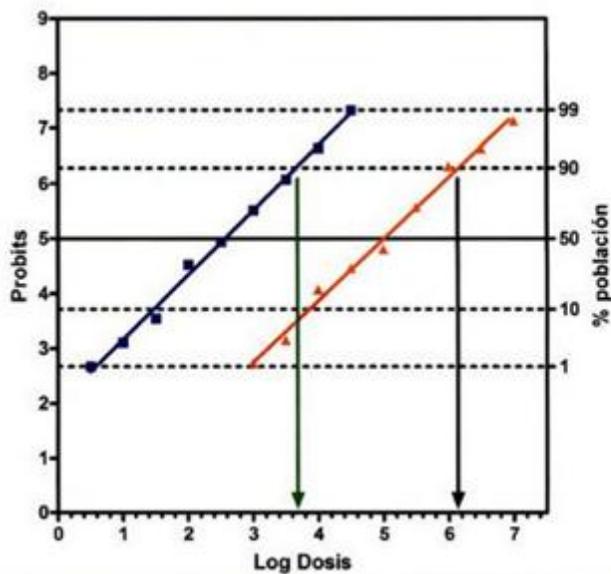
La figura 7 muestra los mismos datos de la figura 5, pero en una escala de probits. Las flechas señalan los logaritmos de las dosis que producen un efecto en 90% de la población. Dichas dosis se conocen como dosis toxica 90 (DE_{90}) y dosis tóxica 90 (DT_{90}). Note que a la DE_{90} (flecha verde en la figura 7) produce también efecto tóxico en un porcentaje bajo (aproximadamente 10%) de la población.

Si calculáramos a la DE_7 , con la que prácticamente cualquier individuo de la población presentaría efecto terapéutico, aproximadamente 30% de la población presentaría el efecto

tóxico. Estas relaciones dependen del fármaco, del uso clínico que se le dé y del efecto tóxico que se esté analizando.

Por ejemplo, si las curvas correspondieran al efecto terapéutico y letal del mismo fármaco, éste difícilmente podría emplearse clínicamente pues se correría un riesgo alto de producir la muerte. ver Ilustración 10 (Patiño, 2008).

Ilustración 10. En una escala de probits, las curvas dosis-respuesta tienden a ser lineales. Facilitando las interpolaciones.



3. Objetivos

3.1 Objetivo general

- Proponer la metodología de una práctica de laboratorio para la asignatura de Farmacología general, que introduzca a los estudiantes a los conceptos básicos de los modelos de estructura química-actividad biológica.

3.1.Objetivos Particulares

- Enfocar y correlacionar la práctica desarrollada con conceptos propios del temario de la asignatura.
- Adaptar el procedimiento a la predicción del índice terapéutico y margen de seguridad de un compuesto perteneciente al grupo de los benzodiazepinas.
- Presentar de manera sintetizada los conceptos básicos de QSAR para ilustrar la forma en que los métodos computacionales y los experimentales pueden complementarse.
- Dar a conocer el uso de software para modelado de estructuras químicas y cálculo de descriptores.

4. Metodología

4.1. Consideraciones previas

Para el diseño de la práctica se tomaron en cuenta las siguientes consideraciones:

| | |
|------------------------------------|--|
| Tiempo | El tiempo de ejecución no debe exceder de dos horas. |
| Software | El software utilizado debe ser de fácil adquisición, manejo intuitivo y de uso libre. |
| Correlación con conceptos previos. | La práctica debe apegarse a conceptos de la asignatura. Tomando en cuenta que los modelos QSAR son de gran utilidad en toxicología, se encontró que la práctica podría asociarse con el margen de seguridad e índice terapéutico; conceptos que se abordan teórica y experimentalmente en la asignatura de Farmacología general. |
| Simplificación | Contrastando cuan extenso es el tema de QSAR, y los conocimientos previos de los alumnos de la asignatura, es pertinente abstraer los conceptos elementales sobre el desarrollo de modelos de predicción, y presentarlos de forma simplificada, con el fin de evitar confusiones o crear obstáculos para su correcta comprensión. Sin embargo, se provee al alumno de bibliografía necesaria para profundizar, en caso de desarrollar interés por el tema. |

4.2. Elección del grupo químico.

Como ya se mencionó en el marco teórico, es preciso que en el desarrollo de modelos de predicción QSAR, se asegure la similitud estructural de los análogos, o estos se categoricen de forma adecuada. Al considerar el tiempo disponible para la presentación, desarrollo y resolución de la práctica, se tomó la decisión de proponer una metodología haciendo uso únicamente de cinco análogos estructurales y se predijera un solo compuesto con el modelo desarrollado.

Se decidió estudiar la posibilidad de que el grupo químico fuese el de las benzodiazepinas. La razón de esta elección es que los compuestos pertenecientes a este grupo farmacológico poseen una elevada similitud estructural, y su mecanismo de acción sobre el receptor GABA ya ha sido ampliamente estudiado. Las benzodiazepinas seleccionadas para el diseño de la práctica son las siguientes:

| Compuesto | Número CAS |
|------------------|------------|
| ESTAZOLAM | 29975-16-4 |
| ALPRAZOLAM | 28981-97-7 |
| QUAZEPAM | 36735-22-5 |
| OXAZEPAM | 604-75-1 |
| FLURAZEPAM | 17617-23-1 |
| TRIAZOLAM | 28911-01-5 |
| LORAZEPAM | 846-49-1 |
| TEMAZEPAM | 846-50-4 |
| CLORAZEPATE | 57109-90-7 |
| MIDAZOLAM | 59467-70-8 |
| CLOBAZAM | 22316-47-8 |
| CHLORDIAZEPOXIDE | 58-25-3 |

El fármaco elegido para la comprobación del modelo y la predicción final es Diazepam.

4.3.Preparación de los datos.

Las estructuras químicas en formato .mol de cada una de las benzodiazepinas, así como los datos de DL₅₀ en rata, se obtuvieron de ChemIDplus, en la biblioteca nacional de medicina de los Estados Unidos en su base de datos TOXNET, a la cual se puede acceder a través de la siguiente liga: <https://toxnet.nlm.nih.gov/>.

ChemIDplus reporta el valor de DL_{50} en mg/kg. Es decir, la cantidad de fármaco en miligramos por cada kilogramo que pese el sujeto de experimentación para provocar un efecto letal en el 50% de la población.

Como ya se ha explicado anteriormente, la transformación de unidades y la reducción numérica es importante en el desarrollo de modelos QSAR. A continuación, se muestra un ejemplo de transformación de la DL_{50} oral en rata en mg/kg a $-\log$ (mol/kg), con el uso del peso molecular. Este último puede obtenerse también de ChemIDplus.

Compuesto: Alprazolam.

Datos:

DL_{50} oral en rata: 1220 mg/kg

Peso molecular: 308.77 g/mol

$$DL_{50} \left(\frac{mol}{Kg} \right) = \frac{1220 \text{ mg}}{Kg} * \frac{mol}{308.77g} * \frac{1 g}{1000 mg} = 0.003951161 \text{ mol/kg}$$

$$-\log_{10} (0.003951161) = 2.40$$

Esta transformación de unidades se realiza para los doce análogos seleccionados. Los resultados se muestran en la Tabla 7.

Tabla 7. Resultados de los doce análogos seleccionados

| COMPUESTO | Número CAS | DL50 -log (mol/kg), | DL50 (mg/kg) | Peso molecular (g/mol) |
|------------------|------------|---------------------|--------------|------------------------|
| ESTAZOLAM | 29975-16-4 | 2.07 | 2500 | 294.7439 |
| ALPRAZOLAM | 28981-97-7 | 2.40 | 1220 | 308.7707 |
| QUAZEPAM | 36735-22-5 | 1.89 | 5000 | 386.7989 |
| OXAZEPAM | 604-75-1 | 1.55 | 8000 | 286.717 |
| FLURAZEPAM | 17617-23-1 | 2.60 | 980 | 387.8837 |
| TRIAZOLAM | 28911-01-5 | 1.66 | 7500 | 343.216 |
| LORAZEPAM | 846-49-1 | 1.85 | 4500 | 321.162 |
| TEMAZEPAM | 846-50-4 | 2.18 | 2000 | 300.7437 |
| CLORAZEPATE | 57109-90-7 | 2.67 | 880 | 408.9219 |
| MIDAZOLAM | 59467-70-8 | 3.18 | 215 | 325.7727 |
| CLOBAZAM | 22316-47-8 | 1.70 | 6000 | 300.7437 |
| CHLORDIAZEPOXIDE | 58-25-3 | 2.88 | 392 | 299.7596 |

4.4. Aseguramiento de la similitud estructural.

Como se ha mencionado en el punto 4.3, las estructuras químicas en formato .mol de los 12 análogos estructurales y el Diazepam, fueron descargadas de Toxnet, en la base de datos ChemIDplus.

Con el uso de M.O.E 2016.08 se generó un archivo .sdf con las 13 estructuras, y se cargó en el software DataWarrior versión 4.5.2. Este software sirvió para evaluar la similitud estructural que existe entre cada uno de los 12 análogos y la sustancia de prueba (diazepam).

El descriptor predeterminado de DataWarrior FragFp es una huella digital binaria basada en un diccionario de fragmento de subestructura similar a las claves MDL. Se basa en un diccionario de 512 fragmentos de estructura predefinidos. Estos fueron seleccionados de la

multitud de posibles fragmentos de estructura mediante la optimización de dos criterios: Todos los fragmentos elegidos deben ocurrir con frecuencia con estructuras de moléculas orgánicas típicas. Cualquiera de los dos fragmentos elegidos debe mostrar poca superposición con respecto a su aparición en diversos conjuntos de compuestos orgánicos. El descriptor FragFp contiene 1 bit por cada fragmento en el diccionario. Un bit se establece como 1 si el fragmento correspondiente está presente en la molécula al menos una vez. En aproximadamente la mitad de los fragmentos, todos los heteroátomos han sido reemplazados por comodines. De esta manera, los reemplazos de un solo átomo solo causan una caída moderada de similitud, lo que refleja una percepción de similitud natural de los químicos.

Además de calcular las similitudes de moléculas, DataWarrior usa el descriptor FragFp para un segundo propósito: la aceleración del filtrado de la subestructura. Como una búsqueda de subestructura es efectivamente un algoritmo de coincidencia de gráficos y, por lo tanto, computacionalmente bastante exigente, DataWarrior emplea un paso de pre-selección que puede excluir rápidamente la mayoría de los compuestos de la comparación de gráficos. En este paso, DataWarrior determina una lista de todos los fragmentos de diccionario, que son parte de la consulta de subestructura. Las moléculas que no contienen todos los fragmentos de la consulta no pueden contener la consulta en sí. Por lo tanto, estos se omiten en la fase de comparación de gráficos (Openmolecules.org)

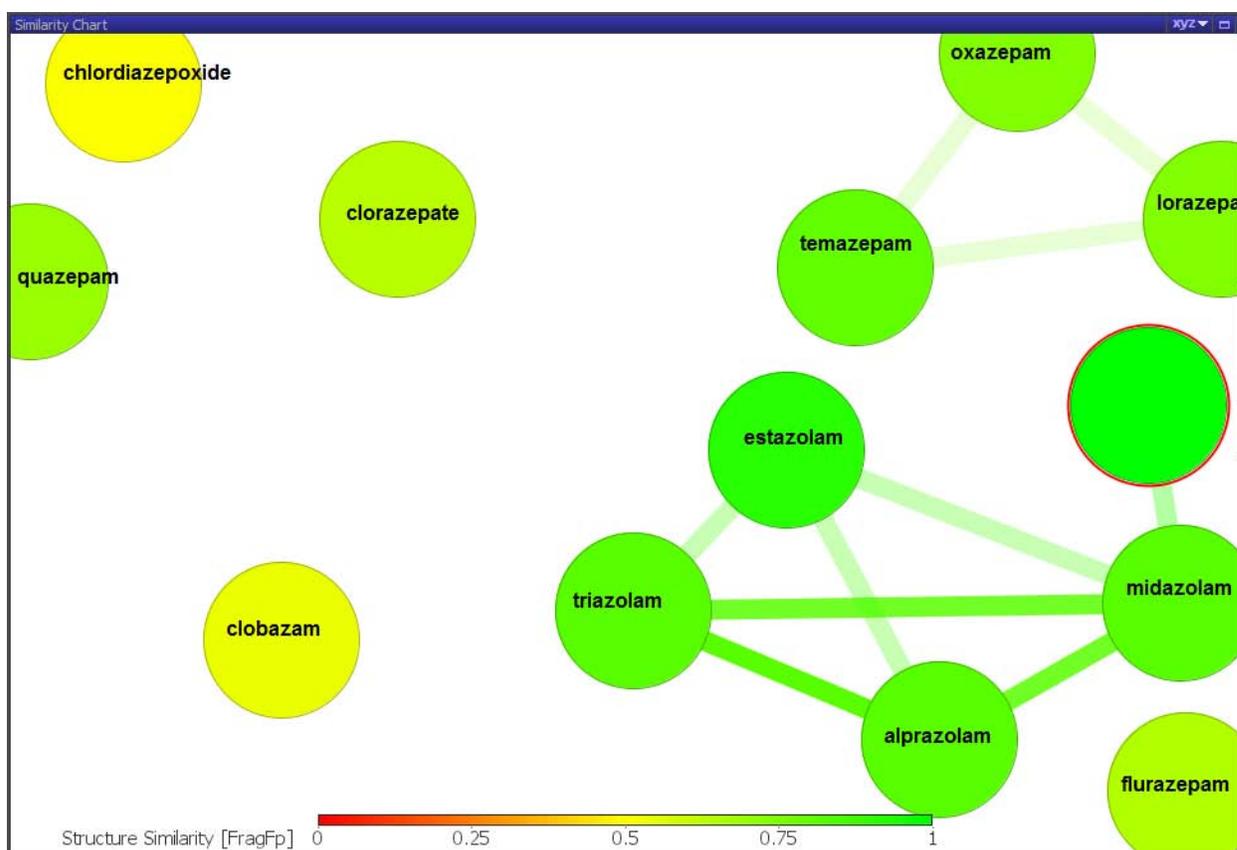
<http://www.openmolecules.org/help/similarity.html>

En la ilustración 11 se puede observar una representación gráfica de la similitud estructural de las 12 benzodiazepinas comparadas con el diazepam, el cual aparece como un círculo con contorno rojo. Se puede apreciar que la mayoría de los compuestos poseen elevada similitud estructural (superior a 0.75), a excepción de clordiazepóxido, clobazam, y flurazepam, por lo

que, aunque pueden utilizarse para la práctica, debe evitarse en lo posible, ya que se complicará la obtención del modelo. Por este motivo, el resultado propuesto es una práctica de 5 análogos, los cuales aparecen interrelacionados en este esquema.

Ilustración 11. Representación gráfica de la similitud estructural de las 12

Benzodiacepinas comparadas con el Diazepam.



5. Resultados

5.1. Material y equipo

- PC con sistema operativo Windows 7 o superior. Procesador i3 o superior. 4 Gb RAM.
- Java virtual machine.
- Microsoft Excel.
- PaDel descriptor
- ChemSketch

5.2. Conceptos previos requeridos para la ejecución de la práctica

- QSAR
- Dosis letal media
- Índice terapéutico y margen de seguridad
- Descriptor
- Regresión lineal múltiple
- Correlación. R^2
- Generalidades sobre el mecanismo de acción de los benzodiazepinas.

5.3. Preparación previa del equipo de cómputo.

- Descargar e instalar chemsketch. Este software tiene una versión gratuita (<http://www.acdlabs.com/resources/freeware/chemsketch/>).
- Descargar e instalar Statgraphics (Cualquier versión)

- Descargar, y descomprimir PaDel descriptor
[\(http://www.yapcwsoft.com/dd/padeldescriptor/\)](http://www.yapcwsoft.com/dd/padeldescriptor/).
- Habilitar el análisis de datos del siguiente modo¹:

1 Haga clic en la pestaña **Archivo**, elija **Opciones** y después haga clic en la categoría **Complementos**.

Si usa Excel 2007, haga clic en el **botón Microsoft Office**  y luego en **Opciones de Excel**.

2 En el cuadro **Administrar**, seleccione **Complementos de Excel** y después haga clic en **Ir**.

Si usa Excel para Mac, en el menú Archivo, vaya a **Herramientas > Complementos de Excel**.

3 En el cuadro **Complementos**, active la casilla **Herramientas para análisis** y después haga clic en **Aceptar**.

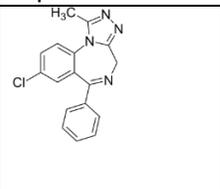
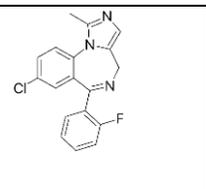
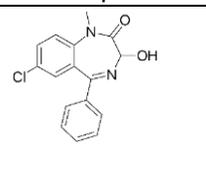
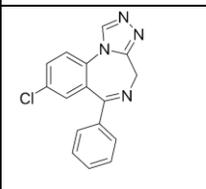
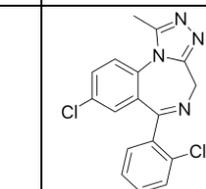
a. Si **Herramientas para análisis** no se enumera en el cuadro **Complementos disponibles**, haga clic en **Examinar** para encontrarlo.

b. Si se le indica que Herramientas para análisis no está instalado actualmente en el equipo, haga clic en **Sí** para instalarlo.

5.4. Modelado de estructuras químicas.

Se modelarán las estructuras químicas de cinco análogos estructurales del diazepam, utilizando el software Chems sketch.

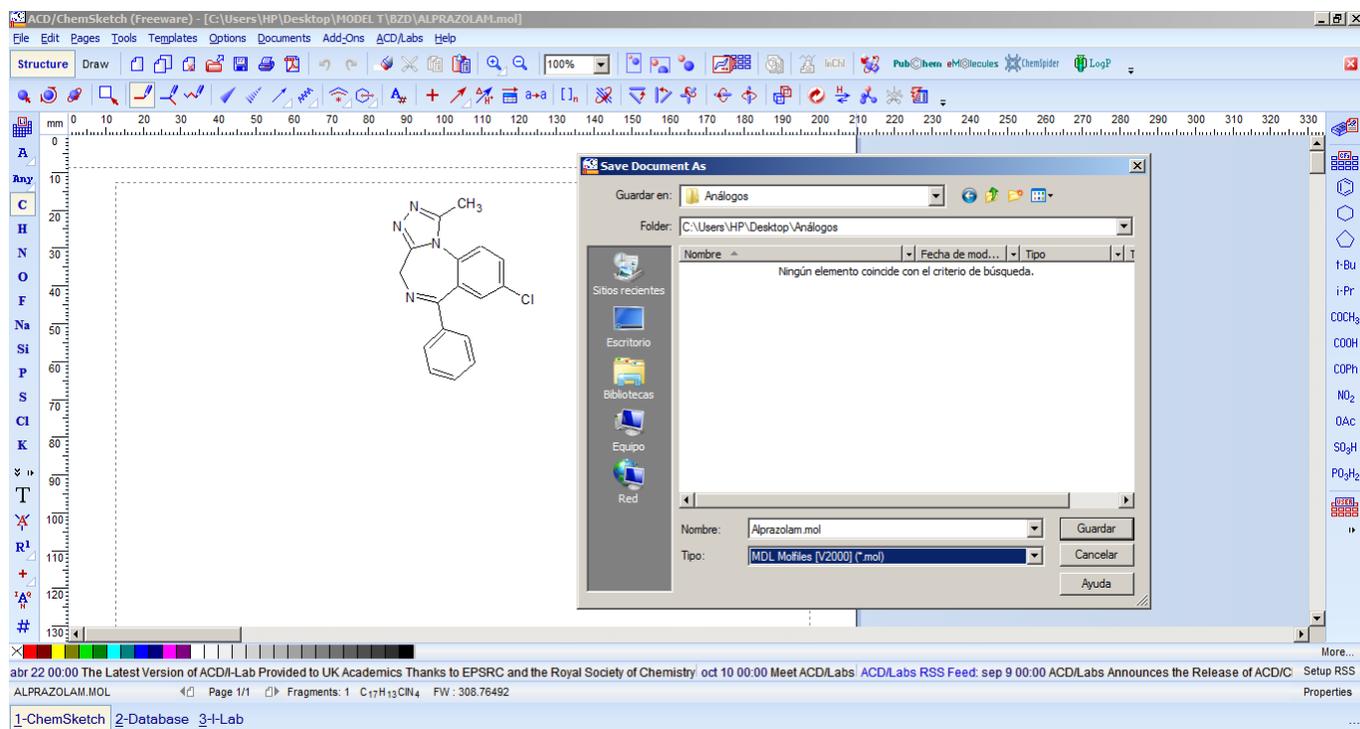
- Alprazolam
- Midazolam
- Temazepam
- Estazolam
- Triazolam.

| Alprazolam | Midazolam | Temazepam | Estazolam | Triazolam |
|--|--|--|---|--|
|  |  |  |  |  |

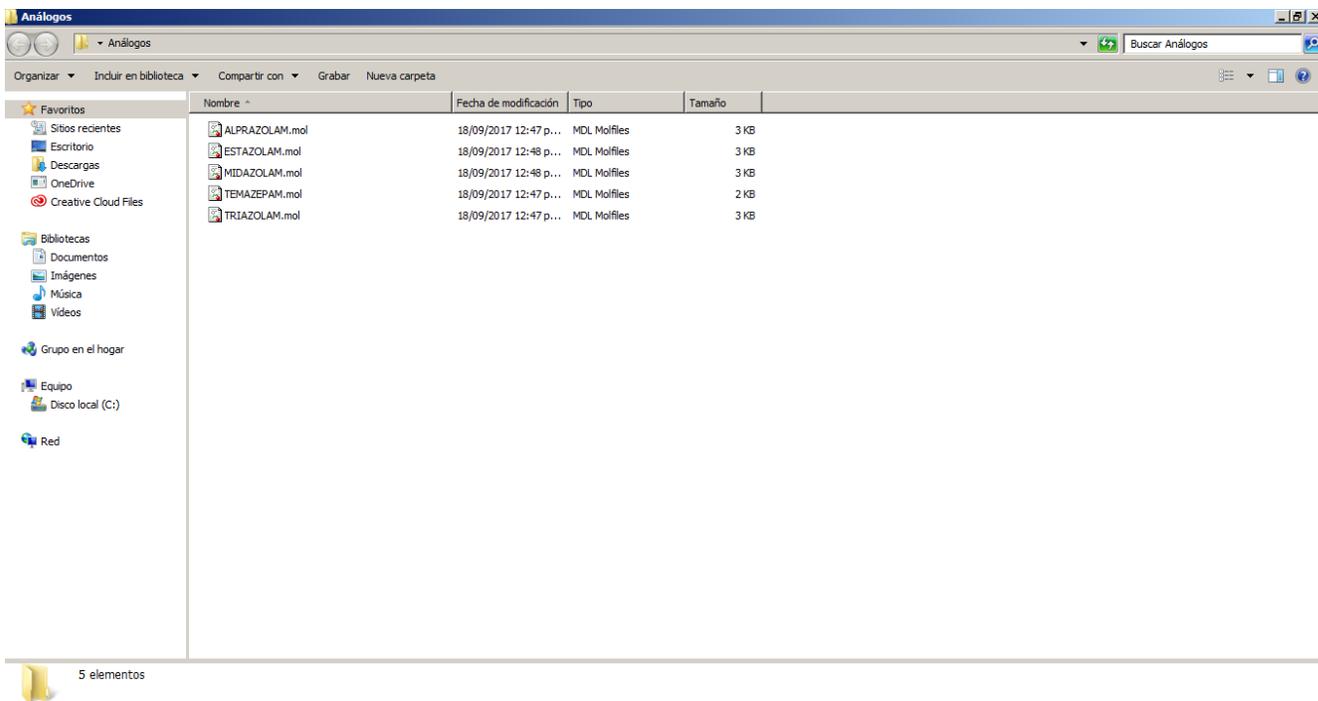
- a) Generar una carpeta llamada “análogos”. Esta puede tener cualquier ubicación. En el presente ejemplo, dicha carpeta se encuentra en el escritorio de Windows.



- b) Abrir el software Chems sketch y dibujar la estructura del Alprazolam. Una vez que se concluye el modelado de la estructura, dar click en: File>Save as>, y nombrar la estructura como “alprazolam”, y escoger el formato MDL Molfiles [V2000] (*.mol). Guardar este archivo en la carpeta previamente generada.

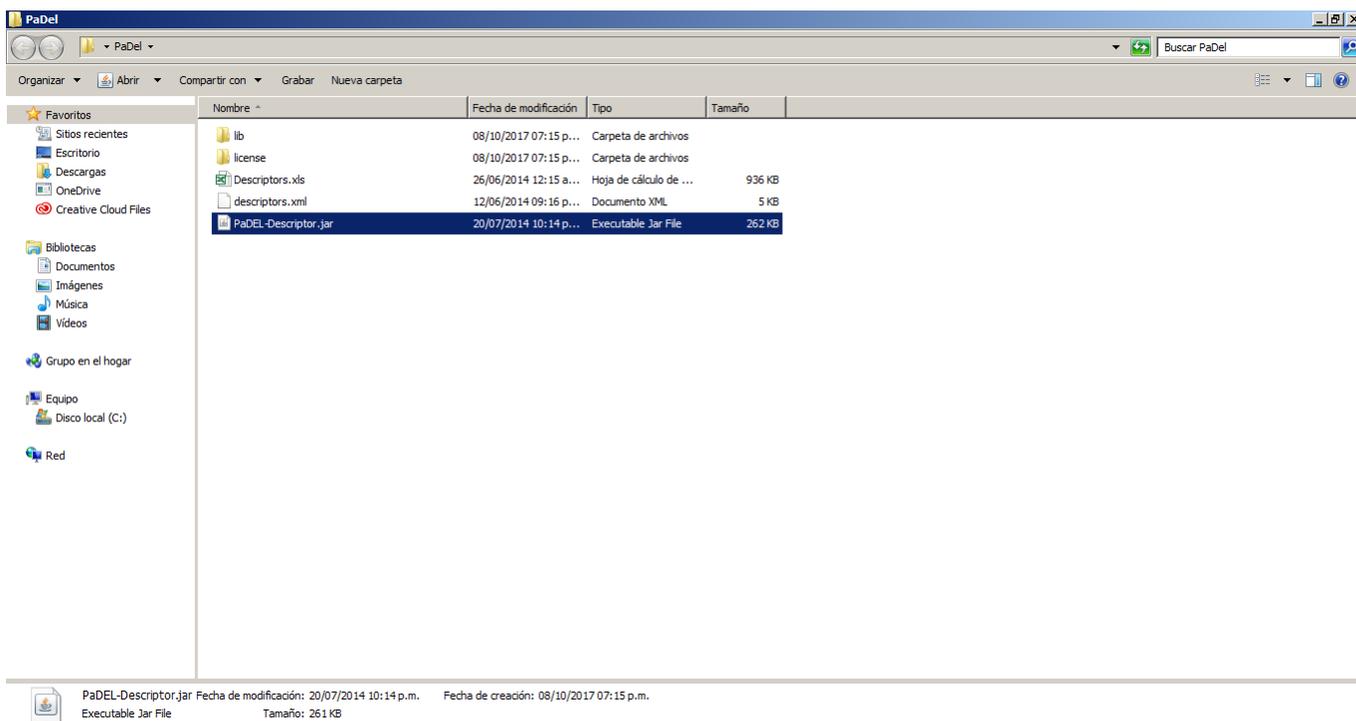


c) Repetir el proceso hasta obtener los archivos .mol de los cinco compuestos modelados.



5.5. Cálculo de descriptores.

a) Dar doble click en el ejecutable de Java: PaDEL-Descriptor.jar



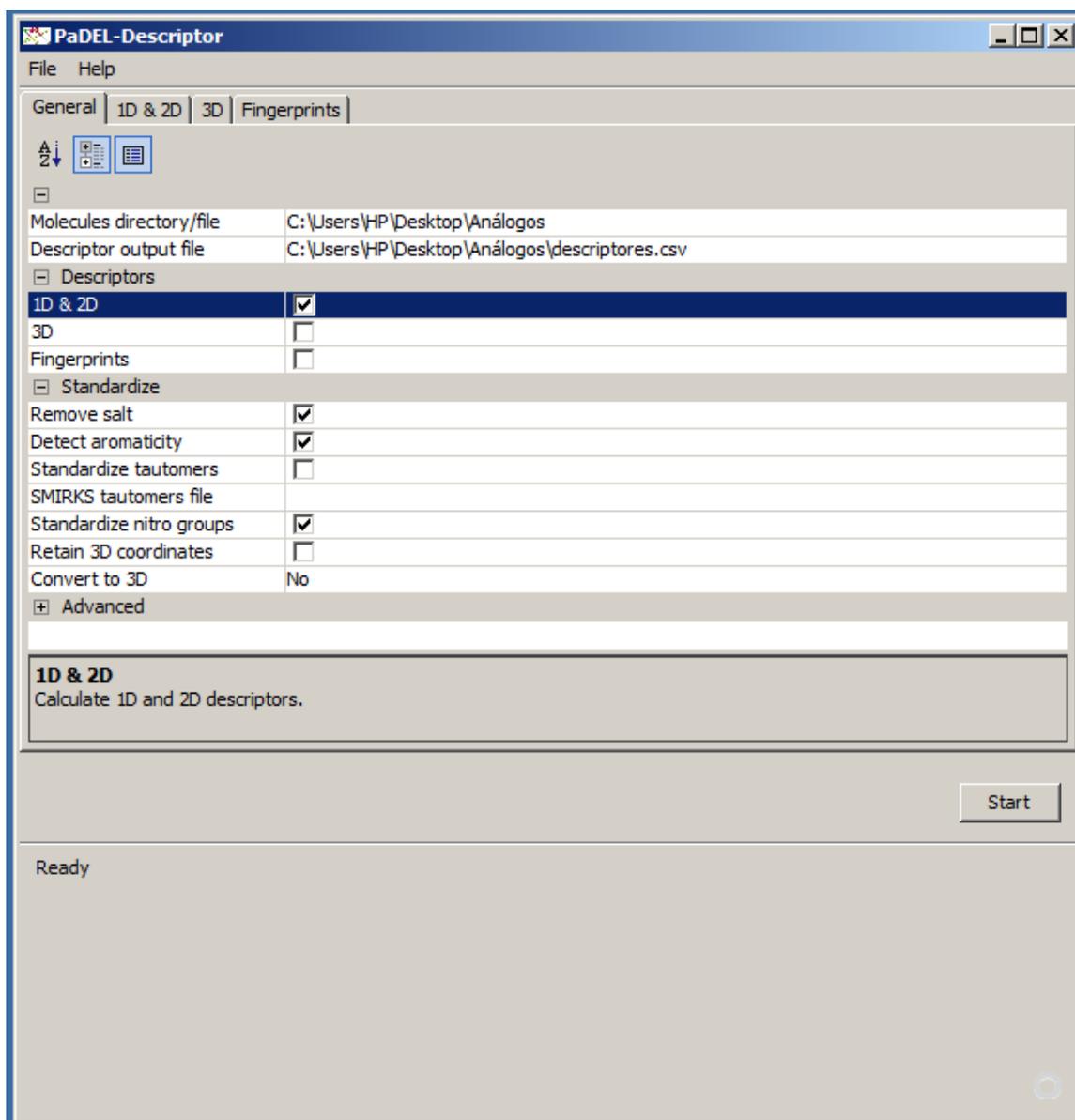
b) En el programa se inserta la siguiente configuración:

Molecules directory/file: (Aquí debe indicarse la ruta de la carpeta donde están los ficheros .mol, en este caso “análogos”, la cual se ubica en el escritorio.

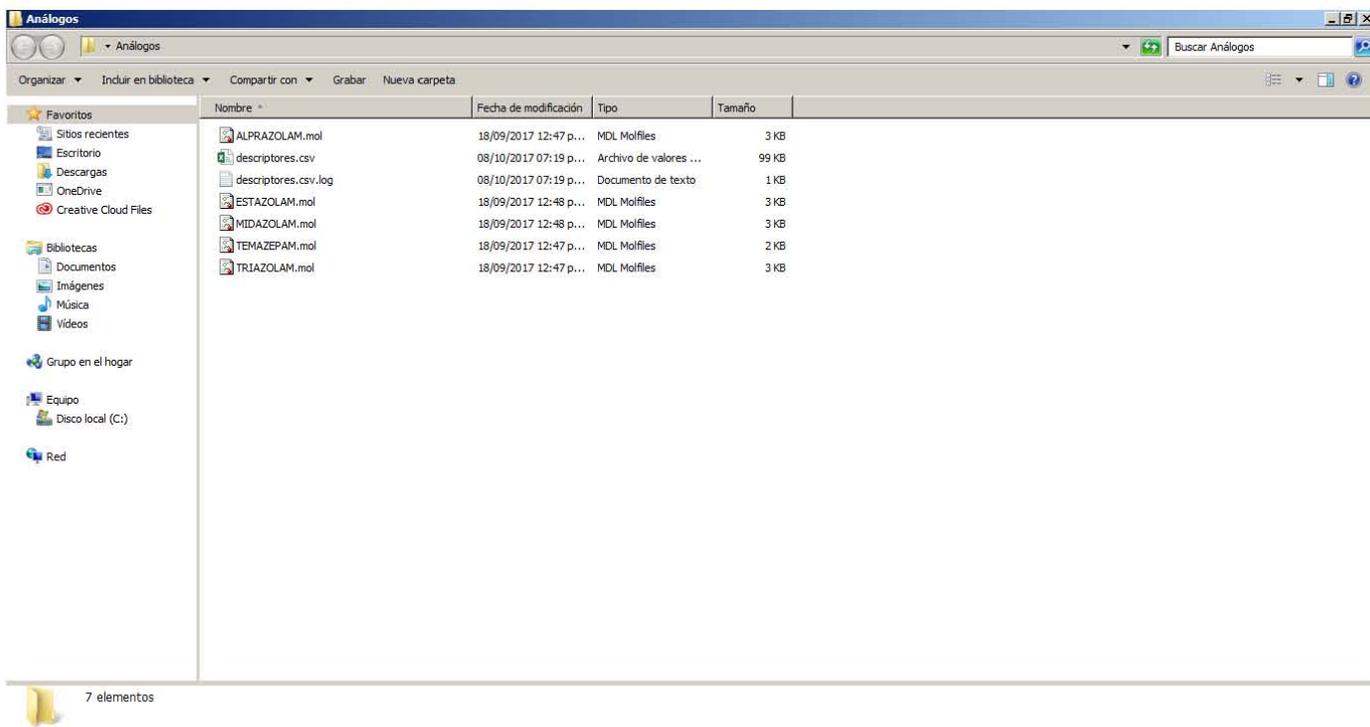
Descriptor output file: (Aquí debe indicarse la carpeta de salida, es decir, aquella donde se genera el archivo con los descriptores calculados. En este caso hemos elegido la misma carpeta donde se encuentran los ficheros mol).

Descriptores: 1D y 2D (Para efectos de la presente práctica, solo se calcularán descriptores 1D y 2D)

Dar click en “START”, para comenzar el cálculo.



- c) Si todo se ha realizado de forma correcta, en la carpeta análogos, aparecerá un documento .csv que contiene los descriptores calculados.



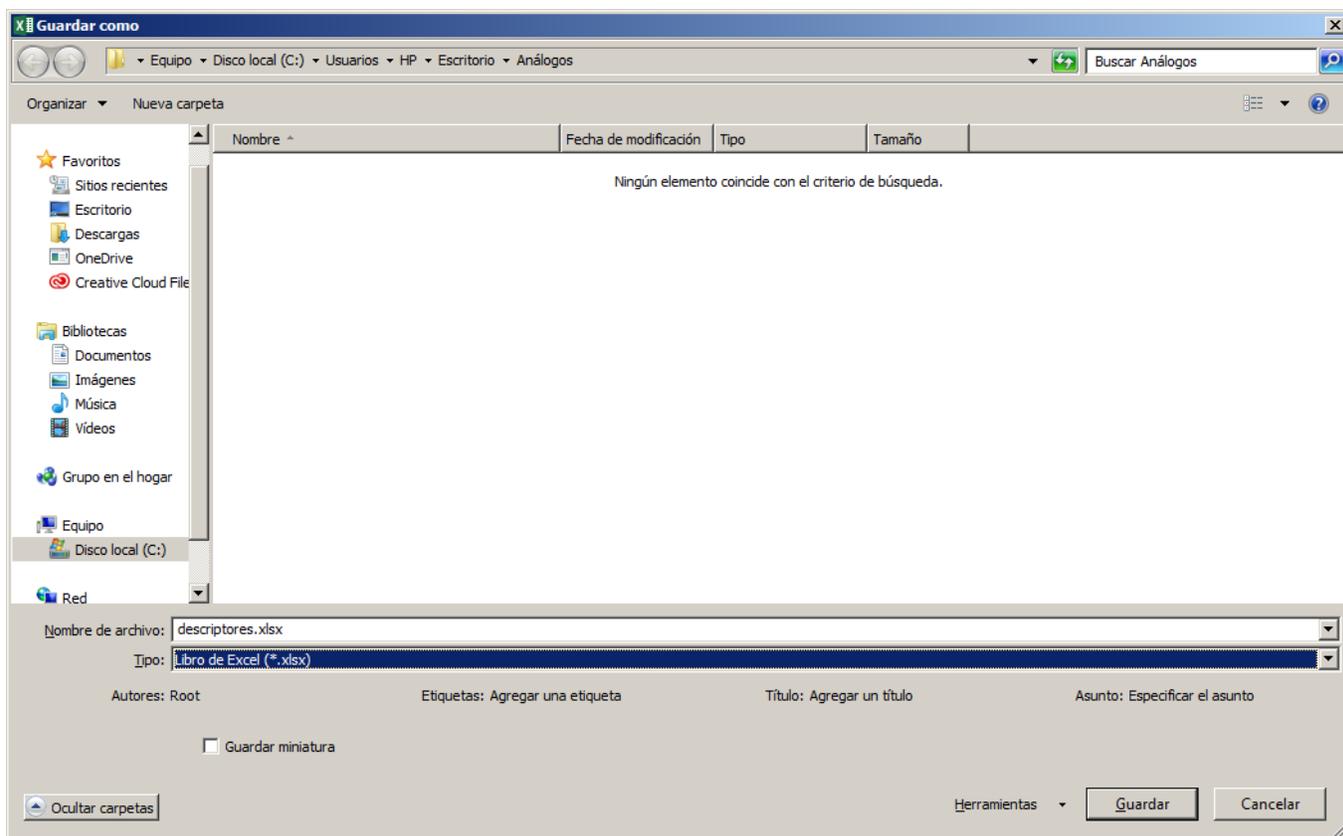
d) Una vez con el archivo abierto, deberá realizarse una “purga”. Es decir, eliminar aquellas columnas que contengan valores de descriptores con cero en todos los

compuestos, o aquellos que contengan un valor constante. La razón de este paso, es que estos descriptores no estarían detectando las variaciones estructurales de cada análogo, y no son útiles para generar el modelo.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|----|----------------|-------|---------|------------|---------|-----------|------------|-----------|-------|------------|----|----|----|----|----|----|
| 1 | Name | nAcid | ALogP | ALogp2 | AMR | apol | naAromAtom | nAromBond | nAtom | nHeavyAtom | nH | nB | nC | nN | nO | nS |
| 2 | ALPRAZOLAM.mol | 0 | 0.4906 | 0.24068836 | 25.5408 | 45.168309 | 17 | 18 | 35 | 22 | 13 | 0 | 17 | 4 | 0 | |
| 3 | ESTAZOLAM.mol | 0 | 0.4393 | 0.19298449 | 19.6421 | 42.074723 | 17 | 18 | 32 | 21 | 11 | 0 | 16 | 4 | 0 | |
| 4 | MIDAZOLAM.mol | 0 | 0.9703 | 0.94148209 | 26.4133 | 46.385309 | 17 | 18 | 36 | 23 | 13 | 0 | 18 | 3 | 0 | |
| 5 | TEMAZEPAM.mol | 0 | -0.2165 | 0.04687225 | 31.1173 | 42.812309 | 12 | 12 | 34 | 21 | 13 | 0 | 16 | 2 | 2 | |
| 6 | TRIAZOLAM.mol | 0 | 1.2349 | 1.52497801 | 31.1564 | 46.681516 | 17 | 18 | 35 | 23 | 12 | 0 | 17 | 4 | 0 | |
| 7 | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | | |
| 22 | | | | | | | | | | | | | | | | |

| | UQ | UR | US | UT | UU | UV | UW | UX | UY | UZ | VA | VB | VC | VD | VE | VF | VZ |
|----|------------|------------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------------|------------|-------|
| 1 | SpMin6_Bhs | SpMin7_Bhs | SpMin8_Bhs | C1SP1 | C2SP1 | C1SP2 | C2SP2 | C3SP2 | C1SP3 | C2SP3 | C3SP3 | C4SP3 | SCH-3 | SCH-4 | SCH-5 | SCH-6 | SCH-7 |
| 2 | 0.90376584 | 0.85029612 | 0.74339306 | 0 | 0 | 2 | 11 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0.09622504 | 0.38992544 | 0.487 |
| 3 | 0.85541516 | 0.7911691 | 0.70285361 | 0 | 0 | 1 | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0.11785113 | 0.32147817 | 0.395 |
| 4 | 0.90413131 | 0.80403222 | 0.75101425 | 0 | 0 | 2 | 12 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0.09622504 | 0.3711967 | 0.560 |
| 5 | 0.85029612 | 0.83104843 | 0.71688925 | 0 | 0 | 1 | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17010345 | 0.205 |
| 6 | 0.8991426 | 0.87923704 | 0.7414664 | 0 | 0 | 2 | 11 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0.09622504 | 0.3711967 | 0.560 |
| 7 | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | | | |
| 22 | | | | | | | | | | | | | | | | | |

e) Una vez que se ha concluido la purga numérica, se deberá cambiar la extensión del archivo dando click en archivo>guardar como> y eligiendo la extensión .xlsx (libro de Excel). Es recomendable que se guarde en la misma carpeta de trabajo “análogos.



5.6. Búsqueda de los mejores descriptores para la generación del modelo:

- a) Abrir el archivo guardado en el paso anterior, y añadir después de la columna “NAME”, otra columna llamada “DL50”, la cual contendrá los valores de dosis letal media (DL50) experimental de cada compuesto en forma logarítmica.

| Name | DL50 | ALogP | ALogp2 | AMR | apol | naAromAtom | nAromBond | nAtom | nHeavyAtom | nH | nC | nN | nCl | nX | ATSOm |
|----------------|------|---------|------------|---------|-----------|------------|-----------|-------|------------|----|----|----|-----|----|--------|
| ALPRAZOLAM.mol | 2.40 | 0.4906 | 0.24068836 | 25.5408 | 45.168309 | 17 | 18 | 35 | 22 | 13 | 17 | 4 | 1 | 1 | 4507.1 |
| ESTAZOLAM.mol | 2.07 | 0.4393 | 0.19298449 | 19.6421 | 42.074723 | 17 | 18 | 32 | 21 | 11 | 16 | 4 | 1 | 1 | 4360.8 |
| MIDAZOLAM.mol | 3.18 | 0.9703 | 0.94148209 | 26.4133 | 46.385309 | 17 | 18 | 36 | 23 | 13 | 18 | 3 | 1 | 2 | 4816.1 |
| TEMAZEPAM.mol | 2.17 | -0.2165 | 0.04687225 | 31.1173 | 42.812309 | 12 | 12 | 34 | 21 | 13 | 16 | 2 | 1 | 1 | 4482.4 |
| TRIAZOLAM.mol | 1.66 | 1.2349 | 1.52497801 | 31.1564 | 46.681516 | 17 | 18 | 35 | 23 | 12 | 17 | 4 | 2 | 2 | 5762.8 |

b) Dar click en la pestaña Datos>Análisis de datos>coeficiente de correlación y dar click en aceptar.

The image shows the 'Datos' ribbon in Excel with the 'Análisis de datos' button circled in red. The 'Análisis de datos' dialog box is open, showing a list of functions for analysis. The 'Coeficiente de correlación' option is selected and circled in red. The 'Aceptar' button is also circled in red.

- c) A continuación, seleccionar todos los datos, a excepción de la primera columna, marcar la casilla “rótulos en la primera columna”, y dar click en aceptar. Dependiendo de la capacidad de procesamiento del equipo, este proceso puede demorar de 1 a 3 minutos.

The screenshot shows an Excel spreadsheet with a table of chemical descriptors. The 'Datos' ribbon is active. A dialog box titled 'Coefficiente de correlación' is open, showing the following settings:

- Entrada: Rango de entrada: SBS1:SAKL56
- Agrupado por: Columnas
- Rótulos en la primera fila
- Opciones de salida: En una hoja nueva

The 'Aceptar' button is highlighted with a red circle. The spreadsheet data is as follows:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|----------------|------|---------|------------|---------|-----------|------------|-----------|-------|------------|----|----|----|-----|----|--------|
| 1 | Name | DL50 | AlogP | Alogp2 | AMR | apol | naAromAtom | nAromBond | nAtom | nHeavyAtom | nH | nC | nN | nCl | nX | AT50m |
| 2 | ALPRAZOLAM.mol | 2.40 | 0.4906 | 0.24068836 | 25.5408 | 45.168309 | 17 | 18 | 35 | 22 | 13 | 17 | 4 | 1 | 1 | 4507.1 |
| 3 | ESTAZOLAM.mol | 2.07 | 0.4393 | 0.19298449 | 19.6421 | 42.074723 | 17 | 18 | 32 | 21 | 11 | 16 | 4 | 1 | 1 | 4360.8 |
| 4 | MIDAZOLAM.mol | 3.18 | 0.9703 | 0.94148209 | 26.4133 | 46.385309 | 17 | 18 | 36 | 23 | 13 | 18 | 3 | 1 | 2 | 4816.3 |
| 5 | TEMAZEPAM.mol | 2.17 | -0.2165 | 0.04687225 | 31.1173 | 42.812309 | 12 | 12 | 34 | 21 | 13 | 16 | 2 | 1 | 1 | 4482.4 |
| 6 | TRIAZOLAM.mol | 1.66 | 1.2349 | 1.52497801 | 31.1564 | 46.681516 | 17 | 18 | 35 | 23 | 12 | 17 | 4 | 2 | 2 | 5762.8 |

- d) El archivo resultante contiene las correlaciones de todos los descriptores con la actividad biológica (DL50). A continuación, añadir una columna adicional al inicio llamada “r2”. En esta columna, se insertará una fórmula para elevar al cuadrado todos los valores de correlación de la DL50 con todos los descriptores. Se deberá arrastrar la fórmula hasta el final para afectar todos los valores.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|-------|------------|-------------|-------------|-------------|-------------|------------|------------|------------|-------------|------------|-------------|------------|------------|------------|------------|-------|
| 1 | r2 | | DL50 | ALogP | ALogp2 | AMR | apol | naAromAtom | nAromBond | nAtom | nHeavyAtom | nH | nC | nN | nCl | nX | ATS |
| 2 | =C2^2 | DL50 | 1 | | | | | | | | | | | | | | |
| 3 | | ALogP | 0.0405631 | 1 | | | | | | | | | | | | | |
| 4 | | ALogp2 | -0.11623262 | 0.90773189 | 1 | | | | | | | | | | | | |
| 5 | | AMR | -0.20417622 | 0.03613795 | 0.40839384 | 1 | | | | | | | | | | | |
| 6 | | apol | 0.23606634 | 0.81408227 | 0.83110243 | 0.45379459 | 1 | | | | | | | | | | |
| 7 | | naAromAtom | 0.12626468 | 0.80235389 | 0.48373027 | -0.51021089 | 0.48552046 | 1 | | | | | | | | | |
| 8 | | nAromBond | 0.12626468 | 0.80235389 | 0.48373027 | -0.51021089 | 0.48552046 | | 1 | | | | | | | | |
| 9 | | nAtom | 0.49457002 | 0.47900356 | 0.54789834 | 0.57819986 | 0.88663725 | 0.14744196 | 0.14744196 | 1 | | | | | | | |
| 10 | | nHeavyAtom | 0.26619725 | 0.88892035 | 0.88784741 | 0.3577743 | 0.98011466 | 0.55901699 | 0.55901699 | 0.82422559 | 1 | | | | | | |
| 11 | | nH | 0.54096669 | -0.18165311 | -0.06363674 | 0.58038174 | 0.4075651 | -0.375 | -0.375 | 0.77407027 | 0.2795085 | 1 | | | | | |
| 12 | | nC | 0.65677944 | 0.7134645 | 0.61529243 | 0.15244778 | 0.87685797 | 0.53452248 | 0.53452248 | 0.86692145 | 0.89642146 | 0.46770717 | 1 | | | | |
| 13 | | nN | -0.31285222 | 0.60854845 | 0.32676877 | -0.4890263 | 0.24962569 | 0.875 | 0.875 | -0.14744196 | 0.2795085 | -0.5625 | 0.13363062 | 1 | | | |
| 14 | | nCl | -0.63348668 | 0.65291645 | 0.83418043 | 0.51480398 | 0.55115195 | 0.25 | 0.25 | 0.22116293 | 0.55901699 | -0.25 | 0.13363062 | 0.375 | 1 | | |
| 15 | | nX | 0.19983519 | 0.849587 | 0.93742251 | 0.3857414 | 0.83522822 | 0.40824829 | 0.40824829 | 0.66212219 | 0.91287093 | 0.10206207 | 0.76376262 | 0.10206207 | 0.61237244 | 1 | |
| 16 | | ATS0m | -0.3820922 | 0.76758193 | 0.94564327 | 0.58179426 | 0.74343628 | 0.29690344 | 0.29690344 | 0.46417231 | 0.75947272 | -0.06207387 | 0.39681672 | 0.28209429 | 0.95585652 | 0.80463681 | |
| 17 | | ATS1m | -0.05306657 | 0.95429386 | 0.96264176 | 0.32015783 | 0.90934186 | 0.63251883 | 0.63251883 | 0.62141338 | 0.93972911 | 0.00300857 | 0.70791583 | 0.48231512 | 0.77690891 | 0.87962509 | 0.893 |
| 18 | | ATS2m | -0.05249977 | 0.91514977 | 0.97480447 | 0.43366245 | 0.92407463 | 0.52881277 | 0.52881277 | 0.67158107 | 0.94635513 | 0.08066391 | 0.70583456 | 0.37220805 | 0.79561885 | 0.90535451 | 0.922 |
| 19 | | ATS3m | -0.10521991 | 0.70555914 | 0.91829722 | 0.72857491 | 0.85438057 | 0.16486719 | 0.16486719 | 0.71421986 | 0.85090791 | 0.25967472 | 0.58604091 | 0.03052352 | 0.80630578 | 0.87772821 | 0.936 |
| 20 | | ATS4m | -0.08654244 | 0.79798518 | 0.95907881 | 0.6284533 | 0.89100261 | 0.30207096 | 0.30207096 | 0.70480923 | 0.90029417 | 0.1885027 | 0.63911616 | 0.15387991 | 0.81744213 | 0.90943369 | 0.949 |
| 21 | | ATS5m | -0.04112841 | 0.78818613 | 0.93981497 | 0.64103206 | 0.92085369 | 0.30344542 | 0.30344542 | 0.75754972 | 0.91799257 | 0.25814115 | 0.67808411 | 0.14597774 | 0.78411457 | 0.89737051 | 0.927 |
| 22 | | ATS6m | -0.13581514 | 0.74625422 | 0.94830944 | 0.67565883 | 0.83561652 | 0.2125362 | 0.2125362 | 0.65842726 | 0.84953813 | 0.16660523 | 0.56858091 | 0.07922506 | 0.83590753 | 0.90021182 | 0.957 |

e) La nueva columna, contiene ahora el coeficiente de determinación (R^2). Este estadístico es importante en el desarrollo de modelos QSAR y aunque en la práctica real, para discriminar entre descriptores útiles se usan más criterios estadísticos, en la presente práctica únicamente consideraremos R^2 para elegir el descriptor con el cual desarrollaremos el modelo. Para seleccionar el descriptor útil, tomaremos en cuenta aquellos descriptores con valores mayores a 0.6 de R^2 . Para ello, podemos agregar un filtro que nos muestre únicamente los descriptores cuyo coeficiente de determinación cumplan con este criterio.

f) Seleccionar la primera columna:

| | A | B | C | D | E |
|----|------------|------------|-------------|-------------|-----------|
| 1 | r2 | | DL50 | ALogP | ALogp2 |
| 2 | 1 | DL50 | 1 | | |
| 3 | 0.00164537 | ALogP | 0.0405631 | 1 | |
| 4 | 0.01351002 | ALogp2 | -0.11623262 | 0.90773189 | |
| 5 | 0.04168793 | AMR | -0.20417622 | 0.03613795 | 0.408393 |
| 6 | 0.05572732 | apol | 0.23606634 | 0.81408227 | 0.831102 |
| 7 | 0.01594277 | naAromAton | 0.12626468 | 0.80235389 | 0.483730 |
| 8 | 0.01594277 | nAromBond | 0.12626468 | 0.80235389 | 0.483730 |
| 9 | 0.2445995 | nAtom | 0.49457002 | 0.47900356 | 0.547898 |
| 10 | 0.07086098 | nHeavyAtom | 0.26619725 | 0.88892035 | 0.887847 |
| 11 | 0.29264496 | nH | 0.54096669 | -0.18165311 | -0.063636 |
| 12 | 0.43135924 | nC | 0.65677944 | 0.7134645 | 0.615292 |
| 13 | 0.09787651 | nN | -0.31285222 | 0.60854845 | 0.326768 |
| 14 | 0.40130537 | nCl | -0.63348668 | 0.65291645 | 0.834180 |
| 15 | 0.0399341 | nX | 0.19983519 | 0.849587 | 0.937422 |
| 16 | 0.14599445 | ATS0m | -0.3820922 | 0.76758193 | 0.945643 |
| 17 | 0.00281606 | ATS1m | -0.05306657 | 0.95429386 | 0.962641 |
| 18 | 0.00275623 | ATS2m | -0.05249977 | 0.91514977 | 0.974804 |
| 19 | 0.01107123 | ATS3m | -0.10521991 | 0.70555914 | 0.918297 |
| 20 | 0.00748959 | ATS4m | -0.08654244 | 0.79798518 | 0.959078 |
| 21 | 0.00169155 | ATS5m | -0.04112841 | 0.78818613 | 0.939814 |
| 22 | 0.01844575 | ATS6m | -0.13581514 | 0.74625422 | 0.948309 |
| 23 | 0.01111656 | ATS7m | -0.1051351 | 0.86614422 | 0.984422 |

g) Seleccionamos la pestaña Datos>Filtro

The screenshot shows the Microsoft Excel interface with the 'Datos' (Data) tab selected in the ribbon. The 'Filtro' (Filter) button is highlighted with a red circle. Below the ribbon, the data table is visible, starting with a header row (row 1) and 22 data rows (rows 2-23). The columns are labeled A through H. The data in the table is as follows:

| | A | B | C | D | E | F | G | H |
|----|------------|------------|-------------|-------------|-------------|-------------|------------|--------|
| 1 | r2 | | DL50 | ALogP | ALogp2 | AMR | apol | naArom |
| 2 | 1 | DL50 | 1 | | | | | |
| 3 | 0.00164537 | ALogP | 0.0405631 | 1 | | | | |
| 4 | 0.01351002 | ALogp2 | -0.11623262 | 0.90773189 | 1 | | | |
| 5 | 0.04168793 | AMR | -0.20417622 | 0.03613795 | 0.40839384 | 1 | | |
| 6 | 0.05572732 | apol | 0.23606634 | 0.81408227 | 0.83110243 | 0.45379459 | 1 | |
| 7 | 0.01594277 | naAromAton | 0.12626468 | 0.80235389 | 0.48373027 | -0.51021089 | 0.48552046 | |
| 8 | 0.01594277 | nAromBond | 0.12626468 | 0.80235389 | 0.48373027 | -0.51021089 | 0.48552046 | |
| 9 | 0.2445995 | nAtom | 0.49457002 | 0.47900356 | 0.54789834 | 0.57819986 | 0.88663725 | 0.1474 |
| 10 | 0.07086098 | nHeavyAtom | 0.26619725 | 0.88892035 | 0.88784741 | 0.3577743 | 0.98011466 | 0.5590 |
| 11 | 0.29264496 | nH | 0.54096669 | -0.18165311 | -0.06363674 | 0.58038174 | 0.4075651 | - |
| 12 | 0.43135924 | nC | 0.65677944 | 0.7134645 | 0.61529243 | 0.15244778 | 0.87685797 | 0.5345 |
| 13 | 0.09787651 | nN | -0.31285222 | 0.60854845 | 0.32676877 | -0.4890263 | 0.24962569 | |
| 14 | 0.40130537 | nCl | -0.63348668 | 0.65291645 | 0.83418043 | 0.51480398 | 0.55115195 | |
| 15 | 0.0399341 | nX | 0.19983519 | 0.849587 | 0.93742251 | 0.3857414 | 0.83522822 | 0.4082 |
| 16 | 0.14599445 | ATS0m | -0.3820922 | 0.76758193 | 0.94564327 | 0.58179426 | 0.74343628 | 0.2969 |
| 17 | 0.00281606 | ATS1m | -0.05306657 | 0.95429386 | 0.96264176 | 0.32015783 | 0.90934186 | 0.6325 |
| 18 | 0.00275623 | ATS2m | -0.05249977 | 0.91514977 | 0.97480447 | 0.43366245 | 0.92407463 | 0.5288 |
| 19 | 0.01107123 | ATS3m | -0.10521991 | 0.70555914 | 0.91829722 | 0.72857491 | 0.85438057 | 0.1648 |
| 20 | 0.00748959 | ATS4m | -0.08654244 | 0.79798518 | 0.95907881 | 0.6284533 | 0.89100261 | 0.3020 |
| 21 | 0.00169155 | ATS5m | -0.04112841 | 0.78818613 | 0.93981497 | 0.64103206 | 0.92085369 | 0.3034 |
| 22 | 0.01844575 | ATS6m | -0.13581514 | 0.74625422 | 0.94830944 | 0.67565883 | 0.83561652 | 0.212 |
| 23 | 0.01111555 | ATS7m | -0.1054251 | 0.86614432 | 0.98443255 | 0.5264087 | 0.88904101 | 0.4146 |

h) Seleccionamos la fecha del filtro, y damos click en Filtros de número>Mayor que

The screenshot shows the Microsoft Excel interface. The 'Datos' ribbon is active, and the 'Filtro' button is highlighted. A dropdown menu is open for cell A1, showing 'Filtros de número' selected. A sub-menu is open for 'Filtros de número', with 'Mayor que...' selected. The spreadsheet data is visible below.

| | A | B | C | D | E | F | G |
|----|------------|-------|-------------|------------|------------|-------------|------------|
| 1 | r2 | | DL50 | ALogP | ALogp2 | AMR | apol |
| | | | | 1 | | | |
| | | | | 0.90773189 | 1 | | |
| | | | | 0.03613795 | 0.40839384 | 1 | |
| | | | | 0.81408227 | 0.83110243 | 0.45379459 | 1 |
| | | | | 0.80235389 | 0.48373027 | -0.51021089 | 0.48552046 |
| | | | | | | .51021089 | 0.48552046 |
| | | | | | | .57819986 | 0.88663725 |
| | | | | | | 0.3577743 | 0.98011466 |
| | | | | | | .58038174 | 0.4075651 |
| | | | | | | .15244778 | 0.87685797 |
| | | | | | | 0.4890263 | 0.24962569 |
| | | | | | | .51480398 | 0.55115195 |
| | | | | | | 0.3857414 | 0.83522822 |
| | | | | | | .58179426 | 0.74343628 |
| | | | | | | .32015783 | 0.90934186 |
| | | | | | | .43366245 | 0.92407463 |
| | | | | | | .72857491 | 0.85438057 |
| | | | | | | 0.6284533 | 0.89100261 |
| | | | | | | .64103206 | 0.92085369 |
| 22 | 0.01844575 | ATS6m | -0.13581514 | 0.74625422 | 0.94830944 | 0.67565883 | 0.83561652 |
| 23 | 0.01111555 | ATS7m | 0.1054251 | 0.85514433 | 0.98443255 | 0.5254087 | 0.88041101 |

i) Escribimos 0.6 y damos click en aceptar.

Autofiltro personalizado ? X

Mostrar las filas en las cuales:

r2

es mayor que 0.6

Y O

Use ? para representar cualquier carácter individual
Use * para representar cualquier serie de caracteres

Aceptar Cancelar

Autoguardado

Archivo Inicio Insertar

Pegar

Portapapeles

Calibri

N K S

Fuente

C239

| | A | B |
|-----|------------|---------|
| 1 | r2 | |
| 2 | | 1 DL50 |
| 62 | 0.60378124 | ATS1s |
| 69 | 0.6715662 | ATS8s |
| 106 | 0.84560632 | AATS0i |
| 110 | 0.74565676 | AATS4i |
| 113 | 0.73592372 | AATS7i |
| 123 | 0.75160934 | AATS8s |
| 164 | 0.63329006 | ATSC4p |
| 169 | 0.70423123 | ATSC0i |
| 170 | 0.78539034 | ATSC1i |
| 176 | 0.71883575 | ATSC7i |
| 193 | 0.69142597 | AATSC6c |
| 227 | 0.73832771 | AATSC4p |
| 232 | 0.69971246 | AATSC0i |
| 233 | 0.77744702 | AATSC1i |
| 239 | 0.63454178 | AATSC7i |
| 285 | 0.80037351 | MATS4p |
| 313 | 0.71801408 | GATS8c |
| 338 | 0.60841904 | GATS1p |
| 352 | 0.71906792 | GATS7i |
| 354 | 0.63212718 | GATS1s |
| 358 | 0.64407418 | GATS5c |

j) A continuación, se mostrarán aquellos descriptores que cumplen con el criterio seleccionado. En este ejemplo, consideraremos el descriptor AATSC1i para averiguar si generando una ecuación con él, es posible predecir la dosis letal media del Diazepam. Para ello, ¿? No termina la oración.

| | A | B |
|-----|------------|---------|
| 1 | r2 | |
| 2 | | 1 DL50 |
| 62 | 0.60378124 | ATS1s |
| 69 | 0.6715662 | ATS8s |
| 106 | 0.84560632 | AATS0i |
| 110 | 0.74565676 | AATS4i |
| 113 | 0.73592372 | AATS7i |
| 123 | 0.75160934 | AATS8s |
| 164 | 0.63329006 | ATSC4p |
| 169 | 0.70423123 | ATSC0i |
| 170 | 0.78539034 | ATSC1i |
| 176 | 0.71883575 | ATSC7i |
| 193 | 0.69142597 | AATSC6c |
| 227 | 0.73832771 | AATSC4p |
| 232 | 0.69971246 | AATSC0i |
| 233 | 0.77744702 | AATSC1i |
| 239 | 0.63454178 | AATSC7i |
| 285 | 0.80037351 | MATS4p |
| 313 | 0.71801408 | GATS8c |
| 338 | 0.60841904 | GATS1p |
| 352 | 0.71906792 | GATS7i |
| 354 | 0.63212718 | GATS1s |

5.7. Generación del modelo de predicción:

- a) Una vez que se tiene el nombre del descriptor que probaremos para generar el modelo, abrimos STATGRAPHICS, y copiaremos la columna que contenga el descriptor anteriormente seleccionado, así como la columna de la respuesta biológica (DL50). Es importante que el documento de statgraphics contenga el nombre de cada columna.

The image shows two overlapping Excel spreadsheets. The top-left spreadsheet is titled 'descriptores.xlsx' and displays a table with columns HX, HY, HZ, and IA. The HY column is highlighted with a green border. The top-right spreadsheet is titled 'DL50' and displays a table with columns DL50, ALogP, ALogP2, and AMR. The DL50 column is highlighted with a green border.

| | HX | HY | HZ | IA |
|-----|------------|-------------|-------------|------------|
| | AATSC0i | AATSC1i | AATSC2i | AATSC3i |
| 121 | 1.67107131 | -0.14257547 | -0.09853988 | 0.26715457 |
| 263 | 1.70650953 | -0.14856826 | 0.05374339 | 0.092959 |
| 482 | 2.21894717 | -0.31126261 | -0.1832795 | 0.0643966 |
| 645 | 1.50663379 | -0.18799446 | 0.13506103 | 0.03129559 |
| 176 | 1.64438376 | -0.13846106 | -0.08163005 | 0.28737871 |

| | DL50 | ALogP | ALogP2 | AMR |
|----------|------|---------|------------|------|
| OLAM.mol | 2.40 | 0.4906 | 0.24068836 | 25.5 |
| LAM.mol | 2.07 | 0.4393 | 0.19298449 | 19.6 |
| ILAM.mol | 3.18 | 0.9703 | 0.94148209 | 26.4 |
| PAM.mol | 2.17 | -0.2165 | 0.04687225 | 31.1 |
| AM.mol | 1.66 | 1.2349 | 1.52497801 | 31.1 |

STATGRAPHICS Centurion - Untitled StatFolio

File Edit Define Measure Analyze Improve Control Forecast SnapStats!! Tools View Window Help

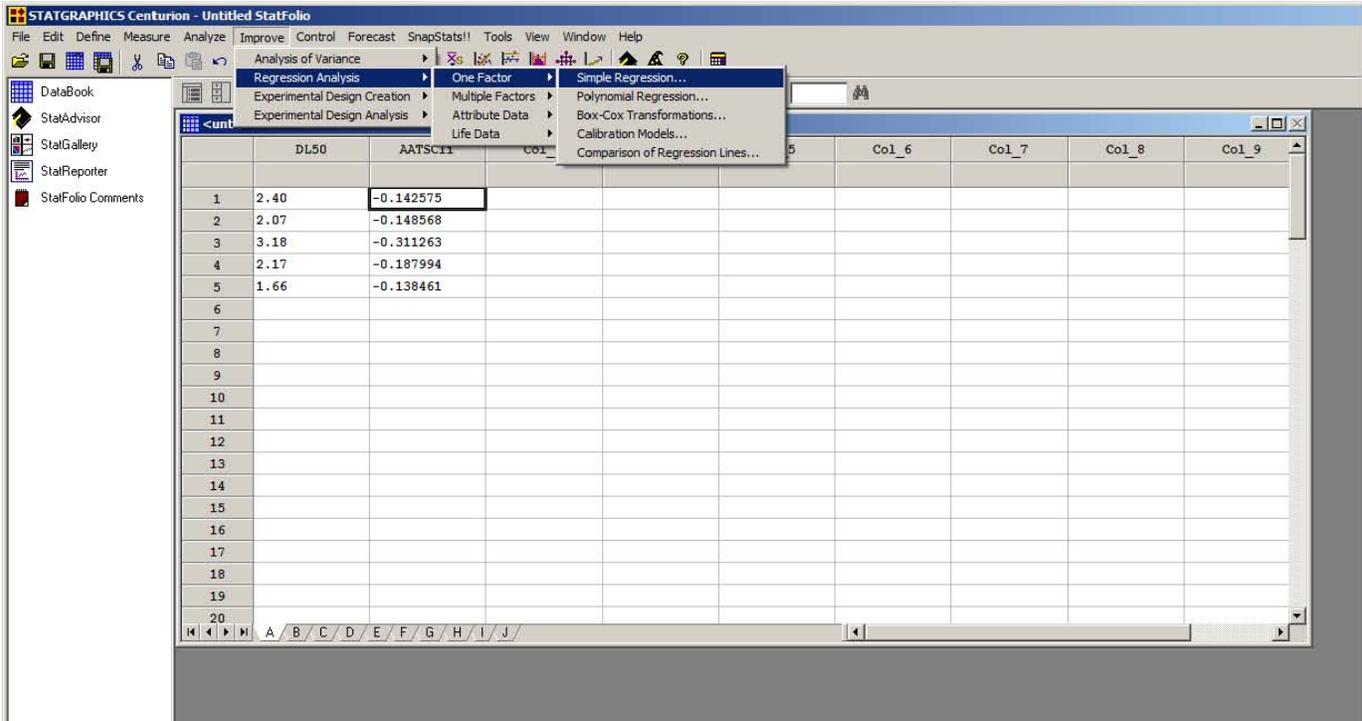
DataBook
StatAdvisor
StatGallery
StatReporter
StatFolio Comments

<untitled>

| | DL50 | AATSC1i | Col_3 | Col_4 |
|----|------|-----------|-------|-------|
| 1 | 2.40 | -0.142575 | | |
| 2 | 2.07 | -0.148568 | | |
| 3 | 3.18 | -0.311263 | | |
| 4 | 2.17 | -0.187994 | | |
| 5 | 1.66 | -0.138461 | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |
| 18 | | | | |
| 19 | | | | |
| 20 | | | | |

A B C D E F G H I J

b) Damos click en la pestaña Improve>Regression Analysis>One factor>Simple Regression



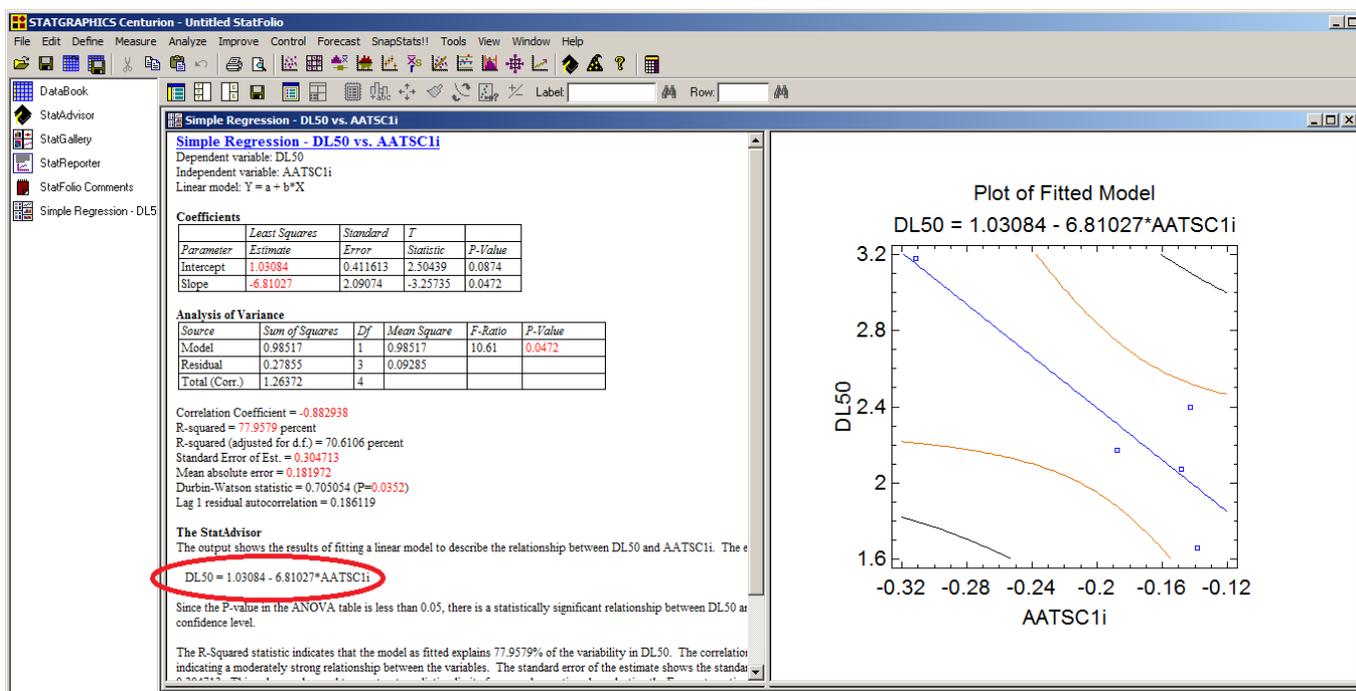
c) En la variable Y, introducimos “DL50”, y en la variable X, el descriptor seleccionado.

Damos click en OK.

The screenshot shows the STATGRAPHICS Centurion interface. The main window displays a data table with columns labeled DL50, AATSC1i, Col_3, Col_4, Col_5, Col_6, Col_7, and Col_8. The first five rows of data are visible, with the AATSC1i column values highlighted in a light blue box. A 'Simple Regression' dialog box is open in the foreground, showing a list of variables on the left with 'AATSC1i' and 'DL50' selected. The 'Y:' field contains 'DL50' and the 'X:' field contains 'AATSC1i'. The '(Select):' field is empty. The 'Sort column names' checkbox is checked. The dialog box has buttons for 'OK', 'Cancel', 'Delete', 'Transform...', and 'Help'.

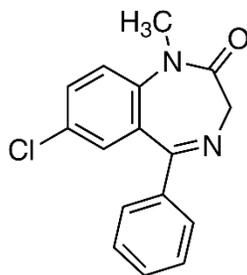
| | DL50 | AATSC1i | Col_3 | Col_4 | Col_5 | Col_6 | Col_7 | Col_8 |
|----|------|-----------|-------|-------|-------|-------|-------|-------|
| 1 | 2.40 | -0.142575 | | | | | | |
| 2 | 2.07 | -0.148068 | | | | | | |
| 3 | 3.18 | -0.311263 | | | | | | |
| 4 | 2.17 | -0.187994 | | | | | | |
| 5 | 1.66 | -0.138461 | | | | | | |
| 6 | | | | | | | | |
| 7 | | | | | | | | |
| 8 | | | | | | | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |
| 11 | | | | | | | | |
| 12 | | | | | | | | |
| 13 | | | | | | | | |
| 14 | | | | | | | | |
| 15 | | | | | | | | |
| 16 | | | | | | | | |
| 17 | | | | | | | | |
| 18 | | | | | | | | |
| 19 | | | | | | | | |
| 20 | | | | | | | | |

d) Entre otras cosas, el análisis de regresión mostrará la ecuación generada.



5.8. Predicción de la DL50 del Diazepam

a) La ecuación generada con statgraphics será utilizada para la predicción de la DL50 del Diazepam. Para ello, es preciso que calculemos el valor del descriptor para este compuesto. Para ello, es preciso modelar la estructura de este compuesto en ChemSketch. Es recomendable guardar el fichero .mol en nueva carpeta llamada “diazepam”.



b) Calculamos los descriptores del diazepam del mismo modo que se hizo para los análogos. Abrimos el archivo de Excel y con la herramienta buscar y seleccionar, localizamos el descriptor que necesitamos escribiendo el nombre de este en el formulario de búsqueda.

The screenshot shows the Excel interface with the 'Inicio' tab selected. The 'Buscar y reemplazar' dialog box is open, and the search term 'AATSC1i' is entered in the 'Buscar:' field. The search results in the spreadsheet are as follows:

| | HW | HX | HY | HZ | IA | IB | IC | ID | IE | IF | IG | IH | II | IJ | IK | IL | IM |
|---|-------------|------------|-------------|-------------|-------------|----------|-------------|------------|------------|-------------|------------|-------------|-----------|-------------|-------------|-------------|---------|
| 1 | AATSC1p | AATSC2p | AATSC3p | AATSC4p | AATSC5p | AATSC6p | AATSC7p | AATSC8p | AATSC0i | AATSC1i | AATSC2i | AATSC3i | AATSC4i | AATSC5i | AATSC6i | AATSC7i | AATSC8i |
| 2 | -0.01043341 | 0.02771421 | -0.01175533 | -0.02899763 | -0.01642844 | 0.020452 | -0.08400425 | 0.00662133 | 1.51571953 | -0.26046756 | 0.14536568 | -0.00407043 | -0.198979 | -0.04330122 | -0.08200493 | -0.13433218 | 0.01 |

- c) Dentro del archivo del mismo archivo de Excel, es posible introducir la ecuación de predicción generada con statgraphics, sustituyendo la variable por el valor numérico del descriptor seleccionado.

The screenshot shows an Excel spreadsheet with the following data table:

| | HW | HX | HY | HZ | IA | IB | IC | ID | IE | IF | IG | IH | II | IJ | IK | IL | IM |
|----|-------------|------------|-------------|-------------|-------------|----------|-------------|------------|------------|-------------|------------|-------------|-----------|-------------|-------------|-------------|----------|
| 1 | AAATSC1p | AAATSC2p | AAATSC3p | AAATSC4p | AAATSC5p | AAATSC6p | AAATSC7p | AAATSC8p | AAATSC0i | AAATSC1i | AAATSC2i | AAATSC3i | AAATSC4i | AAATSC5i | AAATSC6i | AAATSC7i | AAATSC8i |
| 2 | -0.01043341 | 0.02771421 | -0.01175533 | -0.02899763 | -0.01642844 | 0.020452 | -0.08400425 | 0.00662133 | 1.51571953 | -0.26046756 | 0.14536568 | -0.00407043 | -0.198979 | -0.04330122 | -0.08200493 | -0.13433218 | 0.01 |
| 3 | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | | | |
| 22 | | | | | | | | | | | | | | | | | |

The formula bar shows the equation: $=1.03084 - 6.81027*IF2$. In cell I10, the equation $=1.03084 - 6.81027*IF2$ is entered, and in cell I7, the equation $DL50=1.03084 - 6.81027*AAATSC1i$ is entered.

d) En este ejemplo, el valor de predicción de DL50 es de 2.80. Como este es un valor logarítmico, es preciso realizar su conversión a mg/kg. Para ello, es preciso utilizar el peso molecular y una función antilogarítmica. En Excel, es posible hacerlo mediante la fórmula: =POTENCIA(10,-1*M), donde M es el valor logarítmico de la DL50 obtenido en la ecuación de predicción.

El valor obtenido en la fórmula anterior, debe multiplicarse por 1000 y por el peso molecular en g/mol, para obtener la DL50 en mg/kg. El peso molecular del Diazepam es de 284.7 g/mol.

| | | |
|-----------------|---------------------------------|--|
| | DL50= 1.03084 - 6.81027*AATSC1i | |
| DL50 | 2.80469444 | |
| peso molecular | 284.7 | |
| Antilogatirno | 0.00156785 | |
| Antilog*PM*1000 | 446.367977 | |

$$\text{Índice terapéutico} = \frac{\text{Dosis tóxica (Dosis letal)}}{\text{Dosis terapéutica}}$$

$$I.T = \frac{31.2}{30}$$

$$I.T = 1.04$$

$$\text{Margen de seguridad} = \frac{\text{Dosis letal} - \text{Dosis efectiva}}{\text{Dosis letal}} * 100$$

$$M.S = \frac{31.2 - 30}{30} * 100$$

$$M.S = 4$$

6. Discusión

Es preciso considerar que la práctica ha sido pensada como una introducción superficial sobre el tema de QSAR, y que no profundiza en aspectos que son relevantes, aunque son mencionados en el presente trabajo. Considerando el tiempo disponible para ejecutar el trabajo experimental, los estadísticos inherentes a la validación interna y externa no serán calculados, y no se realizará una validación externa con más de un compuesto. Sin embargo, a pesar de estas limitaciones, la práctica puede ampliarse según la pertinencia asociada al tiempo experimental disponible y al tiempo que se tenga para proveer al alumno de conceptos teóricos.

Las variaciones que pueden introducirse son las siguientes:

- Se pueden utilizar compuestos de otro grupo farmacológico, siempre y cuando la similitud estructural sea asegurada.
- Teniendo en cuenta que se han propuesto 12 análogos, es posible utilizar otro grupo diferente al sugerido previamente. No se recomienda utilizar más de 6, en virtud de que para mantener una relación 5:1 entre compuestos y descriptores, el modelo debería incluir dos variables.
- Se pueden usar otros programas para calcular los descriptores, pero debe considerarse que algunos requieren de algún modelado particular de la estructura, la creación de una base de datos de compuestos antes de enviar el cálculo o algún conocimiento en particular sobre lenguajes de programación. Una alternativa gráfica, gratuita y de fácil manejo es cdk, el cual posee una interfaz gráfica en Java, y puede descargarse desde el siguiente enlace: <http://www.rguha.net/code/java/cdkdesc.html>
- Para optimizar el tiempo de práctica, es posible suprimir el modelado de las estructuras, y proveer a los alumnos de los archivos .mol para proceder al cálculo inmediato de descriptores. Se sugiere, que de ser posible, no se elimine este paso, y así dar a conocer una alternativa para el modelado de estructuras químicas, que puede servir de forma concomitante con otras asignaturas como química orgánica
- En la medida de lo posible se puede introducir el cálculo de otros estadísticos importantes, como la Q^2 , o introducir una validación cruzada por leave one out o Y-scrambling.

7. Conclusiones

- Se desarrolló una metodología para una práctica de laboratorio que ilustra los conceptos fundamentales de las relaciones estructura química-actividad biológica a través de una metodología simplificada.
- Replicando la metodología sugerida, se comprobó que la práctica puede ser reproducible, ampliable y perfectible.
- La práctica fue satisfactoriamente adaptada a los conceptos de margen de seguridad de índice terapéutico, mismos que se abordan en el temario de la asignatura.
- El procedimiento consiguió demostrar la predicción de dos parámetros de seguridad para el diazepam.
- Se ilustró el uso de software para cálculo de descriptores y modelado de estructuras químicas.

8 BIBLIOGRAFÍA

- Benfenati E (Ed): Theory, guidance and applications on QSAR and REACH. E-book recuperado de <http://home.deib.polimi.it/gini/papers/orchestra.pdf>. Fecha de consulta: 2 de Enero de 2018
- Crum-Brown, A. and Fraser, T.R. (1868–1869) Trans. R. Soc. Edinburgh, 25, 151.
- Asadollahi, Tahereh. Shayessteh Dadfarnia , Ali Mohammad Haji Shabani , Jahan B. Ghasemi and Maryam Sarkhosh QSAR models for CXCR2 receptor antagonists based on the genetic algorithm for data preprocessing prior to application of the PLS linear regression method and design of the new compounds using in silico virtual screening. *Molecules*. 2011 Feb 25;16(3):1928-55. doi: 10.3390/molecules16031928.
- Devillers J., Genetic Algorithms in Molecular Modeling (Principles of QSAR and Drug Design), Elsevier Science Ltd, Amsterdam, The Netherlands, 1996, 327 pp.
- Devillers, J.,(1996). Neural Networks in QSAR and Drug Design (Principles of QSAR and DrugDesign), Academic Press, London, UK, 284 pp.
- George J, Elias S. Thalidomide and the Titanic: Reconstructing the technology tragedies of the Twentieth Century. *Am J Public Health* 1999; 89(1): 98-101.
- Golbraikh, A.; Tropsha, A. *J Mol Graph Model*. 2002, 20, 269-276.
- Gramatica Paola. (2004) EVALUATION OF DIFFERENT STATISTICAL APPROACHES FOR THE VALIDATION OF QUANTITATIVE STRUCTURE – ACTIVITY RELATIONSHIPS JRC Contract ECVA-CCR.496576-Z.
- Gramatica, Paola. (2014). *A short history of QSAR evolucion* 4-5, recuperado de https://www.researchgate.net/publication/252172555_A_SHORT_HISTORY_OF_QSAR_EVOLUCION Fecha de consulta: 2 de Enero de 2018
- Gramatica, Paola (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701 recuperado de <http://onlinelibrary.wiley.com/doi/10.1002/qsar.200610151/full> Fecha de consulta: 2 de Enero de 2018

- Holliday et al (Holliday, JD., Hu, C-Y. and Willett, P. (2002) *Combinatorial Chemistry and High Throughput Screening* 5, 155-166)
- Leslie H. *Talidomida: Polémica y potencial*. SF AIDS Fdn 1998; 5(98): 1-6.
- Mendoza Patiño, Nicandro.(2008) *Farmacología Médica*, México: editorial Médica Panamericana: UNAM, Facultad de Medicina,
- Powell RJ. (1996) New Roles for thalidomide: A unique anti-inflammatory, but use it only when no alternatives exist. *BMJ* 313(7054): 377-378.
- Roy, K., Kar, S., Das, R. N. (2015). *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. Springer.
- Rouvray, D.H. (1995) Similarity in chemistry: past, present and future. In *Topics in Current Chemistry*, Vol 173. Sen, K. (Ed.) Springer-Verlag: Berlin, pp 1-30
- Todeschini, Roberto. Consonni, Viviana. (2009) *Molecular Descriptors for Chemoinformatics* (2 volumenos), Wiley-VCH,
- Roy, Partha Pratim & Joseph, Thomas & Roy, Kunal. (2008). Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*. 90. 31-42. 10.1016/j.chemolab.2007.07.004. Recuperado de http://www.ifm.liu.se/compchem/msi/doc/life/cerius46/qsar/theory_methods.html
Fecha de consulta: 2 de Enero de 2018
- Veerasamy, Ravichandran & Rajak, Harish & Jain, Avijeet & Sivadasan, Shalini & Christapher, Parayil Varghese & Agrawal, R.K.. (2011). Validation of QSAR Models - Strategies and Importance. *Int J Drug Design and Discov*. 2. 511-519.
- Acuña, (2014). Estudio QSAR para metolacloro y una impureza relacionada Reporte 09030 *Lebenchem México*. pp 1-45
- TEST SOFTWARE (2014) Toxicity Estimation software tool. EPA. USA. Ver 4.2.1.