



**BENEMÉRITA UNIVERSIDAD AUTÓNOMA
DE PUEBLA**

Facultad de Ciencias Químicas

Posgrado en Ciencias Químicas

Área de Bioquímica y Biología Molecular

Maestría en Ciencias Químicas

Título de tesis:

**“Modelo de relación estructura-actividad cuantitativa
para una serie de ligandos sintéticos del receptor de
tropomiosina quinasa A humano asociados a la
supresión del dolor”**

Puebla, Pue. Julio 2020

**Tesis presentada para obtener el grado de maestría en
Ciencias Químicas**

Presenta:

L. F. Kenia Salvador Atonal

Director de tesis:

Dr. Thomas Scior

Resumen

El diseño y desarrollo de nuevos fármacos contra el dolor incluye la identificación de nuevas dianas moleculares dentro de las vías de señalización y percepción del dolor en el sistema nervioso. Uno de los descubrimientos recientes es que el receptor de tropomiosina quinasa A (TrkA) está asociado con el dolor nociceptivo y neuropático al unirse con su ligando natural el factor de crecimiento neuronal (NGF), desencadena la vía de señalización NFG/TrkA y promueve la síntesis de proteínas como el receptor de potencial transitorio V1 (TRPV1), péptidos como la sustancia P y el péptido relacionado con el gen de la calcitonina (CGRP) y sus funciones son la transmisión y modulación del dolor.

En 2017 *Hurzy y colaboradores*, diseñaron y sintetizaron cuarenta nuevas moléculas análogas del tipo –N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il) etanamida, que llamamos "N-PTindoles" y actualmente se sujetan a pruebas experimentales para comprobar si alivian el dolor. Son moléculas que podrían ser posibles candidatos de fármacos para aliviar el dolor, ya que su sitio activo se ubica en la región juxtamembrana y parte del dominio quinasa del receptor TrkA.

En el Laboratorio de Simulaciones moleculares computacionales se estudia la relación cuantitativa entre los descriptores moleculares de los cuarenta análogos inhibidores de TrkA o N-PTindoles y su actividad biológica utilizando la metodología de relación estructura-actividad cuantitativa (**QSAR** por sus siglas en inglés *Quantitative Structure-Activity Relationship*) para predecir la relación de descriptores moleculares cuantitativos (**Qsar**) de 40 inhibidores análogos de N-PTindoles (**qsar**) con su actividad biológica del receptor TrkA (**qsAr**), por medio de una regresión lineal múltiple (MLR) (**qsaR**).

Índice

ABREVIATURAS Y GLOSARIO	5
1. INTRODUCCIÓN	8
1.1 Dolor	8
1.2 Receptor tropomiosina quinasa (Trks)	12
1.3 Receptor tropomiosina quinasa A (TrkA)	12
1.4 Inhibidores relacionados con la vía de señalización NFG/TrkA	14
1.5 Modelo relación estructura actividad cuantitativa (QSAR)	16
1.5.1 Descriptores moleculares	17
1.5.2 Regresión lineal múltiple en QSAR	18
2. ANTECEDENTES	28
3. HIPÓTESIS	30
4. OBJETIVO GENERAL	31
5. OBJETIVOS PARTICULARES	31
6. METODOLOGÍA	32
6.1 Búsqueda del PDB.	32
6.2 Construcción de moléculas.	32
6.3 Cálculo de los descriptores	35
6.4 El cálculo de la regresión lineal	36
7. RESULTADOS Y DISCUSIÓN DE RESULTADOS	37
7.1 Búsqueda de PDB y Construcción de moléculas en 3 dimensiones (3D)	38
7.2 Construcción de moléculas 2 Dimensiones (2D)	45
7.3 Calculo de descriptores	62
7.4 Selección de descriptores	65
7.5 Regresión lineal múltiple	77
7.6 Generación de nuevas moléculas	96

8. CONCLUSIÓN	100
9. REFERENCIAS	102
10. SUPLEMENTO	107

ABREVIATURAS Y GLOSARIO

TrkA: Receptor de tropomiosina quinasa A.

NGF: Factor de crecimiento neuronal.

TRPV1: Receptor de potencial transitorio V1.

CGRP: Péptido relacionado con el gen de la calcitonina.

N-PTindoles: Moléculas análogas del tipo $\text{-N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il)etanamida}$.

QSAR: Relación estructura-actividad cuantitativa o *Quantitative Structure-Activity Relationship* en inglés.

SAR: Relación estructura-actividad o *Structure-Activity Relationship* en inglés.

MLR: Regresión lineal múltiple.

TNF- α : Factor de necrosis tumoral α .

CB1: Canabinoide tipo 1.

CB2: Canabinoide tipo 2.

ASIC: Canales iónicos sensibles al ácido.

Trks: Receptores de tropomiosina quinasa o receptores quinasa relacionados con la tropomiosina.

TrkB: Receptor de tropomiosina quinasa B.

TrkC: Receptor de tropomiosina quinasa C.

BDNF: Factor neurotrófico derivado del cerebro.

NT-4/5: Neurotrofina 4/5.

NT-3: Neurotoxina 3.

DEG: Triada de aminoácidos ácido aspártico, ácido glutámico y glicina.

IC50: Concentración inhibitoria 50.

DL50: Dosis letal 50.

DV: Actividad biológica representada como IC50, es la variable dependiente.

IDV: Descriptores moleculares, variables independientes.

β_0 : Es el intercepto.

$\beta_1, 2, 3 \dots n$: Pendientes.

e_i : Residuo o error.

RSS: Suma residual de cuadrados o *Residual Sum of Squares*.

R^2 : Coeficiente de determinación.

2D: Bidimensional.

3D: Tridimensional.

PDB: Banco de datos de proteínas cristalizadas o *Protein Data Bank* en inglés.

BS: Sitio de unión o *binding site* en inglés es el lugar del acoplamiento de ligandos

SLR: Regresión simple lineal.

K_i : Constante de inhibición.

M0: Modelo nulo.

C: Carbono.

H: Hidrógeno.

DG o LBE: Energía libre de interacción.

ESOL: Estimación de la solubilidad acuosa directamente de la estructura molecular.

Log P: Coeficiente de partición lípido/agua.

Log SW: Logaritmo negativo de la solubilidad en agua del compuesto.

FractionCsp3: Relación de los carbonos hibridados sp3 sobre el recuento total de carbono de la molécula.

LogKp: Permeación de la piel.

Se: Suma total de las electronegatividades de la estructura.

Ss: Suma de estados electrológicos.

Sv: Suma total de volúmenes atómicos de Van Der Waals.

POL: Número de polaridad.

IAC: Información total de la composición atómica.

UNIP: unipolaridad.

IDM: Información de la magnitud de la distancia.

HOMT: Índice de aromaticidad.

HBD o H-Don: Número de átomos donadores de hidrógeno unidos a nitrógeno y oxígeno.

Log PC: Logaritmo negativo de la distribución del componente en octano/agua utilizando el método de radio aromático de Ghose/Crippen.

MW: Masa molecular.

MR: Refractividad molar.

PSA: Área de superficie polar.

VOL: Volumen.

TPSA: Área de superficie polar total.

1. INTRODUCCIÓN

1.1 Dolor

El dolor según la Asociación Internacional para el Estudio del Dolor se define como “una experiencia sensorial y emocional desagradable que se relaciona con un daño tisular”. [1-3]

El dolor se clasifica por origen, duración e intensidad véase **Figura 1**.

Por el origen es nociceptivo y neuropático: El dolor nociceptivo o fisiológico es causado por una lesión tisular donde ocurre una activación de los nociceptores o receptores del dolor que emiten señales hacia el sistema nervioso central. Los nociceptores pueden ser mecanorreceptores, quimiorreceptores o termorreceptores. [4]

El dolor neuropático es causado por un daño estructural o una enfermedad del sistema nervioso central o periférico y se denomina dolor patológico. [4]

Tanto el dolor nociceptivo como el neuropático activan diferentes vías de señalización ya que al momento de ocurrir un daño de origen fisiológico u origen patológico ocasiona una liberación de sustancias endógenas que pueden segregarse desde las células inflamatorias en el tejido lesionado o por células de Schwann en los nervios dañados como NGF (factor de crecimiento neuronal), las interleucinas y TNF- α (factor de necrosis tumoral- α). [4]

Por su intensidad es leve, moderado o severo. Para conocer el tipo de intensidad de dolor de un paciente se establece una medición del dolor desde el punto de vista clínico, con información verbal o escrita dada por el mismo y basada en la escala analógica visual. [3]

El dolor según su duración puede ser agudo y crónico. El dolor agudo tiene una duración corta y se expresa como respuesta a un trauma tisular que se disminuye progresivamente hasta desaparecer.

El crónico se caracteriza por la persistencia de estímulos nocivos de una enfermedad o exacerbación de una lesión, este puede durar meses e incluso años y está asociado con el dolor nociceptivo y neuropático. [1, 2, 4]



Figura 1: Tipos de dolor: Esquema que representa las tres clasificaciones del dolor por origen, duración e intensidad.

Aproximadamente el 30% de la población tiene dolor crónico generado por diferentes enfermedades como: artritis, cáncer, drepanocitosis y dolor lumbar crónico véase **Tabla 1**. [1, 4-6]

Los analgésicos utilizados para el dolor neuropático según las guías de práctica clínica son medicamentos antiinflamatorios no esteroideos, gabapentinoides, antidepresivos tricíclicos y opioides. Sin embargo, se ha reportado un alivio inadecuado del dolor e intolerancia de los efectos secundarios en los pacientes tratados con estos medicamentos. [2, 7, 8]

Tabla 1: Enfermedades asociadas al dolor crónico.

Enfermedad	Causa del dolor crónico
Artritis	Inflamación de las articulaciones que provoca sensibilización periférica y central.
Cáncer	Crecimiento tumoral y los procedimientos diagnósticos o terapéuticos de los paciente, como efectos secundarios generados por las quimioterapias.
Drepanocitosis	Asociada a las afecciones producidas por la falciformación de los eritrocitos, que promueven la necrosis a vascular por la dificultad del flujo sanguíneo y mala oxigenación, provocando dolor crónico en los miembros y articulaciones.
Dolor lumbar crónico	Degeneración de los discos intervertebrales y las articulaciones posteriores: asociado a dolor ciático.

En investigaciones recientes reportan nuevos blancos moleculares para aliviar el dolor como el receptor Canabinoide tipo 1 y 2 (CB1, CB2), bloqueadores de los canales de sodio dependientes de voltaje, canales iónicos sensibles al ácido (ASIC) y tropomiosina quinasa A (TrkA). [5, 9-11]. En la **Tabla 2** se muestra un panorama general de los analgésicos que se encuentran en el mercado y el receptor en el que actúan, se incluye una lista de los nuevos receptores mencionados anteriormente y además se hace hincapié en si existen estudios de relación estructura actividad cuantitativa (QSAR) en cada uno de ellos. El color verde representa los fármacos que cuentan con estudios tipos QSAR [10-13], el rojo los que no tienen estudios tipo QSAR y en naranja las moléculas análogas $-N$ -(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il) etanamida inhibidoras del receptor TrkA [5] las cuales se utilizarán en este trabajo y tendrán un modelo QSAR.

Tabla 2: Fármaco-receptor y relación estructura actividad cuantitativa.

FÁRMACOS	Opioides	Antiinflamatorios no esteroideos AINES	Corticoesteroide	Antidepresivo tricíclico	Canabinoides	Bloqueadores de los canales de sodio dependientes de voltaje	Bloqueadores de canales iónicos sensibles al ácido ASIC	Inhibidores de TrkA
	Morfina Oxicodona Tramadol Hidrocodona Hidromorfona Fentanilo Tapentadol Metadona	Ácido Acetil Salicílico Diflunisal Acetaminofeno Ibuorofeno	Dexametasona	Desipramina Nortriptilina	Canabinoides	Carbamazepina Oxcarbazepina Acetato de Eslicarbazepina	Amilorida	-N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il)etanamida N-PTíndoles
RECEPTOR	Opioide μ y kappa	Ciclooxigenasa 1 y 2 (Cox 1 y 2)	Glucocorticoide (GRC)	Ácido gamma aminobutírico GABA	Canabinoide tipo 1 y 2 CB1 Y CB 2	Canales de Na ⁺ NAV1.3, 1.9, 1.7 Y 1.8	Canales iónicos sensibles al ácido ASIC	TrkA Tropomiosina quinasa A
QSAR	Si	Si	Si	No	No	No	No	No

1.2 Receptor tropomiosina quinasa (Trks)

Los receptores de tropomiosina quinasa o receptores quinasa relacionados con la tropomiosina (Trks) son enzimas transmembrana unidas a glicoproteínas que participan en la señalización celular. La familia de los Trks está constituida por TrkA, TrkB y TrkC. [14] Todas las isoformas de Trks tienen una parte extracelular que contiene tres repeticiones ricas de leucina, dos repeticiones de cisteína, dominios de inmunoglobulina (Ig), una región transmembrana y una parte intracelular con una región juxtamembrana y un dominio tirosina quinasa. [15, 16]

El ligando endógeno de TrkA es el factor de crecimiento neuronal (NGF), los de TrkB son factor de crecimiento neuronal (NGF) y el factor neurotrófico derivado del cerebro (BDNF), mientras que los de TrkC son la neurotrofina 4/5 (NT-4/5) y la neurotoxina 3 (NT-3). [4, 6, 15, 17]

La activación de los receptores Trks ocurre cuando el ligando endógeno respectivo se une con el dominio extracelular de TrkA, TrkB y TrkC, en el segundo dominio Ig dando como resultado la dimerización del receptor con activación del centro catalítico, la autofosforilación de residuos de tirosina en el interior y exterior del dominio quinasa e iniciando la vía de señalización intracelular regulada por Shc (ERK) o fosfolipasa C- γ (PLC- γ) que termina en la diferenciación, supervivencia y crecimiento neuronal. [14, 17-18]

1.3 Receptor tropomiosina quinasa A (TrkA)

El receptor de tropomiosina quinasa A (TrkA) es una proteína codificada por el gen NTRK1. Está constituida por un dominio extracelular de unión al ligando, un dominio transmembrana, una región juxtamembrana y un dominio quinasa intracelular. Se expresa en el sistema nervioso periférico,

central, el tejido inmunitario, tracto digestivo, corteza suprarrenal, próstata, útero, riñón y piel. [4, 7, 18]

El ligando endógeno de TrkA es el factor de crecimiento neuronal (NFG), es secretado por las células inflamatorias y se une al TrkA ubicado en la membrana celular de las terminaciones neuronales sensoriales. El NFG es absorbido por endosomas y transportado de manera retrograda a través del axón hasta los cuerpos celulares de los ganglios, donde se produce la traducción de señales intracelulares de la vía de señalización NFG/TrkA. [4, 7]

La vía NFG/TrkA se induce por la unión de NFG al interactuar con el tercer motivo rico de leucina, el dominio 2-cisteína y el dominio 2-inmunoglobulina de la parte extracelular, lo que promueve la dimerización del TrkA. Posteriormente en la región juxtamembrana y dominio quinasa la tríada de aminoácidos DEG sufre un cambio conformacional y da paso a la apertura del sitio activo donde se une el ATP y transfiere un grupo fosfato (forma activa). La forma activa de TrkA induce autofosforilación de tirosina 670, 674 y 675 en el dominio citoplasmático que actúan como sitios efectores acoplados en la vía de señalización NFG/TrkA y promoviendo finalmente la síntesis de varios tipos de proteínas como la TRPV1, péptidos como la sustancia P y CGRP, los cuales están asociados a la señalización del dolor **Figura 2.** [4, 7, 14, 19 -22]

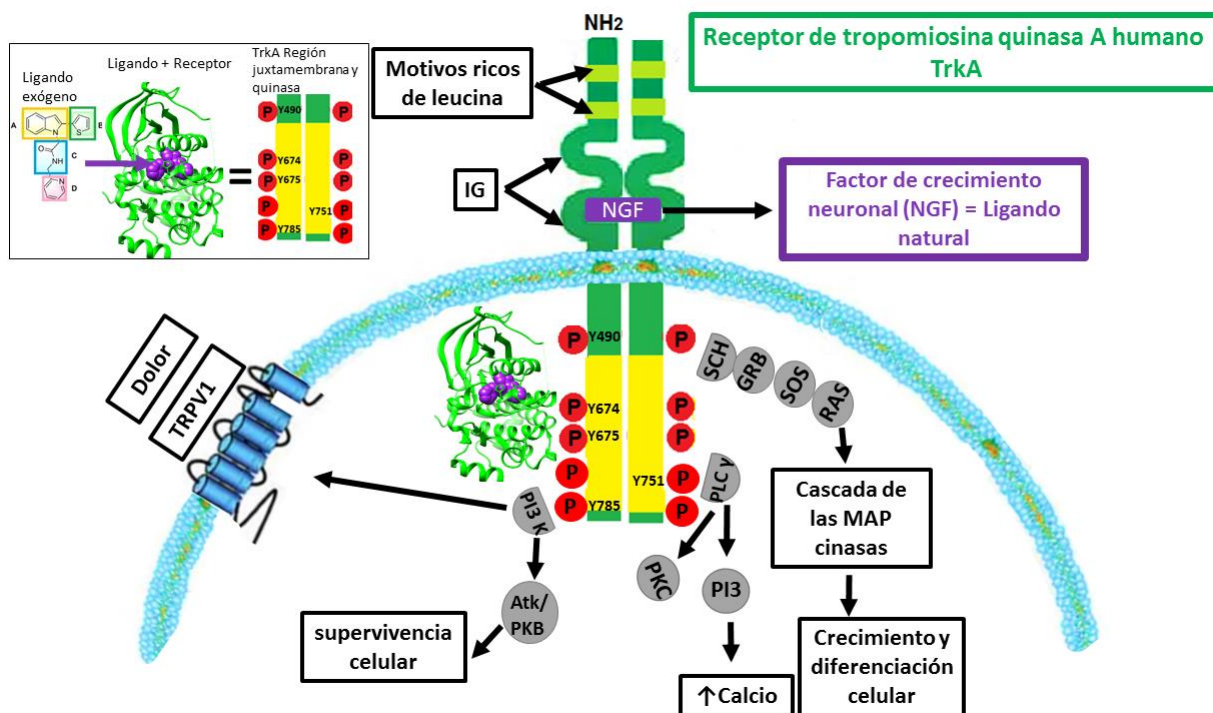


Figura 2: Vía de señalización NGF/TrkA y sitio activo de N-PTindoles en la región juxtamembrana y quinasa.

1.4 Inhibidores relacionados con la vía de señalización NFG/TrkA

Durante décadas se han buscado nuevas terapias contra el dolor y en los últimos años se han descrito nuevos analgésicos relacionados con la inhibición de la vía de señalización NFG/TrkA.

Los inhibidores del receptor TrkA se dividen en tres tipos, según su mecanismo de acción. En la **Tabla 3**, se muestran los mecanismos de acción y ejemplos de cada uno de estos inhibidores que ya se encuentran en el mercado farmacéutico y otros en investigación. [4, 7]

Los inhibidores tipo tres se subdividen de acuerdo a las interacciones que realicen con el dominio quinasa del TrkA:

3a) Son inhibidores que se encuentran en el sitio activo de la bisagra del dominio quinasa.

3b) Inhibidores que se encuentran dentro del sitio activo de la bisagra del dominio quinasa y además en su diseño son inhibidores capaces de extender su interacción a otros residuos aminoacídicos que no son propios del sitio activo de la enzima para aumentar su selectividad.

3c) Son inhibidores que se unen fuera del sitio activo.

3d) Inhibidores que se unen con otros dominios de la enzima. [7]

Tabla 3: Tipos de inhibidores de la vía de señalización NFG/TrkA

Tipo de inhibidor	Mecanismo de acción de inhibidores relacionados con TrkA	Ejemplos
Tipo 1	Capturan NFG	Anticuerpo NGF (ABT-110, alpha-D11, AMG403, fulranumab, Medi-578, muMab911, (MNAC13), ALE0540, PD90780, tanezumad.
Tipo 2	Inhiben la unión de NGF y TrkA	Anti-TrkA (MNAC13), ALE0540, PD90780.
Tipo 3	Inhibidores de la actividad enzimática de TrkA	ARRY-470, ARRY-872 CT-327, CT-335, CT-340, IPTRK342, K252a, NMS-P626.

1.5 Modelo relación estructura actividad cuantitativa (QSAR)

El modelo relación estructura actividad cuantitativa QSAR por sus siglas en inglés (*Quantitative Structure-Activity Relationship*), es un modelo teórico que se ha utilizado principalmente para la optimización de un fármaco líder, el descubrimiento de nuevos medicamentos, la predicción de la actividad de nuevas moléculas estructuralmente similares y estudiar mecanismos de acción. [23]

En la química farmacéutica es ampliamente utilizada para el diseño de nuevos fármacos, por ejemplo: oseltamivir, captopril y losartan. Históricamente el QSAR surgió en 1963 por *Corwin Hansh* con la interpretación sistemática de la respuesta biológica versus las estructuras químicas. [23, 24]

El QSAR se basa en la observación de estructuras similares, pequeñas y activas, cuyas propiedades moleculares o descriptores se cuantifican y se relaciona con la actividad biológica. Este modelo parte de la idea “la estructura determina la actividad” y varias moléculas con estructura similar tendrán una actividad similar. El modelo se expresa como un modelo matemático multivariable que intenta capturar la relación entre la actividad biológica (variable dependiente) con los descriptores moleculares de las estructuras en términos cuantitativos (variables independientes) mediante el uso de herramientas de la estadística como la regresión lineal múltiple. [23, 25-27]

Para que se lleve a cabo esta metodología es necesario contar con dos principales entradas: la primera es una serie de estructuras químicas similares (qSAR) y la segunda es actividades biológicas de cada una de ellas, esta actividad biológica puede estar representada como una concentración inhibitoria 50 o IC50, dosis letal 50 o LD50, constante de inhibición o K_i , entre otras (qsAR). Durante la realización del QSAR se requiere que las estructuras químicas sean interpretadas por

números como descriptores moleculares (**Qsar**) y al final se relacionan mediante una regresión lineal simple o múltiple (**qsaR**). [23, 25, 27]

En los siguientes apartados se explicará con mayor detalle los descriptores moleculares y la regresión lineal múltiple.

1.5.1 Descriptores moleculares

Los descriptores moleculares son representaciones matemáticas de una propiedad perteneciente a una molécula, donde la información química relacionada con la estructura se simboliza de manera numérica [22, 24]. Por ejemplo, en las estructuras N-PTindoles se puede cuantificar el número de nitrógenos, su masa atómica, superficie polar, coeficiente de partición, etc.

Los descriptores moleculares tienen un papel fundamental en el desarrollo de modelos de la química farmacéutica, toxicología, investigación de la salud, protección ambiental, entre otras ciencias. [22, 29]

Los descriptores moleculares se derivan aplicando diferentes teorías de la química cuántica, química orgánica, fisicoquímica, física, farmacodinamia, farmacocinética y toxicología. [23, 26, 28, 29] Estos descriptores pueden ser calculados de forma experimental o teórica.

Los descriptores moleculares se pueden dividir como:

- Descriptores topológicos o conectividad molecular
- Descriptores electrónicos
- Descriptores estéricos
- Descriptores hidrofóbicos

Los descriptores topológicos o de conectividad molecular son los que comienzan por la representación de los átomos de la misma mediante puntos (llamados vértices) y la de los enlaces por segmentos (llamados ejes). Estas estructuras denominadas grafos químicos, no son un simple ordenamiento de puntos y segmentos sino un diagrama de la estructura molecular y una representación de la topología molecular. Los índices topológicos pueden ser sensibles a una o más características estructurales de la molécula, como el tamaño, la forma, la simetría, la ramificación y así también pueden codificar información química sobre el tipo de átomo y la multiplicidad de enlaces. [30-34]

Los descriptores electrónicos son propiedades fisicoquímicas donde se utilizan para su obtención los efectos inductivos (F) y la resonancia (R) son estudiados mediante la ecuación de *Hammett*. Un ejemplo es el momento bipolar. [23, 31, 33]

Los descriptores estéricos son aquellos que se relacionan con el parámetro estérico (volumen de la molécula) de una estructura y la interacción del sitio activo; sus conformaciones, interacciones intermoleculares con los ligandos, efectos polares influyen sobre la interacción directa con la cavidad un ejemplo es la refractividad molar. [23, 31]

Los descriptores hidrófobos como su nombre lo indica miden la hidrofobicidad de una molécula por medio del coeficiente de partición octano/agua o LogP. Parámetros que son fundamentales para relacionar la cantidad de inhibidor que puede entrar al organismo. [23, 31, 34]

1.5.2 Regresión lineal múltiple en QSAR

La regresión lineal múltiple (MLR) es un modelo matemático estadístico cuyo objetivo es relacionar a la variable dependiente (Y o DV) con dos o más variables independientes (X o IDVs) para explicar cómo influye las IDVs con la DV. Esta metodología es una extensión de la regresión

lineal simple (SLR) con la excepción que utiliza más de un descriptor en un espacio hiperdimensional relacionados con una sola DV. [35-38]

Los modelos de regresión lineal se emplean para predecir el valor de la DV o para evaluar la influencia que tienen los predictores sobre ella (este último se debe analizar cuidadosamente para no tener una mala interpretación causa-efecto). [35, 38]

Tanto en la ecuación de SLR como en MLR hay un intercepto, una pendiente o pendientes respectivamente y un coeficiente de determinación, cuyos valores son desconocidas al inicio del método y se resuelven al encontrar la mejor relación. El intercepto y las pendientes son conocidas como coeficientes parciales de regresión. La magnitud de cada coeficiente parcial de regresión es dependiente de la unidad en la que se mida la unidad del descriptor molecular a la que corresponda por lo tanto la magnitud no se debe asociar a la importancia de cada descriptor molecular en la ecuación final. Para la determinación del impacto de cada IDV en el modelo QSAR, se emplean los coeficientes parciales estandarizados (coeficientes β) que se obtienen al sustraer la media y dividir entre la desviación estándar de las IDVs previo al ajuste del modelo. Los coeficientes β representan la variación que experimenta la DV cuando la IDV varía en una unidad y todas las demás parecen constantes. [23, 35-38]

Otro parámetro que se obtiene por la regresión lineal múltiple es el R cuadrado, indica el porcentaje de la varianza de la variable dependiente explicado por el conjunto de variables independientes. Cuanto mayor sea la R-cuadrado más explicativo y mejor es el modelo causal. [35, 38]

Los modelos lineales múltiples para un modelo QSAR siguen la siguiente **ecuación 1** [23, 35-38]:

$$DV = (\beta_0 + \beta_1 IDV_{1i} + \beta_2 IDV_{2i} + \dots + \beta_n IDV_{ni}) + e_i \quad \text{ecuación 1}$$

- DV: Actividad biológica representada como IC50, es la variable dependiente
- IDV 1, 2, 3...n: descriptores moleculares, variables independientes (IDV)
- β_0 : es el intercepto u ordenada en el origen, el valor de la variable dependiente DV cuando todos los predictores son cero.
- $\beta_1, 2, 3 \dots n$: son las pendientes dadas por el efecto promedio que tiene el incremento en una unidad del descriptor molecular "IDV 1, 2, 3...n" sobre la DV, manteniéndose constantes el resto de variables. Indica la intensidad y la dirección de la relación entre esa variable independiente (IDV) y la variable dependiente (DV). El signo de la pendiente indica la dirección (signo positivo indica que al aumentar los valores de la IDV aumentan los valores de la DV; signo menos indica que al aumentar los valores de la IDV, los valores de la DV descienden).
- e_i : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Existe una serie de condiciones para la elaboración de una MLR que son las mismas y otras condiciones adicionales que utilizan los modelos lineales simples. A continuación, se presentan algunas de las condiciones de MLR para obtener un buen modelo QSAR.

La no colinealidad o multicolinealidad: Indica que los IDVs deben ser independientes, por lo que no deben tener colinealidad entre ellos. En otras palabras, se refiere a cuando un descriptor molecular esta linealmente relacionado con uno o varios descriptores moleculares del modelo. [35, 38]

Un ejemplo es colocar en una ecuación de MLR dos IDVs que representen al coeficiente de reparto solo por obtenerse diferentes métodos, ya que la consecuencia inmediata es la colinealidad que no permite identificar de forma precisa el efecto individual que tiene cada IDV con respecto a la

actividad biológica y esto finalmente se traduce en el incremento de la varianza de los coeficientes de regresión estimados y resulta prácticamente imposible establecer su significancia estadística.

No existe un método estadístico concreto para determinar la existencia de colinealidad o multicolinealidad entre los descriptores moleculares en un modelo QSAR, sin embargo, se han desarrollado numerosas “reglas” prácticas que tratan de determinar la medida en la que afecta la estimación y contraste de un modelo. Las reglas para descartar una colinealidad o multicolinealidad son [25, 35, 36, 38]:

- Calcular el coeficiente de determinación (R^2) entre los descriptores, si el R^2 es alto (cercano a 1) y ninguno de los descriptores es significativo, hay indicios de colinealidad.
- Realizar una matriz de correlación en la que se analice la relación lineal entre cada par de descriptores moleculares o IDVs. Cabe mencionar que, a pesar de no obtenerse ningún coeficiente de correlación alto, esto no asegura la ausencia de multicolinealidad. Se puede dar el caso de tener una relación lineal casi perfecta entre tres o más variables y que las correlaciones simples entre pares de estas mismas variables no sean mayores que 0.5.
- Determinar un modelo de regresión lineal simple entre cada uno de los IDVs frente al resto. Si en alguno de los modelos el coeficiente de determinación R^2 es alto, estaría señalando una posible colinealidad.

En caso de encontrar colinealidad entre los IDVs, existen dos posibles soluciones. La primera es excluir uno de los IDVs redundantes e intentar conservar el descriptor que, a juicio del investigador, este influyendo con la respuesta biológica o la DV. El IDV redundante excluido no suele tener mucho impacto en el modelo en cuanto a su capacidad predictiva ya que, al existir colinealidad, la información que aporta uno de los descriptores moleculares es la misma en presencia del otro por

lo que no hay una pérdida de información causal. La segunda opción consiste en combinar las variables colineales en un único IDV, aunque con el riesgo de perder su interpretación. En las dos soluciones se puede cambiar por otro IDV que aporte mejor información al modelo. [35, 38] En ejemplo anterior del coeficiente de reparto si se elimina una IDV colineal, la nueva ecuación no sufrirá una pérdida de información ya que el IDV eliminado era redundante.

En la metodología QSAR se recomienda evitar la colinealidad, ya que puede llevar a conclusiones erróneas, haciendo creer que una variable es la causa cuando en realidad es otra la que está influenciando sobre ese descriptor. Lo que resulta en una mala predicción.

Parsimonia: Indica que el mejor modelo QSAR es aquel capaz de explicar con mayor precisión la variabilidad observada en la VD empleando el menor número de descriptores moleculares, por lo tanto, con menos asunciones. Se recomienda utilizar descriptores representativos y fáciles de identificar de manera experimental. [25, 35, 38]

Relación lineal entre los predictores numéricos y la variable respuesta: Cada descriptor molecular numérico tiene que estar linealmente relacionado con la actividad biológica mientras los demás IDVs se mantienen constantes, de lo contrario no se puede introducir en el modelo. La forma más recomendable de comprobarlo es representando los residuos del modelo frente a cada uno de los descriptores moleculares. Si la relación es lineal, los residuos se distribuyen de forma aleatoria entorno a cero. Estos análisis son solo aproximados, ya que no hay forma de saber si realmente la relación es lineal cuando el resto de descriptores se mantienen constantes. [35]

Distribución normal de los residuos: Los residuos obtenidos del modelo se deben distribuir de forma normal con media cero. Para comprobarlo se recurre a histogramas, a los cuartiles normales o a test de hipótesis de normalidad. [35-38]

Variabilidad constante de los residuos (homocedasticidad): La varianza de los residuos debe de ser constante en todo el rango de observaciones (moléculas). Para comprobarlo se representan los residuos y si la varianza es constante, se distribuye de forma aleatoria manteniendo una misma dispersión y sin ningún patrón específico existe homocedasticidad, Una distribución cónica en la gráfica es un claro ejemplo de falta de homocedasticidad. También se puede recurrir a contrastes de homocedasticidad como el test de *Breusch-Pagan*. [25, 35, 38]

Valores atípicos, con alto leverage o influyentes: Es recomendable identificar moléculas que sean atípicas o que puedan estar influenciando al modelo QSAR. La forma más fácil de detectarlas es a través de los residuos. [25, 35-38]

Tamaño de la muestra: Para que la MLR tenga significado estadístico debe haber suficientes moléculas (datos) para generar un QSAR, si no se dispone de suficientes moléculas, los descriptores que no son realmente influyentes podrían parecerlo. En el libro *Handbook of biological statistics* recomiendan que el número de observaciones sea como mínimo entre 10 y 20 veces el número de descriptores del modelo. Y en trabajos anteriores se ha descrito que para elaborar un buen modelo QSAR se necesita cerca de la sexta parte de descriptores del conjunto de moléculas (estadísticamente los datos “n” que se utilizan para generar la ecuación). Por ejemplo, en una ecuación múltiple lineal que contiene 40 moléculas estructuralmente similares se necesitan de siete a ocho descriptores para generar la ecuación que describa la mejor relación de la actividad biológica. [25, 35-38]

La gran mayoría de condiciones se verifican utilizando los residuos, por lo tanto, se suele generar primero el modelo y posteriormente validar las condiciones. De hecho, el ajuste de un modelo debe verse como un proceso iterativo en el que se ajusta el modelo, se evalúan sus residuos y se mejora hasta llegar a un modelo óptimo o el mejor modelo.

Para la selección de descriptores en un modelo QSAR es necesario dar las siguientes definiciones estadísticas:

Bias o sesgo: es el error que se introduce al intentar explicar la relación existente entre variables del mundo real en este caso las variables de la naturaleza mediante un modelo matemático más simple lo cual es extremadamente complicado. El bias de un modelo es independiente de los datos empleados para crearlo, depende únicamente del tipo de modelo utilizado para intentar explicar la relación entre variables. [25, 35-38]

Varianza: es la variación que tiene un modelo en este caso el QSAR dependiendo de las moléculas que se emplean para ajustarlo (*training data*). [25, 35]

El modelo de regresión lineal múltiple se ajusta por mínimos cuadrados y donde debemos tener en cuenta dos puntos fundamentales, el error de predicción y la Interpretación del modelo QSAR.

El error de predicción: Para que un modelo de regresión lineal múltiple generado por mínimos cuadrados tenga una buena capacidad predictiva (*test error rate*) se necesita que la relación existente entre los IDVs y la actividad biológica (DV) sea aproximadamente lineal, ya que así se asegura que el método de mínimos cuadrados tenga un bias o sesgo pequeño. Además, se requiere que el número de moléculas con las que se ajusta el modelo sea mayor al número de descriptores ($n \gg p$) y así las estimaciones tendrán poca varianza. Por el contrario, si existe un número de estructuras mucho mayor que los IDVs, la varianza aumenta ($p > n$) y por lo tanto el método de mínimos cuadrados no debe ser utilizado. [25,38]

Interpretabilidad del modelo: Cuantos más IDVs se utilicen en el modelo QSAR utilizando la MLR más compleja es su interpretación. Por lo tanto, se recomienda limitar al modelo con IDVs que tengan una influencia importante sobre la respuesta biológica (DV), excluyendo aquellos descriptores irrelevantes y que solo incorporan complejidad innecesaria. Además, la

interpretabilidad se asocia a descriptores que puedan justificar adecuadamente una actividad biológica, de forma teórica pero también que estos puedan utilizarse para cambios de manera experimental. [25, 28, 35, 38]

Cuando se dispone de pocas moléculas o muchos descriptores, es necesario emplear un método de regresión que nos ayude a identificar los IDVs más relevantes y excluir a los irrelevantes. A este proceso se le conoce como variable de selección. Existen tres métodos más utilizados para elegir a los mejores IDVs y son "subset selection", "shrinkage/regularization" y "dimension reduction" [35-38]. Cabe mencionar que todos estos métodos reducen la varianza de las estimaciones, pero siguen siendo modelos lineales. En este trabajo solo de mencionar el *subset selection*, ya que es el método que se utilizara en esta investigación.

Subset Selection: Este método tiene como propósito identificar y seleccionar de entre todos los IDVs calculados, aquellos que están más relacionados con la actividad biológica y así crear el mejor modelo. Sin embargo, hay que tener en cuenta que al final el investigador debe seleccionar los descriptores que permitan una mayor interpretabilidad al método. [35-38]

El procedimiento general de los métodos de *subset selection* son:

Crear un conjunto de modelos, todos los posibles (*best subset*) o bien un conjunto considerable de ellos (*stepwise*), mediante diferentes combinaciones de los descriptores moleculares disponibles.

Dentro del Subset selection existe una subdivisión: best subset selection y stepwise selection (forward, backward e hybrid). Este último es la combinación de los métodos *forward* y *backward selection*.

Best Subset Selection: Este proceso consiste en evaluar todos los posibles modelos creando combinaciones de los IDVs disponibles. Primero se genera un modelo sin ningún IDV al que se le conoce como un modelo nulo (M0). Después se generan todos los posibles modelos que contienen solo un IDV y se selecciona el que tiene menor suma residual de cuadrados o *Residual Sum of Squares (RSS)* o mayor R^2 . Al modelo seleccionado se denomina (M1). Posteriormente se repite el paso anterior para modelos con dos predictores y así sucesivamente hasta llegar al modelo con todos los IDVs (Mk). [35-38] Este tipo de selección de descriptores se produce por programas estadísticos con algoritmos muchas veces desconocido para el usuario y por lo que el mejor modelo por este método es el que numéricamente tiene un R^2 cercano a 1 y no considera si el descriptor tiene interpretabilidad biológica y química, por lo tanto, el investigador no debe perder el objetivo de su estudio ya que él conoce la parte biológica, fisicoquímica, toxicológica, etc.

Otra cuestión que se debe tener en cuenta es que después de identificar y seleccionar el mejor modelo se tiene que verificar que esto sea correcto por lo que se propone hacer una validación cruzada o R^2 ajustado.

Forward Stepwise Selection: Esta metodología no evalúa todas las combinaciones posibles de los IDVs, solo realiza subconjuntos de las mismas. Los pasos son generar un modelo nulo y después crear todos los posibles modelos que se pueden generar al introducir un IDV más al M0. Luego entre todos estos modelos con un solo IDV se selecciona el mejor por medio de la RSS o R^2 (M1). Se repite el paso anterior partiendo del último modelo seleccionado y así sucesivamente hasta tener el modelo con todos los descriptores moleculares que se relacionan mejor con la actividad biológica. Se seleccionan los mejores modelos para cada número de predictores (M0, M1, M2..., Mk) y se identifica el mejor, esta vez empleando una validación cruzada o R^2 ajustado como en el caso de Best subset. [35-38]

Backward Stepwise Selection: este método a diferencia del *forward stepwise selection* inicia con un modelo donde están contenidos todos los posibles IDVs conocido como *full model*. Y después

en cada iteración se generan todos los modelos que se pueden crear eliminando un IDV a la vez y se selecciona el que tiene menor RSS o mayor R^2 . El proceso se repite hasta llegar al modelo nulo sin predictores. De entre los mejores modelos seleccionados para cada número de descriptores ($M_0, M_1, M_2, \dots, M_k$) se identifica el mejor, empleando R^2 ajustado. [35-38]

2 ANTECEDENTES

Actualmente existen moléculas inhibidoras de la vía de señalización NFG/TrkA las cuales se clasifican en cuatro tipos véase **Tabla 2** y de los cuales los inhibidores que se unen con otros dominios de la enzima en este caso la región juxtamembrana y quinasa son de interés para esta investigación. [4, 5, 7]

En 2017 *Hurzy y colaboradores* diseñaron, sintetizaron y calcularon la concentración inhibitoria 50 (IC50) de 40 nuevas moléculas con estructuras similares a N-TPindoles véase **Figura 3** la plantilla del ligando cuyo sitio activo se encuentra en el dominio intracelular de la juxtamembrana y dominio quinasa entre los residuos 440-497 del receptor TrkA representada en la **Figura 2**. [5] Las 40 estructuras análogas de N-PTindoles se muestran en el suplemento y en la **Tabla 3**. Cabe mencionar que una plantilla de un ligando se refiere a la molécula en ya existente que es similar a las estructuras que se quieren construir, además se encuentra en una base de datos y a partir de ella se generan las demás moléculas.

El diseño de las 40 moléculas se obtuvo por el método ALIS o sistema de selección de afinidad-espectrometría de masas para el descubrimiento y caracterización de interacciones proteína-ligando utilizando los residuos 498–796 de la región juxtamembrana y dominio quinasa. La actividad biológica se evaluó en función de la actividad de la quinasa basada en células mediante el método de *DiscoverX PathHunter*. [5, 7]

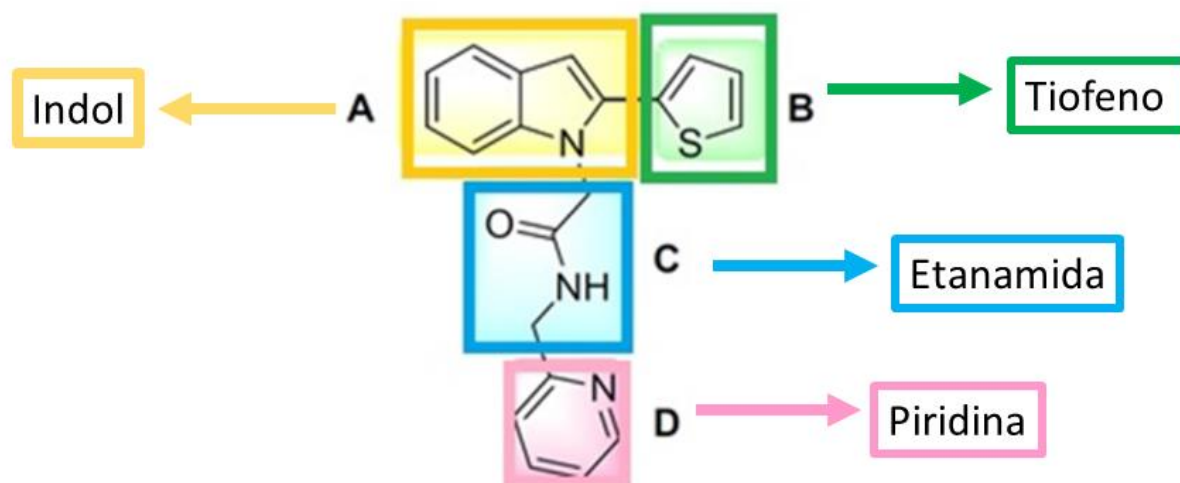


Figura 3: Plantilla del ligando referencia de N-PTindoles, se observa en colores las modificaciones de las estructuras que realizaron *Hurzy y colaboradores*. En color amarillo “A” se muestra el grupo indol donde se realizaron ocho modificaciones, en color verde “B” el grupo tiofeno donde se realizaron once modificaciones, en color azul “C” el grupo etanamida donde se realizaron siete modificaciones y en color rosa “D” el grupo 2-piridil donde se realizaron catorce modificaciones.

3 HIPÓTESIS

Es posible realizar un modelo de relación estructura-actividad cuantitativa de 40 inhibidores del receptor de tropomiosina quinasa A humano análogos de -N-PTindoles asociada a la supresión del dolor, a partir de una publicación científica que sólo presenta SAR.

4 OBJETIVO GENERAL

Realizar un modelo de relación estructura-actividad QSAR de 40 inhibidores de TrkA análogos de –N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il) etanamida del receptor de tropomiosina quinasa A humano asociada a la supresión del dolor.

5 OBJETIVOS PARTICULARES

a) Búsqueda en la base de datos de proteínas PDB, el receptor TrkA y análisis de las estructuras cristalizadas que contengan el dominio juxtamembrana y quinasa.

b) Generación de 40 moléculas análogas de –N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il) etanamida bidimensional (2D) y tridimensional (3D).

c) Superposición de las 40 moléculas análogas de –N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il) etanamida en 3D.

d) Cálculo de los descriptores moleculares o IDVs para la realización del QSAR y la selección de estos con los análogos de –N-(piridin-2-ilmetil)-2-(2-tiofen-2-ilindol-1-il).

e) Cálculo de la regresión lineal múltiple de la actividad biológica y los descriptores.

f) Análisis de los modelos QSAR.

6 METODOLOGÍA

6.1 Búsqueda del PDB.

Se realizó una búsqueda en el banco de proteínas para encontrar ligandos en 3D cristalizados con su sitio activo de la proteína TrkA en la región quinasa y juxtamembrana. Esta búsqueda se realizó por tres motivos la primera es encontrar una plantilla que ayude a la construcción de los análogos de N-PTindoles (plantilla del ligando), la segunda es visualizar dentro del sitio de unión al ligando (*binding site*) los aminoácidos que están involucrados y en tercera tener una plantilla de la proteína para poder utilizarla en el acoplamiento molecular y generación de otras moléculas.

6.2 Construcción de moléculas.

Existen tres formas de construir moléculas en 3D [23]:

I. Construcción a partir de fragmentos de moléculas ya conocidas que se encuentran en las librerías de los programas.

II. Construcción de la estructura sin tener una estructura base o fragmento; dibujando cada uno de sus átomos.

III. Construcción a partir de una plantilla de un ligando en 3D o molécula similar que proviene de una estructura cristalográfica.

La construcción de las 40 moléculas análogas del N-(2- metilpiridina)-2-(2-tiofen-2-ilindol-1-il) etanamida o N-PTindoles se realizó a partir de una plantilla de un ligando 3D llamado 6UF que se

encuentra en la proteína cristalizada con código PDB 5KMJ en resultados se explica detalladamente como se eligió esta plantilla [5].

Las modificaciones para la construcción de cada una de las moléculas se realizaron en VegaZZ. [39]. Para la construcción de una molécula en VegaZZ [39] de la forma II requiere conocer el tipo de elemento, hibridación y tipo de enlace, ya que al adicionar un átomo en la molécula deseada debe ser el correcto. Por ejemplo, la molécula C20 **Figura 4** está conformada por un indol, un tiofeno, una N-metiletanamida y un 2- metilpiridina. El indol es un heterociclo con estructura bicíclica constituida por un benceno y un pirrol; por lo tanto, el elemento para la construcción del benceno son los elementos carbono (C) e hidrógeno (H), la hibridación del carbono del benceno es sp^2 y el tipo de enlace carbono-carbono será doble y simple alternativamente para los seis carbonos que constituyen al benceno. El enlace hidrógeno-carbono será un tipo de enlace simple e hibridación sp^2 para cada elemento. En el caso del pirrol los elementos involucrados son carbono, hidrógeno y nitrógeno; la hibridación es sp^2 para cada uno de los elementos y el tipo de enlace de carbono-carbono es simple, simple y doble; los enlaces carbono hidrógeno son simples y el enlace carbono nitrógeno es simple. Después un carbono del pirrol se une a un carbono del tiofeno con una interacción carbono-carbono con enlace simple e hibridación sp^2 . El tiofeno construido es un anillo plano constituido por cuatro carbonos y un azufre con hibridación sp^2 y tipo de enlace doble, simple, doble, simple, simple y 3 hidrógenos con hibridación sp^2 con tipo de enlace simple por lo que es un compuesto aromático al igual que el indol. El carbono de la N-metiletanamida está unida al nitrógeno del pirrol en una hibridación sp^3 y un tipo de enlace simple, los elementos de N-metiletanamida son carbono, oxígeno, hidrógeno y nitrógeno; las interacciones carbono-carbono de la amida son con enlace simple y el carbono tiene una hibridación sp^2 , la interacción carbono-oxígeno enlace doble, la interacción carbono- nitrógeno es un enlace simple e hibridación sp^3 y los hidrógenos tienen una hibridación sp^3 con enlaces simples. El nitrógeno de la amida se une con el carbono de 2- metilpiridina con un enlace simple y una hibridación del carbono sp^3 para cada uno de los carbonos unidos a la piridina junto con el nitrógeno son hibridación sp^2 y enlaces alternados dobles y simples; mientras que los hidrógenos son enlaces simples.

Después de haber construido la molécula se prosigue con una optimización de la molécula, en este caso se utilizó el campo de fuerza *TRIPOS* el cual otorga el tipo de átomo a la molécula y para la carga se utilizó *GASTEIGER*.

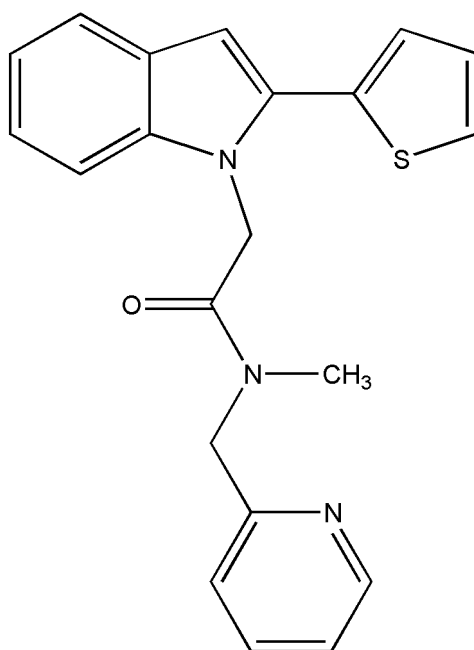


Figura 4: Representación 2D de la molécula C20 de *Hurzy y colaboradores*.

Los tipos de átomos más utilizados en el área biológica en la parametrización del campo de fuerza *TRIPOS* están definidos en la **Tabla 4**. Los tipos de átomos no son los elementos químicos, son la relación de los átomos con su hibridación, geometría e interacción con otros elementos.

Existe un sistema final de clasificación de átomos, existen 68 tipos de átomos básicos. Es integral para los elementos que se encuentran comúnmente en las moléculas orgánicas (C, H, N, O, S, P, halógenos) y también incluye metales y gases nobles. Está diseñado de tal manera que cada átomo presente en la molécula coincidirá con uno y solo un tipo de átomo, eliminando así la ambigüedad del sistema de tipificación. [40]

En el ejemplo anterior la molécula C20 los carbonos tanto del indol, el tiofeno y la piridina son hibridación sp^3 sin embargo para un tipo de átomo la clasificación correcta es C.ar.

Tabla 4: Lista resumida de los tipos de átomos en *TRIPOS*

Tipo de átomo	Descripción
C.3	Carbono con hibridación sp^3
C.2	Carbono con hibridación sp^2
C.1	Carbono con hibridación sp^1
C.ar	Carbono aromático
N.3	Nitrógeno con hibridación sp^3
N.2	Nitrógeno con hibridación sp^2
N.1	Nitrógeno con hibridación sp^1
N.ar	Nitrógeno aromático
N.4	Nitrógeno con hibridación sp^3 cargado positivamente
N.am	Nitrógeno de la amida
O.2	Oxígeno con hibridación sp^2
O.3	Oxígeno con hibridación sp^3
O.ar	Oxígeno aromático
S.3	Azufre aromático

6.3 Cálculo de los descriptores

Para el cálculo de algunos descriptores se utilizó un formato conocido como Smiles y MOL2 con extensión “.smi“ y “.mol2“ respectivamente. La extensión “.smi“ es una especificación de introducción lineal molecular simplificada que describe sin ambigüedades la estructura de una molécula usando cadenas ASCII cortas las cuales se pueden importar a diferentes softwares para su conversión en dibujos bidimensionales o tridimensionales. Mientras que la extensión “.mol2“ guarda las coordenadas, cargas y topologías de las moléculas.

Para el cálculo de los descriptores se utilizaron los programas *VegaZZ* [39], *E-dragon* [41] y *Swiss ADME* [42].

VegaZZ es un programa gratuito en el cual se pueden calcular descriptores topológicos, estéricos y electrónicos de las moléculas. Además de otras funciones como la construcción y optimización de las moléculas. [39]

E-Dragon es un programa en línea gratuito que calcula diferentes tipos de descriptores moleculares tanto topológicos y geométricos. Fue desarrollado por *Todeshini y colaboradores*. Proporciona más de 1600 descriptores moleculares o IDVs. Puede analizar máximo 149 moléculas y en cada una de ellas un máximo de 150 átomos. Los datos de entrada que se requieren para este programa son extensiones SMILES, SDF (MDL) o MOL2 (Sybyl) files. [41]

Swiss ADME es una herramienta computacional en línea, que integra otras herramientas de la web con el fin de tener resultados de descriptores fisicoquímicos, farmacocinéticas y toxicológicos. [42]

En el caso de VegaZZ [39] y E-dragon [41] se utilizó la estructura 3D en formato “. mol2” para obtener los descriptores mientras que para Swiss ADME [42] se utilizó la extensión “.smi”

6.4 El cálculo de la regresión lineal

Se realizó en el software Statistics [43] y R software [44]. El primer programa requiere de una licencia y es fácil de utilizar mientras que el segundo programa es gratis y se requiere conocer sobre comandos y en algunas ocasiones de paquetes especiales.

7 RESULTADOS Y DISCUSIÓN DE RESULTADOS

7.1 Búsqueda de PDB y Construcción de moléculas en 3 dimensiones (3D)

Para la construcción de las estructuras en tres dimensiones (3D) se encontraron 11 cristalografías de TrkA de la región juxtamembrana y quinasa, de las cuales solo 6 incluían un ligando en el sitio activo véase **Tabla 5**.

La elección del cristal se basó en el análisis del ligando con su receptor. Se revisó que la estructura sea lo más parecida a los análogos de N-PTindoles **Figura 3** para utilizarla como plantilla de un ligando en la construcción de los N-PTindoles.

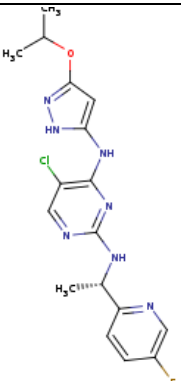
La construcción a partir de una plantilla de un ligando se propuso por tres motivos:

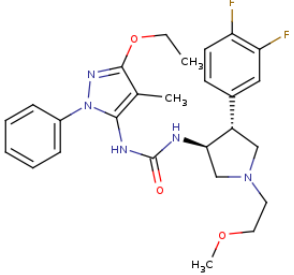
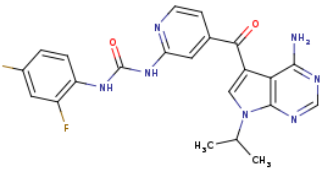
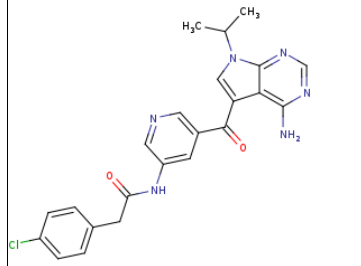
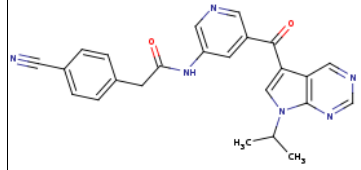
1. Facilitar la construcción de las moléculas tridimensionalmente.

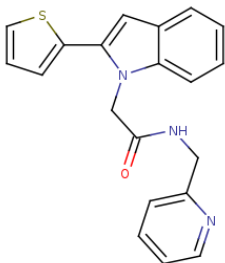
2. Visualizar a N-PTindoles en su sitio activo, conocer los aminoácidos que se encuentran participando en las interacciones ligando-receptor y utilizar esta información para la elección de algunos descriptores moleculares en el modelo QSAR de estas estructuras.

3. Como parte de la validación de nuestro modelo, construyendo nuevas moléculas con estructuras similares a las de *Hurzy y colaboradores* y comprobar la predicción del modelo QSAR obtenido.

Tabla 5: Lista de proteínas cristalizadas de TrkA con o sin ligando encontradas en PDB.

ID PDB	Resolución (Å)	Ligando	Organismo	Literatura
4AOJ	2.75		<i>Homo Sapiens</i>	[45]
1WWA	2.5	NO HAY	<i>Homo Sapiens</i>	[46]
4GT5	2.398	NO HAY	<i>Homo Sapiens</i>	[47]
4F0I	2.302	NO HAY	<i>Homo Sapiens</i>	[48]

1WWW	2.2	NO HAY	<i>Homo Sapiens</i>	[49]
5H3Q	2.1		<i>Homo Sapiens</i>	[50]
5JFS	2.07		<i>Homo Sapiens</i>	[17]
1HE7	2	No hay	<i>Homo Sapiens</i>	[51]
5JFV	1.59		<i>Homo Sapiens</i>	[17]
5JFW	1.52		<i>Homo sapiens</i>	[17]

5KMJ	2.04		<i>Homo Sapiens</i>	[7]
------	------	---	---------------------	-----

Inicialmente se propuso la construcción de los 40 análogos N-Ptindoles con el ligando 7HF proveniente del código PDB 5H3Q [50]. Su andamio contiene varias coincidencias estructurales con N-PTindoles. La urea contenida en la parte central de 7HF es casi la estructura de la etanamida de N-PTindoles y además esta bisustituida casi simétricamente lo que se asemeja a la molécula D28. Sin embargo, al estar bisustituida casi simétricamente es difícil proponer donde se construye la parte indol y tiofeno de los N-Ptindoles en 7HF. Por lo tanto, se planteó la construcción del indol y el tiofeno en el 3,4-bis(fluranil)fenil del ligando 7HF véase la **Figura 5**. Para garantizar que la región en la que se decidió construir es correcta, se realizó una superposición y un acoplamiento molecular, el resultado se observa en la **Figura 6**, demostrando así que la forma de la construcción sugerida por nosotros era correcta.

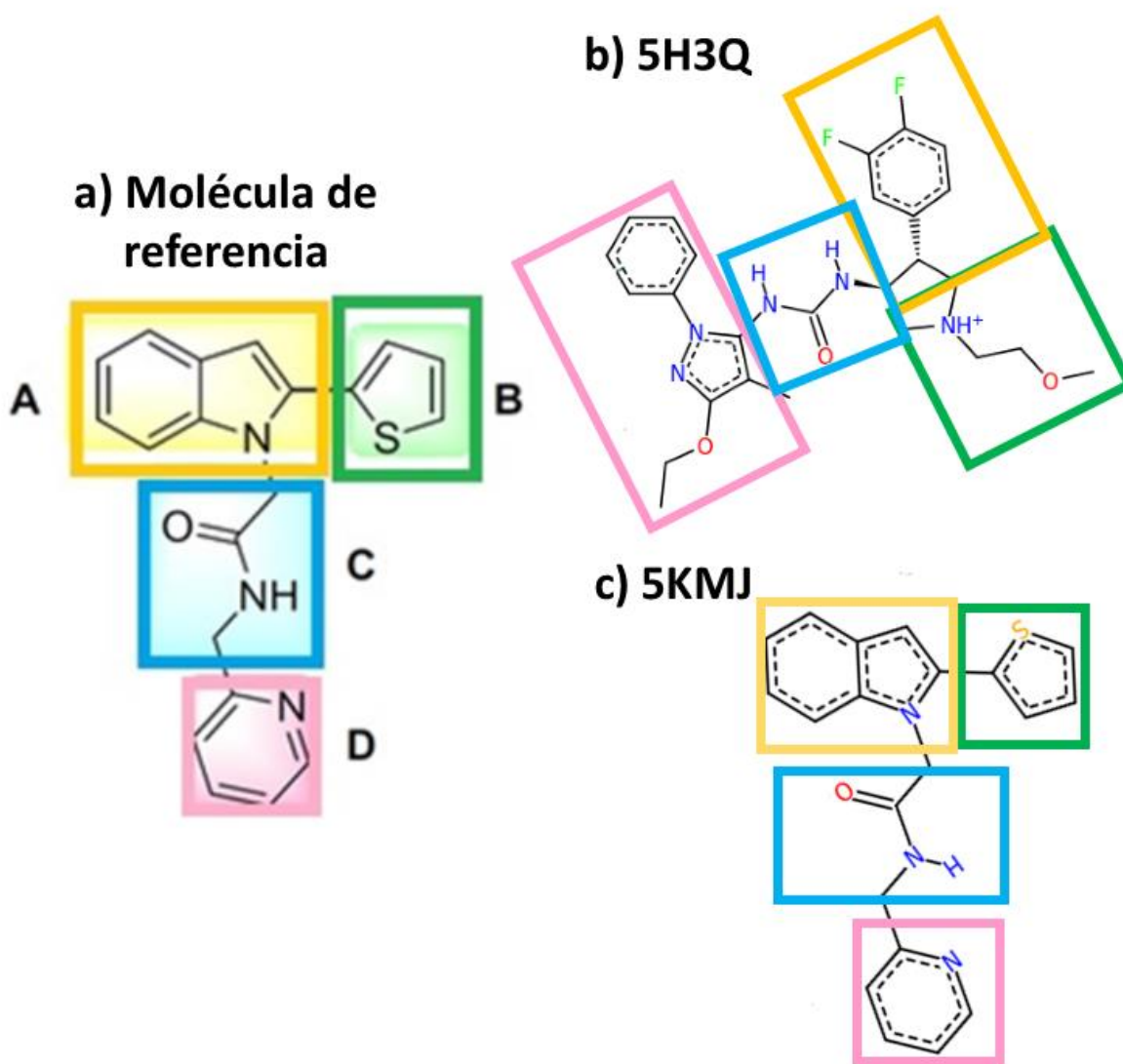


Figura 5: Comparación de los PDBs 5H3Q y 5KMJ candidatos para la construcción en 3D. A) Molécula de referencia con incisos y colores de las regiones que son esenciales para la construcción de las 40 moléculas análogas de N-Ptindoles. B) Ligando 7HF del código PDB 5H3Q en colores donde se indica las regiones donde se propone realizar la construcción de N-Ptindoles. C) Ligando 6UF en el PDB 5KMJ en colores donde se indica las regiones donde se propone realizar la construcción de N-Ptindoles.

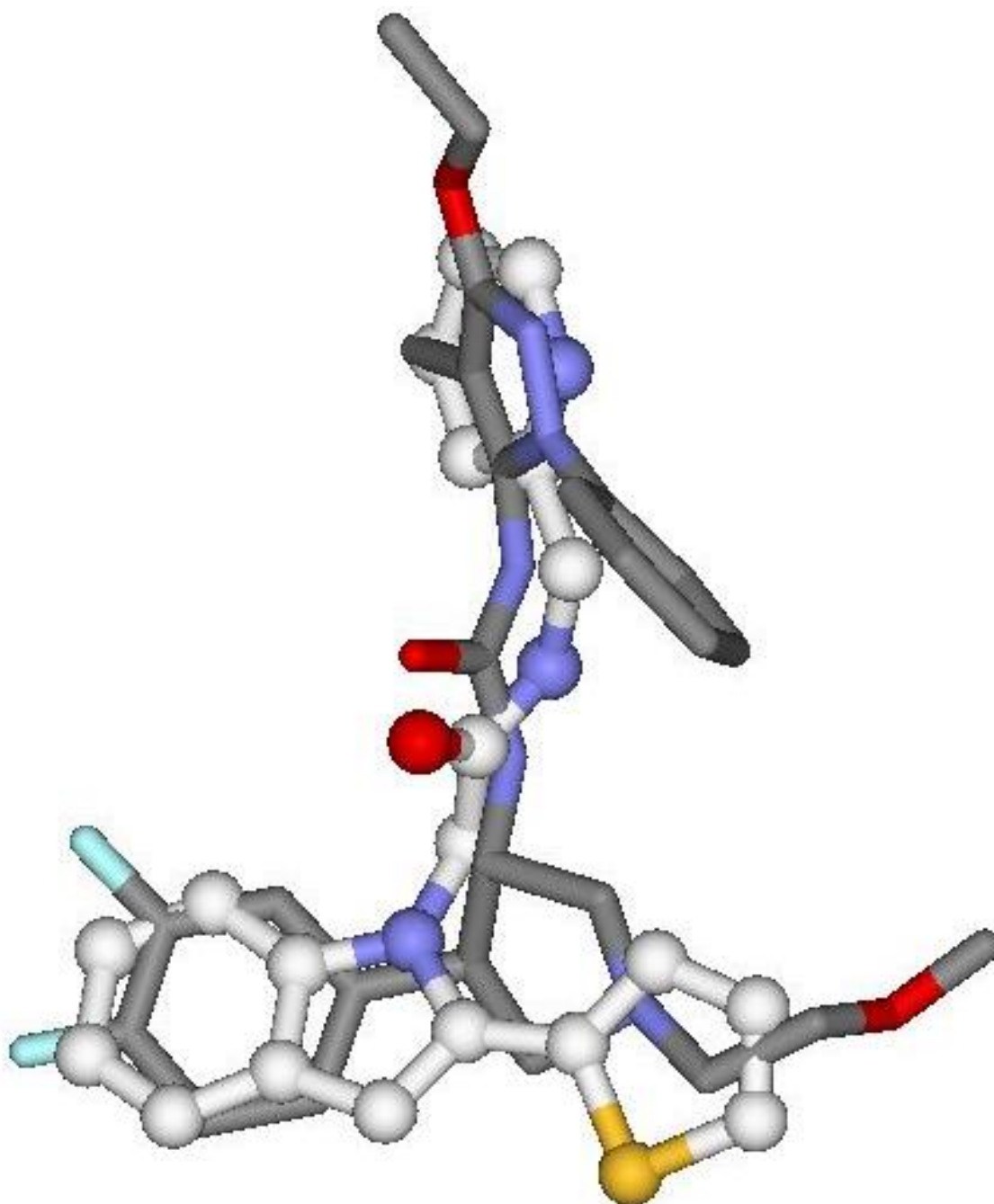


Figura 6: Superposición de los ligandos: ligando 6UF (código PDB 5KMJ) Color blanco y ligando 7HF (código PDB5H3Q) color gris.

Posteriormente se encontró el cristal código PDB 5KMJ [7] el cual muestra un complejo entre nuestra enzima TrkA y el ligando (6UF) se observó que el ligando cristalizado es muy parecido a la estructura C20 de los datos de entrada publicados e informados en 2017 por *Hurzy y colaboradores* véase las 40 estructuras en la **Tabla 6**.

Comparando todos los ligandos encontrados el que presento mayor similitud estructural de todas las estructuras cristalinas conocidas hasta ahora de complejos TrkA de la región juxtamembrana y quinasa (última inspección PDB 15.12.2018) fue 6UF, por lo que la construcción de las moléculas análogas de N-PTindoles se realizaron a partir de su plantilla.

En la **Figura 5** se muestra el ligando 6UF indicando las similitudes estructurales con la molécula N-PTindol de referencia de *Hurzy y colaboradores*, se puede apreciar que 6UF tiene un mayor parecido que el ligando propuesto anteriormente 7HF, ya que en su estructura contiene un indol, tiofeno, etanamida y una piridina.

La construcción de las moléculas 3D en VegaZZ [39] se observan en la **Figura 7** donde se aprecia las pequeñas modificaciones de *Hurzy y colaboradores* además de mostrar la superposición de las cuarenta moléculas análogas de N-PT indoles manteniendo la misma posición en el sitio activo.

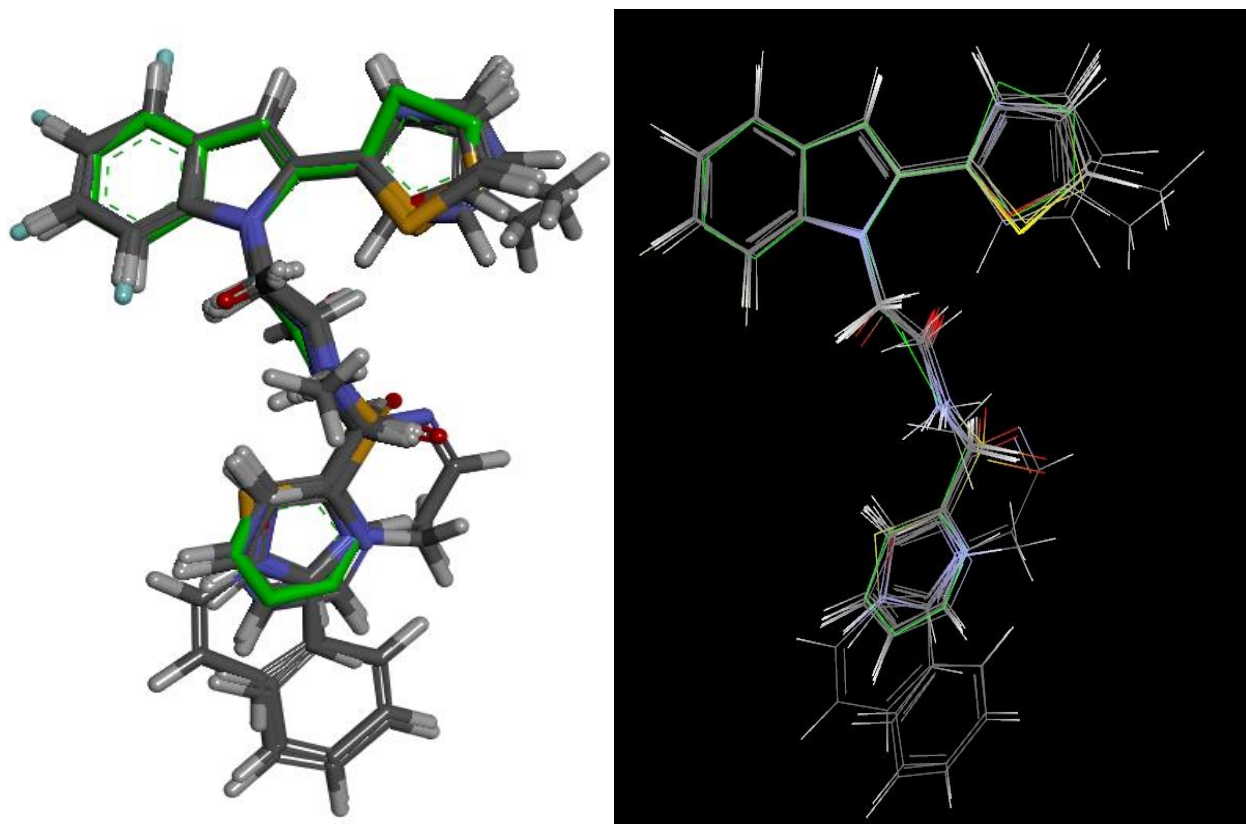


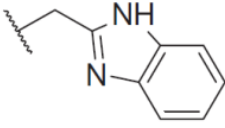
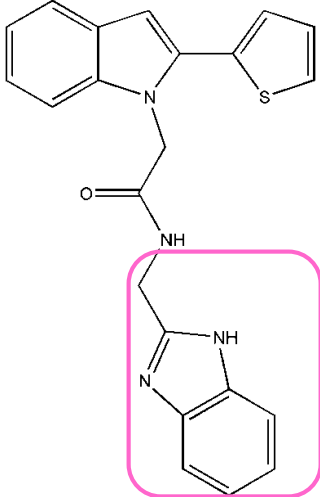
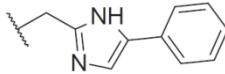
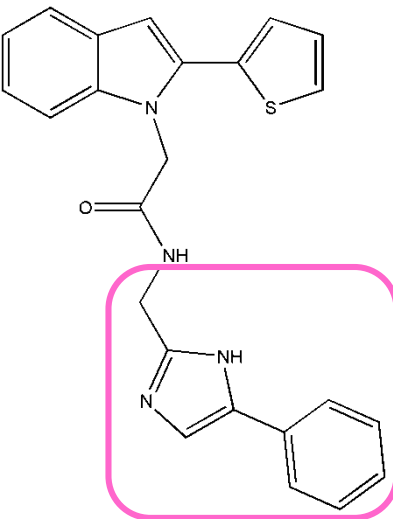
Figura 7: Representación de las 40 moléculas de la referencia [5], construidas a partir de la plantilla del ligando 6UF del código PDB 5KMJ. Código de colores en gris las 40 moléculas construidas y en color verde la molécula de referencia.

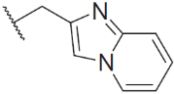
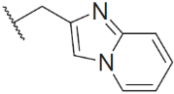
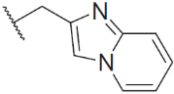
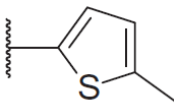
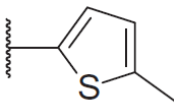
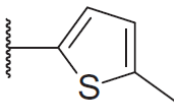
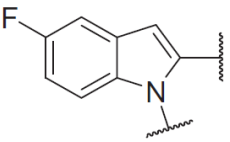
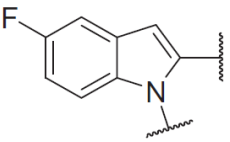
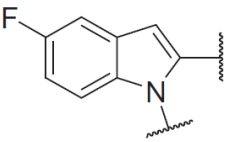
7.2 Construcción de moléculas 2 Dimensiones (2D)

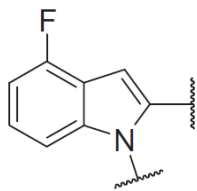
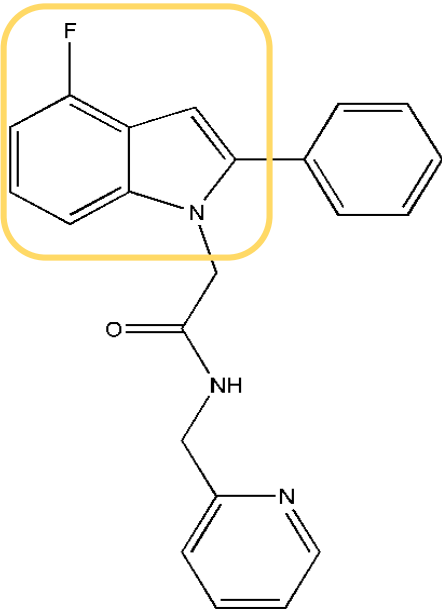
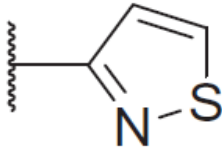
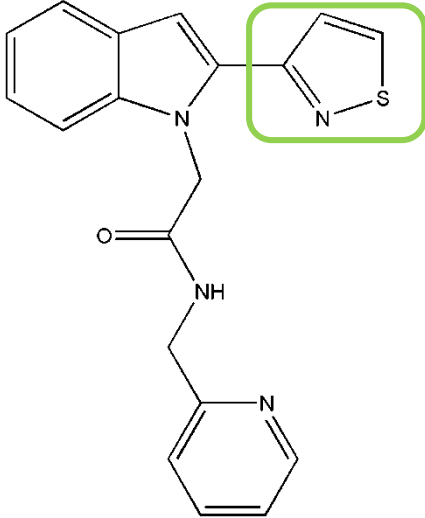
Se realizó la construcción de las moléculas 2D y se guardó en formato “.smiles” véase la **Tabla 6** con las 40 moléculas análogas de N-PTindoles. En la primera columna de la tabla se muestra la posición de las estructuras en una lista de menor a mayor IC50, la segunda columna indica el nombre y número de identificación de la estructura de acuerdo al artículo de *Hurzy y colaboradores*, en la tercera columna la estructura del análogo correspondiente de N-PTindoles junto con un rectángulo que representa la modificación realizada según la **Figura 3** y en la última columna los valores de IC50 y su código smiles.

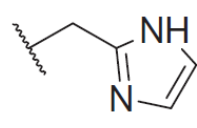
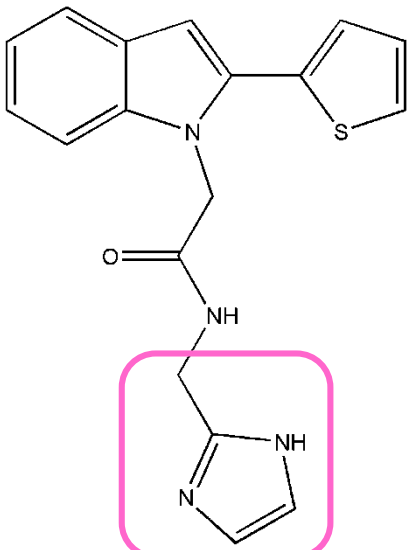
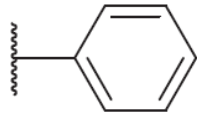
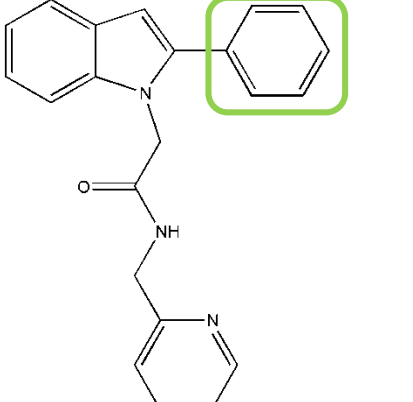
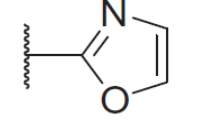
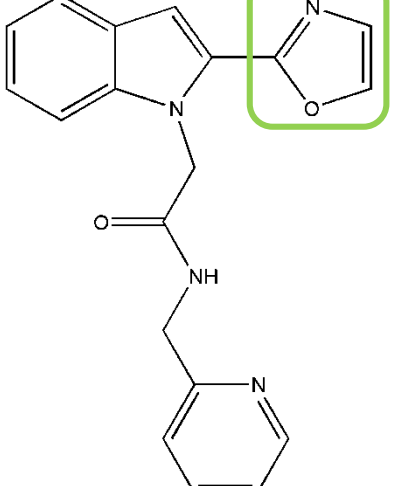
La tercera columna nos permitió analizar cada uno de los análogos de N-PTindoles de acuerdo a las modificaciones que sufrieron en A, B, C y D. Al comparar las estructuras con modificaciones en el mismo sitio, fue curioso encontrar que las IC50 varían de manera diferente al modificar ligeramente la estructura, por ejemplo, en el caso de las modificaciones en D las estructuras D29 y D27 presentan una diferencia significativa con respecto a sus IC50 experimentales cuyos valores son 70 nM y de >81,010 nM respectivamente, donde la única diferencia estructural se encuentra en el imidazol, uno está alquilado (D27) mientras el otro tiene un benceno unido como biciclo (D29) la posible explicación es que el benceno le da mayor peso molecular, volumen, aromaticidad, hidrofobicidad lo que tal vez esté relacionado con la actividad biológica. Otro ejemplo es B35 que tiene un IC50 de 662 nM mientras que B43 tiene un IC50 de >81,010 nM la diferencia se encuentra en que B35 tiene un tiofeno alquilado y B43 una piridina, nuevamente se ve reflejada la implicación de efecto estérico y electrónico de las moléculas en la actividad por lo tanto serán puntos clave para la toma de decisión en las elecciones de los descriptores moleculares que se utilizarán para el modelo QSAR.

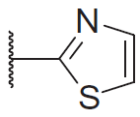
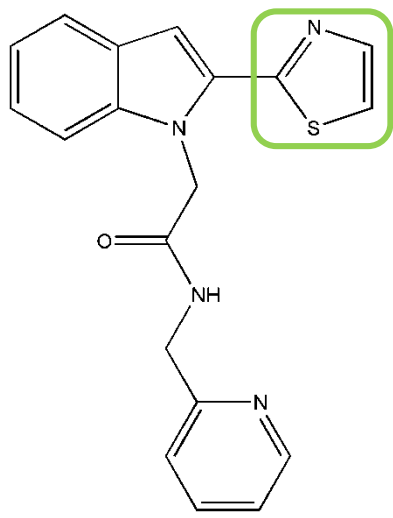
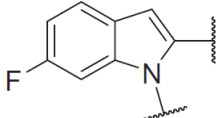
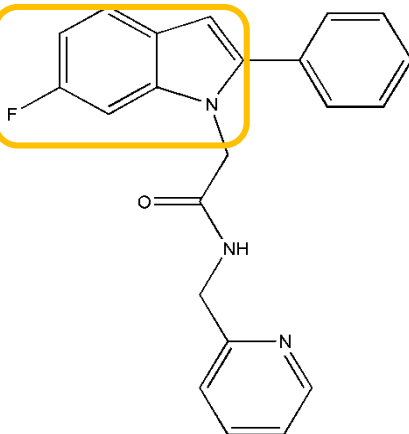
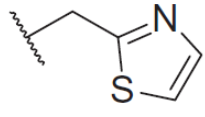
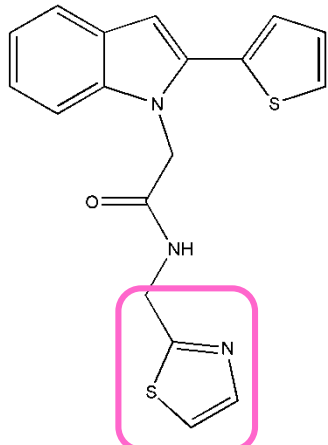
Tabla 6: Lista de las 40 estructuras inhibitoras de TrkA y actividades biológicas.

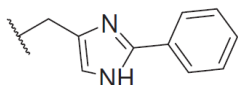
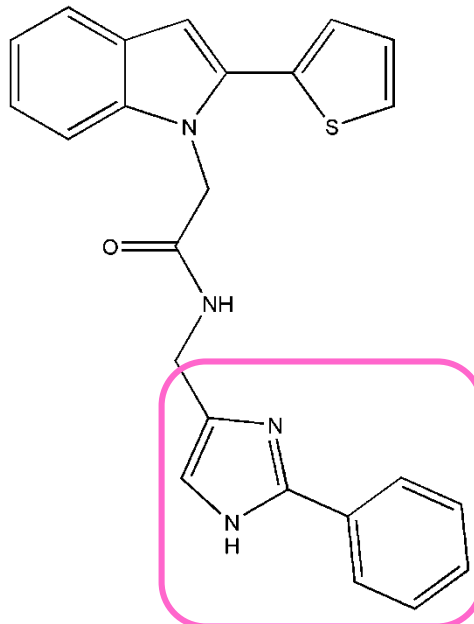
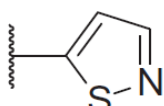
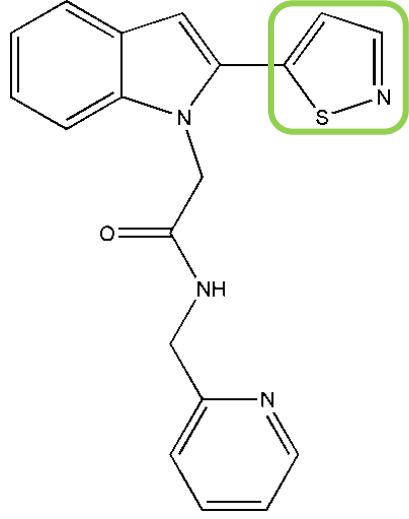
Posición de mayor a menor actividad	Clase de modificación por letra (y su número de identificación en [5])	Estructura	TrkA IC50 (nM) y SMILES
1	 <p>D29</p>		70 <chem>O=C(NCC4=NC(C=CC=C5)=C5N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem>
2	 <p>D28</p>		210 <chem>O=C(NCC4=NC=C(C5=CC=CC=C5)N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem>

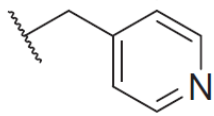
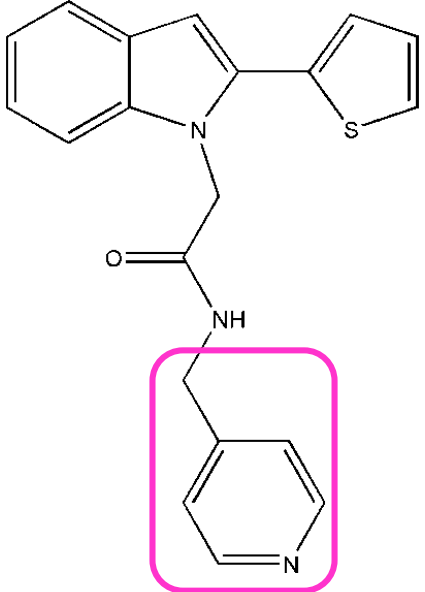
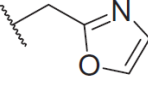
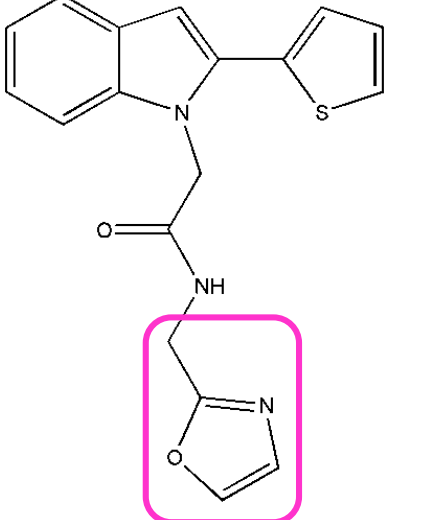
<p>3</p> <p>D34</p> 			<p>218</p> <chem>O=C(NCC4=CN(C=CC=C5)C5=N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem>
<p>4</p> <p>B35</p> 			<p>662</p> <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CC=C(C)S3</chem>
<p>5</p> <p>A51</p> 			<p>880</p> <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=C(F)C=C1C=C2C3=CC=CC=C3</chem>

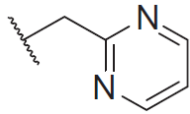
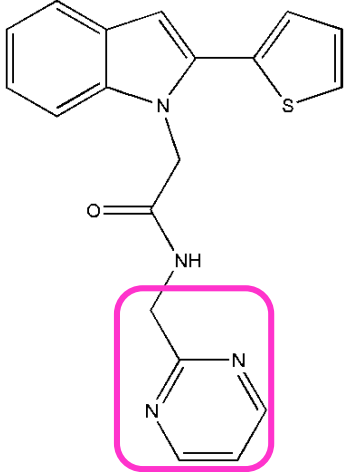
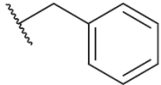
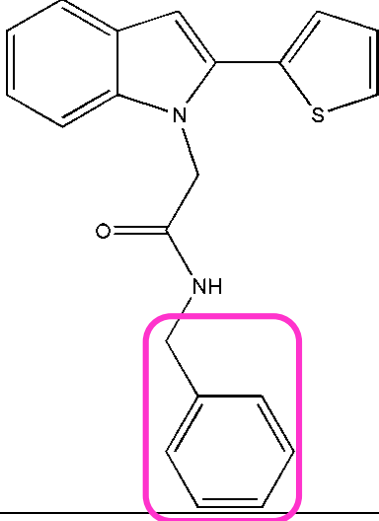
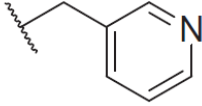
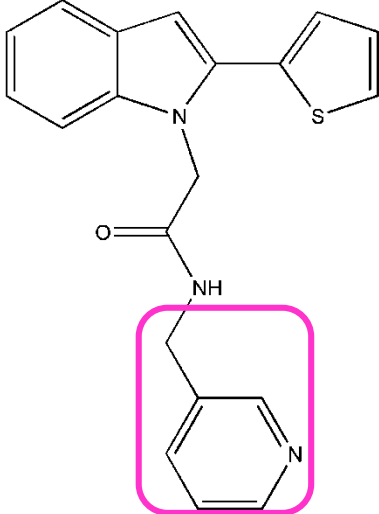
6	 <p>A50</p>		960 <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC(F)=C1C=C2C3=CC=CC=C3</chem>
7	 <p>B40</p>		961 <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=NSC=C3</chem>

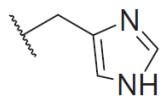
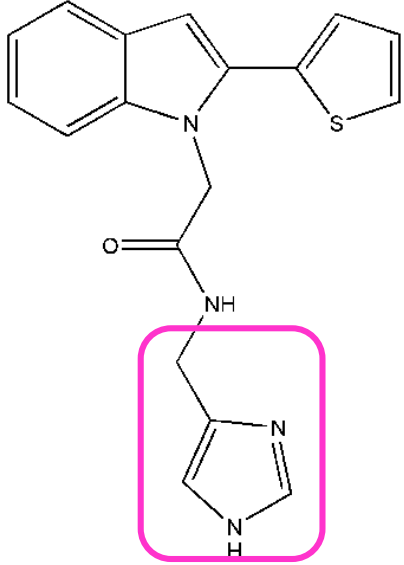
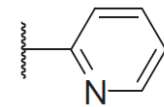
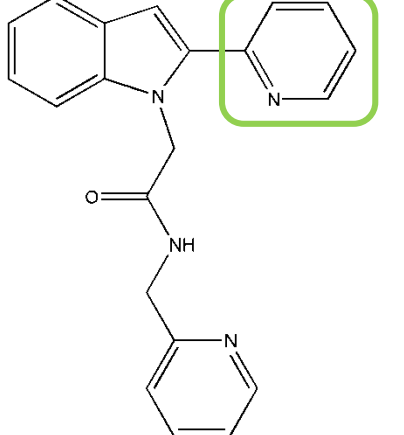
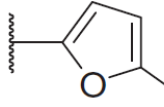
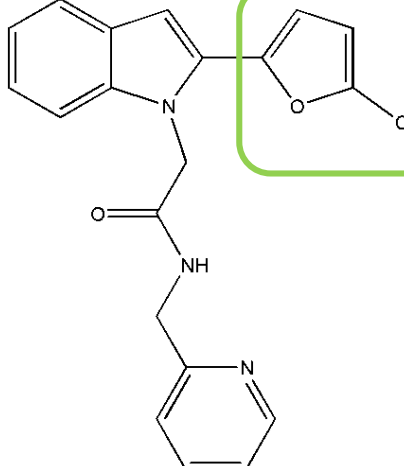
<p>8</p> <p>D26</p> 		<p>1300</p> <p><chem>O=C(NCC4=NC=CN4)C</chem> <chem>N2C1=CC=CC=C1C=C2</chem> <chem>C3=CC=CS3</chem></p>
<p>9</p> <p>B45</p> 		<p>1424</p> <p><chem>O=C(NCC4=CC=CC=N4)C</chem> <chem>CN2C1=CC=CC=C1C=C2</chem> <chem>2C3=CC=CC=C3</chem></p>
<p>10</p> <p>B41</p> 		<p>2114</p> <p><chem>O=C(NCC4=CC=CC=N4)C</chem> <chem>CN2C1=CC=CC=C1C=C2</chem> <chem>2C3=NC=CO3</chem></p>

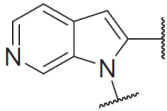
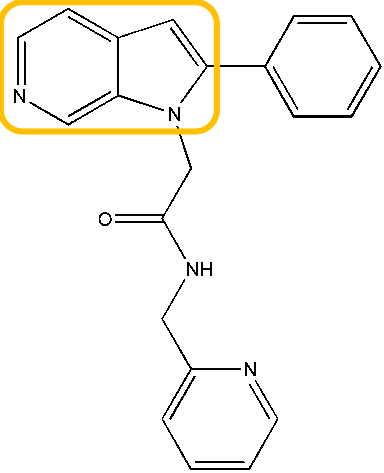
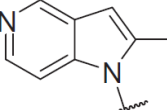
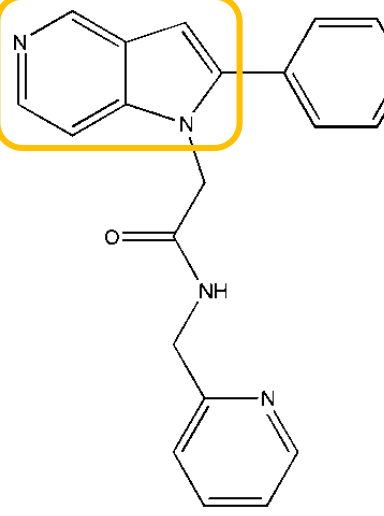
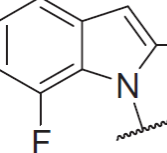
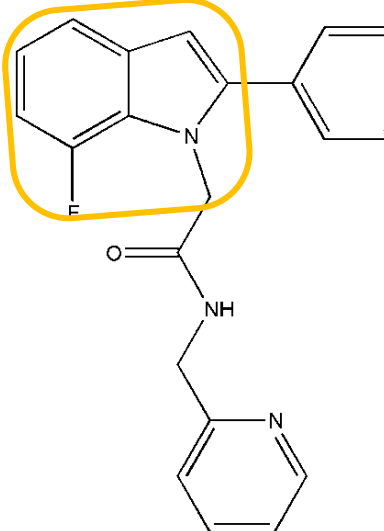
11	 <p>B39</p>		<p>2400</p> <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=NC=CS3</chem>
12	 <p>A52</p>		<p>2552</p> <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC(F)=CC=C1C=C2C3=CC=CC=C3</chem>
13	 <p>D30</p>		<p>3163</p> <chem>O=C(NCC4=NC=CS4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem>

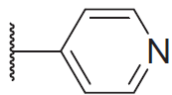
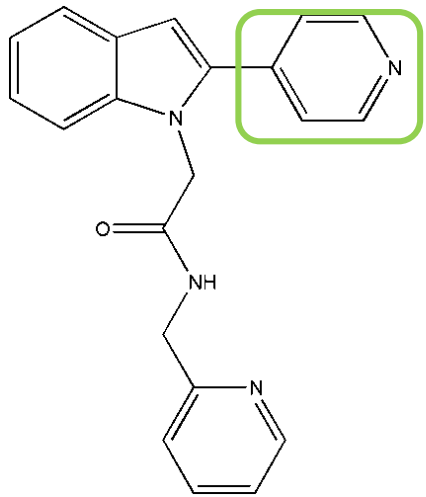
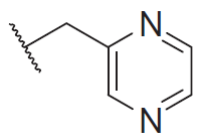
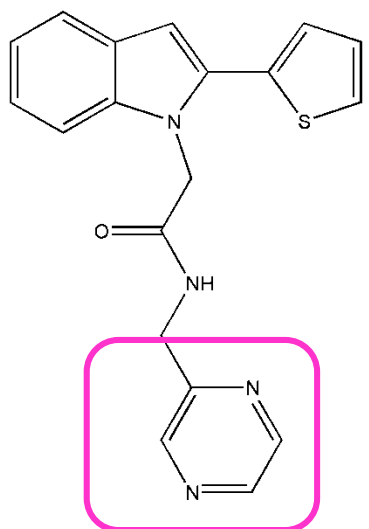
<p>14</p>	<p>D33</p> 		<p>3388</p> <chem>O=C(NCC4=CNC(C5=CC=CC=C5)=N4)CN2C1=CC=CC=C1C=C2C3=CC=CC=C3</chem>
<p>15</p>	<p>B37</p> 		<p>3944</p> <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CC=CC=N3</chem>

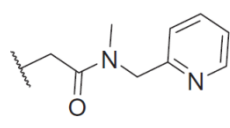
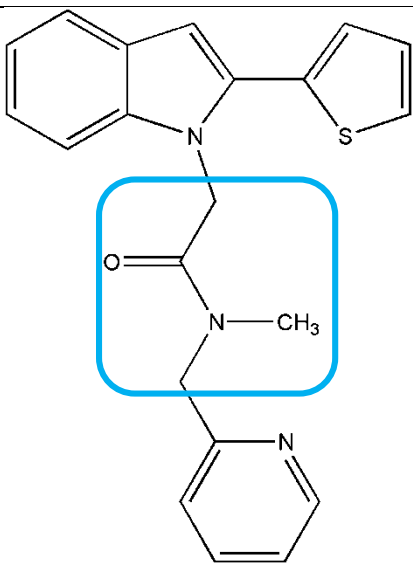
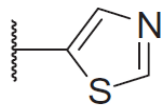
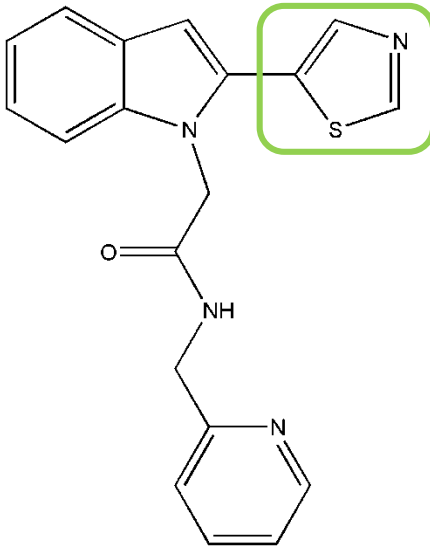
<p>16</p> <p>D22</p> 		<p>4394</p> <p><chem>O=C(NCC4=CC=NC=C4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>
<p>17</p> <p>D31</p> 		<p>7310</p> <p><chem>O=C(NCC4=NC=CO4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>

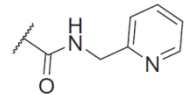
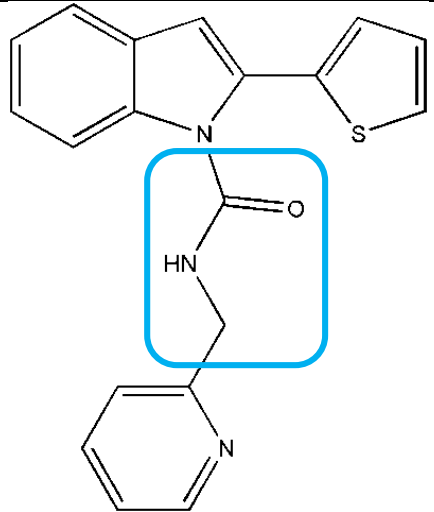
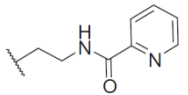
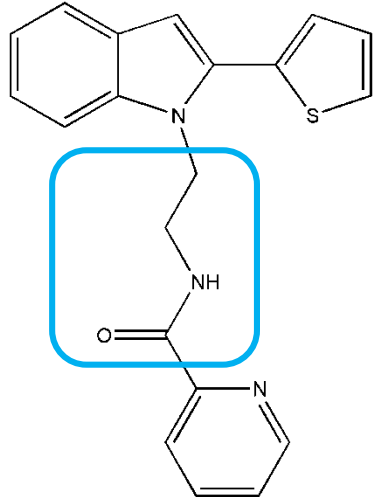
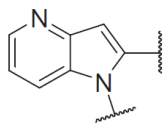
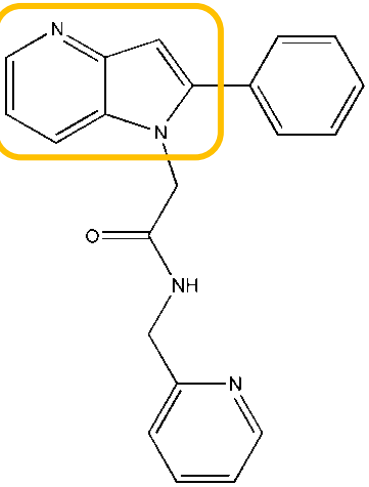
18	<p>D23</p> 		<p>8829 <chem>O=C(NCC4=NC=CC=N4)</chem> <chem>CN2C1=CC=CC=C1C=C</chem> <chem>2C3=CC=CS3</chem></p>
19	<p>D25</p> 		<p>>9001 <chem>O=C(NCC4=CC=CC=C4)</chem> <chem>CN2C1=CC=CC=C1C=C</chem> <chem>2C3=CC=CS3</chem></p>
20	<p>} D21</p> 		<p>9085 <chem>O=C(NCC4=CC=CN=C4)</chem> <chem>CN2C1=CC=CC=C1C=C</chem> <chem>2C3=CC=CS3</chem></p>

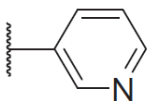
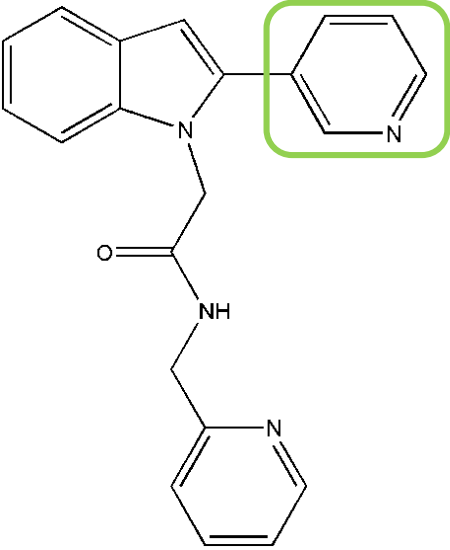
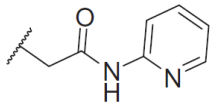
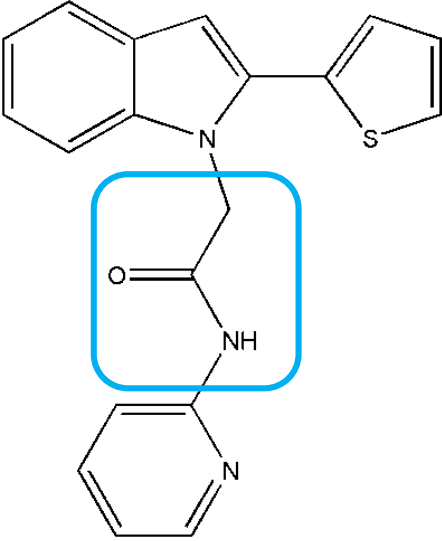
<p>21</p> <p>D32</p> 		<p>9400</p> <chem>O=C(NCC4=CNC=N4)C</chem> <chem>N2C1=CC=CC=C1C=C2</chem> <chem>C3=CC=CS3</chem>
<p>22</p> <p>B42</p> 		<p>11,890</p> <chem>O=C(NCC4=CC=CC=N4)C</chem> <chem>CN2C1=CC=CC=C1C=C2</chem> <chem>2C3=NC=CC=C3</chem>
<p>23</p> <p>B36</p> 		<p>12,970</p> <chem>O=C(NCC4=CC=CC=N4)C</chem> <chem>CN2C1=CC=CC=C1C=C2</chem> <chem>2C3=CC=C(C)O3</chem>

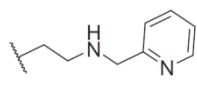
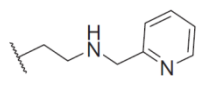
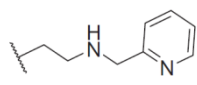
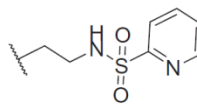
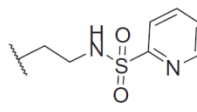
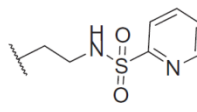
24	 <p>A48</p>		14,000 <chem>O=C(NCC4=CC=CC=N4)CN2C1=CN=CC=C1C=C2C3=CC=CC=C3</chem>
25	 <p>A47</p>		22,000 > <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=NC=C1C=C2C3=CC=CC=C3</chem>
26	 <p>A53</p>		>27,000 <chem>O=C(NCC4=CC=CC=N4)CN2C1=C(F)C=CC=C1C=C2C3=CC=CC=C3</chem>

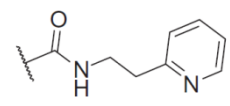
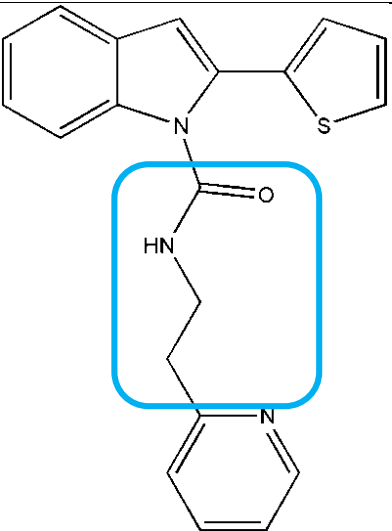
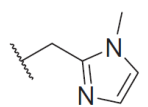
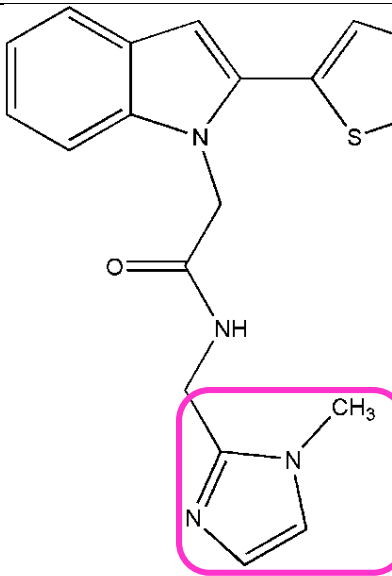
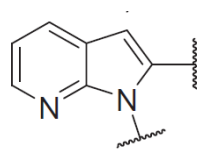
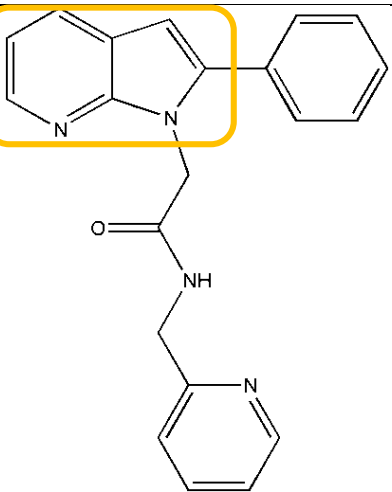
27	 <p>B44</p>		<p>30,870 <chem>O=C(NCC4=CC=CC=N4)</chem> <chem>CN2C1=CC=CC=C1C=C</chem> <chem>2C3=CC=NC=C3</chem></p>
28	 <p>D24</p>		<p>36,600 <chem>O=C(NCC4=CN=CC=N4)</chem> <chem>CN2C1=CC=CC=C1C=C</chem> <chem>2C3=CC=CS3</chem></p>

<p>29</p>	 <p>C20</p>		<p>47,300 <chem>O=C(N(C)CC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>
<p>30</p>	 <p>B38</p>		<p>49,000 <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>

<p>31</p>	 <p>C15</p>		<p>64,320 <chem>O=C(NCC4=CC=CC=N4)</chem> <chem>N2C1=CC=CC=C1C=C2</chem> <chem>C3=CC=CS3</chem></p>
<p>32</p>	 <p>C17</p>		<p>74,020 <chem>O=C(C4=CC=CC=N4)NC</chem> <chem>CN2C1=CC=CC=C1C=C2</chem> <chem>2C3=CC=CS3</chem></p>
<p>33</p>	 <p>A46</p>		<p>>81,010 <chem>O=C(NCC4=CC=CC=N4)</chem> <chem>CN2C1=CC=CN=C1C=C2</chem> <chem>2C3=CC=CC=C3</chem></p>

<p>34</p>	<p>B43</p> 		<p>>81,010 <chem>O=C(NCC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CN=CC=C3</chem></p>
<p>35</p>	<p>C14</p> 		<p>>81,010 <chem>O=C(NC4=CC=CC=N4)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>

<p>36</p> <p>C16</p> 			<p>>81,010</p> <chem>C12=CC=CC=C1C=C(C4=CC=CS4)N2CCNCC3=C</chem> <chem>C=CC=N3</chem>
<p>37</p> <p>C18</p> 			<p>>81,010</p> <chem>O=S(NCCN2C1=CC=CC=C1C=C2C3=CC=CS3)(C4=CC=CC=N4)=O</chem>

<p>38</p>  <p>C19</p>		<p>>81,010 <chem>O=C(NCCC4=CC=CC=N4)N2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>
<p>39</p>  <p>D27</p>		<p>>81,010 <chem>O=C(NCC4=NC=CN4C)CN2C1=CC=CC=C1C=C2C3=CC=CS3</chem></p>
<p>40</p>  <p>A49</p>		<p>99,820 <chem>O=C(NCC4=CC=CC=N4)CN2C1=NC=CC=C1C=C2C3=CC=CC=C3</chem></p>

7.3 Calculo de descriptores

Se obtuvieron 1732 descriptores los cuales 28 descriptores se calcularon con el programa Vega ZZ [39], 40 descriptores de Swiss ADME [42] y 1664 de E-dragon [41] véase suplemento.

Se realizó un acoplamiento molecular de los 40 análogos de N-PTindoles con el programa Autodock 4 [52] para obtener la energía de interacción y utilizar este parámetro como un descriptor molecular para los modelos QSAR en la **Tabla 7** se observan los resultados.

El acoplamiento molecular nos proporcionó los aminoácidos que se mantienen presentes en el sitio activo de todas las moléculas análogas de N-PTindoles y esta información nos dio un mejor panorama para la toma de decisiones de los IDVs que se utilizaran en la ecuación final.

Por otro lado, al analizar los datos del acoplamiento molecular se observaron unas discrepancias en el método, ya que al contar con los datos experimentales se esperaba observar algo similar con la predicción de las energías de interacciones dadas por el programa. Con otras palabras, las moléculas con menor IC₅₀ debían tener las menores energías de interacción y constantes de inhibición. Sin embargo, no en todas las moléculas se observó, por ejemplo, en la molécula A51 su IC₅₀ experimental es de 880 nM y cuenta con la energía de interacción teórica más alta de todos los análogos de N-PTindoles con un valor de -12.16 kcal/mol y una constante de inhibición de 1.23 nM lo que sugiere que teóricamente la molécula A51 es la mejor molécula inhibidora de TrkA, pero experimentalmente es D29 con un IC₅₀ de 70nM. Otro ejemplo es A41 que tiene un IC₅₀ experimental de 99820 nM, una energía de interacción de -11.92 kcal/mol y una constante de inhibición de 1.82nM lo que indicaría teóricamente que es un buen inhibidor, sin embargo, es la peor molécula experimentalmente.

Otro punto que se observó al analizar el acoplamiento molecular es que las estructuras que tienen modificaciones en A cuentan con una energía de interacción similar lo que indica que el programa en su logaritmo pudiera estar tomando además de las interacciones y parámetros electrónicos también estéricos. Esto se debe comprobar con otras investigaciones ya que no es la finalidad de esta investigación.

Tabla 7: Acoplamiento molecular de las 40 moléculas análogas de N-PTindoles.

No. Posición por actividad	ID de referencia	TrkA IC 50 (nM)	TrkAIC50 (μ M)	DG o LBE (kcal/mol)	KI (nM)
1	D29	70	0.07	-11.81	2.22
2	D28	210	0.21	-10.74	13.42
3	D34	218	0.218	-11.45	4.03
4	B35	662	0.662	-12.12	1.3
5	A51	880	0.88	-12.16	1.23
6	A50	960	0.96	-11.77	2.35
7	B40	961	0.961	-11.48	3.83
8	D26	1300	1.3	-10.71	14.07
9	B45	1424	1.424	-11.65	2.88
10	B41	2114	2.114	-10.7	14.41
11	B39	2400	2.4	-11.11	7.17
12	A52	2552	2.552	-11.78	2.33
13	D30	3163	3.163	-10.42	23.09
14	D33	3388	3.388	-10.22	32.37
15	B37	3944	3.944	-11.06	7.75
16	D22	4394	4.394	-10.55	18.44
17	D31	7310	7.31	-10.12	37.89
18	D23	8829	8.829	-10.95	9.45

19	D25	9001	9.001	-11.06	7.77
20	D21	9085	9.085	-10.58	17.48
21	D32	9400	9.4	-10.38	24.45
22	B42	11890	11.89	-11.47	3.92
23	B36	12970	12.97	-10.65	15.5
24	A48	14000	14	-11.16	6.63
25	A47	22000	22	-11.41	4.31
26	A53	27000	27	-11.35	4.78
27	B44	30870	30.87	-11.26	5.58
28	D24	36600	36.6	-11.05	7.99
29	C20	47300	47.3	-10.77	12.83
30	B38	49000	49	-11.09	7.38
31	C15	64320	64.32	-9.36	138.45
32	C17	74020	74.02	-10.82	11.62
33	A46	81010	81.01	-11.23	5.87
34	B43	81010	81.01	-11.59	3.22
35	C14	81010	81.01	-10.21	32.72
36	C16	81010	81.01	-10.19	34.17
37	C18	81010	81.01	-7.4	3.74
38	C19	81010	81.01	-10.65	15.74
39	D27	81010	81.01	-9.92	53.26
40	A49	99820	99.82	-11.92	1.82
Referencia	Referencia			-11.16	6.56

7.4 Selección de descriptores

La elección de los IDVs para modelos QSAR de 9 y 8 descriptores de los 1733 descriptores fue por justificación teórica y por utilización de métodos estadísticos.

Se tomó en cuenta que los IDVs seleccionados deben tener una justificación teórica que garantice la relación con la actividad biológica con su receptor, ya que son estructuras que actuaran como posibles fármacos inhibidores del dolor. Los ligandos análogos de los N-PTindoles tienen un sitio activo en la región intracelular del TrkA lo biológicamente se refiere a que las moléculas deben tener propiedades lipofílicas para poder atravesar membranas por lo que un buen descriptor biológico es el LogP o coeficiente lípido/ agua, otra consideración es que los aminoácidos involucrados con el sitio activo tienen interacciones de tipo enlaces de hidrógeno, fuerzas electrostáticas y acido-base por lo tanto un descriptor que ayude a la predicción de la actividad biológica son Donadores de Hidrógeno, la polaridad, carga, etc. [53-55]

Los descriptores moleculares seleccionados para un modelo QSAR tienen que ser fáciles de manipular experimentalmente, ya que al final este modelo nos permite decidir si aumentar o disminuir un valor de un descriptor la actividad biológica puede aumentar. Los descriptores que se pueden interpretar y modificar químicamente de forma más sencilla son masa molecular, volumen, donadores de hidrógeno, polaridad, pKa, hidrofobicidad, entre otros IDVs fisicoquímicos, mientras que los descriptores cuánticos son difíciles de manipular de forma experimental.

Para la selección se revisó cada uno de los descriptores obtenidos de VegaZZ [39], Swiss ADME [42] y E- dragon [41] y se eliminaron los que no proporcionaban ninguna información representativa la ecuación. Se eligieron 30 descriptores moleculares de los 1733 véase la **Tabla 8**

donde se resume la información de cada uno de ellos y en la **Tabla 9** se muestran los resultados obtenidos de los diferentes programas.

Tabla 8: Descriptores moleculares

Descriptor molecular/IDV	Información	Programa utilizado
DG o LBE	Energía libre de interacción obtenida después de realizar acoplamiento molecular.	Autodock 4 [52]
ESOL	Estimación de la solubilidad acuosa directamente de la estructura molecular.	Swiss ADME [42]
Log P	Log P parámetro que representa la medida en que una molécula prefiere un medio hidrofílico (índice bajo de lipofilidad) o hidrofóbico (alto índice de lipofilidad). Este parámetro determina la capacidad de un fármaco de alcanzar su objetivo y cuánto tiempo permanecerá activo en el organismo. Es calculado usando el método silicos-IT.	Swiss ADME [42]
Log SW	Es el logaritmo negativo de la solubilidad en agua del compuesto (log Sw). Solubilidad calculada utilizando el método silicos-IT.	Swiss ADME [42]

FractionCsp3	Relación de los carbonos hibridados sp ³ sobre el recuento total de carbono de la molécula	Swiss ADME [42]
LogKp	Permeación de la piel	Swiss ADME [42]
Se	Suma total de las electronegatividades de la estructura.	E-DRAGON [41]
Ss	Suma de estados electrofílicos.	E-DRAGON [41]
Sv	Suma total de volúmenes atómicos de Van Der Waals.	E-DRAGON [41]
POL	Número de Polaridad	E-DRAGON [41]
IAC	Información total de la composición atómica.	E-DRAGON [41]
UNIP	unipolaridad	E-DRAGON [41]
IDM	Información de la magnitud de la distancia.	E-DRAGON [41]
HOMT	HOMA TOTAL Modelo de oscilador armónico del índice de aromaticidad Indica la importancia de la aromaticidad del andamio	E-DRAGON [41]

HBD o H-Don	Número de átomos donadores de hidrógeno unidos a nitrógeno y oxígeno.	VegaZZ [39]
Log PC	Logaritmo negativo de la distribución del componente en octano/agua utilizando el método de radio aromático de Ghose/Crippen.	VegaZZ [39]
MW	Masa molecular	VegaZZ [39]
MR	Polarizabilidad total de un mol de sustancia	VegaZZ [39]
PSA	Área de superficie polar	VegaZZ [39]
Dipolo	Momento dipolar	VegaZZ [39]
VOL	Volumen	VegaZZ [39]
TPSA	Área de superficie polar total	VegaZZ [39]

Tabla 9: Valores de los 30 IDVs más representativos para la generación de los modelos QSAR con 9 y 8 descriptores

ID de referencia	TrkAIC50	LBE	MW	VOL	Sv	HOMT
D29	0.07	-11.81	386.5	336.4	31.76	18.974
D28	0.21	-10.74	412.5	374.7	34.36	18.987
D34	0.22	-11.45	386.1	338.6	31.76	18.178
B35	0.66	-12.12	361.1	331.2	30.36	15.583
A51	0.88	-12.16	359.4	324	30.39	18.521
A50	0.96	-11.77	359.4	324.6	30.39	18.515
B40	0.96	-11.48	348.1	306.7	28.16	15.865
D26	1.3	-10.71	336.4	292.1	27.16	13.99
B45	1.42	-11.65	341.4	317.8	30.28	18.425
B41	2.11	-10.7	332.4	298.1	27.59	15.299
B39	2.4	-11.11	348.4	310.7	28.16	15.738
A52	2.55	-11.78	359.4	323.4	30.39	18.433
D30	3.16	-10.42	353.5	296.9	27.26	12.961
D33	3.39	-10.22	412.5	370.3	34.36	18.812
B37	3.94	-11.06	348.4	308.5	28.16	15.782
D22	4.39	-10.55	347.4	315	28.77	15.583
D31	7.31	-10.12	337.4	288.9	26.68	12.052
D23	8.83	-10.95	348.4	310.1	28.16	15.572
D25	9	-11.06	346.4	318.9	29.37	15.589
D21	9.09	-10.58	347.4	313.7	28.77	15.58
D32	9.4	-10.38	336.4	294.5	27.16	13.987
B42	11.89	-11.47	342.4	314.6	29.67	18.488
B36	12.97	-10.65	345.4	320.5	29.79	15.173
A48	14	-11.16	342.4	314.5	29.67	19.636

A47	22	-11.41	342.4	314.5	29.67	18.562
A53	27	-11.35	359.4	323.8	30.39	18.454
B44	30.87	-11.26	342.4	314.8	29.67	18.403
D24	36.6	-11.05	348.4	310.2	28.16	15.576
C20	47.3	-10.77	361.5	330.3	30.36	15.583
B38	49	-11.09	348.4	308.6	28.16	16.551
C15	64.32	-9.36	333.4	304.5	27.17	17.126
C17	74.02	-10.82	347.4	315.3	28.77	14.926
A46	81.01	-11.23	342.4	314.6	29.67	19.742
B43	81.01	-11.59	342.4	313.9	29.67	18.401
C14	81.01	-10.21	333.4	302.7	27.17	15.612
C16	81.01	-10.19	333.5	310.6	28.85	15.583
C18	81.01	-7.4	383.5	327.2	29.37	15.583
C19	81.01	-10.65	347.4	315	28.77	17.126
D27	81.01	-9.92	350.4	310.4	28.76	13.99
A49	99.82	-11.92	342.4	315.4	29.67	18.565

ID de referencia	Ilogp	TPSA	ESOL	ALOGPS	SA	AliSol
D29	2.64	90.95	4.67E-03	4.32	3.14	1.14E-03
D28	2.19	90.95	2.77E-03	4.53	3.41	6.88E-04
D34	3.07	79.57	3.16E-03	4.3	3.17	1.04E-03
B35	2.11	75.16	9.67E-03	3.96	3.19	4.07E-03
A51	2.54	46.92	9.90E-03	3.75	2.77	1.62E-02
A50	2.99	46.92	9.90E-03	3.64	2.93	1.62E-02
B40	2.7	88.05	4.61E-02	3.16	2.92	2.35E-02
D26	1.28	90.95	6.96E-02	2.99	2.91	3.03E-02
B45	2.82	46.92	1.34E-02	4.06	2.68	1.96E-02

B41	2.54	72.95	1.42E-01	2.34	2.94	2.20E-01
B39	2.3	88.05	4.82E-02	2.71	2.99	2.53E-02
A52	2.98	46.92	9.90E-03	3.76	2.86	1.62E-02
D30	2.59	103.4	1.55E-02	3.56	3.06	2.03E-03
D33	2.96	90.95	2.77E-03	4.73	3.34	6.88E-04
B37	2.68	88.05	4.61E-02	2.77	2.98	2.35E-02
D22	2.56	75.16	2.01E-02	3.38	2.95	1.09E-02
D31	2.66	88.3	4.59E-02	3.17	2.99	1.77E-02
D23	2.82	88.05	4.82E-02	2.87	2.96	2.53E-02
D25	3.19	62.27	4.31E-03	4.91	2.96	1.58E-03
D21	2.68	75.16	2.01E-02	3.34	2.93	1.09E-02
D32	1.97	90.95	6.96E-02	3.01	2.9	3.03E-02
B42	2.7	59.81	5.97E-02	3.15	2.76	1.26E-01
B36	2.8	60.06	2.82E-02	3.45	3.2	3.47E-02
A48	2.6	59.81	6.24E-02	2.72	2.73	1.36E-01
A47	2.6	59.81	6.24E-02	2.66	2.66	1.36E-01
A53	3.11	46.92	9.90E-03	3.66	2.82	1.62E-02
B44	2.55	59.81	6.24E-02	2.83	2.61	1.36E-01
D24	2.24	88.05	8.99E-02	2.74	2.98	7.06E-02
C20	2.95	66.37	1.31E-02	3.95	3.08	1.03E-02
B38	2.42	88.05	4.82E-02	2.56	3.16	2.53E-02
C15	2.59	75.16	1.24E-02	3.92	3.02	5.12E-03
C17	1.9	75.16	1.30E-02	3.88	2.9	5.34E-03
A46	2.57	59.81	5.47E-02	3.08	2.75	1.10E-01
B43	2.51	59.81	6.24E-02	2.8	2.86	1.36E-01
C14	2.57	75.16	1.07E-02	3.84	2.95	4.03E-03
C16	3.18	58.09	1.95E-02	4.3	3.02	1.93E-02
C18	2.49	100.61	1.42E-02	3.03	3.33	3.61E-03
C19	2.89	75.16	6.68E-03	4.08	3.11	1.78E-03

D27	2.22	80.09	6.64E-02	3.08	3.02	5.88E-02
A49	2.78	59.81	3.86E-02	3.09	2.68	6.17E-02

ID de referencia	Pol	nHDon	IAC	ISIZ	Se	MLOGP2
D29	41	2	72.9	254.1	46	8.7
D28	43	2	77.7	282.2	49.9	12.8
D34	41	1	72.9	254.1	46	8.7
B35	37	1	69.4	247.1	44.8	7.7
A51	40	1	69.2	247.1	45.2	9.2
A50	41	1	69.2	247.1	45.2	9.2
B40	35	1	66.9	219.7	41.1	2.4
D26	32	2	65.8	212.9	40.1	4.8
B45	38	1	63.6	247.1	44.7	7.1
B41	35	1	64.9	219.7	41.4	2.5
B39	35	1	66.9	219.7	41.1	2.4
A52	40	1	69.2	247.1	45.2	9.2
D30	32	1	65.7	206.1	39.1	4.7
D33	43	2	77.7	282.2	49.9	12.8
B37	35	1	66.9	219.7	41.1	2.4
D22	35	1	65.8	226.5	41.9	6.6
D31	32	1	65.7	206.1	39.3	4.8
D23	35	1	66.9	219.7	41.1	5.6
D25	35	1	64	233.3	42.7	12.9
D21	35	1	65.8	226.5	41.9	6.6
D32	32	2	65.8	212.9	40.1	4.8
B42	38	1	64.9	240.2	43.9	3.7
B36	37	1	67.4	247.1	45	4
A48	38	1	64.9	240.2	43.9	2.7

A47	38	1	64.9	240.2	43.9	2.7
A53	41	1	69.2	247.1	45.2	9.2
B44	38	1	64.9	240.2	43.9	2.7
D24	35	1	66.9	219.7	41.1	2.4
C20	38	0	69.4	247.1	44.8	7.7
B38	35	1	66.9	219.7	41.1	2.4
C15	35	1	62.1	206.1	39	7.5
C17	36	1	65.8	226.5	41.9	8
A46	38	1	64.9	240.2	43.9	2.7
B43	38	1	64.9	240.2	43.9	2.7
C14	34	1	62.1	206.1	39	11.6
C16	33	1	61.4	233.3	42.5	9.2
C18	39	1	74.4	233.3	43.3	4.6
C19	36	1	65.8	226.5	41.9	8.8
D27	35	1	69.6	233.3	43	5.9
A49	38	1	64.9	240.2	43.9	6.1

ID de referencia	WLOGP	logPC	MR	FCsp3	logKp	ALOGPS_logS
D29	4.41	3.4406	113.27	0.09	-5.87	-4.42
D28	4.92	3.8128	121.2	0.08	-5.86	-4.77
D34	4.18	4.2195	112.37	0.09	-5.68	-4.55
B35	4.24	3.2782	106.38	0.14	-5.89	-4.34
A51	4.43	3.918	103.49	0.09	-5.89	-5.3
A50	4.43	3.918	103.49	0.09	-5.89	-5.23
B40	3.32	3.32	99.21	0.11	-6.53	-4
D26	3.26	2.3573	95.76	0.11	-6.58	-3.87
B45	3.87	3.7599	103.53	0.09	-5.85	-4.92
B41	2.86	1.7859	93.59	0.11	-6.89	-3.68

B39	3.32	2.3519	99.21	0.11	-6.55	-4.03
A52	4.43	3.918	103.49	0.09	-5.89	-5.26
D30	3.99	2.7227	99.29	0.11	-6.05	-4.08
D33	4.92	4.4561	121.2	0.08	-5.86	-4.77
B37	3.32	3.32	99.21	0.11	-6.53	-4.03
D22	3.93	2.7936	101.41	0.1	-6.11	-4.21
D31	3.52	2.1567	93.68	0.11	-6.39	-3.65
D23	3.32	2.2683	99.21	0.11	-6.55	-4.05
D25	4.53	4.1307	103.62	0.1	-5.35	-4.33
D21	3.93	2.7936	101.41	0.1	-6.11	-4.21
D32	3.26	2.13	95.76	0.11	-6.58	-3.89
B42	3.26	2.7286	101.33	0.1	-6.59	-4.49
B36	3.77	2.7122	100.77	0.14	-6.23	-4.04
A48	3.26	2.4228	101.33	0.1	-6.61	-4.53
A47	3.26	2.4228	101.33	0.1	-6.61	-4.51
A53	4.43	3.918	103.49	0.09	-5.89	-5.2
B44	3.26	2.4228	101.33	0.1	-6.61	-4.51
D24	3.32	1.8801	99.21	0.11	-6.86	-4.05
C20	4.27	3.1528	106.31	0.14	-6.04	-4.08
B38	3.32	2.0763	99.21	0.11	-6.55	-4.05
C15	4.37	3.9858	96.99	0.05	-5.81	-4.22
C17	4.19	3.5654	101.63	0.1	-5.9	-4.2
A46	3.26	2.7286	101.33	0.1	-6.55	-4.48
B43	3.26	2.4228	101.33	0.1	-6.61	-4.51
C14	4.21	3.4412	98.14	0.05	-5.74	-4.11
C16	4.4	3.6881	101.21	0.15	-5.96	-4.23
C18	4.82	4.82	104.53	0.11	-6.34	-3.8
C19	4.57	4.3903	101.8	0.1	-5.57	-4.37
D27	3.27	2.8407	100.66	0.16	-6.7	-3.48

A49	3.26	3.1673	101.33	0.1	-6.38	-4.52
-----	------	--------	--------	-----	-------	-------

ID de referencia	IDM	Ss	HA	AROM	AMR	XLOGP3	RB
D29	8.311	59.17	28	0.783	111.843	3.92	6
D28	8.5	63.17	30	0.785	120.901	4.16	7
D34	8.311	59	28	0.78	111.551	4.19	6
B35	8.109	55	26	0.765	105.768	3.68	6
A51	8.222	63.17	27	0.943	102.309	3.67	6
A50	8.225	63.17	27	0.943	102.309	3.67	6
B40	7.993	54.33	25	0.791	99.606	2.67	6
D26	7.879	51.83	24	0.756	95.887	2.49	6
B45	8.11	55.5	26	0.943	102.092	3.57	6
B41	7.993	56	25	0.931	92.279	2.02	6
B39	7.993	54.33	25	0.755	98.722	2.64	6
A52	8.225	63.17	27	0.943	102.309	3.67	6
D30	7.879	51.17	24	0.693	100.085	3.39	6
D33	8.5	63.17	30	0.785	120.652	4.16	7
B37	7.993	54.33	25	0.775	99.855	2.67	6
D22	7.993	53.33	25	0.765	101.299	3.25	6
D31	7.879	52.83	24	0.754	93.642	2.77	6
D23	7.993	54.33	25	0.758	97.28	2.64	6
D25	7.993	52.33	25	0.772	103.455	4.32	6
D21	7.993	53.33	25	0.765	101.299	3.25	6
D32	7.879	51.83	24	0.756	96.038	2.49	6
B42	8.11	56.5	26	0.933	99.563	2.53	6
B36	8.109	56.67	26	0.936	99.324	3.07	6
A48	8.11	56.5	26	0.941	99.935	2.5	6
A47	8.11	56.5	26	0.932	99.935	2.5	6

A53	8.226	63.17	27	0.943	102.309	3.67	6
B44	8.11	56.5	26	0.931	99.935	2.5	6
D24	7.993	54.33	25	0.758	98.77	2.21	6
C20	8.117	54.83	26	0.765	105.823	3.47	6
B38	7.993	54.33	25	0.769	99.123	2.64	6
C15	7.887	51.83	24	0.766	96.215	3.55	5
C17	7.99	53.33	25	0.766	100.927	3.55	6
A46	8.11	56.5	26	0.942	99.563	2.59	6
B43	8.11	56.5	26	0.931	99.935	2.5	6
C14	7.888	51.83	24	0.766	96.777	3.65	5
C16	7.862	46.17	24	0.765	100.697	3.34	6
C18	8.112	59.58	26	0.765	105.189	3.24	6
C19	7.991	53.33	25	0.766	100.765	4.01	6
D27	7.998	53.33	25	0.756	100.784	2.45	6
A49	8.11	56.5	26	0.933	99.888	2.83	6

7.5 Regresión lineal múltiple

La regresión lineal múltiple se realizó en el programa Statistics [43, 45] y R software. Primero se evaluó el histograma de la concentración inhibitoria de los 40 análogos N-PTindoles para observar la distribución de los inhibidores. La **Figura 8** contiene el histograma generado donde se muestra gráficamente que los 40 análogos N-PTindoles tienen una frecuencia mayor en una concentración de 0 a 20 μM .

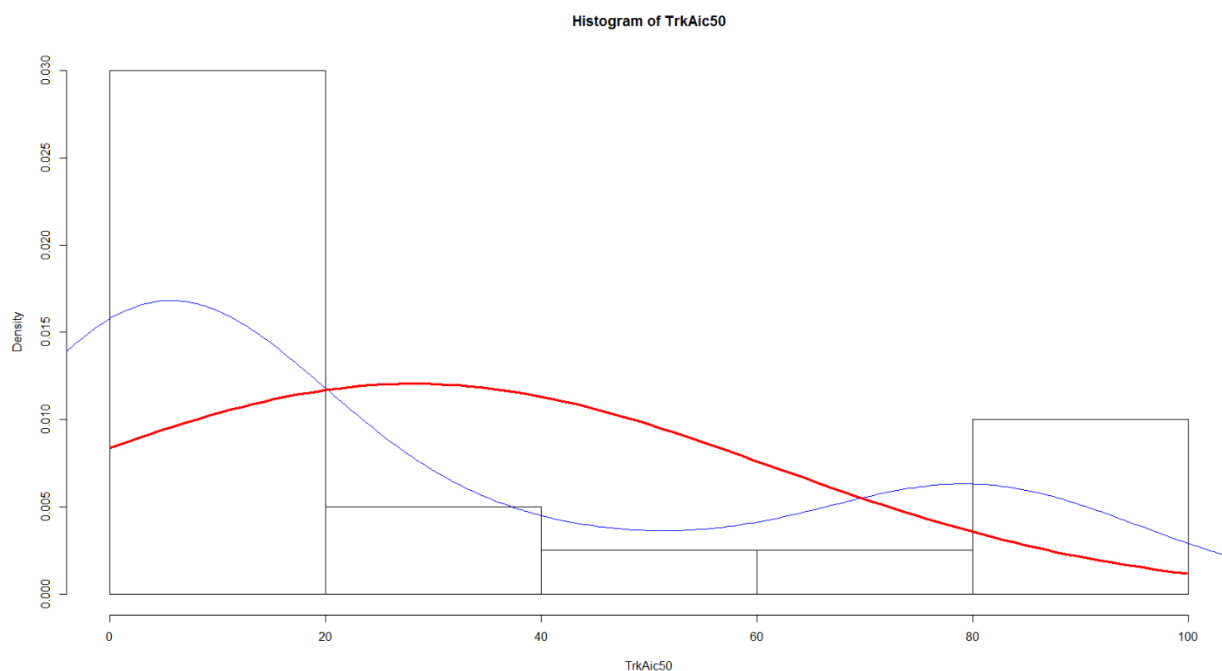


Figura 8: Histograma de la concentración inhibitoria de los 40 análogos N-PTindoles. En color azul se muestra la línea de tendencia de la IC50 de las moléculas análogas de N-PTindoles y en color rojo la normalización. Histograma obtenido del programa R.

Para realizar el modelo de regresión lineal múltiple se analizaron las relaciones entre las 30 IDVs seleccionadas para la identificación de los mejores predictores del modelo, descartando las IDVs que tienen relaciones de tipo no lineal con la DV y para determinar la existencia colinialidad entre IDVs con IDVs. Para la comprensión de este análisis se realizó una matriz de correlación **Figura 9** y en el suplemento (**Figura S1, S2, S3**) muestran los resultados de los cálculos obtenidos con el programa R, en estas figuras se observan los resultados de la correlación de manera numérica y grafica que existe entre cada una de las IDVs con la DV y con las otras IDVs. Los valores numéricos presentados en los cálculos de la correlación **Figura S1**, indican que entre más cercano a uno (positivo o negativo) hay una mayor correlación. En el caso de la correlación de IDVs con DV se traduce a que son buenos descriptores moleculares. para el modelo, sin embargo, en el caso de la correlación IDV con IDV indica que los descriptores moleculares son colineales.

Los resultados numéricos obtenidos de la correlación entre las IDVs versus DV están en un rango de 0.35 a 0.003, por lo tanto, si fuera un modelo QSAR utilizando una regresión lineal simple la ecuación no tendría una buena predicción puesto que la justificación máxima de la relación con la DV sería del 35%.

Los IDVs que tienen una mayor correlación con la DV son LBE, HBD, IAC, MW, AMR mientras que ESOL es el que está menos correlacionado con la actividad. Además, estos resultados se pueden corroborar por medio de las gráficas contenidas en las **Figuras S1, S2, S3**. En la **Figura 9** se muestra la matriz de correlación resultante donde se observa que según la intensidad del color y el tamaño de los círculos hay mayor o menor correlación. El color rojo indica una correlación negativa mientras que el azul una positiva. Se analizaron los descriptores y los que se relacionaron linealmente entre si se descartaron para estar en un mismo modelo y así evitar la colinealidad.

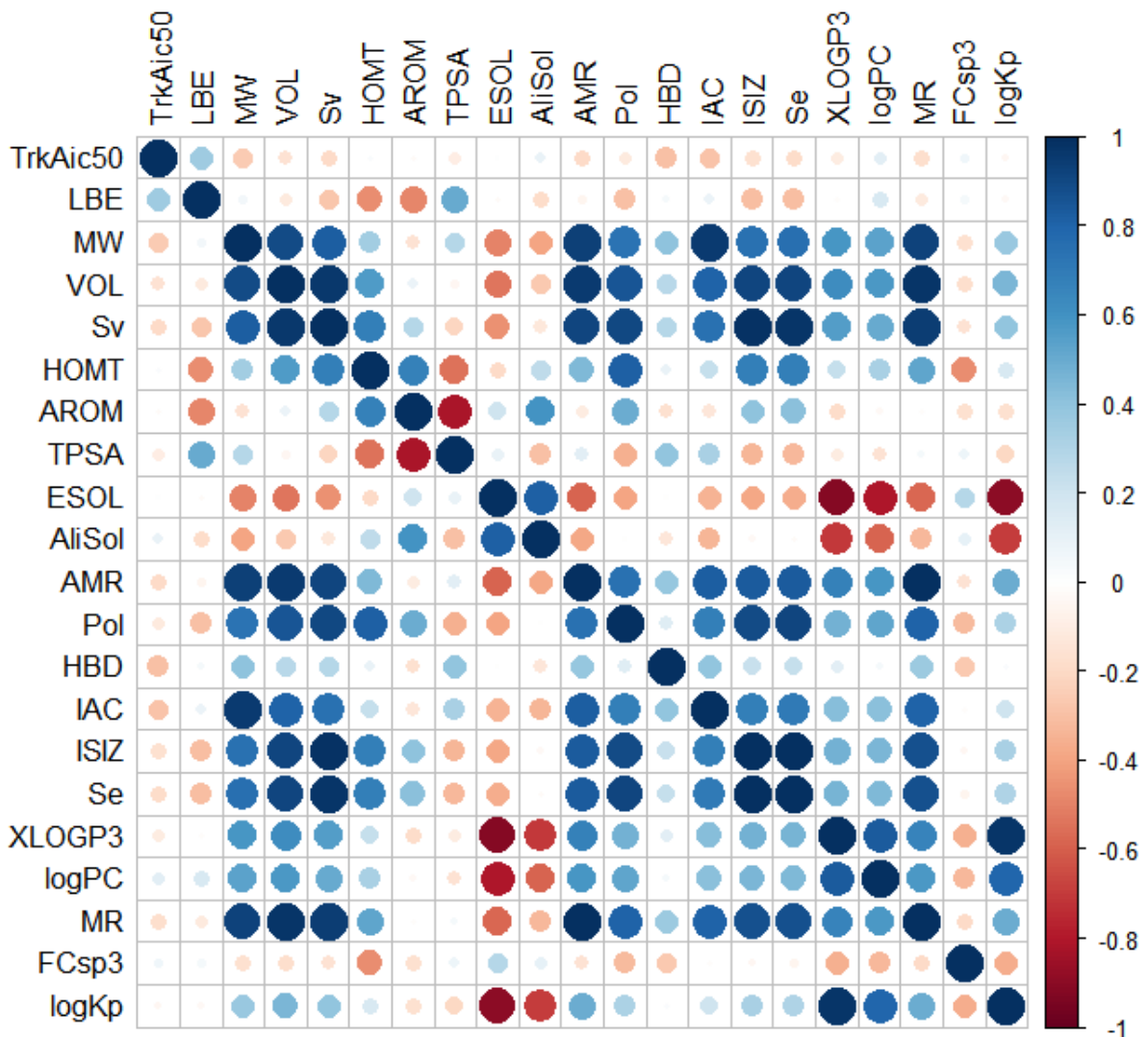


Figura 9: Matriz de correlación de 19 descriptores moleculares con la actividad biológica. En color azul y rojo intenso se observan las correlaciones más fuertes con valor cercano a 1 o -1, Mientras que las correlaciones más bajas se muestran con un color menos intenso.

Las ecuaciones finales del QSAR se realizaron con 8 y 9 predictores, estos se obtuvieron de diferentes métodos, uno consistió en el análisis de los resultados de la matriz de correlación donde manualmente se eligieron conjuntos de 9 y 8 IDVs que se relacionaban mejor con la DV, que no

fueran colineales y que su significación molecular no se repitiera para poder realizar la MLR y el segundo método fue utilizando la selección de descriptores por stepwise en el programa de Statistica y de R.

Las ecuaciones obtenidas por el método manual fueron las siguientes:

La ecuación 1 con 9 IDVs y una R^2 de 0.54 indica que el modelo explica cerca del 54 % de la varianza de la VD (actividad biológica). Por lo tanto, el 50% de la justificación de la actividad biológica es explicada por DG, LOG p, logSw, Sv, Ss, Pol, UNIP, IDE, IDM. Donde se puede observar que los descriptores que más contribuyen a la relación con la actividad biológica son los que están representados en color rojo véase la **Figura 10** contiene la ecuación y los resultados obtenidos del programa Statistics.

Interpretando la ecuación 1, nos indica que para conocer la DV de una molécula necesitamos conocer los IDVs descritos anteriormente ya que estos influyen en la actividad biológica de manera diferente. En el caso de la DG, Log P, IDE, IDM estos son negativos lo que se interpreta que al aumentar sus valores de estos descriptores moleculares la actividad biológica desciende. Mientras que los IDVs positivos como logSw, Sv, Ss, Pol, UNIP al aumentar sus valores aumentan los valores de la DV.

Modelo No.	IDVs No.	Ecuación QSAR	R²	R²adj
1	9	$IC_{50} = 8885.55 + (-0.44 * DG) + (-18.92 * \text{LogP}) + (10.7 * \text{LogSW}) + (41.76 * Sv) + (4.16 * Ss) + (27.72 * \text{Pol}) + (5.87 * \text{UNIP}) + (-306.73 * \text{IDE}) + (-1319.92 * \text{IDM})$	0.54	0.40

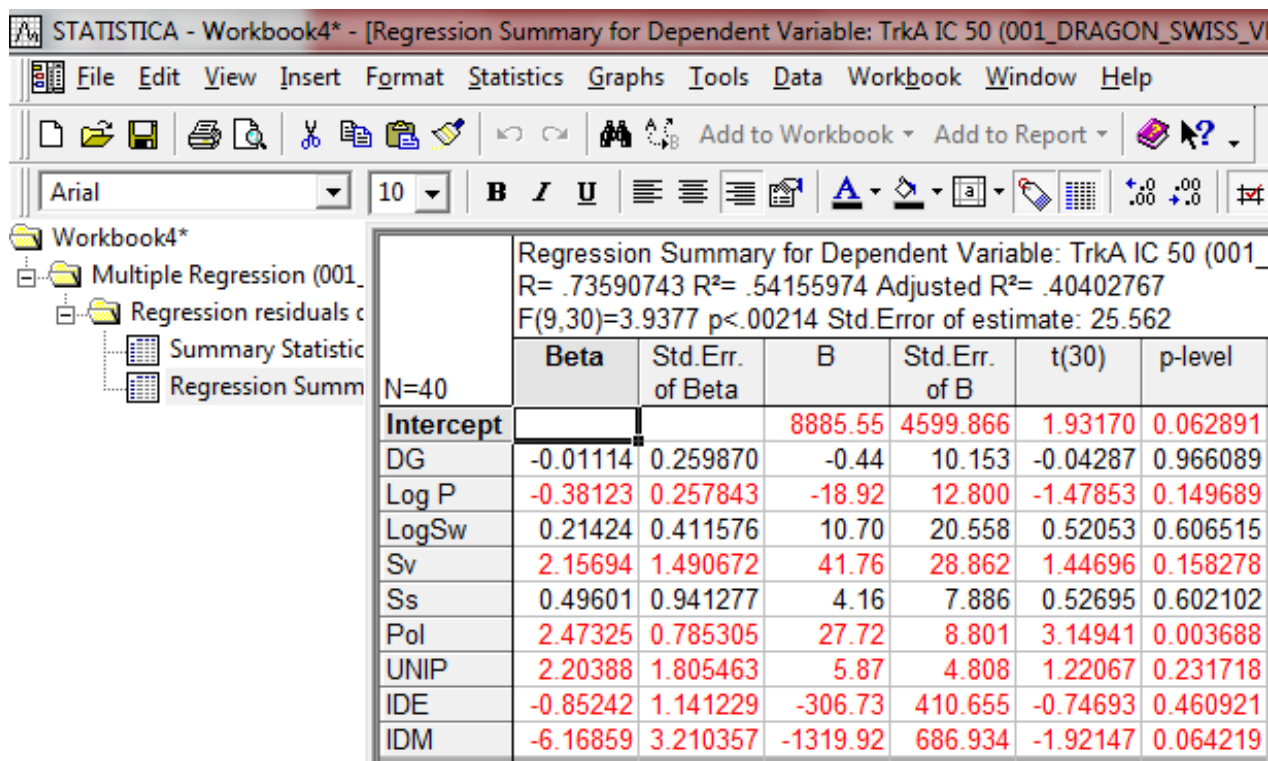


Figura 10: Imagen obtenida en Statistica de los datos resultantes del modelo 1, además se muestra en color rojo las IDVs que tienen mayor contribución a la ecuación del modelo obtenido.

La validación del modelo 1 se basó en una representación gráfica de dispersión siendo el eje X el valor de IC50 experimental mientras que el eje Y representa el IC50 teórico. Se observó que los valores teóricos y los experimentales coinciden en algunas moléculas véase la **Figura 11**.

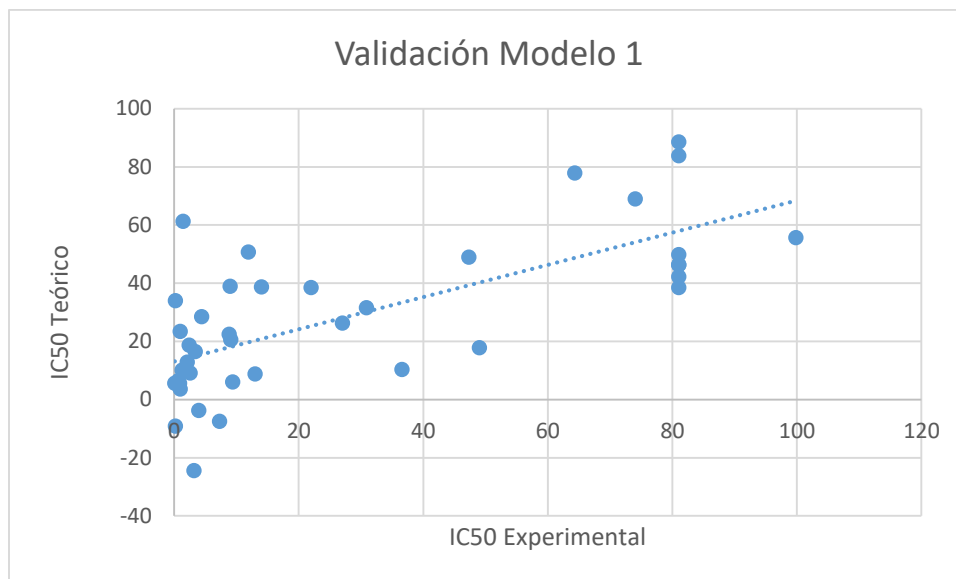


Figura 11: Gráfico de dispersión modelo 1. Los valores teóricos de la IC50 se encuentran en el eje Y mientras que los valores experimentales en el eje X. Los valores de los IC50 teóricas de las moléculas N-PTindoles son ligeramente semejantes a los experimentales ya que coinciden en algunos puntos en la gráfica.

En el caso del modelo 2 se eliminó Ss del primer modelo ya que este IDV es un poco difícil de interpretarse, visualizarse experimentalmente y su contribución en el modelo era baja. Los resultados de la ecuación 2 con 8 IDVs fue con una R^2 de 0.54 lo que indica que el modelo explica cerca del 54 % de la varianza de la VD (actividad biológica) y se confirma que Ss es un IDV que no contribuye mucho al modelo. La justificación de la actividad biológica se explica por DG, LOG p, logSw, Sv, Pol, UNIP, IDE, IDM. Donde se puede observar que los descriptores que más contribuyen a la relación con la actividad biológica son los que están representados en color rojo véase la **Figura 12** de la ecuación 2.

Modelo No.	IDVs No.	Ecuación QSAR	R ²	R ² adj
2	8	$\text{IC50} = 6761.075 + (3.033 \cdot \text{DG}) + (-16.03 \cdot \text{LogP}) + (6.794 \cdot \text{LogSW}) + (28.013 \cdot \text{Sv}) + (26.572 \cdot \text{Pol}) + (4.233 \cdot \text{UNIP}) + (-236.102 \cdot \text{IDE}) + (-985.011 \cdot \text{IDM})$	0.54	0.42

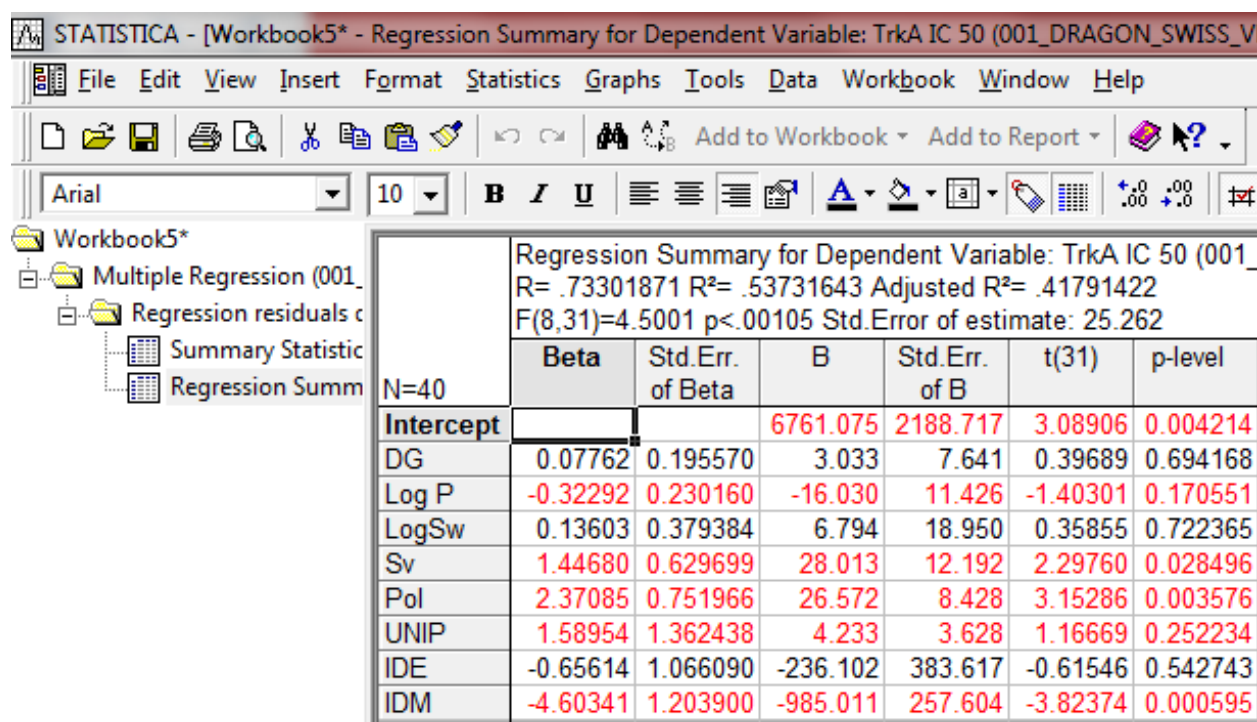


Figura 12: Imagen obtenida en Statistica de los datos resultantes del modelo 2, además se muestra en color rojo las IDVs que tienen mayor contribución a la ecuación del modelo obtenido.

La validación del modelo 2 se basó en una representación gráfica de dispersión siendo el eje X el valor de IC50 experimental mientras que el eje Y representa el IC50 teórico. Se observó que los valores teóricos y los experimentales coinciden en algunas moléculas véase la **Figura 3**.

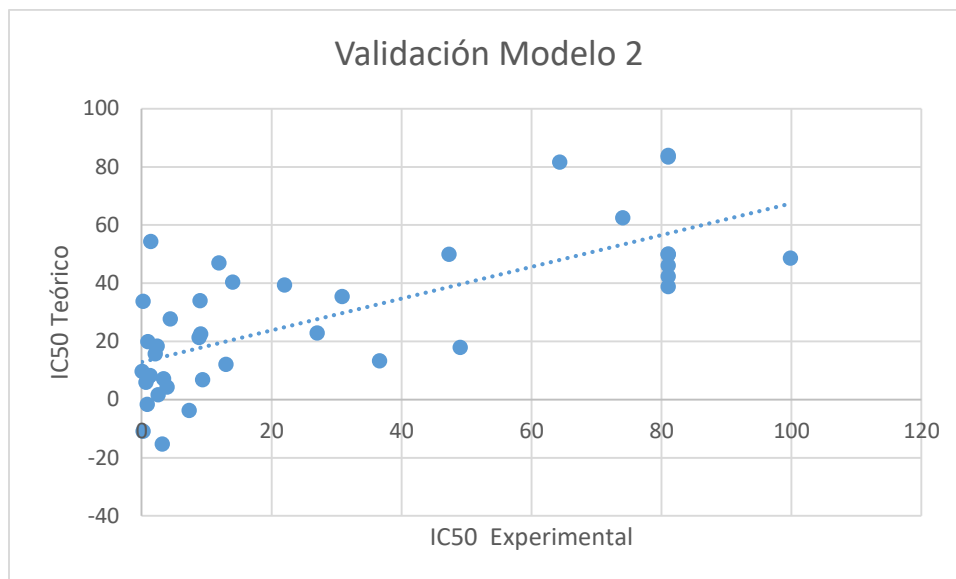


Figura 13: Gráfico de dispersión modelo 2. Los valores teóricos de la IC50 se encuentran en el eje Y mientras que los valores experimentales en el eje X. Los valores de los IC50 teóricas de las moléculas N-PTindoles son ligeramente semejantes a los experimentales ya que coinciden en algunos puntos en la gráfica. En comparación con la gráfica del modelo 1 no hay gran diferencia.

Después se seleccionó un conjunto diferente de 8 IDVs, algunos representaban una misma propiedad fisicoquímica como el caso de LOG p y Log PC solo con la diferencia del método de obtención. El modelo 3 que se obtuvo tiene una R^2 de 0.52 lo que indica que el modelo explica cerca del 52 % de la varianza de la VD (actividad biológica). Por lo tanto, la actividad biológica de este modelo se explica por DG, Log PC, H-Don, IAC, Ss, Pol, IDM, ESOL log S. Los descriptores que más contribuyen a la relación con la actividad biológica son los que están representados en color rojo véase la **Figura 14** del modelo 3. Aunque la R^2 no vario mucho a comparación de la ecuación 1 los descriptores de esta ecuación son más fáciles de interpretar y de poder utilizar. Como es el caso de H-Don que es positiva y por lo tanto al aumentar los donadores de hidrógenos en la molécula líder de los análogos de N-PTindoles aumenta la actividad biológica y molecularmente es más fácil de modificar en la estructura a un parámetro cuántico.

Modelo No.	IDVs No.	Ecuación QSAR	R ²	R ² adj
3	8	IC50=2368.040 + (9.663*DG) + (6.612*LogPC) + (2.315*H-Don) + (1.243*IAC) + -7.859*Ss) + (26.403* Pol) + (-343.114*IDM) + (24.803*ESOL logS)	0.51	0.39

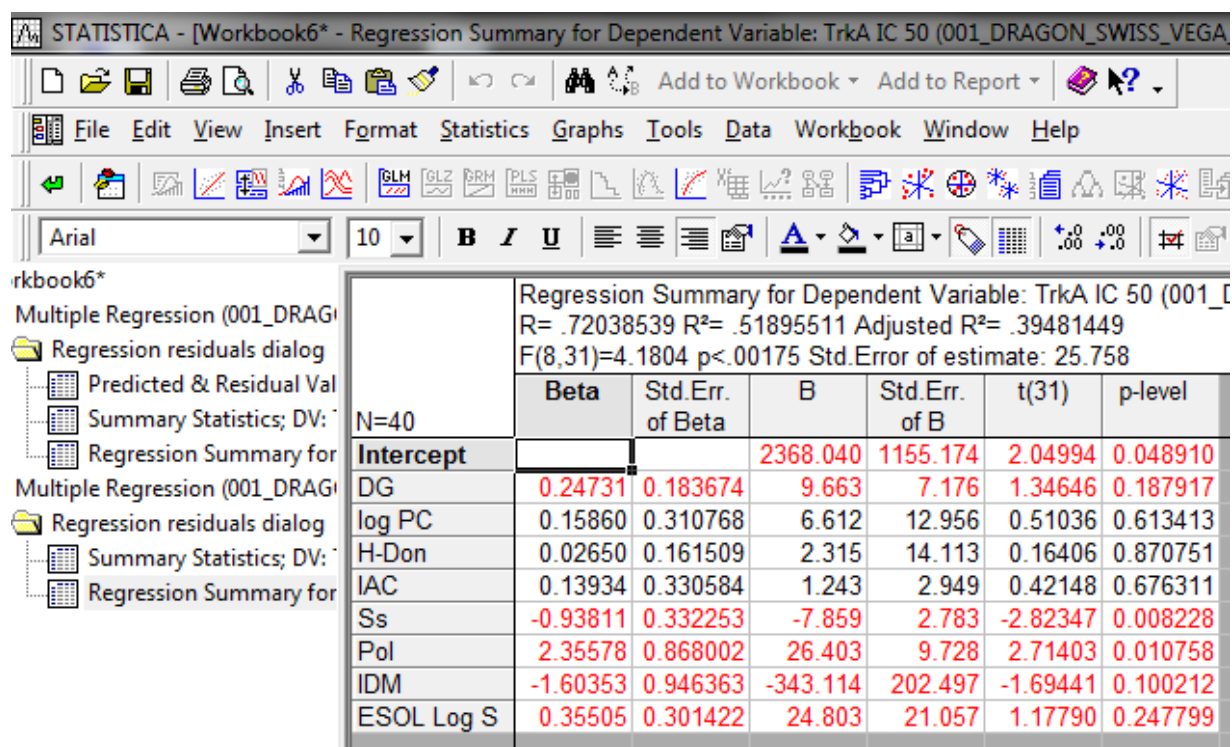


Figura 14: Imagen obtenida en Statistica de los datos resultantes del modelo 3, además se muestra en color rojo las IDVs que tienen mayor contribución a la ecuación del modelo obtenido.

La validación del modelo 3 se basó en una representación gráfica de dispersión siendo el eje X el valor de IC50 experimental mientras que el eje Y representa el IC50 teórico. Se observó que los valores teóricos y los experimentales coinciden en algunas moléculas véase la **Figura 15**.

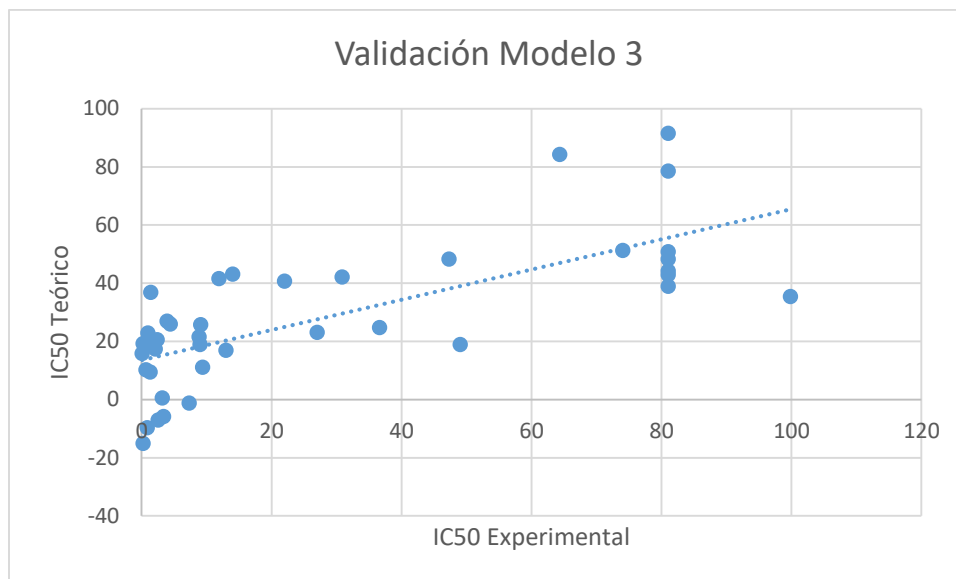


Figura 15: Gráfico de dispersión modelo 3. Los valores teóricos de la IC50 se encuentran en el eje Y mientras que los valores experimentales en el eje X. Los valores de los IC50 teóricas de las moléculas N-PTindoles son ligeramente semejantes a los experimentales ya que coinciden en algunos puntos en la gráfica. En comparación con la gráfica del modelo 1 y 2 no hay gran diferencia.

Para la ecuación 4 se seleccionaron diferentes IDVs y se obtuvo el coeficiente de terminación más alto de todos los modelos con una R^2 de 0.66 lo cual indica que el modelo explica cerca del 66 % de la varianza de la VD (actividad biológica). Por lo tanto, el 66% de la justificación de la actividad biológica es explicada por LBE, Dipolo, LogPC, PSA-based-HLB, MW, NoAromatic, FractionCsp3, H-Don, logKp-cmPs. Interpretando la ecuación 4, los IDVs negativos como la LBE, dipole, MW, H-Don y logkp-cmPs significa que al aumentar sus valores de estos descriptores moleculares la actividad biológica desciende. Mientras que los IDVs positivos como Log PC, PSA-based- HLB, No-Aromatic y FractionCsp3 al aumentar sus valores aumentan los valores de la DV.

En la obtención de la ecuación 5 con 8 descriptores se modificó la ecuación 4 únicamente eliminando No-Aromatic y cambiando a Dipolo por Mol-refract. El resultado fue una R^2 de 0.65 lo que explica el 65% de la justificación de la actividad biológica por LBE, Mol-refrac, LogPC, PSA-based-HLB, MW, FractionCsp3, H-Don, logKp-cmPs. Este es el mejor modelo ya que la variación de la R^2 con el modelo 4 es de 0.01 lo que significa a nivel biológico que representan casi la misma relación. Sin embargo, la ecuación 5 explica con mayor precisión la variabilidad observada en la DV porque emplea el menor número de IDV.

Modelo No.	IDVs No.	Ecuación QSAR	R^2	R^2_{adj}
4	9	$IC_{50} = - 682.579 + (- 5.2372 * LBE) + (- 0.0619 * Dipolo) + (66.7041 * LogPC) + (18.8262 * PSA\text{-based-HLB}) + (- 1.3818 * MW) + (13.586 * NoAromatic) + (472.0407 * FractionCsp3) + (- 26.2775 * HBD) + (- 86.5182 * logKp\text{-}cmPs)$	0.6571 0.66	0.5028
5	8	$IC_{50} = - 654.959 + (- 0.6138 * LBE) + (0.5662 * PSA) + (64.4082 * LogPC) + (2.291 * Mol\text{-}refract) + (- 1.191 * MW) + (465.5126 * FractionCsp3) + (- 19.4002 * HBD) + (- 94.6691 * logKp\text{-}cmPs)$	0.6467 0.65	0.512

La validación del modelo 5 está representado por una gráfica de dispersión siendo el eje X el valor de IC50 experimental mientras que el eje Y representa el IC50 teórico. Se observa en la **Figura 16** que los valores teóricos obtenidos son semejantes a los experimentales.

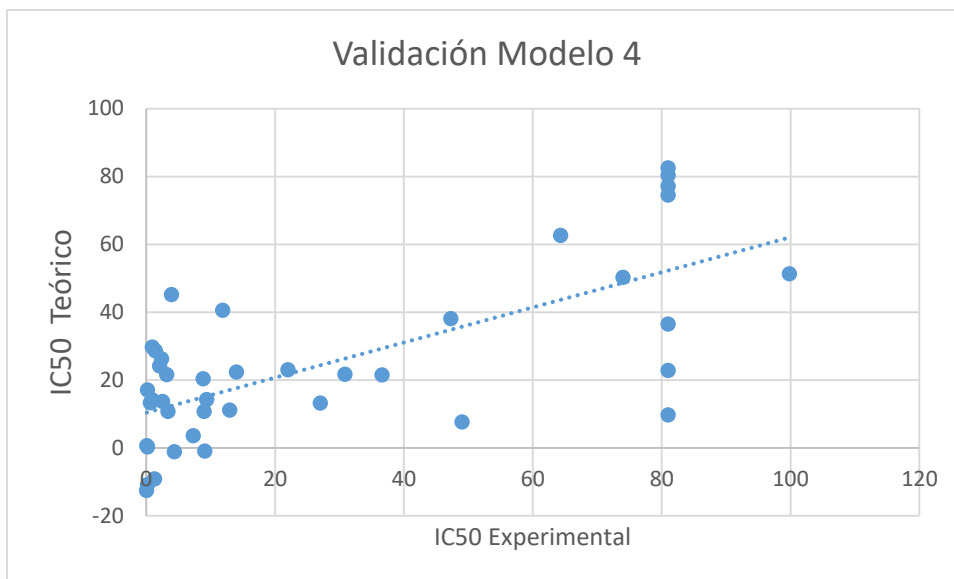


Figura 16: Gráfico de dispersión modelo 4. Los valores teóricos de la IC50 se encuentran en el eje Y mientras que los valores experimentales en el eje X. Los valores de los IC50 teóricas de las moléculas N-PTindoles son ligeramente semejantes a los experimentales ya que coinciden en algunos puntos en la gráfica. En comparación con la gráfica del modelo 1, 2 y 3 se observa una mejor relación ya que hay una mayor coincidencia en las moléculas y sus IC50.

Para la validación del modelo 5 se realizó una representación gráfica de dispersión siendo el eje X el valor de IC50 experimental mientras que el eje Y representa el IC50 teórico. Se observa en la Figura 17 que los valores teóricos obtenidos son semejantes a los experimentales.

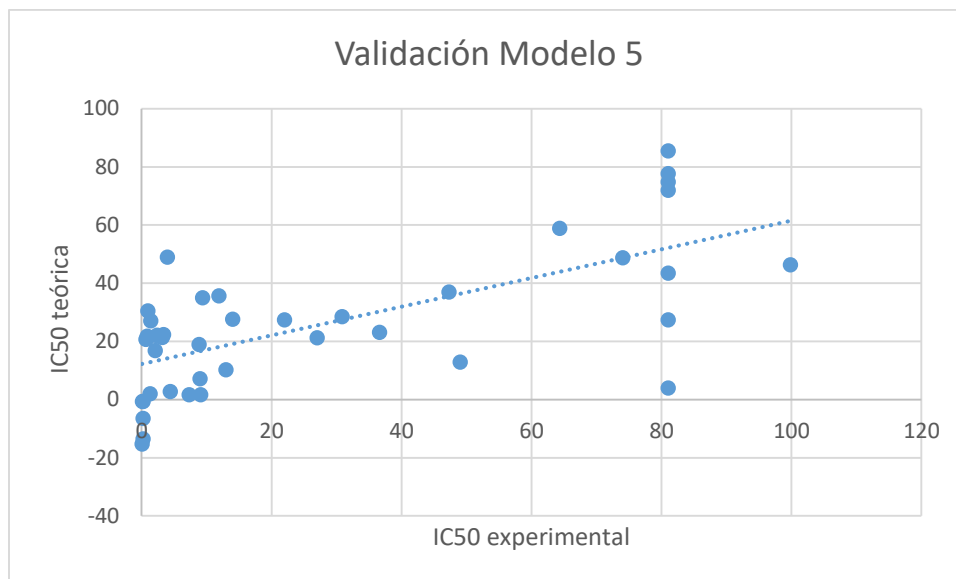


Figura 17: Gráfico de dispersión modelo 5. Los valores teóricos de la IC50 se encuentran en el eje Y mientras que los valores experimentales en el eje X. Los valores de los IC50 teóricas de las moléculas N-PTindoles son ligeramente semejantes a los experimentales ya que coinciden en algunos puntos en la gráfica.

La ecuación 6 se realizó con el método paso a paso (*stepwise*) mixto utilizando el programa R. Primero se analizó la relación entre las variables véase **Figura S1, S2, S3** en el suplemento, donde se observa que las IDVs con mayor relación lineal con la DV son: LBE, MW, AMR, HBD y Sv siendo la correlación más grande de 0.35. Para disminuir las 30 IDVs se realizó una matriz de correlación de la IC50 versus los siete parámetros del coeficiente de partición obtenidas por diferentes métodos véase **Figura 17** al analizar el método con mayor correlación fue el Log PC el cual se utilizó para los pasos siguientes.

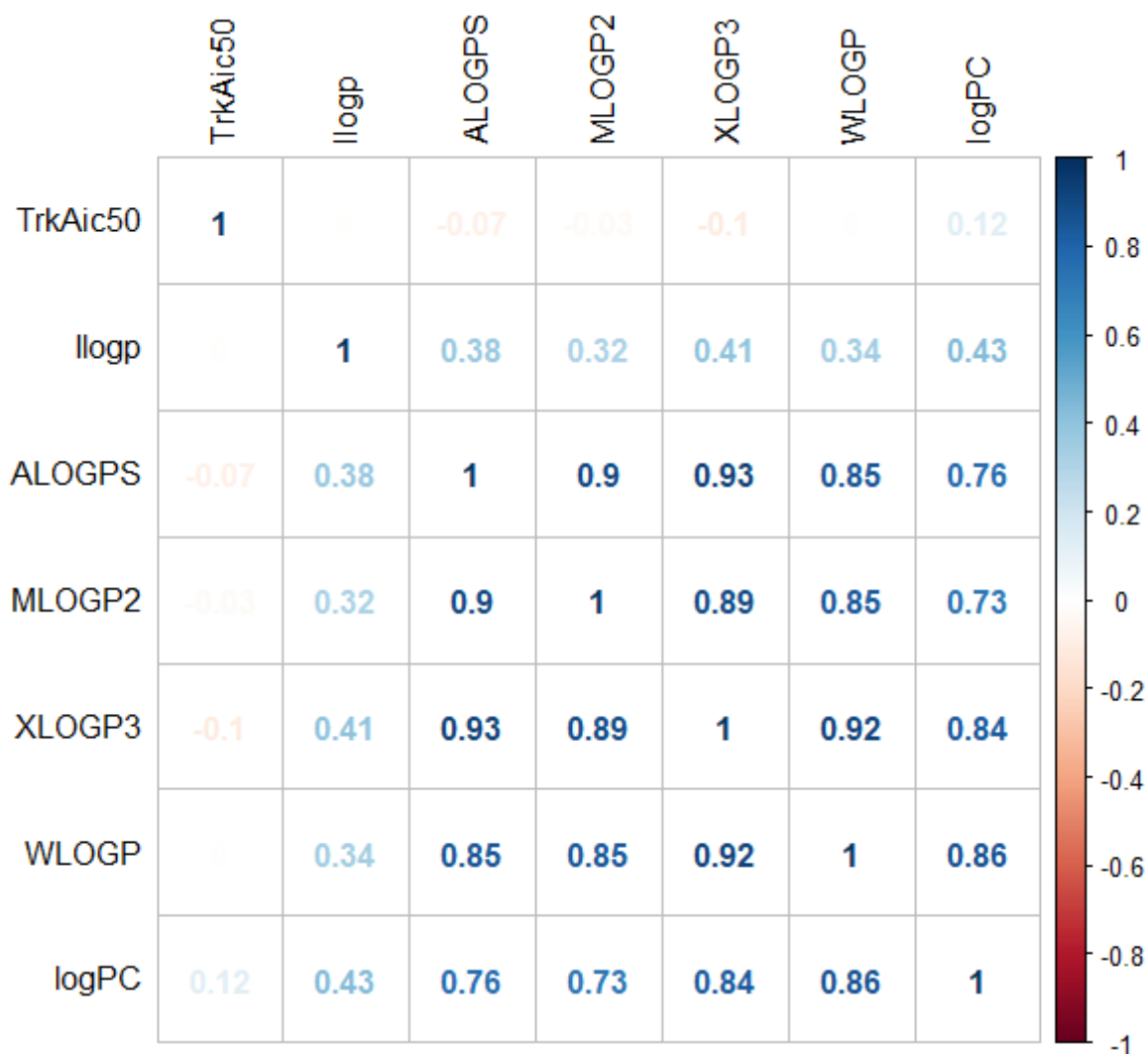


Figura 17: Matriz de correlación de la IC50 (DV) y los IDVs del coeficiente de partición lípido/agua.

El segundo paso para emplear el método mixto consiste en generar un modelo con todas las IDVs como predictores véase **Figura 18**. El modelo resultante tiene un R^2 de 0.62, es capaz de explicar el 62% de la variabilidad observada en la actividad biológica. Los coeficientes parciales de regresión son distintos de 0 pero algunos de ellos no son significativos, lo que es un indicativo de que podrían no contribuir al modelo.

```

Archivo  Editar  Misc  Paquetes  Ventanas  Ayuda

Residual standard error: 29.24 on 19 degrees of freedom
Multiple R-squared:  0.6201,    Adjusted R-squared:  0.2202
F-statistic: 1.551 on 20 and 19 DF,  p-value: 0.172

> ModelR <- lm (formula = TrkAic50 ~ LBE + MW + VOL + Sv + HOMT + A
> summary(ModelR)

Call:
lm(formula = TrkAic50 ~ LBE + MW + VOL + Sv + HOMT + AROM + TPSA +
    ESOL + AliSol + AMR + Pol + HBD + IAC + ISIZ + Se + XLOGP3 +
    logPC + MR + FCsp3 + logKp, data = TRKA30IDV)

Residuals:
    Min       1Q   Median       3Q      Max
-31.985 -13.252  -1.198   9.932  54.439

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -47643.041  30833.898  -1.545  0.139
LBE           4.338     16.599   0.261  0.797
MW          -46.987     29.806  -1.576  0.131
VOL          -1.887      2.524  -0.748  0.464
Sv          -403.340    251.190  -1.606  0.125
HOMT         18.486     13.981   1.322  0.202
AROM        -1720.413   1266.301  -1.359  0.190
TPSA         3.996       3.379   1.183  0.252
ESOL        -1725.350   1400.897  -1.232  0.233
AliSol       632.107     561.627   1.125  0.274
AMR          -1.229     15.600  -0.079  0.938
Pol           9.400     16.893   0.556  0.584
HBD          -39.348     41.797  -0.941  0.358
IAC          -44.746     40.135  -1.115  0.279
ISIZ        -39.243     51.954  -0.755  0.459
Se           553.023     505.068   1.095  0.287
XLOGP3       5229.733   3334.567   1.568  0.133
logPC        13.435      19.269   0.697  0.494
MR           30.564     22.844   1.338  0.197
FCsp3       -2367.969   1991.950  -1.189  0.249
logKp       -7460.150   4728.190  -1.578  0.131

Residual standard error: 29.24 on 19 degrees of freedom
Multiple R-squared:  0.6201,    Adjusted R-squared:  0.2202
F-statistic: 1.551 on 20 and 19 DF,  p-value: 0.172

```

Figura 18: Modelo con todas las IDVs generado como primer paso de *stepwise* mixto con los siguientes IDVs LBE, MW, VOL, Sv, HOMT, AROM, TPSA, ESOL, AliSol, AMR, Pol, HBD, IAC, ISIZ, Se, XLOGP3, logPC, MR, FCsp3 y logKp. Realizado con el programa R.

Para la selección de los descriptores se utiliza la estrategia de *stepwise mixto* véase el resultado de cada uno en suplemento **Figura S5**. El valor matemático empleado para determinar la calidad del modelo va a ser *Akaike(AIC)* la cual se basa en el desorden de información de los IDVs: se ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. Para la interpretación el modelo preferido es el que tiene el valor mínimo en el AIC los resultados se muestran en el suplemento. Los mejores IDVs para el modelo fueron MW, Sv, HOMT, AROM, TPSA, ESOL, AliSol, MR, FCsp3 y logKp como se muestra en la **Figura 19**.

step (object = ModelR, direction = "both", trace = 1)

```
Call:
lm(formula = TrkAic50 ~ MW + Sv + HOMT + AROM + TPSA + ESOL +
    AliSol + HBD + IAC + ISIZ + Se + XLOGP3 + logPC + MR + FCsp3 +
    logKp, data = TRKA30IDV)

Coefficients:
(Intercept)      MW          Sv          HOMT          AROM          TPSA          ESOL          AliSol
-41429.248    -40.891   -416.237    19.781   -1446.841    3.265   -1830.070    678.891
      MR      FCsp3      logKp
 35.930  -2164.229  -6392.984
```

Figura 19: IDVs resultantes del método *stepwise* mixto.

Después se realizó la regresión lineal múltiple con los IDVs de la **Figura 19**, pero el coeficiente de determinación fue de 0.25 por lo tanto un modelo totalmente automatizado con programas estadísticos no da la mejor solución en esta metodología y se hace hincapié que para la elaboración de un QSAR se debe tener toda la justificación biológica y química. Se analizó los modelos anteriores y junto con los datos de *stepwise* mixto se obtuvo como mejor modelo la ecuación 6.

El **R² de la ecuación 6 fue de 0.47 lo cual** indica que el modelo explica cerca del 47 % de la varianza de la VD (actividad biológica). Por lo tanto, el 47% de la justificación de la actividad biológica es explicada por MW, Sv, HBD, Se, logPC, MR, FCsp3 y logKp. Los IDVs negativos

como MW, Sv, HBD, FractionCsp3 logkp-cmPs y LogKp indican que al aumentar sus valores la actividad biológica desciende. Mientras que los IDVs positivos como Se, LogPC y MR al aumentar sus valores aumentan los valores de la DV véase **Figura 20**.

Ecuación No.	IDVs No.	Ecuación QSAR	R ²	R ² adj
6	8	$IC50 = -298.3251 + (-3.3195 * MW) + (-207.5418 * Sv) + (-38.1686 * HBD) + (79.7350 * Se) + (29.4424 * LogPC) + (36.1866 * MR) + (-846.3803 * FCsp3) + (-77.5730 * LogKp)$	0.47	0.34

```

> ModelR6 <-lm(formula = TrkAic50 ~ MW + Sv + HBD + Se + logPC + MR + FCsp3 +
+   logKp, data = TRKA30IDV)
>
> summary(ModelR6)

Call:
lm(formula = TrkAic50 ~ MW + Sv + HBD + Se + logPC + MR + FCsp3 +
    logKp, data = TRKA30IDV)

Residuals:
    Min       1Q   Median       3Q      Max
-43.323 -14.163  -4.527  11.436  59.992

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -298.3251   178.3138  -1.673  0.104384
MW           -3.3195    0.8548   -3.883  0.000505 ***
Sv          -207.5418   82.1451   -2.527  0.016840 *
HBD          -38.1686   15.7693   -2.420  0.021554 *
Se           79.7350   35.2285    2.263  0.030762 *
logPC        29.4424   10.8699    2.709  0.010900 *
MR           36.1866   11.5186    3.142  0.003682 **
FCsp3       -846.3803  435.7174   -1.942  0.061209 .
logKp        -77.5730   22.4424   -3.457  0.001610 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.96 on 31 degrees of freedom
Multiple R-squared:  0.4729,    Adjusted R-squared:  0.3369
F-statistic: 3.477 on 8 and 31 DF,  p-value: 0.005659

```

Figura 20: Resultados del modelo 6, realizado en el programa R.

Se realizó un gráfico Cuantil-Cuantil (QQ plots) para comprobar la distribución normal de los residuos de los datos del modelo 6 véase **Figura21** donde gráficamente se observa que existe una distribución de la probabilidad normal.

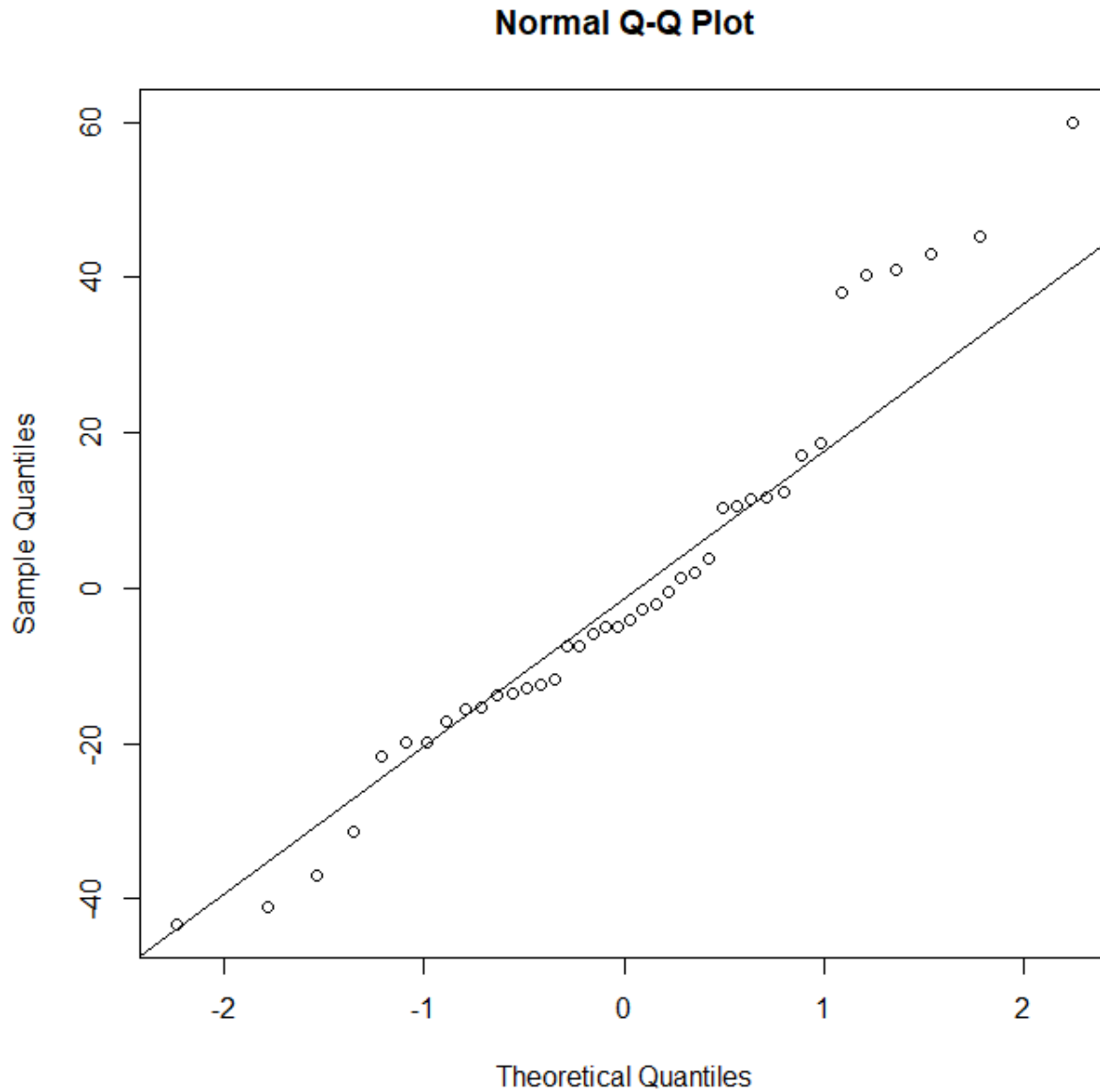


Figura21: Gráfico Cuantil-Cuantil (QQ plots), de los residuos del modelo 6.

Para comprobar la variabilidad constante de los residuos (homocedasticidad), se representaron los residuos frente a los valores ajustados por el modelo, los primeros se distribuyen de forma aleatoria en torno a cero, manteniendo aproximadamente la misma variabilidad excepto en el rango de 20 a 40 μM a lo largo del eje X, pero esto se observó desde la curva de normalidad véase **Figura22**.

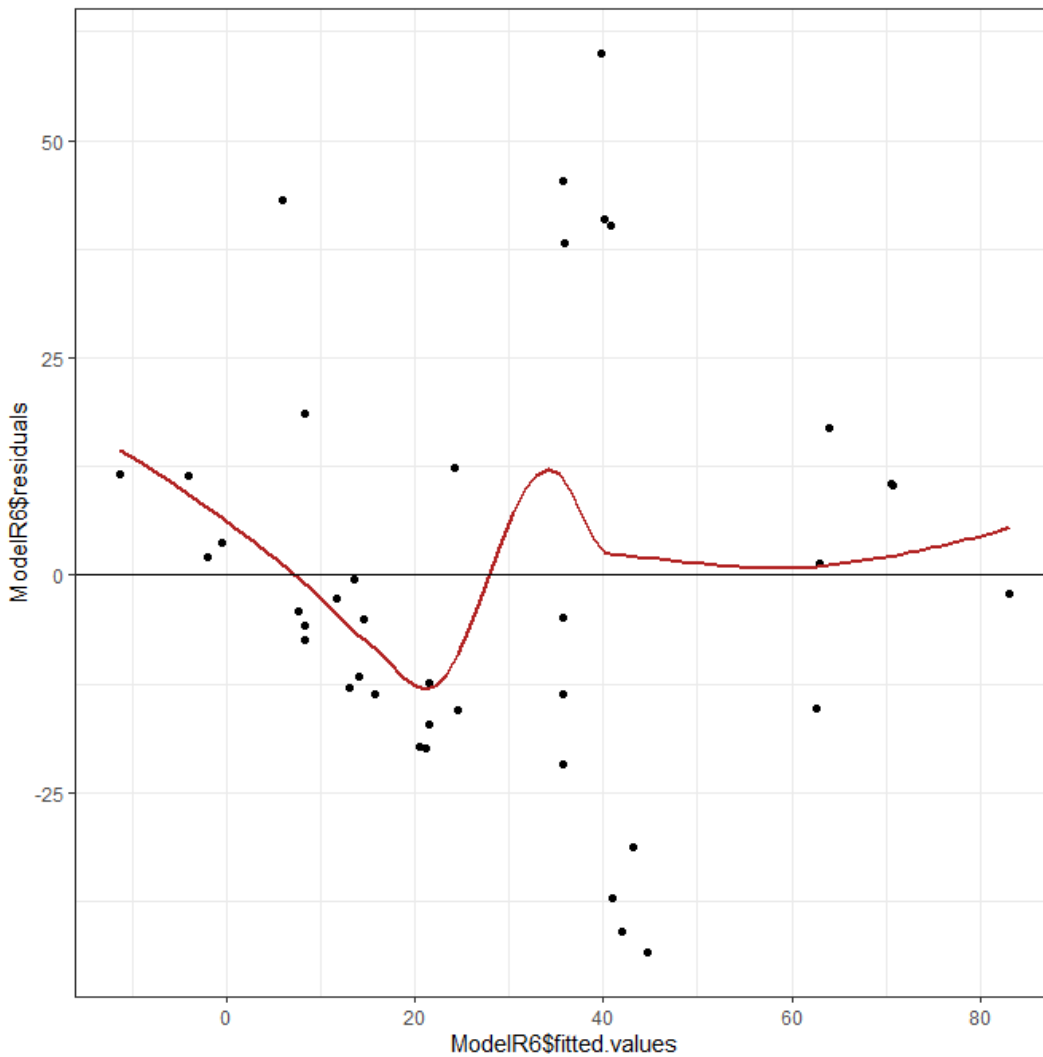


Figura22: Representación de la variabilidad constante de los residuos (homocedasticidad) en el eje X se representa los valores del modelo 6 y en eje Y los residuos.

7.6 Generación de nuevas moléculas

Se generaron 4 nuevas moléculas para la validación e interpretación del modelo 5. Se tomó la molécula D29 como base para las modificaciones de las nuevas moléculas ya que experimentalmente es la que tiene una mayor inhibición con el receptor TrkA.

El modo de unión de la molécula D29 con el receptor TrkA en su sitio activo se explican a continuación: dos anillos del grupo indol interaccionan con sus electrones pi en el anillo de la histidina 648, el hidrógeno de la isoleucina 490 hace una interacción de hidrógeno con el oxígeno de la etanamida y además la mesomería del grupo oxo hace una interacción pi-pi con el anillo del benceno del grupo indol. En el caso del tiol hay dos interacciones pi-pi con leucina 564 y valina 573. La isoleucina 490 hace un enlace de hidrógeno con el oxígeno de la etanoamida. La asparagina 668 hace interacción con sus oxígenos del grupo carboxilato en el carbono 1 de la etanamida y con el imidazol de la 1H-Indol-3-amina. La fenilalanina hace una interacción pi-pi con el anillo del imidazol de la 1H-Indol-3-amina. La glicina 488 hace una interacción de hidrógeno con el hidrógeno de la amina secundaria del imidazol de la 1H-Indol-3-amina. En el caso del aminoácido fenilalanina 521, leucina 485 e histidina 489 hacen interacción pi-pi con los electrones del anillo bencénico de 1H-Indol-3-amina, véase **Figura S4** y **Figura 23**. Esta última imagen muestra además la cavidad con el sitio activo comparándose con el ligando 6UF del cristal 5KMJ.

Después de analizar las interacciones se prosiguió con las modificaciones de la molécula D29 según lo que plantea el modelo 5. El peso molecular y los donadores de hidrógeno son inversamente proporcional a la IC50 por lo tanto al aumentar estos dos IDVs se disminuye el IC50, por otro lado, al disminuir la PSA baja la DV y al realizar un acoplamiento molecular se espera que el LBE sea menor.

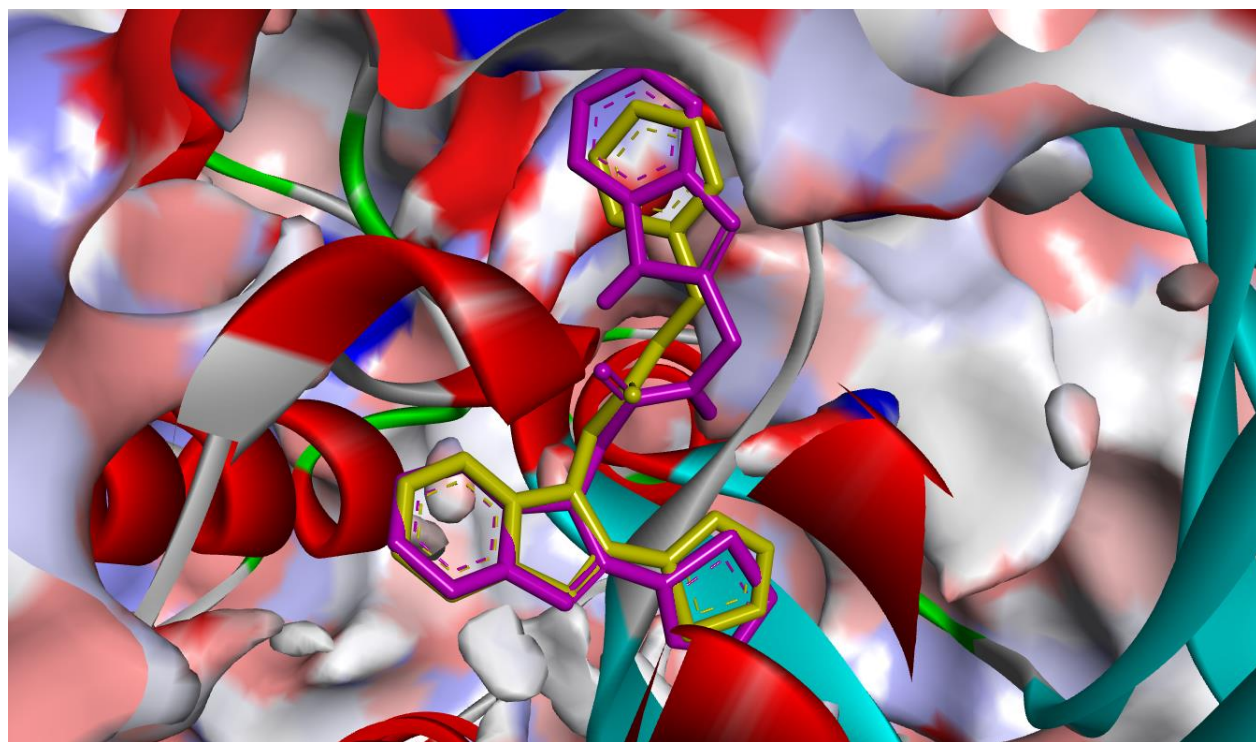


Figura 23: Imagen obtenida de Discovery Studio de la cavidad del sitio activo y modo de unión de los aminoácidos del receptor TrkA en la región juxtamembrana y quinasa con D29 y el ligando 6UF. Se muestra en color magenta la molécula D29 y en color amarillo la plantilla del ligando 6UF. La superficie de la región juxtamembrana y quinasa es de acuerdo a la carga de los átomos contenidos en los aminoácidos el color azul indica que son positivos, en color rojo negativos y en blanco zonas neutras, además se puede apreciar en color rojo algunas alfa hélices y en turquesa betas plegadas que están conformando a la proteína.

Las modificaciones fueron mínimas para conservar la similitud de los inhibidores análogos de N-PTindoles ya que es un requisito para un modelo QSAR. En el caso de la molécula 1 y 2 sufrieron una modificación en el anillo bencénico de 1H-Indol-3-amina cambiando por un metilo y etilo respectivamente en el carbono número 4 del imidazol. La molécula 3 es similar a la 1 con la diferencia de un cambio de un hidrógeno por un flúor en el anillo aromático del grupo indol en la posición -para y la molécula 4 se modificó cambiando el 1H-Indol-3-amina por un indol para hacer simétrica la molécula véase **Figura 24**.

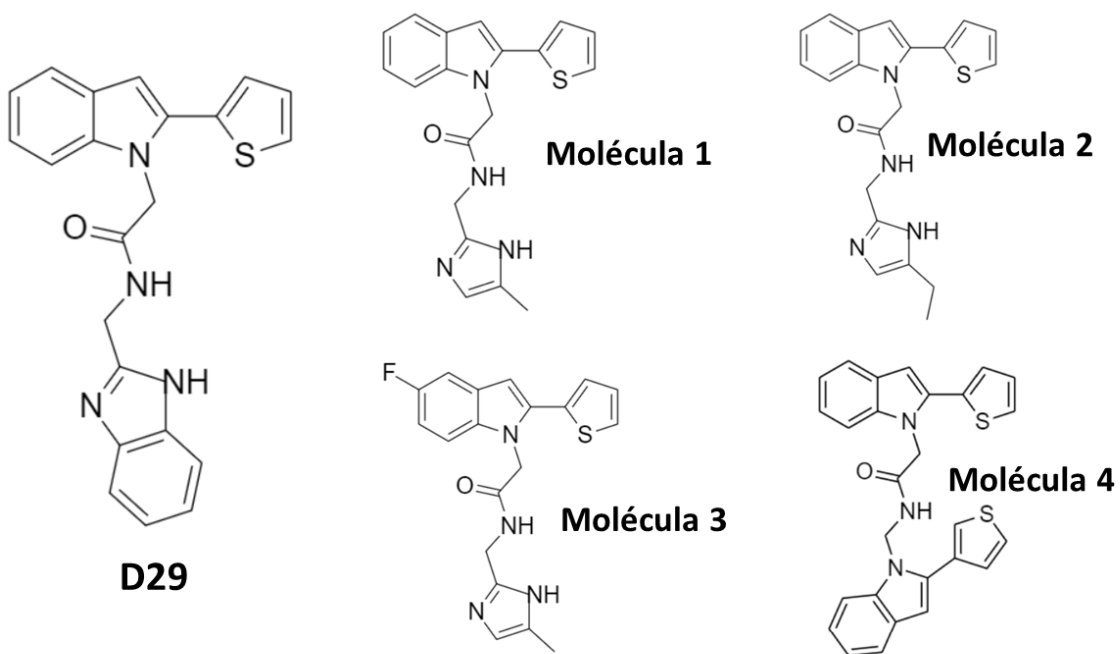


Figura 24: Nuevas moléculas análogas de N-PTindoles construidas a partir de la molécula 29.

Los resultados del acoplamiento molecular en la cavidad oral con las mejores energías de interacción de cada uno de las estructuras se muestran en la **Figura 25** mostrando el modo de unión en color magenta la molécula D29, en verde la molécula 1, en amarillo la molécula 2, en azul la molécula 3 y en naranja la molécula 4.

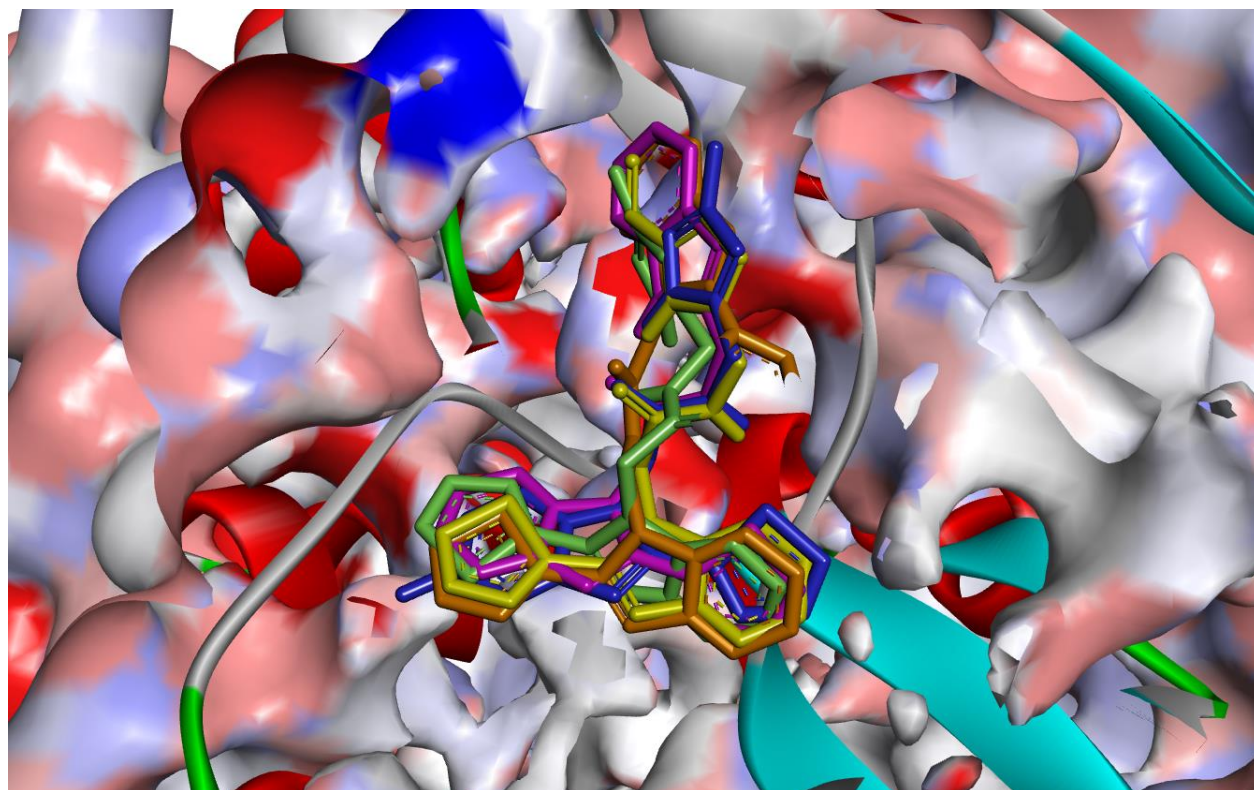


Figura 25: Imagen de la cavidad del sitio activo de TrkA en la región juxtamembrana y quinasa con D29 y las nuevas moléculas análogas de N-PTindoles. Código de colores en magenta la molécula D29, en verde la molécula 1, en amarillo la molécula 2, en azul la molécula 3 y en naranja la molécula 4. La superficie de la región juxtamembrana y quinasa es de acuerdo a la carga de los átomos contenidos en los aminoácidos el color azul indica que son positivos, en color rojo negativos y en blanco zonas neutrales, además se puede apreciar en color rojo algunas alfa hélices y en turquesa betas plegadas que están conformando a la proteína.

Al modificar la molécula D29 se obtuvo los siguientes resultados véase **Tabla 10** donde se resumen los valores de cada IDV de las moléculas nuevas análogas de N-PTindoles y de D29. Para la obtención de los nuevos valores y se utilizó la ecuación del modelo 5:

$$IC_{50} = - 654.959 + (- 0.6138 * LBE) + (0.5662 * PSA) + (64.4082 * \text{LogPC}) + (2.291 * \text{Mol-refract}) + (- 1.191 * MW) + (465.5126 * \text{FractionCsp3}) + (- 19.4002 * HBD) + (- 94.6691 * \text{logKp-cmPs})$$

Para la obtención de los valores teóricos de IC50 se sustituyen los valores de los IDVs de la **Tabla 10** los cuales dieron valores de IC50 teóricos similares a los que se obtuvieron en las 40 moléculas, nótese en los valores de IDVs que no varían mucho ya que se hizo cambios mínimos y en el único cambio grande es la molécula 4 la cual tiene mayores valores de PSA, LogP, Mol-refract y MW. En el suplemento se incluyen las figuras de los datos obtenidos por swissADME de las nuevas moléculas N-PTindoles. La molécula que teóricamente sería un buen inhibidor de TrkA análogo de N-PTindoles para la supresión del dolor es la molécula 4. Se sugiere hacer una mejora en el diseño ya que las moléculas propuestas solo se realizaron para utilizar el modelo 5 de QSAR y no para mejorar su efecto inhibitorio.

Tabla 10: Valores de los IDVs de las nuevas 4 moléculas.

Nombre	LBE	PSA	LogPC	Mol- refract	MW	FractionCsp3	HBD	logKp- cmPs
D29	-11.81	99.3	3.4406	111.805	386.47	0.09	2	-5.87
Molécula 1	-10.84	88.7	2.1456	102.1173	350.437	0.16	2	-6.39
Molécula 2	-10.61	88.7	2.7693	106.7315	364.464	0.2	2	-6.16
Molécula 3	-10.54	88.7	2.3037	102.6241	368.428	0.16	2	-6.16
Molécula 4	-11.49	111.3	5.8966	136.7565	467.605	0.07	1	-6.16

8 CONCLUSIÓN

La construcción de nuevos modelos QSAR para una serie de ligandos inhibidores de TrkA nos proporcionó la siguiente información.

- Se obtuvieron 6 modelos de QSAR, de los cuales el modelo 4 y 5 son los mejores con una R^2 de 0.66 y 0.65 respectivamente, lo que explica el 66% y 65% de la justificación de la actividad biológica por LBE, Mol-refrac, LogPC, PSA-based-HLB, MW, FractionCsp3, H-Don, logKp-cmPs, No-Aromatic y Mol-refract. Los dos modelos QSAR presentan la mayor relación lineal existente de los IDVs y la DV por lo tanto tienen menor error de predicción. Además, la R^2 al no tener valores tan altos en el ajuste de la recta se incluye una menor discriminación de las 40 moléculas análogas de N-PTindoles, en consecuencia, no habrá pérdida de la información que se puede utilizar para posibles candidatos a fármacos.
- Los descriptores que tienen mayor contribución a la relación con la actividad biológica de acuerdo al modelo QSAR 4 y 5 son la LogPC, FractionCsp3 y logKp-cmPs ya que se alejan más de cero. Estos IDVs experimentalmente pueden contribuir a la mejora de un diseño de un nuevo inhibidor análogo de N-PTindoles.
- Los programas utilizados Statistica y R software nos permitieron generar los modelos con R^2 de 0.47 a 0.66 con diferentes IDVs. También se realizaron los cálculos del modelo 1 al 5 con Excel para comparar las variaciones de los resultados de los coeficientes beta y el coeficiente de determinación en Statistics puesto que es un programa con licencia, los valores fueron parecidos por lo que no se descarta el uso de Excel. En el caso del R software es un programa con bastantes paquetes y gratuito por lo que para este tipo de modelos se recomienda.

- En los 6 modelos obtenidos se pudo evidenciar que un conjunto de IDVs para un modelo QSAR determinado influye de manera diferente cuando se encuentra en otra combinación de predictores.
- Los Programas estadísticos son un apoyo para la elaboración de un modelo tipo QSAR, pero el uso irracional y completamente automático se relaciona con un mal modelo predictivo.
- Los valores obtenidos de todos los modelos nos permitieron corroborar que el QSAR es sólo un modelo predictivo en una serie de análogos en este caso los N-TPindoles, lo cual nos llevó a revelar un patrón de IDVs que pueden explicar un poco la relación con una actividad biológica lo que ayuda al descubrimiento de un objetivo terapéutico para la disminución del dolor o analgésico. Pero el valor del coeficiente de correlación no es tan alto lo que afirma que hay complejidades de los procesos biológicos que no se pueden explicar totalmente con una ecuación matemática y estos se pueden deber a mecanismos moleculares, celulares de los organismos lo que se traduce a que no exista una extrapolabilidad del QSAR al 100%. Evidenciando que la linealidad sólo es una aproximación y no muestra el comportamiento natural de los datos (moléculas) que puede ser un “requisito matemático, pero no una ley de la naturaleza”.
- Las nuevas moléculas generadas a partir de la molécula D29 análoga de N-PTindoles cumplieron con la interpretación del modelo 5 de QSAR. La molécula 4 obtuvo una energía de interacción muy parecida a la estructura de D29 por lo que podría ser un buen candidato para la supresión del dolor.

9 REFERENCIAS

- [1] Scoping Document for WHO Guidelines for the pharmacological treatment of persisting pain in adults with medical illnesses. (2012). Recuperado de https://www.who.int/medicines/areas/quality_safety/Scoping_WHO_GLs_PersistPainAdults_webversion.pdf?ua=1.
- [2] Scoping Document for WHO Treatment Guideline on Non-malignant Pain in Adults Adopted in WHO Steering Group on Pain Guidelines. (2008). Recuperado de https://www.who.int/medicines/areas/quality_safety/Scoping_WHOGuide_non_malignant_pain_adults.pdf.
- [3] García Romero, J.; Fernández, A.; Jiménez Romeroc, M.; Fernández-Abascal Puented, A.; Sánchez Carrillo, F.; Gil Fernández, M. (2002). La medición del dolor: una puesta al día. *Med Integral*. 39(7): 317-320.
- [4] Hirose, M.; Kuroda, Y.; Murata, E. (2016). NGF/TrkA signaling as a therapeutic target for pain. *Pain Pract*. 16(2): 175-182.
- [5] Hurzy, D. M.; Henze, D. A.; Cabalu, T. D.; Narayan, K.; Heller, A.; Cooke, A. J. Design, synthesis and SAR of substituted indoles as selective TrkA Inhibitors. (2017). *Bioorg Med Chem Lett*. 27: 2695-2701.
- [6] Luo, T.; Zhang, S. G.; Zhu, L. F.; Zhang, F. X.; Li, W.; Zhao, K.; Wen, X. X.; Yu, M.; Zhan, Y. Q.; Chen, H.; Ge, C. H.; Gao, H. Y.; Wang, L.; Yang, X. M.; Li, C. Y. (2019). A selective c-Met and Trks inhibitor Indo5 suppresses hepatocellular carcinoma growth. *J Exp Clin Cancer Res*. 38(1): 130.
- [7] Sua, H. P.; Rickertb, K.; Burleinc, C.; Narayanc, K.; Bukhtiyarovac, M.; Hurzya, D. M.; Stumpa, C. A.; Zhanga, X.; Reida, J.; Krasowska-Zoladekd, A.; Tummalab, S. (2017). Structural characterization of nonactive site, TrkA-selective kinase inhibitors. *Proc. Natl. Acad. Sci. U S A*. 114(3): E297-E306.
- [8] Pezet, S.; McMahon, S. B. (2006). Neurotrophins: Mediators and modulators of pain. *Annu Rev Neurosci*. 29: 507-538.
- [9] Chayasirisobhon, S. Cannabis and Neuropsychiatric Disorders: An Updated Review. (2019). *Acta Neurol Taiwan*. 28(2): 27-39.
- [10] Alcántara Montero, A.; Sánchez Carnerero, C. (2018). Role of voltage-gated sodium channel blockers in the treatment of chronic pain: Potential uses in clinical practice based on available evidence. *Rev Esp Anestesiología y Reanimación (English Edition)*. 65(5): 275-283.
- [11] Rajathej, D. M.; Parthasarathy, S.; Selvaraj, S. (2019). QSAR Analysis of Multimodal Antidepressants Vortioxetine Analogs Using Physicochemical Descriptors and MLR Modeling. *Curr Comput Aided Drug Des*. 15 (4): 294 – 307.

- [12] Pan, C.; Meng, H.; Zhang, S.; Zuo, Z.; Shen, Y.; Wang, L.; Chang, K. J. (2019) Homology modeling and 3D-QSAR study of benzhydrylpiperazine δ opioid receptor agonists. *Comput Biol Chem.* 83: 107109.
- [13] Dobričić, V.; Jačević, V.; Vučićević, J.; Nikolic, K. Vladimirov, S.; Čudina, O. (2017). Evaluation of Biological Activity and Computer-Aided Design of New Soft Glucocorticoids. *Arch. Pharm. Chem. Life Sci.* 350(5): e1600383
- [14] Obiany, O.; Ye, K. (2013). Novel Small Molecule Activators of the Trk Family of Receptor N Tyrosine Kinases. *Biochim Biophys Acta.* 1834(10): 2213-2218.
- [15] Deinhardt, K.; Chao, M. V. (2014). Trk receptors. *Handb Exp Pharmacol.* 220: 103-19.
- [16] Tammiku-Taul, J.; Park, R.; Jaanson, K.; Luberg K.; Dobchev, D. A.; Kananovich, D.; Noole, A.; Mandel, M.; Kaasik, A.; Lopp, M.; Timmusk, T.; Karelson, M. (2016). Indole-like Trk receptor antagonists. *Eur J Med Chem.* 121: 541-552.
- [17] Skerratt, S. E.; Sharan, M. A.; Bagal, K.; Bilsland, J.; Brown, D.; Bungay, P. J.; Cole, S.; Gibson, K. R.; Jones, R.; Morao, I.; Nedderman, A.; Omoto, K.; Robinson, C.; Ryckmans, T.; Skinner, K.; Stuppel, P.; Waldron, G. (2016). Discovery of a Potent, Selective, and Peripherally Restricted PanTrk Inhibitor (PF-06273340) for the Treatment of Pain. *J. Med. Chem.* 59: 10084–10099.
- [18] Leopold, A. V.; Chernov, K. G.; Shemetov, A. A.; Verkhusha, V. V. (2019). Neurotrophin receptor tyrosine kinases regulated with near-infrared light. *Nat Commun.* 10: 1129.
- [19] Shahin, R.; Mansi, I.; Swellmeen, L.; Alwidyan, T.; Al-Hashimi, N.; Al-Qarar'h, Y.; Shaheen, O. (2018) Ligand-based computer aided drug design reveals new tropomyosin receptor kinase a (TrkA) inhibitors. *J Mol Graph Model.* 80: 327-352.
- [20] Stachel, S. J.; Sanders, J. M.; Henze, D. A.; Rudd, M. T.; Su, H. P.; Li, Y.; Nanda, K. K.; Egbertson, M. S.; Manley, P. J.; Jones, K. L.; Brnardic, E. J.; Green, A.; Grobler J. A.; Hanney, B.; Leitl, M.; Lai, M. T.; Munshi, V.; Murphy, D.; Rickert, K.; Riley, D.; Krasowska-Zoladek, A.; Daley, C.; Zuck, P.; Kane, S. A.; Bilodeau, M. T. (2014). Maximizing diversity from a kinase screen: identification of novel and selective pan-Trk inhibitors for chronic pain. *J Med Chem.* 57(13): 5800-5816.
- [21] Herbrich, S. M.; Kannan, S.; Nolo, R. M.; Hornbaker, M.; Chandra, J.; Zweidler-McKay, P. A. (2018). Characterization of TRKA signaling in acute myeloid leukemia. *Oncotarget.* 9(53): 30092–30105.
- [22] Hubbard, S. R.; Miller, W. T. (2007). Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr Opin Cell Biol.* 19(2): 117-123.
- [23] Avendaño, M. C. (1994). Introducción a la Química Farmacéutica. Primera edición. España: MC Graw Hill. 47-94.
- [24] Kubinyi, H. (2002) From Narcosis to Hyperspace: The History of QSAR. *Quant. Struct. -Act. Relat.* 21: 348-356.

- [25] Scior, T.; Medina-Franco, J. L.; QT, D.; Martínez-Mayorga, K.; Yunes Rojas, J. A.; Bernard, P. (2009). How to recognize and workaround pitfalls in QSAR studies: A Critical Review. *Current Medicinal Chemistry*, 16 (1): 4297-4313.
- [26] Sadeghian-Rizi, S.; Sakhteman, A.; Hassanzadeh, F. (2016) A quantitative structure-activity relationship (QSAR) study of some diaryl urea derivatives of B-RAF inhibitors. *Research in Pharmaceutical Sciences*. 11(6): 445-453.
- [27] Lozano Aponte, J.; Scior, T. (2012). ¿Qué sabe Ud. acerca de...QSAR?. *Rev Mex Cienc Farm*. 43(2): 82-84.
- [28] Polanski, J.; Bak, A.; Gieleciak, R.; Magdziarz. (2006). Modeling Robust QSAR. *J. Chem. Inf. Model*. 46(6): 2310-2318.
- [29] Toropova, A. A.; Toropovaa, A. P.; Beegb, M.; Gobbib, M.; Salmonac, M. (2017). QSAR model for blood-brain barrier permeation. *J Pharm Toxic Methods*. 88: 7-18.
- [30] Mauri, A. Consonni, V. Todeschini. R. *Molecular Descriptors*. En: Lesczynsky et al. (2017). *Handbook of Computational Chemistry*, First edition. Switzerland: Springer International Publishing. 2066-2090.
- [31] Todeschini, R.; Consonni, V. (2002). *Handbook of Molecular Descriptors*. Wiley-VCH. 1: 29-93.
- [32] Diudea, M. V.; Gutman, I.; Lorentz, J. (2001). *Molecular Topology*. Babes-Bolyai. 101-151.
- [33] Leach, A. R.; Gillet, V. J. (2007). *An Introduction to Chemoinformatics*. First edition. The Netherlands: Springer. 54-61.
- [34] Wildman, S. A.; Crippen, G. M. (1999). Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci*. 39: 868-873.
- [35] Geoffrey, R. N.; Streiner, D. L. (1998). *Bioestadística*. Segunda edición. España: Harcourt Brace. 100-118.
- [36] Gareth, J.; Witten, D.; Hastie, T.; Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer Science. 59-120
- [37] Horton, N. J.; Kleinman, K. (2015). *Using Rand RStudio for Data Management, Statistical Analysis, and Graphics* Second Edition. 67-76.
- [38] Amat, J. R. (2016) *Introducción a la Regresión Lineal Múltiple*, última consulta 23 febrero. Recuperado de https://rpubs.com/Joaquin_AR/226291
- [39] Pedretti, A. Vistoli, G. (2015). VegaZZ: Drug Desing Laboratoty. https://www.ddl.unimi.it/cms/index.php?Software_projects:VEGA_ZZ
- [40] Hippe, Z. (1991). *Artificial Intelligence in Chemistry: Structure Elucidation and Simulation of Organic Reactions*. 1st Edition. 73: 190-198.
- [41] Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. (2006). DRAGON software: An easy approach to molecular descriptor calculations. 56: 237-248.

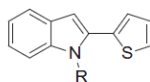
- [42] The Molecular Modelling group. (2019). Swiss ADME: Swiss Institute of Bioinformatics. <http://www.swissadme.ch/index.php>
- [43] STATISTICA, data analysis software system. (2011). StatSoft, Inc. version 10. www.statsoft.com.
- [44] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- [45] Wang, T.; Lamb, M. L.; Block, M. H.; Davies, A. M.; Han, Y.; Hoffmann, E.; Ioannidis, S.; Josey, J.A.; Liu, Z.; Lyne, P. D.; Macintyre, T.; Mohr, P. J.; Omer, C. A.; Sjogren, T.; Thress, K.; Wang, B.; Wang, H.; Yu, D.; Zhang, H. (2012). Discovery of Disubstituted Imidazo[4,5-B] Pyridines and Purines as Potent Trka Inhibitors. *Acs Med.Chem. Lett.* 3: 705.
- [46] Ultsch, M. H.; Wiesmann, C.; Simmons, L. C.; Henrich, J.; Yang, M.; Reilly, D.; Bass, S. H.; de Vos, A. M. (1999). Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J.Mol.Biol.* 290: 149-159.
- [47] Artim, S. C.; Mendrola, J. M.; Lemmon, M. A. (2012). Assessing the range of kinase autoinhibition mechanisms in the insulin receptor family. *Biochem. J.* 448: 213-220.
- [48] Bertrand, T.; Kothe, M.; Liu, J.; Dupuy, A.; Rak, A.; Berne, P. F.; Davis, S.; Gladysheva, T.; Valtre, C.; Crenne, J. Y.; Mathieu, M. (2012). The Crystal Structures of TrkA and TrkB Suggest Key Regions for Achieving Selective Inhibition. *J.Mol.Biol.* 423: 439-453.
- [49] Wiesmann, C.; Ultsch, M. H.; Bass, S. H.; de Vos, A. M. (1999). Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor. *Nature.* 401: 184-188.
- [50] Furuya, N.; Momose, T.; Katsuno, K.; Fushimi, N.; Muranaka, H.; Handa, C.; Ozawa, T.; Kinoshita, T. (2017). The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity. *Bioorg. Med. Chem. Lett.* 27: 1233-1236.
- [51] Robertson, A. G. S.; Banfield, M. J.; Allen, S. J.; Dando, J. A.; Mason, G. G. F.; Tyler, S. J.; Bennett, G. S.; Brain, S. D.; Clarke, A. R.; Naylor, R. L.; Wilcock, G. K.; Brady, R. L.; Dawbarn, D. (2001). Identification and Structure of the Nerve Growth Factor Binding Site on Trka. *Biochem Biophys Res Commun.* 282: 131.
- [52] Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. (2009). Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry.* 16: 2785-91.
- [53] Wojtulewski, S.; Grabowski, S. J. (2003). Different donors and acceptors for intramolecular hydrogen bonds. *Chemical Physics Letters.* 378(3): 388-394.
- [53] Comini, L. R.; Nuñez, S. C.; Arguello, G.; Cabrera, J. L. (2006). Determinación del coeficiente de partición (log P) para derivados antraquinónicos aislados de *Heterophyllaea pustulata* Hook. f. (Rubiáceas). *Acta Farm. Bonaerense.* 25(2): 252-255.
- [54] Padrón, R. J. A.; Carrasco, R.; Pellón, R. F. (2002). Molecular descriptor based on a molar refractivity partition using type graph – theoretical invariant. *J. Pharm Pharmaceut Sci.* 5(3): 258-265.

[55] Konai, M. M.; Samaddar, S.; Bocchinfuso, G.; Santucci, V.; Stella, L.; Haldar, J. (2018). Selectively targeting bacteria by tuning the molecular design of membrane-active peptidomimetic amphiphiles. *The Royal Society of Chemistry*. 56(1): 1-4.

10 SUPLEMENTO

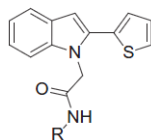
Estructuras de N-PTindoles extraídos de [5]

Table 1
SAR of linker region C.



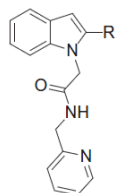
Compound	R	TrkA IC ₅₀ (nM)	TrkC IC ₅₀ (nM)	Selectivity (C/A)
14		>81,010	>9001	-
15		64,320	81,160	1.2
16		>81,010	25,140	-
17		74,020	80,020	1.1
18		>81,010	>81,010	-
19		>81,010	80,270	-
20		47,300	>81,010	1.7

Table 2
SAR of 2-pyridyl region D.



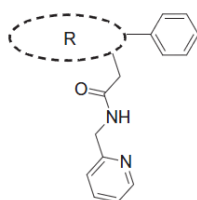
Compound	R	TrkA IC ₅₀ (nM)	TrkC IC ₅₀ (nM)	Selectivity (C/A)
21		9085	>81,010	9
22		4394	>81,010	18
23		8829	58,440	7
24		36,600	>27,000	-
25		>9001	>3000	-
26		1300	>81,010	>62
27		>81,010	>81,010	-
28		210	>81,010	>386
29		70	>81,010	>1157
30		3163	>81,010	>25
31		7310	38,500	5
32		9400	120,000	13
33		3388	>81,010	>24
34		218	71,720	328

Table 3
SAR of thiophene region B.



Compound	R	TrkA IC ₅₀ (nM)	TrkC IC ₅₀ (nM)	Selectivity (C/A)
35		662	47,860	72
36		12,970	73,070	5.6
37		3944	70,240	17
38		49,000	>81,010	>1.6
39		2400	>81,010	>34
40		961	>81,010	>84
41		2114	53,490	25
42		11,890	>81,010	>7
43		>81,010	>81,010	-
44		30,870	-	-
45		1424	25,960	18

Table 4
SAR of indole region A.



Compound	R	TrkA IC ₅₀ (nM)	TrkC IC ₅₀ (nM)	Selectivity (C/A)
46		>81,010	>81,010	-
47		22,000	>81,010	>4

Table 4 (continued)

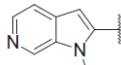
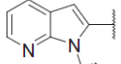
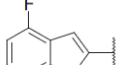
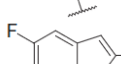
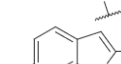
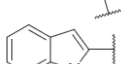
Compound	R	TrkA IC ₅₀ (nM)	TrkC IC ₅₀ (nM)	Selectivity (C/A)
48		14,000	51,000	4
49		99,820	38,550	–
50		960	>81,010	>84
51		880	>81,010	>92
52		2552	>81,010	>32
53		>27,000	>9001	–

Tabla S1: Descriptores moleculares obtenidos por Vega ZZ

Posición por actividad	ID de referencia	TrkA IC 50 (μM)	TrkA		MW (Da)	volumen (\AA^3)	log P crippen	MR	TPSA	DIPOLE	PSA	Carga
			IC 50 (nM)									
1	D29	0.07	70	386.47	336.4	3.4406	11.805	645.3	1.2932	98.8	0	
2	D28	0.21	210	412.51	374.7	3.8128	119.059	714.8	1.2916	96.9	0	
3	D34	0.218	218	386.12	338.6	4.2195	106.926	643.6	3.1676	79.4	0	
4	B35	0.662	662	361.12	331.2	3.2782	104.796	646.8	3.157	72.7	0	
5	A51	0.88	880	359.396	324	3.918	101.496	628.4	20.400	57.5	0	
6	A50	0.96	960	359.396	324.6	3.918	101.496	626.2	4.8638	57.5	0	
7	B40	0.961	961	348.104	306.7	3.32	99.2	606	3.5758	118.6	0	
8	D26	1.3	1300	336.411	292.1	2.3573	93.913	579.7	1.304	107.4	0	
9	B45	1.424	1424	341.406	317.8	3.7599	100.989	618	3.3584	57.5	0	
10	B41	2.114	2114	332.356	298.1	1.7859	93.9601	596.9	3.0583	81.7	0	
11	B39	2.4	2400	348.422	310.7	2.3519	99.5817	606.1	3.2198	97.4	0	
12	A52	2.552	2552	359.396	323.4	3.918	101.496	630.1	2.7386	57.5	0	
13	D30	3.163	3163	353.461	296.9	2.7227	100.780	576	2.3031	115.1	0	
14	D33	3.388	3388	412.507	370.3	4.4561	118.317	701.8	2.9162	101.1	0	
15	B37	3.944	3944	348.422	308.5	3.32	99.21	602.3	4.1199	110	0	
16	D22	4.394	4394	347.433	315	2.7936	100.587	607	2.7566	88	0	
17	D31	7.31	7310	337.396	288.9	2.1567	95.1585	567	1.8897	100.3	0	
18	D23	8.829	8829	348.422	310.1	2.2683	100.673	615.9	2.3178	95.8	0	
19	D25	9.001	9001	346.445	318.9	4.1307	102.187	614.5	2.1457	62.3	0	
20	D21	9.085	9085	347.433	313.7	2.7936	100.587	609.2	3.387	88	0	
21	D32	9.4	9400	336.411	294.5	2.13	94.82	577.6	3.0802	122.1	0	
22	B42	11.89	11890	342.394	314.6	2.7286	100.606	616.2	2.7994	59.5	0	
23	B36	12.97	12970	345.394	320.5	99.1744	2.7122	626.9	2.8377	55.8	0	
24	A48	14	14000	342.394	314.5	2.4228	99.3888	616.1	2.7987	83.8	0	
25	A47	22	22000	342.394	314.5	2.4228	99.3888	613.2	3.5437	82.5	0	
26	A53	27	27000	359.396	323.8	3.918	101.4963	622.8	2.871	57.5	0	
27	B44	30.87	30870	342.394	314.8	2.4228	99.3888	615	4.027	85.2	0	
28	D24	36.6	36600	348.422	310.2	1.8801	97.7661	605.8	2.2396	104.2	0	

29	C20	47.3	47300	361.46	330.3	3.1528	105.6434	623.2	3.4127	58.3	0
30	B38	49	49000	348.422	308.6	2.0763	99.5509	600.9	3.757	106.9	0
31	C15	64.32	64320	333.407	304.5	3.9858	93.6734	606.3	2.3763	91.1	0
32	C17	74.02	74020	347.433	315.3	3.5654	98.7606	619.5	3.9025	98.7	0
33	A46	81.01	81010	342.394	314.6	2.7286	100.6068	614.1	4.3427	81.8	0
34	B43	81.01	81010	342.394	313.9	2.4228	99.3888	615.4	3.3606	80.4	0
35	C14	81.01	81010	333.407	302.7	3.4412	95.9193	599.9	2.5943	77.3	0
36	C16	81.01	81010	333.45	310.6	3.6881	99.4734	612.6	1.4753	59	0
37	C18	81.01	81010	383.487	327.2	4.82		635.8	7.192	126.3	0
38	C19	81.01	81010	347.433	315	4.3903	98.2876	621.1	2.5254	95.5	0
39	D27	81.01	81010	350.437	310.4	2.8407	98.8643	612.8	1.1281	76.6	0
40	A49	99.82	99820	342.394	315.4	3.1673	98.4293	619.4	2.8757	73.7	0

Tabla S2: Descriptores obtenidos por Swiss ADME

Molecule	ID de referenc	Canonical SM	Formula	MW	#Heavy atom	#Aromatic he	Fraction Csp3	#Rotatable bo	#H-bond acce	#H-bond donc	MR	TPSA	llogp
Molecule 1	D29	O=C(Cn1c(cc2	C22H18N4OS	386.47	28	23	0.09	6	2	2	113.27	90.95	2.64
Molecule 2	D28	O=C(Cn1c(cc2	C24H20N4OS	412.51	30	25	0.08	7	2	2	121.2	90.95	2.19
Molecule 3	D34	O=C(Cn1c(cc2	C22H18N4OS	386.47	28	23	0.09	6	2	1	112.37	79.57	3.07
Molecule 4	B35	Cc1ccc(s1)c1c	C21H19N3OS	361.46	26	20	0.14	6	2	1	106.38	75.16	2.11
Molecule 5	A51	Fc1ccc2c(c1)c	C22H18FN3O	359.4	27	21	0.09	6	3	1	103.49	46.92	2.54
Molecule 6	A50	O=C(Cn1c(cc2	C22H18FN3O	359.4	27	21	0.09	6	3	1	103.49	46.92	2.99
Molecule 7	B40	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.7
Molecule 8	D26	O=C(Cn1c(cc2	C18H16N4OS	336.41	24	19	0.11	6	2	2	95.76	90.95	1.28
Molecule 9	B45	O=C(Cn1c(cc2	C22H19N3O	341.41	26	21	0.09	6	2	1	103.53	46.92	2.82
Molecule 10	B41	O=C(Cn1c(cc2	C19H16N4O2	332.36	25	20	0.11	6	4	1	93.59	72.95	2.54
Molecule 11	B39	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.3
Molecule 12	A52	O=C(Cn1c(cc2	C22H18FN3O	359.4	27	21	0.09	6	3	1	103.49	46.92	2.98
Molecule 13	D30	O=C(Cn1c(cc2	C18H15N3OS	353.46	24	19	0.11	6	2	1	99.29	103.4	2.59
Molecule 14	D33	O=C(Cn1c(cc2	C24H20N4OS	412.51	30	25	0.08	7	2	2	121.2	90.95	2.96
Molecule 15	B37	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.68
Molecule 16	D22	O=C(Cn1c(cc2	C20H17N3OS	347.43	25	20	0.1	6	2	1	101.41	75.16	2.56
Molecule 17	D31	O=C(Cn1c(cc2	C18H15N3O2	337.4	24	19	0.11	6	3	1	93.68	88.3	2.66
Molecule 18	D23	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.82
Molecule 19	D25	O=C(Cn1c(cc2	C21H18N2OS	346.45	25	20	0.1	6	1	1	103.62	62.27	3.19
Molecule 20	D21	O=C(Cn1c(cc2	C20H17N3OS	347.43	25	20	0.1	6	2	1	101.41	75.16	2.68
Molecule 21	D32	O=C(Cn1c(cc2	C18H16N4OS	336.41	24	19	0.11	6	2	2	95.76	90.95	1.97
Molecule 22	B42	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.7
Molecule 23	B36	Cc1ccc(o1)c1c	C21H19N3O2	345.39	26	20	0.14	6	3	1	100.77	60.06	2.8
Molecule 24	A48	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.6
Molecule 25	A47	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.6
Molecule 26	A53	O=C(Cn1c(cc2	C22H18FN3O	359.4	27	21	0.09	6	3	1	103.49	46.92	3.11
Molecule 27	B44	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.55
Molecule 28	D24	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.24
Molecule 29	C20	CN(C(=O)Cn1c	C21H19N3OS	361.46	26	20	0.14	6	2	0	106.31	66.37	2.95
Molecule 30	B38	O=C(Cn1c(cc2	C19H16N4OS	348.42	25	20	0.11	6	3	1	99.21	88.05	2.42
Molecule 31	C15	O=C(n1c(cc2c	C19H15N3OS	333.41	24	20	0.05	5	2	1	96.99	75.16	2.59
Molecule 32	C17	O=C(c1ccccc1	C20H17N3OS	347.43	25	20	0.1	6	2	1	101.63	75.16	1.9
Molecule 33	A46	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.57
Molecule 34	B43	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.51
Molecule 35	C14	O=C(Cn1c(cc2	C19H15N3OS	333.41	24	20	0.05	5	2	1	98.14	75.16	2.57
Molecule 36	C16	c1ccc(nc1)CN	C20H19N3S	333.45	24	20	0.15	6	2	1	101.21	58.09	3.18
Molecule 37	C18	O=S(=O)(c1cc	C19H17N3O2	383.49	26	20	0.11	6	4	1	104.53	100.61	2.49
Molecule 38	C19	O=C(n1c(cc2c	C20H17N3OS	347.43	25	20	0.1	6	2	1	101.8	75.16	2.89
Molecule 39	D27	O=C(Cn1c(cc2	C19H18N4OS	350.44	25	19	0.16	6	2	1	100.66	80.09	2.22
Molecule 40	A49	O=C(Cn1c(cc2	C21H18N4O	342.39	26	21	0.1	6	3	1	101.33	59.81	2.78

Tabla S2: Descriptores obtenidos por Swiss ADME

Molecule	ID de referenc	XLOGP3	WLOGP	MLOGP	Silicos-IT Log	Consensus Lo	ESOL Log S	ESOL Solubilit	ESOL Solubilit	ESOL Class	Ali Log S	Ali Solubility (i	Ali Solubility (i	Ali Class	Silicos-IT Log ^c	Silicos-IT Solu
Molecule 1	D29	3.92	4.41	2.41	5.04	3.69	-4.92	4.67E-03	1.21E-05	Moderately sc	-5.53	1.14E-03	2.96E-06	Moderately sc	-8.02	3.71E-06
Molecule 2	D28	4.16	4.92	2.26	5.58	3.82	-5.17	2.77E-03	6.71E-06	Moderately sc	-5.78	6.88E-04	1.67E-06	Moderately sc	-8.84	6.00E-07
Molecule 3	D34	4.19	4.18	2.14	3.96	3.51	-5.09	3.16E-03	8.17E-06	Moderately sc	-5.57	1.04E-03	2.69E-06	Moderately sc	-7.19	2.48E-05
Molecule 4	B35	3.68	4.24	2.22	5.03	3.46	-4.57	9.67E-03	2.67E-05	Moderately sc	-4.95	4.07E-03	1.13E-05	Moderately sc	-7.54	1.05E-05
Molecule 5	A51	3.67	4.43	2.78	4.3	3.54	-4.56	9.90E-03	2.75E-05	Moderately sc	-4.34	1.62E-02	4.52E-05	Moderately sc	-8.16	2.47E-06
Molecule 6	A50	3.67	4.43	2.78	4.3	3.63	-4.56	9.90E-03	2.75E-05	Moderately sc	-4.34	1.62E-02	4.52E-05	Moderately sc	-8.16	2.47E-06
Molecule 7	B40	2.67	3.32	0.94	3.95	2.72	-3.88	4.61E-02	1.32E-04	Soluble	-4.17	2.35E-02	6.75E-05	Moderately sc	-6.79	5.68E-05
Molecule 8	D26	2.49	3.26	1.15	4.03	2.44	-3.68	6.96E-02	2.07E-04	Soluble	-4.05	3.03E-02	9.02E-05	Moderately sc	-6.38	1.41E-04
Molecule 9	B45	3.57	3.87	2.4	3.89	3.31	-4.41	1.34E-02	3.91E-05	Moderately sc	-4.24	1.96E-02	5.74E-05	Moderately sc	-7.89	4.35E-06
Molecule 10	B41	2.02	2.86	0.8	2.71	2.18	-3.37	1.42E-01	4.27E-04	Soluble	-3.18	2.20E-01	6.62E-04	Soluble	-6.74	6.01E-05
Molecule 11	B39	2.64	3.32	1.21	3.95	2.68	-3.86	4.82E-02	1.38E-04	Soluble	-4.14	2.53E-02	7.25E-05	Moderately sc	-6.79	5.68E-05
Molecule 12	A52	3.67	4.43	2.78	4.3	3.63	-4.56	9.90E-03	2.75E-05	Moderately sc	-4.34	1.62E-02	4.52E-05	Moderately sc	-8.16	2.47E-06
Molecule 13	D30	3.39	3.99	1.57	5.15	3.34	-4.36	1.55E-02	4.40E-05	Moderately sc	-5.24	2.03E-03	5.75E-06	Moderately sc	-6.43	1.33E-04
Molecule 14	D33	4.16	4.92	2.53	5.58	4.03	-5.17	2.77E-03	6.71E-06	Moderately sc	-5.78	6.88E-04	1.67E-06	Moderately sc	-8.84	6.00E-07
Molecule 15	B37	2.67	3.32	0.94	3.95	2.71	-3.88	4.61E-02	1.32E-04	Soluble	-4.17	2.35E-02	6.75E-05	Moderately sc	-6.79	5.68E-05
Molecule 16	D22	3.25	3.93	1.99	4.52	3.25	-4.24	2.01E-02	5.79E-05	Moderately sc	-4.5	1.09E-02	3.15E-05	Moderately sc	-7.16	2.40E-05
Molecule 17	D31	2.77	3.52	1.15	3.9	2.8	-3.87	4.59E-02	1.36E-04	Soluble	-4.28	1.77E-02	5.25E-05	Moderately sc	-6.38	1.40E-04
Molecule 18	D23	2.64	3.32	1.34	3.95	2.82	-3.86	4.82E-02	1.38E-04	Soluble	-4.14	2.53E-02	7.25E-05	Moderately sc	-6.79	5.68E-05
Molecule 19	D25	4.32	4.53	3.07	5.08	4.04	-4.91	4.31E-03	1.24E-05	Moderately sc	-5.34	1.58E-03	4.55E-06	Moderately sc	-7.53	1.01E-05
Molecule 20	D21	3.25	3.93	1.99	4.52	3.27	-4.24	2.01E-02	5.79E-05	Moderately sc	-4.5	1.09E-02	3.15E-05	Moderately sc	-7.16	2.40E-05
Molecule 21	D32	2.49	3.26	1.15	4.03	2.58	-3.68	6.96E-02	2.07E-04	Soluble	-4.05	3.03E-02	9.02E-05	Moderately sc	-6.38	1.41E-04
Molecule 22	B42	2.53	3.26	1.35	3.32	2.63	-3.76	5.97E-02	1.74E-04	Soluble	-3.43	1.26E-01	3.69E-04	Soluble	-7.52	1.03E-05
Molecule 23	B36	3.07	3.77	1.39	3.78	2.96	-4.09	2.82E-02	8.15E-05	Moderately sc	-4	3.47E-02	1.00E-04	Soluble	-7.49	1.11E-05
Molecule 24	A48	2.5	3.26	1.35	3.32	2.61	-3.74	6.24E-02	1.82E-04	Soluble	-3.4	1.36E-01	3.97E-04	Soluble	-7.52	1.03E-05
Molecule 25	A47	2.5	3.26	1.35	3.32	2.61	-3.74	6.24E-02	1.82E-04	Soluble	-3.4	1.36E-01	3.97E-04	Soluble	-7.52	1.03E-05
Molecule 26	A53	3.67	4.43	2.78	4.3	3.66	-4.56	9.90E-03	2.75E-05	Moderately sc	-4.34	1.62E-02	4.52E-05	Moderately sc	-8.16	2.47E-06
Molecule 27	B44	2.5	3.26	1.35	3.32	2.6	-3.74	6.24E-02	1.82E-04	Soluble	-3.4	1.36E-01	3.97E-04	Soluble	-7.52	1.03E-05
Molecule 28	D24	2.21	3.32	0.94	3.95	2.53	-3.59	8.99E-02	2.58E-04	Soluble	-3.69	7.06E-02	2.03E-04	Soluble	-6.79	5.68E-05
Molecule 29	C20	3.47	4.27	2.22	4.46	3.47	-4.44	1.31E-02	3.63E-05	Moderately sc	-4.55	1.03E-02	2.85E-05	Moderately sc	-6.82	5.43E-05
Molecule 30	B38	2.64	3.32	0.94	3.95	2.66	-3.86	4.82E-02	1.38E-04	Soluble	-4.14	2.53E-02	7.25E-05	Moderately sc	-6.79	5.68E-05
Molecule 31	C15	3.55	4.37	2.85	4.14	3.5	-4.43	1.24E-02	3.71E-05	Moderately sc	-4.81	5.12E-03	1.54E-05	Moderately sc	-6.76	5.73E-05
Molecule 32	C17	3.55	4.19	2.26	4.52	3.28	-4.43	1.30E-02	3.74E-05	Moderately sc	-4.81	5.34E-03	1.54E-05	Moderately sc	-7.16	2.40E-05
Molecule 33	A46	2.59	3.26	1.62	3.32	2.67	-3.8	5.47E-02	1.60E-04	Soluble	-3.49	1.10E-01	3.20E-04	Soluble	-7.52	1.03E-05
Molecule 34	B43	2.5	3.26	1.35	3.32	2.59	-3.74	6.24E-02	1.82E-04	Soluble	-3.4	1.36E-01	3.97E-04	Soluble	-7.52	1.03E-05
Molecule 35	C14	3.65	4.21	2.44	4.14	3.4	-4.49	1.07E-02	3.21E-05	Moderately sc	-4.92	4.03E-03	1.21E-05	Moderately sc	-6.76	5.73E-05
Molecule 36	C16	3.34	4.4	2.51	4.99	3.68	-4.23	1.95E-02	5.86E-05	Moderately sc	-4.24	1.93E-02	5.79E-05	Moderately sc	-7.63	7.89E-06
Molecule 37	C18	3.24	4.82	1.54	3.41	3.1	-4.43	1.42E-02	3.70E-05	Moderately sc	-5.03	3.61E-03	9.42E-06	Moderately sc	-7.11	2.97E-05
Molecule 38	C19	4.01	4.57	3.07	4.52	3.81	-4.72	6.68E-03	1.92E-05	Moderately sc	-5.29	1.78E-03	5.12E-06	Moderately sc	-7.16	2.40E-05
Molecule 39	D27	2.45	3.27	1.38	3.5	2.56	-3.72	6.64E-02	1.89E-04	Soluble	-3.78	5.88E-02	1.68E-04	Soluble	-5.93	4.11E-04
Molecule 40	A49	2.83	3.26	2.03	3.32	2.84	-3.95	3.86E-02	1.13E-04	Soluble	-3.74	6.17E-02	1.80E-04	Soluble	-7.52	1.03E-05

Tabla S2: Descriptores obtenidos en Swiss ADME

Silicos-IT Solu	Silicos-IT clas:	GI absorption	BBB permean	Pgp substrate	CYP1A2 inhibi	CYP2C19 inhibi	CYP2C9 inhibi	CYP2D6 inhibi	CYP3A4 inhibi	log Kp (cm/s)	Lipinski #violat	Ghose #violat	Veber #violat	Egan #violatic	Muegge #viol:	Bioavailability
9.59E-09	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.87	0	0	0	0	0	0.55
1.46E-09	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.86	0	0	0	0	0	0.55
6.41E-08	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.68	0	0	0	0	0	0.55
2.90E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.89	0	0	0	0	0	0.55
6.87E-09	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.89	0	0	0	0	0	0.55
6.87E-09	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.89	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.53	0	0	0	0	0	0.55
4.18E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.58	0	0	0	0	0	0.55
1.27E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.85	0	0	0	0	0	0.55
1.81E-07	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.89	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.55	0	0	0	0	0	0.55
6.87E-09	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.89	0	0	0	0	0	0.55
3.75E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.05	0	0	0	0	0	0.55
1.46E-09	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.86	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.53	0	0	0	0	0	0.55
6.90E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.11	0	0	0	0	0	0.55
4.16E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.39	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.55	0	0	0	0	0	0.55
2.92E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.35	0	0	0	0	0	0.55
6.90E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.11	0	0	0	0	0	0.55
4.18E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.58	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.59	0	0	0	0	0	0.55
3.21E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.23	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.61	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.61	0	0	0	0	0	0.55
6.87E-09	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.89	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.61	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.86	0	0	0	0	0	0.55
1.50E-07	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.04	0	0	0	0	0	0.55
1.63E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.55	0	0	0	0	0	0.55
1.72E-07	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.81	0	0	0	0	0	0.55
6.90E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.9	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.55	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.61	0	0	0	0	0	0.55
1.72E-07	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.74	0	0	0	0	0	0.55
2.37E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.96	0	0	0	0	0	0.55
7.74E-08	Poorly soluble	High	No	No	Yes	Yes	Yes	Yes	Yes	-6.34	0	0	0	0	0	0.55
6.90E-08	Poorly soluble	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.57	0	0	0	0	0	0.55
1.17E-06	Moderately sol	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.7	0	0	0	0	0	0.55
3.01E-08	Poorly soluble	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-6.38	0	0	0	0	0	0.55

Tabla S3: Descriptores obtenidos por E- Dragon

No.	Id de referencia	MW	AMW	Sv	Se	Sp	Ss	Mv	Me	Mp	Ms	nAT	nSK	nBT	
1	D29	386.51	8.4	31.76	46	33.45	59.17	0.69		1	0.73	2.11	46	28	50
2	D28	412.55	8.25	34.36	49.88	36.22	63.17	0.69		1	0.72	2.11	50	30	54
3	D34	386.51	8.4	31.76	46	33.45	59	0.69		1	0.73	2.11	46	28	50
4	B35	361.5	8.03	30.36	44.78	32.21	55	0.67		1	0.72	2.12	45	26	48
5	A51	359.43	7.99	30.39	45.21	31.5	63.17	0.68		1	0.7	2.34	45	27	48
6	A50	359.43	7.99	30.39	45.21	31.5	63.17	0.68		1	0.7	2.34	45	27	48
7	B40	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
8	D26	336.45	8.41	27.16	40.11	28.69	51.83	0.68		1	0.72	2.16	40	24	43
9	B45	341.44	7.59	30.28	44.7	31.56	55.5	0.67	0.99		0.7	2.13	45	26	48
10	B41	332.39	8.11	27.59	41.36	28.5	56	0.67	1.01		0.7	2.24	41	25	44
11	B39	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
12	A52	359.43	7.99	30.39	45.21	31.5	63.17	0.68		1	0.7	2.34	45	27	48
13	D30	353.5	9.06	27.26	39.09	29.34	51.17	0.7		1	0.75	2.13	39	24	42
14	D33	412.55	8.25	34.36	49.88	36.22	63.17	0.69		1	0.72	2.11	50	30	54
15	B37	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
16	D22	347.47	8.27	28.77	41.89	30.45	53.33	0.68		1	0.72	2.13	42	25	45
17	D31	337.43	8.65	26.68	39.34	28.14	52.83	0.68	1.01		0.72	2.2	39	24	42
18	D23	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
19	D25	346.48	8.06	29.37	42.68	31.2	52.33	0.68	0.99		0.73	2.09	43	25	46
20	D21	347.47	8.27	28.77	41.89	30.45	53.33	0.68		1	0.72	2.13	42	25	45
21	D32	336.45	8.41	27.16	40.11	28.69	51.83	0.68		1	0.72	2.16	40	24	43
22	B42	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
23	B36	345.43	7.68	29.79	45.03	31.02	56.67	0.66		1	0.69	2.18	45	26	48
24	A48	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
25	A47	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
26	A53	359.43	7.99	30.39	45.21	31.5	63.17	0.68		1	0.7	2.34	45	27	48
27	B44	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
28	D24	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
29	C20	361.5	8.03	30.36	44.78	32.21	54.83	0.67		1	0.72	2.11	45	26	48
30	B38	348.46	8.5	28.16	41.11	29.69	54.33	0.69		1	0.72	2.17	41	25	44
31	C15	333.44	8.55	27.17	39.01	28.69	51.83	0.7		1	0.74	2.16	39	24	42
32	C17	347.47	8.27	28.77	41.89	30.45	53.33	0.68		1	0.72	2.13	42	25	45
33	A46	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
34	B43	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47
35	C14	333.44	8.55	27.17	39.01	28.69	51.83	0.7		1	0.74	2.16	39	24	42
36	C16	333.49	7.76	28.85	42.45	30.76	46.17	0.67	0.99		0.72	1.92	43	24	46
37	C18	383.53	8.92	29.37	43.3	31.55	59.58	0.68	1.01		0.73	2.29	43	26	46
38	C19	347.47	8.27	28.77	41.89	30.45	53.33	0.68		1	0.72	2.13	42	25	45
39	D27	350.48	8.15	28.76	43	30.45	53.33	0.67		1	0.71	2.13	43	25	46
40	A49	342.43	7.78	29.67	43.92	30.81	56.5	0.67		1	0.7	2.17	44	26	47

Tabla S3: Descriptores obtenidos por E- Dragon

nBO	nBM	SCBO	ARR	nCIC	nCIR	RBN	RBF	nDB	nTB	nAB	nH	nC	nN
32	26	45.5	0.781	5	7	5	0.1	1	0	25	18	22	4
34	27	48	0.765	5	6	6	0.111	1	0	26	20	24	4
32	26	45.5	0.781	5	7	5	0.1	1	0	25	18	22	4
29	22	40.5	0.724	4	5	5	0.104	1	0	21	19	21	3
30	23	42	0.733	4	5	5	0.104	1	0	22	18	22	3
30	23	42	0.733	4	5	5	0.104	1	0	22	18	22	3
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
27	21	38	0.741	4	5	5	0.116	1	0	20	16	18	4
29	23	41	0.759	4	5	5	0.104	1	0	22	19	22	3
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
30	23	42	0.733	4	5	5	0.104	1	0	22	18	22	3
27	21	38	0.741	4	5	5	0.119	1	0	20	15	18	3
34	27	48	0.765	5	6	6	0.111	1	0	26	20	24	4
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
28	22	39.5	0.75	4	5	5	0.111	1	0	21	17	20	3
27	21	38	0.741	4	5	5	0.119	1	0	20	15	18	3
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
28	22	39.5	0.75	4	5	5	0.109	1	0	21	18	21	2
28	22	39.5	0.75	4	5	5	0.111	1	0	21	17	20	3
27	21	38	0.741	4	5	5	0.116	1	0	20	16	18	4
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
29	22	40.5	0.724	4	5	5	0.104	1	0	21	19	21	3
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
30	23	42	0.733	4	5	5	0.104	1	0	22	18	22	3
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
28	22	39.5	0.75	4	5	5	0.114	1	0	21	16	19	4
27	22	38.5	0.778	4	5	3	0.071	1	0	21	15	19	3
28	22	39.5	0.75	4	5	5	0.111	1	0	21	17	20	3
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4
27	22	38.5	0.778	4	5	4	0.095	1	0	21	15	19	3
27	21	37.5	0.778	4	5	6	0.13	0	0	21	19	20	3
29	23	41.5	0.724	4	5	6	0.13	2	0	21	17	19	3
28	22	39.5	0.75	4	5	4	0.089	1	0	21	17	20	3
28	21	39	0.714	4	5	5	0.109	1	0	20	18	19	4
29	23	41	0.759	4	5	5	0.106	1	0	22	18	21	4

Tabla S3: Descriptores obtenidos por E- Dragon

ZM1	ZM1V	ZM2	ZM2V	Qindex	SNar	HNar	GNar	Xt	Dz	Ram	PoI	LPRS	VDA	MSD	
	154	362.444	184	394.667	24	22.364	2.154	2.223	0.211	59	9	41	140.904	156	0.242
	162	380.444	192	410.667	24	23.75	2.143	2.207	0.205	63	9	43	155.696	182.733	0.248
	154	364.444	184	396.667	24	22.364	2.154	2.223	0.211	59	9	41	140.904	156	0.242
	138	322.444	162	337.333	20	19.879	2.053	2.148	0.224	54.5	8	37	126.802	134.077	0.241
	142	388	166	389	20	20.572	2.051	2.142	0.22	58	8	40	133.317	142.593	0.237
	142	388	167	390	20	20.572	2.051	2.142	0.22	58	8	41	133.049	141.111	0.234
	132	330.444	155	341.333	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	128	312.444	151	327.667	19	18.781	2.118	2.187	0.231	51	7	32	113.404	115.167	0.243
	136	332	159	355	19	20.167	2.108	2.172	0.223	54.5	7	38	126.818	134.154	0.241
	132	366	155	381	19	19.474	2.113	2.179	0.227	54	7	35	120.296	125.52	0.245
	132	330.444	155	343.667	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	142	388	166	389	20	20.572	2.051	2.142	0.22	58	8	40	133.174	141.778	0.235
	128	296.889	151	304.333	19	18.781	2.118	2.187	0.231	50.5	7	32	113.404	115.167	0.243
	162	380.444	192	411.667	24	23.75	2.143	2.207	0.205	63	9	43	155.696	182.733	0.248
	132	330.444	155	337	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	132	314.444	155	327.667	19	19.474	2.113	2.179	0.227	52.5	7	35	120.296	125.52	0.245
	128	332.444	151	341.667	19	18.781	2.118	2.187	0.231	51.5	7	32	113.404	115.167	0.243
	132	330.444	155	343.667	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	132	298.444	155	315.667	19	19.474	2.113	2.179	0.227	52	7	35	120.296	125.52	0.245
	132	314.444	155	327.667	19	19.474	2.113	2.179	0.227	52.5	7	35	120.296	125.52	0.245
	128	312.444	151	326.667	19	18.781	2.118	2.187	0.231	51	7	32	113.404	115.167	0.243
	136	348	159	369	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	138	358	162	380	20	19.879	2.053	2.148	0.224	55.5	8	37	126.802	134.077	0.241
	136	348	159	367	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	136	348	159	367	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	142	388	167	390	20	20.572	2.051	2.142	0.22	58	8	41	132.896	140.296	0.232
	136	348	159	367	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	132	330.444	155	341.667	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	138	324.444	163	340.667	20	19.879	2.053	2.148	0.224	54.5	8	38	126.148	130.538	0.233
	132	330.444	155	341.667	19	19.474	2.113	2.179	0.227	53	7	35	120.296	125.52	0.245
	128	310.444	152	331.667	19	18.781	2.118	2.187	0.231	50.5	7	35	112.666	111.667	0.235
	132	314.444	156	333.667	19	19.474	2.113	2.179	0.227	52.5	7	36	120.622	126.96	0.248
	136	348	159	369	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	136	348	159	367	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241
	128	310.444	151	329.667	19	18.781	2.118	2.187	0.231	50.5	7	34	112.813	112.25	0.236
	126	266.444	148	293.667	18	19.068	2.182	2.213	0.229	49.5	6	33	114.6	121	0.258
	140	334.889	166	291	21	19.761	2.039	2.138	0.225	55.5	8	39	126.616	132.615	0.238
	132	314.444	156	335.667	19	19.474	2.113	2.179	0.227	52.5	7	36	120.179	125.04	0.244
	134	322.444	159	339.667	20	19.186	2.055	2.154	0.228	53	8	35	120.002	123.92	0.241
	136	348	159	369	19	20.167	2.108	2.172	0.223	55	7	38	126.818	134.154	0.241

Tabla S3: Descriptores obtenidos por E- Dragon

X5sol	XMOD	RDCHI	RDSQ	ISIZ	IAC	AAC	IDE	IDM	IDDE	IDDM	IDET	IDMT	IVDE	IVDM
7.852	89.583	4.133	251.378	254.084	72.918	1.585	3.588	8.311	4.164	4.782	1356.127	18151.053	1.108	4.769
7.942	95.629	4.331	271.105	282.193	77.715	1.554	3.75	8.5	4.507	4.881	1631.178	23299.4	1.075	4.87
7.852	89.749	4.133	251.378	254.084	72.918	1.585	3.588	8.311	4.164	4.782	1356.127	18151.053	1.108	4.769
7.093	82.327	3.879	220.958	247.133	69.429	1.543	3.475	8.109	4.547	4.67	1129.22	14134.62	1.239	4.65
6.766	82.11	3.951	232.44	247.133	69.212	1.538	3.505	8.222	4.236	4.723	1230.408	15828.116	1.218	4.706
6.766	82.211	3.932	233.676	247.133	69.212	1.538	3.483	8.225	4.31	4.724	1222.551	15667.719	1.218	4.706
6.484	81.335	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
5.936	77.876	3.739	198.736	212.877	65.818	1.645	3.346	7.879	4.002	4.555	923.529	10889	1.099	4.546
6.766	78.881	3.904	219.753	247.133	63.56	1.412	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.154	77.243	3.828	208.712	219.66	64.949	1.584	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.549	80.876	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.766	82.11	3.945	232.803	247.133	69.212	1.538	3.487	8.225	4.208	4.723	1223.762	15742.92	1.218	4.706
6.232	81.964	3.739	198.736	206.131	65.714	1.685	3.346	7.879	4.002	4.555	923.529	10889	1.099	4.546
7.942	95.629	4.331	271.105	282.193	77.715	1.554	3.75	8.5	4.507	4.881	1631.178	23299.4	1.075	4.87
6.549	80.922	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.549	80.468	3.828	208.712	226.477	65.797	1.567	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
5.936	78.331	3.739	198.736	206.131	65.714	1.685	3.346	7.879	4.002	4.555	923.529	10889	1.099	4.546
6.549	80.876	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.549	79.968	3.828	208.712	233.329	64.032	1.489	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.549	80.468	3.828	208.712	226.477	65.797	1.567	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
5.936	77.922	3.739	198.736	212.877	65.818	1.645	3.346	7.879	4.002	4.555	923.529	10889	1.099	4.546
6.766	79.335	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.539	79.061	3.879	220.958	247.133	67.429	1.498	3.475	8.109	4.547	4.67	1129.22	14134.62	1.239	4.65
6.766	79.381	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.766	79.381	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.766	82.211	3.924	234.183	247.133	69.212	1.538	3.472	8.226	4.208	4.724	1218.697	15580.73	1.218	4.706
6.766	79.381	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.549	80.922	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.717	83.092	3.841	223.143	247.133	69.429	1.543	3.417	8.117	4.316	4.672	1110.566	13775.197	1.239	4.65
6.549	80.922	3.828	208.712	219.66	66.949	1.633	3.429	7.993	4.084	4.614	1028.673	12541.254	1.078	4.606
6.512	77.486	3.69	201.576	206.131	62.062	1.591	3.297	7.887	4.168	4.555	909.849	10568.822	1.099	4.546
6.382	80.523	3.834	208	226.477	65.797	1.567	3.44	7.99	4.164	4.617	1032.019	12680.587	1.078	4.606
6.766	79.335	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.766	79.381	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663
6.388	77.376	3.701	200.723	206.131	62.062	1.591	3.3	7.888	4.002	4.556	910.666	10625.116	1.099	4.546
6.224	77.528	3.816	194.598	233.329	61.425	1.428	3.444	7.862	4.054	4.555	950.498	11415.718	0.811	4.56
7.207	91.07	3.837	222.822	233.329	74.377	1.73	3.433	8.112	4.239	4.675	1115.575	13984.43	1.354	4.642
6.555	80.486	3.817	209.606	226.477	65.797	1.567	3.428	7.991	4.134	4.613	1028.281	12490.588	1.078	4.606
6.16	80.546	3.782	210.764	233.329	69.56	1.618	3.378	7.998	4.164	4.616	1013.455	12388.468	1.26	4.592
6.766	79.335	3.904	219.753	240.215	64.918	1.475	3.475	8.11	4.393	4.67	1129.295	14144.076	1.058	4.663

Tabla S3: Descriptores obtenidos por E- Dragon

qmax	Qpos	Qneg	Qtot	Qmean	Q2	RPCG	RNCG	SPP	TE1	TE2	PCWTe	LDI	Ui	Hy
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.755	0.212
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.807	0.177
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.755	-0.358
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.382
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.382
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	0.296
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.408
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.382
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	-0.301
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.807	0.177
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.356
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	-0.301
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.396
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.356
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	0.296
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.382
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.798
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.342
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.356
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.342
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	-0.383
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.294
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.524	-0.356
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.459	-0.317
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4.585	-0.37

Tabla S3: Descriptores obtenidos por E- Dragon

AMR	TPSA(NO)	TPSA(Tot)	MLOGP	MLOGP2	ALOGP	ALOGP2	LAI	GVWAI-80	GVWAI-50	Inflammat-80	Inflammat-50	Depressant-80	Depressant-50	Psychotic-80	Psychotic-50
111.843	62.71	90.95	2.953	8.719	4.872	23.732	0	1	0	0	0	0	0	1	0
120.901	62.71	90.95	3.576	12.785	5.053	25.533	0	1	0	0	0	0	0	1	0
111.551	51.33	79.57	2.953	8.719	4.68	21.904	0	1	0	0	0	0	0	1	0
105.768	46.92	75.16	2.782	7.741	4.373	19.121	0	1	0	1	0	1	0	1	0
102.309	46.92	46.92	3.036	9.22	4.479	20.064	0	1	0	1	0	1	0	1	0
102.309	46.92	46.92	3.036	9.22	4.479	20.064	0	1	0	1	0	1	0	1	0
99.606	59.81	88.05	1.544	2.384	3.715	13.798	0	1	1	1	0	1	0	1	0
95.887	62.71	90.95	2.198	4.83	3.243	10.518	0	1	1	1	1	1	0	1	0
102.092	46.92	46.92	2.661	7.078	4.274	18.265	0	1	0	1	0	1	0	1	0
92.279	72.95	72.95	1.584	2.509	2.578	6.645	0	1	1	1	0	1	0	1	0
98.722	59.81	88.05	1.544	2.384	3.136	9.834	0	1	1	1	0	1	0	1	0
102.309	46.92	46.92	3.036	9.22	4.479	20.064	0	1	0	1	0	1	0	1	0
100.085	46.92	103.4	2.173	4.723	3.812	14.528	0	1	1	1	0	1	0	0	0
120.652	62.71	90.95	3.576	12.785	5.19	26.937	0	1	0	0	0	0	0	1	0
99.855	59.81	88.05	1.544	2.384	3.578	12.799	0	1	1	1	0	1	0	1	0
101.299	46.92	75.16	2.559	6.55	3.799	14.43	0	1	1	1	0	1	0	1	0
93.642	60.06	88.3	2.198	4.83	3.253	10.584	0	1	1	1	1	1	0	0	0
97.28	59.81	88.05	2.356	5.551	3.272	10.707	0	1	1	1	0	1	0	1	0
103.455	34.03	62.27	3.59	12.888	4.949	24.496	0	1	0	0	0	0	0	1	0
101.299	46.92	75.16	2.559	6.55	3.799	14.43	0	1	1	1	0	1	0	1	0
96.038	62.71	90.95	2.198	4.83	3.265	10.66	0	1	1	1	1	1	0	1	0
99.563	59.81	59.81	1.922	3.694	3.552	12.615	0	1	1	1	0	1	0	1	0
99.324	60.06	60.06	1.995	3.978	3.815	14.55	0	1	1	1	0	1	0	1	0
99.935	59.81	59.81	1.654	2.736	3.123	9.754	0	1	1	1	0	1	0	1	0
99.935	59.81	59.81	1.654	2.736	3.123	9.754	0	1	1	1	0	1	0	1	0
102.309	46.92	46.92	3.036	9.22	4.479	20.064	0	1	0	1	0	1	0	1	0
99.935	59.81	59.81	1.654	2.736	3.123	9.754	0	1	1	1	0	1	0	1	0
98.77	59.81	88.05	1.544	2.384	3.077	9.465	0	1	1	1	0	1	0	1	0
105.823	38.13	66.37	2.782	7.741	4.433	19.652	0	1	0	1	0	1	0	1	0
99.123	59.81	88.05	1.544	2.384	3.021	9.125	0	1	1	1	0	1	0	1	0
96.215	46.92	75.16	2.738	7.495	4.734	22.411	0	1	0	0	0	1	0	0	0
100.927	46.92	75.16	2.827	7.993	4.227	17.869	0	1	0	1	0	1	0	1	0
99.563	59.81	59.81	1.654	2.736	3.552	12.615	0	1	1	1	0	1	0	1	0
99.935	59.81	59.81	1.654	2.736	3.123	9.754	0	1	1	1	0	1	0	1	0
96.777	46.92	75.16	3.412	11.64	4.331	18.757	0	1	0	1	0	1	0	0	0
100.697	29.85	58.09	3.035	9.211	4.557	20.764	0	1	0	0	0	1	0	1	0
105.189	63.99	100.61	2.146	4.603	4.235	17.938	0	1	0	1	0	0	0	1	0
100.765	46.92	75.16	2.965	8.793	4.769	22.741	0	1	0	0	0	1	0	1	0
100.784	51.85	80.09	2.43	5.905	3.449	11.896	0	1	1	1	0	1	0	1	0
99.888	59.81	59.81	2.466	6.082	3.183	10.129	0	1	1	1	0	1	0	1	0

Tabla S3: Descriptores obtenidos por E- Dragon

Hypertens-80	Hypertens-50	Hypnotic-80	Hypnotic-50	Neoplastic-80	Neoplastic-50	Infective-80	Infective-50	BLTF96	BLTD48	BLTA96	ALOGPS_logP	ALOGPS_logS
0	0	0	0	0	0	1	0	-3.9	-4.13	-4.18	4.32	-4.42
0	0	0	0	0	0	1	0	-4.43	-4.72	-4.81	4.53	-4.77
0	0	0	0	1	0	1	0	-3.9	-4.13	-4.18	4.3	-4.55
1	0	0	0	1	0	1	0	-3.75	-3.96	-4.01	3.96	-4.34
1	0	0	0	1	0	1	0	-3.97	-4.2	-4.27	3.75	-5.3
1	0	0	0	1	0	1	0	-3.97	-4.2	-4.27	3.64	-5.23
1	0	0	0	1	0	1	1	-2.7	-2.79	-2.77	3.16	-4
1	1	1	0	1	1	1	1	-3.26	-3.41	-3.43	2.99	-3.87
1	0	0	0	1	0	1	0	-3.65	-3.85	-3.89	4.06	-4.92
1	1	1	0	1	1	1	1	-2.74	-2.82	-2.81	2.34	-3.68
1	1	0	0	1	1	1	1	-2.7	-2.79	-2.77	2.71	-4.03
1	0	0	0	1	0	1	0	-3.97	-4.2	-4.27	3.76	-5.26
1	0	0	0	1	0	1	0	-3.24	-3.38	-3.4	3.56	-4.08
0	0	0	0	0	0	0	0	-4.43	-4.72	-4.81	4.73	-4.77
1	0	0	0	1	1	1	1	-2.7	-2.79	-2.77	2.77	-4.03
1	0	0	0	1	0	1	1	-3.57	-3.75	-3.79	3.38	-4.21
1	1	1	0	1	1	1	1	-3.26	-3.41	-3.43	3.17	-3.65
1	1	0	0	1	1	1	1	-3.39	-3.56	-3.59	2.87	-4.05
0	0	0	0	0	0	1	0	-4.44	-4.73	-4.82	4.91	-4.33
1	0	0	0	1	0	1	1	-3.57	-3.75	-3.79	3.34	-4.21
1	1	1	0	1	1	1	1	-3.26	-3.41	-3.43	3.01	-3.89
1	0	0	0	1	1	1	0	-3.02	-3.15	-3.15	3.15	-4.49
1	0	0	0	1	0	1	0	-3.09	-3.21	-3.22	3.45	-4.04
1	1	0	0	1	1	1	0	-2.8	-2.89	-2.88	2.72	-4.53
1	1	0	0	1	1	1	0	-2.8	-2.89	-2.88	2.66	-4.51
1	0	0	0	1	0	1	0	-3.97	-4.2	-4.27	3.66	-5.2
1	1	0	0	1	1	1	0	-2.8	-2.89	-2.88	2.83	-4.51
1	1	0	0	1	1	1	1	-2.7	-2.79	-2.77	2.74	-4.05
1	0	0	0	1	0	1	0	-3.75	-3.96	-4.01	3.95	-4.08
1	1	0	0	1	1	1	1	-2.7	-2.79	-2.77	2.56	-4.05
0	0	0	0	0	0	1	0	-3.72	-3.92	-3.97	3.92	-4.22
1	0	0	0	1	0	1	0	-3.79	-4.01	-4.06	3.88	-4.2
1	0	0	0	1	1	1	0	-2.8	-2.89	-2.88	3.08	-4.48
1	1	0	0	1	1	1	0	-2.8	-2.89	-2.88	2.8	-4.51
1	0	0	0	1	0	1	0	-4.29	-4.56	-4.64	3.84	-4.11
0	0	0	0	1	0	1	0	-3.97	-4.2	-4.27	4.3	-4.23
1	0	0	0	1	0	1	0	-3.21	-3.36	-3.38	3.03	-3.8
0	0	0	0	0	0	1	0	-3.91	-4.14	-4.2	4.08	-4.37
1	0	0	0	1	1	1	0	-3.46	-3.63	-3.66	3.08	-3.48
1	1	0	0	1	1	1	0	-3.49	-3.66	-3.7	3.09	-4.52

Resultados de los parámetros estadísticos, mínimos, máximos, mediana de los 30 IDVs en R software

R Console (64-bit)

Archivo Editar Misc Paquetes Ventanas Ayuda

ID.de.referencia	TrkAic50	LBE	MW	
A46 : 1	Min. : 0.070	Min. :-12.16	Min. :332.4	
A47 : 1	1st Qu.: 2.328	1st Qu.: -11.46	1st Qu.:342.4	
A48 : 1	Median : 9.245	Median : -11.05	Median :347.4	
A49 : 1	Mean :28.228	Mean :-10.91	Mean :352.4	
A50 : 1	3rd Qu.:52.830	3rd Qu.: -10.57	3rd Qu.:359.4	
A51 : 1	Max. :99.820	Max. : -7.40	Max. :412.5	
(Other):34				
VOL	Sv	HOMT	AROM	
Min. :288.9	Min. :26.68	Min. :12.05	Min. :0.6930	
1st Qu.:309.7	1st Qu.:28.16	1st Qu.:15.58	1st Qu.:0.7650	
Median :314.6	Median :29.37	Median :15.82	Median :0.7735	
Mean :316.9	Mean :29.32	Mean :16.64	Mean :0.8251	
3rd Qu.:323.5	3rd Qu.:30.30	3rd Qu.:18.46	3rd Qu.:0.9323	
Max. :374.7	Max. :34.36	Max. :19.74	Max. :0.9430	
Ilogp	TPSA	ESOL	ALOGPS	
Min. :1.280	Min. : 46.92	Min. :0.00277	Min. :2.340	
1st Qu.:2.505	1st Qu.: 59.81	1st Qu.:0.00990	1st Qu.:2.960	
Median :2.600	Median : 75.16	Median :0.02010	Median :3.360	
Mean :2.599	Mean : 73.35	Mean :0.03442	Mean :3.452	
3rd Qu.:2.820	3rd Qu.: 88.05	3rd Qu.:0.05595	3rd Qu.:3.928	
Max. :3.190	Max. :103.40	Max. :0.14200	Max. :4.910	
SA	AliSol	AMR	Pol	
Min. :2.610	Min. :0.000688	Min. : 92.28	Min. :32.0	
1st Qu.:2.860	1st Qu.:0.004857	1st Qu.: 99.27	1st Qu.:35.0	
Median :2.955	Median :0.018500	Median :100.01	Median :36.5	
Mean :2.966	Mean :0.039848	Mean :101.62	Mean :36.8	
3rd Qu.:3.065	3rd Qu.:0.040725	3rd Qu.:102.31	3rd Qu.:38.0	
Max. :3.410	Max. :0.220000	Max. :120.90	Max. :43.0	
HBD	IAC	ISIZ	Se	MLOGP2
Min. :0.0	Min. :61.4	Min. :206.1	Min. :39.00	Min. : 2.400
1st Qu.:1.0	1st Qu.:64.9	1st Qu.:219.7	1st Qu.:41.10	1st Qu.: 2.700
Median :1.0	Median :65.8	Median :233.3	Median :43.15	Median : 6.000
Mean :1.1	Mean :67.2	Mean :234.0	Mean :43.07	Mean : 6.263
3rd Qu.:1.0	3rd Qu.:69.2	3rd Qu.:247.1	3rd Qu.:44.80	3rd Qu.: 8.725
Max. :2.0	Max. :77.7	Max. :282.2	Max. :49.90	Max. :12.900
XLOGP3	WLOGP	logPC	MR	
Min. :2.020	Min. :2.860	Min. :1.786	Min. : 93.59	
1st Qu.:2.575	1st Qu.:3.268	1st Qu.:2.423	1st Qu.: 99.21	
Median :3.245	Median :3.900	Median :3.160	Median :101.33	
Mean :3.152	Mean :3.854	Mean :3.141	Mean :102.33	
3rd Qu.:3.670	3rd Qu.:4.402	3rd Qu.:3.839	3rd Qu.:103.49	

Figura S1: Resultados numéricos de la correlación entre IDV vs DV y de IDS vs IDVs obtenida de R *software*.

R Console (64-bit)

Archivo Editar Misc Paquetes Ventanas Ayuda

	TrkAic50	LBE	MW	VOL	Sv
TrkAic50	1.000000000	0.35233640	-0.25348264	-0.15609032	-0.1935181
LBE	0.352336399	1.00000000	0.05911844	-0.11062997	-0.2743865
MW	-0.253482635	0.05911844	1.00000000	0.89844296	0.8295094
VOL	-0.156090315	-0.11062997	0.89844296	1.00000000	0.9609830
Sv	-0.193518062	-0.27438647	0.82950943	0.96098296	1.0000000
HOMT	0.025141486	-0.46436985	0.34302870	0.56711045	0.6813876
AROM	-0.022272219	-0.48665899	-0.15432117	0.08572605	0.2861319
TPSA	-0.098016815	0.50270190	0.28723202	-0.04791295	-0.2174852
ESOL	0.003326543	-0.02265716	-0.49943180	-0.53116337	-0.4599285
SA	-0.065698790	0.46113815	0.65436482	0.49240956	0.3050951
AliSol	0.092660314	-0.18943779	-0.39209020	-0.26650361	-0.1256095
AMR	-0.193993615	-0.05651664	0.93928896	0.95531538	0.9134146
Pol	-0.117741696	-0.29488658	0.73569236	0.85367291	0.9082940
HBD	-0.298038844	0.04327659	0.40121343	0.27879579	0.2830682
IAC	-0.281353196	0.08779058	0.95418988	0.80723603	0.7418170
ISIZ	-0.168045716	-0.30994737	0.74300849	0.91363455	0.9821921
Se	-0.185998046	-0.30717485	0.75606084	0.91108682	0.9799269
XLOGP3	-0.102530370	-0.01981076	0.58450759	0.62614215	0.5592174
logPC	0.122787742	0.16570801	0.53591167	0.57030893	0.5078283
MR	-0.179516126	-0.11802240	0.92071390	0.97148057	0.9466028
FCsp3	0.064221512	0.04706440	-0.16119874	-0.17879743	-0.1528716
logKp	-0.041848692	-0.03909750	0.37226676	0.45012060	0.3939764
	HOMT	AROM	TPSA	ESOL	SA
TrkAic50	0.02514149	-0.02227222	-0.09801681	0.003326543	-0.06569879
LBE	-0.46436985	-0.48665899	0.50270190	-0.022657158	0.46113815
MW	0.34302870	-0.15432117	0.28723202	-0.499431805	0.65436482
VOL	0.56711045	0.08572605	-0.04791295	-0.531163368	0.49240956
Sv	0.68138759	0.28613187	-0.21748517	-0.459928477	0.30509514
HOMT	1.00000000	0.67122101	-0.54151965	-0.190429782	-0.25839450
AROM	0.67122101	1.00000000	-0.81215903	0.206081892	-0.60570198
TPSA	-0.54151965	-0.81215903	1.00000000	0.098618054	0.61325275
ESOL	-0.19042978	0.20608189	0.09861805	1.00000000	-0.37144532
SA	-0.25839450	-0.60570198	0.61325275	-0.371445321	1.00000000
AliSol	0.25541955	0.59353415	-0.29315108	0.814315699	-0.53056264
AMR	0.44672816	-0.10782290	0.12761978	-0.582891763	0.56001696
Pol	0.81228247	0.49640691	-0.35669513	-0.395971722	0.13747485
HBD	0.09127234	-0.16002977	0.39775263	-0.003025015	0.27410378
IAC	0.23143746	-0.13148309	0.32450594	-0.347544072	0.65578323
ISIZ	0.68519491	0.40121122	-0.33340965	-0.382746958	0.22056834
Se	0.68448548	0.41650228	-0.32617113	-0.366965624	0.22556802
XLOGP3	0.23046916	-0.18015265	-0.10233053	-0.911151945	0.44500400
logPC	0.32577584	-0.03867679	-0.15863576	-0.801587152	0.34455952
MR	0.52713049	-0.01845892	0.04459036	-0.573526757	0.50298047
FCsp3	-0.46343984	-0.16052855	0.07758313	0.280732608	0.12609659
logKp	0.16250491	-0.16154055	-0.20070048	-0.893796863	0.31568972

MR	0.52713049	-0.01845892	0.04459036	-0.573526757	0.50298047
FCsp3	-0.46343984	-0.16052855	0.07758313	0.280732608	0.12609659
logKp	0.16250491	-0.16154055	-0.20070048	-0.893796863	0.31568972
	AliSol	AMR	Pol	HBD	IAC
TrkAic50	0.092660314	-0.19399362	-0.117741696	-0.298038844	-0.28135320
LBE	-0.189437792	-0.05651664	-0.294886582	0.043276593	0.08779058
MW	-0.392090200	0.93928896	0.735692363	0.401213428	0.95418988
VOL	-0.266503614	0.95531538	0.853672906	0.278795789	0.80723603
Sv	-0.125609515	0.91341459	0.908294049	0.283068195	0.74181704
HOMT	0.255419554	0.44672816	0.812282470	0.091272338	0.23143746
AROM	0.593534147	-0.10782290	0.496406915	-0.160029769	-0.13148309
TPSA	-0.293151079	0.12761978	-0.356695129	0.397752627	0.32450594
ESOL	0.814315699	-0.58289176	-0.395971722	-0.003025015	-0.34754407
SA	-0.530562639	0.56001696	0.137474852	0.274103782	0.65578323
AliSol	1.000000000	-0.38358581	-0.009623855	-0.137208845	-0.33333979
AMR	-0.383585814	1.00000000	0.747363446	0.381347080	0.82872919
Pol	-0.009623855	0.74736345	1.000000000	0.132843162	0.68413883
HBD	-0.137208845	0.38134708	0.132843162	1.000000000	0.39542808
IAC	-0.333339792	0.82872919	0.684138831	0.395428080	1.00000000
ISIZ	-0.030203667	0.83846402	0.897653570	0.229666879	0.68284388
Se	-0.022183949	0.83160445	0.912622064	0.234127726	0.70854579
XLOGP3	-0.707912366	0.67162066	0.470443481	0.122232156	0.42271675
logPC	-0.585138015	0.58172010	0.526091767	0.040908747	0.41797763
MR	-0.327849537	0.99204836	0.800334092	0.363777726	0.80372814
FCsp3	0.101891197	-0.15648352	-0.318985675	-0.263770394	-0.01850627
logKp	-0.692479214	0.49048659	0.319515270	0.023337921	0.20016773
	ISIZ	Se	XLOGP3	logPC	MR
TrkAic50	-0.16804572	-0.18599805	-0.10253037	0.12278774	-0.17951613
LBE	-0.30994737	-0.30717485	-0.01981076	0.16570801	-0.11802240
MW	0.74300849	0.75606084	0.58450759	0.53591167	0.92071390
VOL	0.91363455	0.91108682	0.62614215	0.57030893	0.97148057
Sv	0.98219206	0.97992691	0.55921740	0.50782834	0.94660280
HOMT	0.68519491	0.68448548	0.23046916	0.32577584	0.52713049
AROM	0.40121122	0.41650228	-0.18015265	-0.03867679	-0.01845892
TPSA	-0.33340965	-0.32617113	-0.10233053	-0.15863576	0.04459036
ESOL	-0.38274696	-0.36696562	-0.91115195	-0.80158715	-0.57352676
SA	0.22056834	0.22556802	0.44500400	0.34455952	0.50298047
AliSol	-0.03020367	-0.02218395	-0.70791237	-0.58513801	-0.32784954
AMR	0.83846402	0.83160445	0.67162066	0.58172010	0.99204836
Pol	0.89765357	0.91262206	0.47044348	0.52609177	0.80033409
HBD	0.22966688	0.23412773	0.12223216	0.04090875	0.36377773
IAC	0.68284388	0.70854579	0.42271675	0.41797763	0.80372814
ISIZ	1.00000000	0.99792726	0.47478395	0.45125338	0.87930948
Se	0.99792726	1.00000000	0.46234620	0.44705118	0.87220368
XLOGP3	0.47478395	0.46234620	1.00000000	0.83673395	0.66947513
logPC	0.45125338	0.44705118	0.83673395	1.00000000	0.57387044

MR	0.52713049	-0.01845892	0.04459036	-0.573526757	0.50298047
FCsp3	-0.46343984	-0.16052855	0.07758313	0.280732608	0.12609659
logKp	0.16250491	-0.16154055	-0.20070048	-0.893796863	0.31568972
	AliSol	AMR	Pol	HBD	IAC
TrkAic50	0.092660314	-0.19399362	-0.117741696	-0.298038844	-0.28135320
LBE	-0.189437792	-0.05651664	-0.294886582	0.043276593	0.08779058
MW	-0.392090200	0.93928896	0.735692363	0.401213428	0.95418988
VOL	-0.266503614	0.95531538	0.853672906	0.278795789	0.80723603
Sv	-0.125609515	0.91341459	0.908294049	0.283068195	0.74181704
HOMT	0.255419554	0.44672816	0.812282470	0.091272338	0.23143746
AROM	0.593534147	-0.10782290	0.496406915	-0.160029769	-0.13148309
TPSA	-0.293151079	0.12761978	-0.356695129	0.397752627	0.32450594
ESOL	0.814315699	-0.58289176	-0.395971722	-0.003025015	-0.34754407
SA	-0.530562639	0.56001696	0.137474852	0.274103782	0.65578323
AliSol	1.000000000	-0.38358581	-0.009623855	-0.137208845	-0.33333979
AMR	-0.383585814	1.000000000	0.747363446	0.381347080	0.82872919
Pol	-0.009623855	0.74736345	1.000000000	0.132843162	0.68413883
HBD	-0.137208845	0.38134708	0.132843162	1.000000000	0.39542808
IAC	-0.333339792	0.82872919	0.684138831	0.395428080	1.000000000
ISIZ	-0.030203667	0.83846402	0.897653570	0.229666879	0.68284388
Se	-0.022183949	0.83160445	0.912622064	0.234127726	0.70854579
XLOGP3	-0.707912366	0.67162066	0.470443481	0.122232156	0.42271675
logPC	-0.585138015	0.58172010	0.526091767	0.040908747	0.41797763
MR	-0.327849537	0.99204836	0.800334092	0.363777726	0.80372814
FCsp3	0.101891197	-0.15648352	-0.318985675	-0.263770394	-0.01850627
logKp	-0.692479214	0.49048659	0.319515270	0.023337921	0.20016773
	ISIZ	Se	XLOGP3	logPC	MR
TrkAic50	-0.16804572	-0.18599805	-0.10253037	0.12278774	-0.17951613
LBE	-0.30994737	-0.30717485	-0.01981076	0.16570801	-0.11802240
MW	0.74300849	0.75606084	0.58450759	0.53591167	0.92071390
VOL	0.91363455	0.91108682	0.62614215	0.57030893	0.97148057
Sv	0.98219206	0.97992691	0.55921740	0.50782834	0.94660280
HOMT	0.68519491	0.68448548	0.23046916	0.32577584	0.52713049
AROM	0.40121122	0.41650228	-0.18015265	-0.03867679	-0.01845892
TPSA	-0.33340965	-0.32617113	-0.10233053	-0.15863576	0.04459036
ESOL	-0.38274696	-0.36696562	-0.91115195	-0.80158715	-0.57352676
SA	0.22056834	0.22556802	0.44500400	0.34455952	0.50298047
AliSol	-0.03020367	-0.02218395	-0.70791237	-0.58513801	-0.32784954
AMR	0.83846402	0.83160445	0.67162066	0.58172010	0.99204836
Pol	0.89765357	0.91262206	0.47044348	0.52609177	0.80033409
HBD	0.22966688	0.23412773	0.12223216	0.04090875	0.36377773
IAC	0.68284388	0.70854579	0.42271675	0.41797763	0.80372814
ISIZ	1.00000000	0.99792726	0.47478395	0.45125338	0.87930948
Se	0.99792726	1.00000000	0.46234620	0.44705118	0.87220368
XLOGP3	0.47478395	0.46234620	1.00000000	0.83673395	0.66947513
logPC	0.45125338	0.44705118	0.83673395	1.00000000	0.57387044

```

TrkAic50 -0.16804572 -0.18599805 -0.10253037 0.12278774 -0.17951613
LBE      -0.30994737 -0.30717485 -0.01981076 0.16570801 -0.11802240
MW       0.74300849 0.75606084 0.58450759 0.53591167 0.92071390
VOL      0.91363455 0.91108682 0.62614215 0.57030893 0.97148057
Sv       0.98219206 0.97992691 0.55921740 0.50782834 0.94660280
HOMT    0.68519491 0.68448548 0.23046916 0.32577584 0.52713049
AROM    0.40121122 0.41650228 -0.18015265 -0.03867679 -0.01845892
TPSA    -0.33340965 -0.32617113 -0.10233053 -0.15863576 0.04459036
ESOL    -0.38274696 -0.36696562 -0.91115195 -0.80158715 -0.57352676
SA       0.22056834 0.22556802 0.44500400 0.34455952 0.50298047
AliSol  -0.03020367 -0.02218395 -0.70791237 -0.58513801 -0.32784954
AMR     0.83846402 0.83160445 0.67162066 0.58172010 0.99204836
Pol     0.89765357 0.91262206 0.47044348 0.52609177 0.80033409
HBD     0.22966688 0.23412773 0.12223216 0.04090875 0.36377773
IAC     0.68284388 0.70854579 0.42271675 0.41797763 0.80372814
ISIZ    1.00000000 0.99792726 0.47478395 0.45125338 0.87930948
Se      0.99792726 1.00000000 0.46234620 0.44705118 0.87220368
XLOGP3  0.47478395 0.46234620 1.00000000 0.83673395 0.66947513
logPC   0.45125338 0.44705118 0.83673395 1.00000000 0.57387044
MR      0.87930948 0.87220368 0.66947513 0.57387044 1.00000000
FCsp3   -0.04482350 -0.05228348 -0.35757335 -0.32198269 -0.19129172
logKp   0.32275354 0.30451479 0.97064452 0.79737507 0.49362708

          FCsp3          logKp
TrkAic50 0.06422151 -0.04184869
LBE      0.04706440 -0.03909750
MW       -0.16119874 0.37226676
VOL      -0.17879743 0.45012060
Sv       -0.15287158 0.39397644
HOMT    -0.46343984 0.16250491
AROM    -0.16052855 -0.16154055
TPSA    0.07758313 -0.20070048
ESOL    0.28073261 -0.89379686
SA       0.12609659 0.31568972
AliSol  0.10189120 -0.69247921
AMR     -0.15648352 0.49048659
Pol     -0.31898568 0.31951527
HBD     -0.26377039 0.02333792
IAC     -0.01850627 0.20016773
ISIZ    -0.04482350 0.32275354
Se      -0.05228348 0.30451479
XLOGP3  -0.35757335 0.97064452
logPC   -0.32198269 0.79737507
MR      -0.19129172 0.49362708
FCsp3   1.00000000 -0.36273932
logKp   -0.36273932 1.00000000

```

Figura S2: Resultados numéricos y gráficos de la correlación entre IDV vs DV y de IDS vs IDVs obtenida de R software.

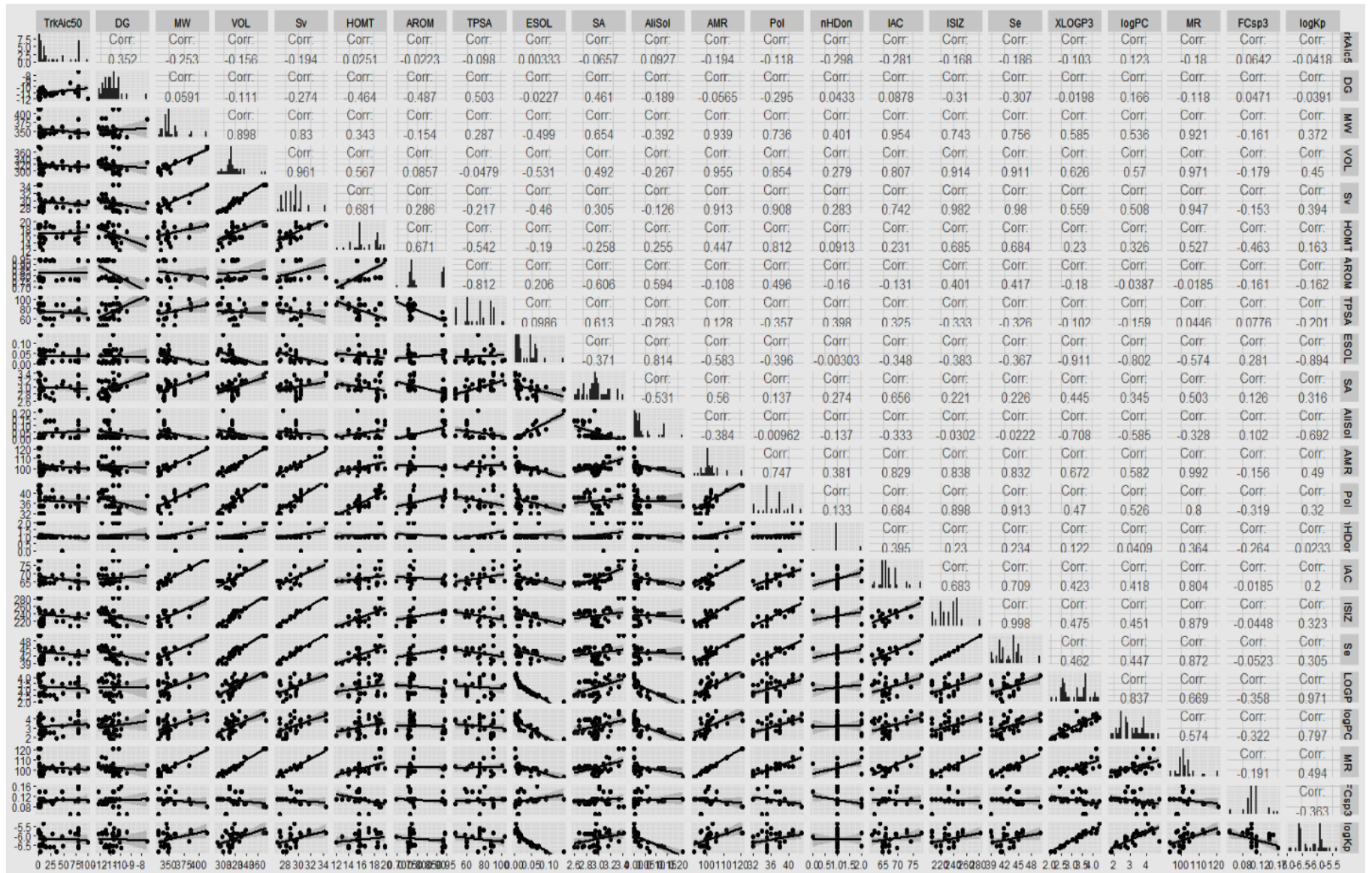


Figura 3. Representación gráfica de la correlación entre IDV vs DV

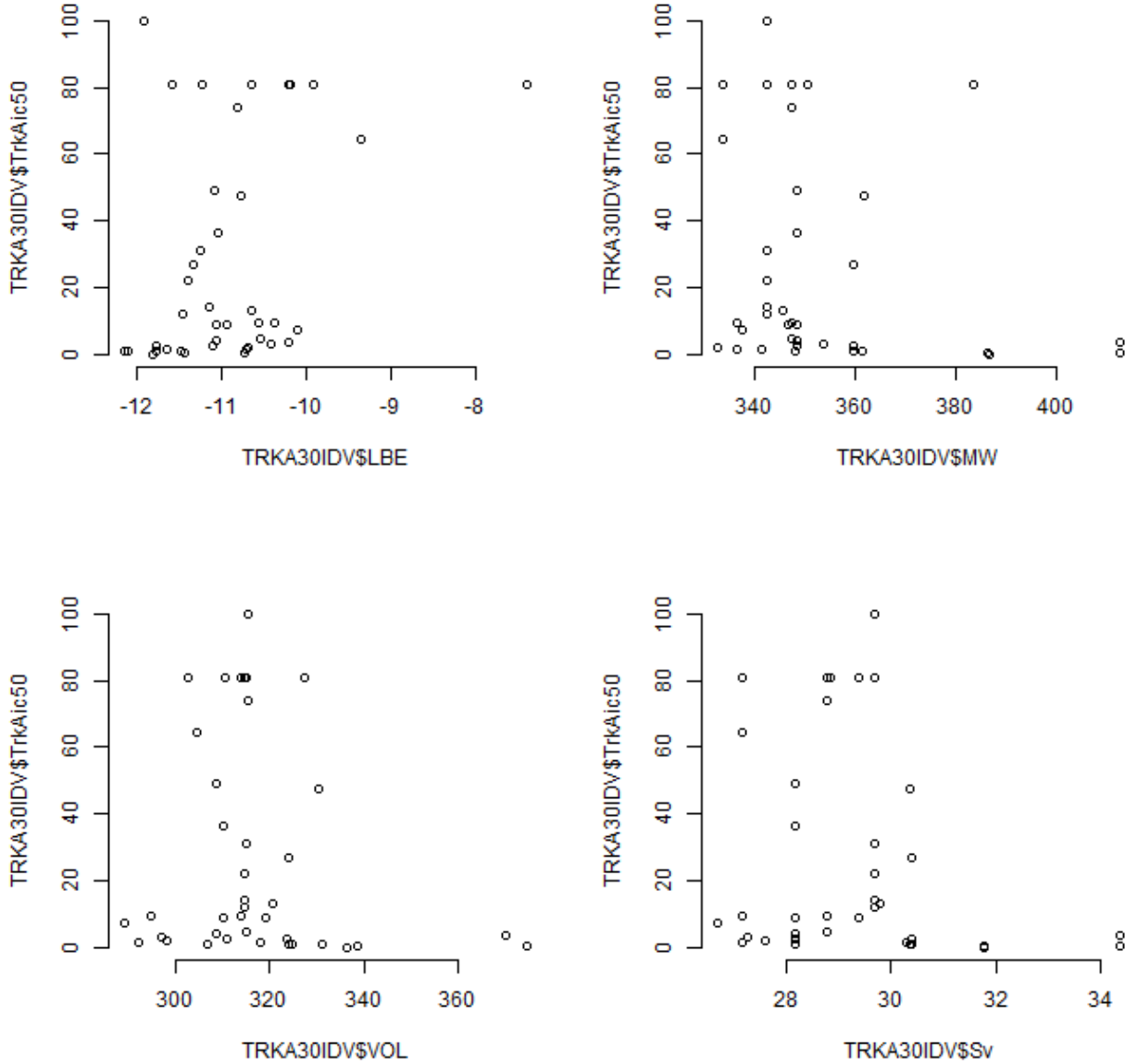


Figura 3. Representación gráfica de la correlación entre IDV vs DV

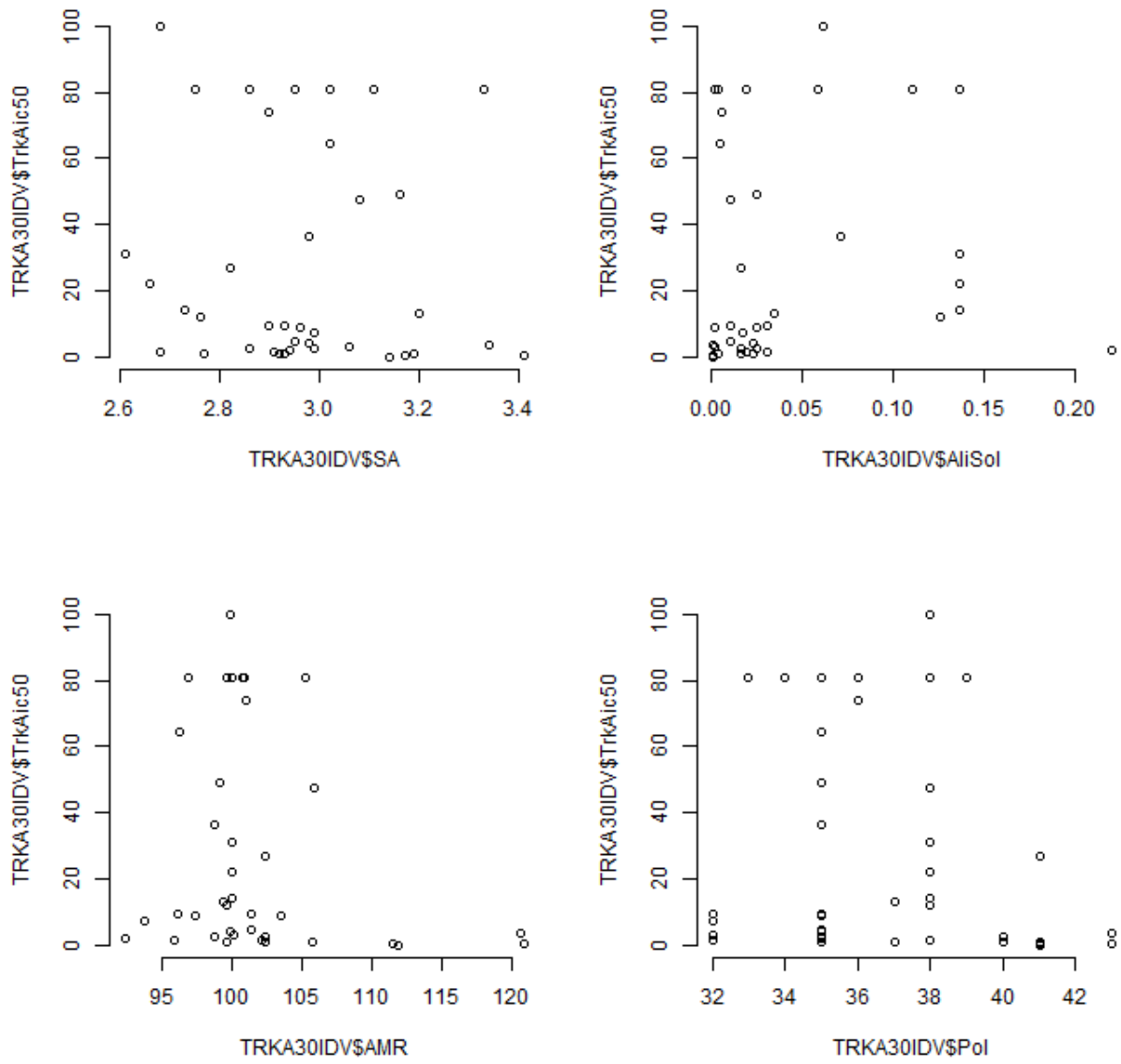


Figura 3. Representación gráfica de la correlación entre IDV vs DV

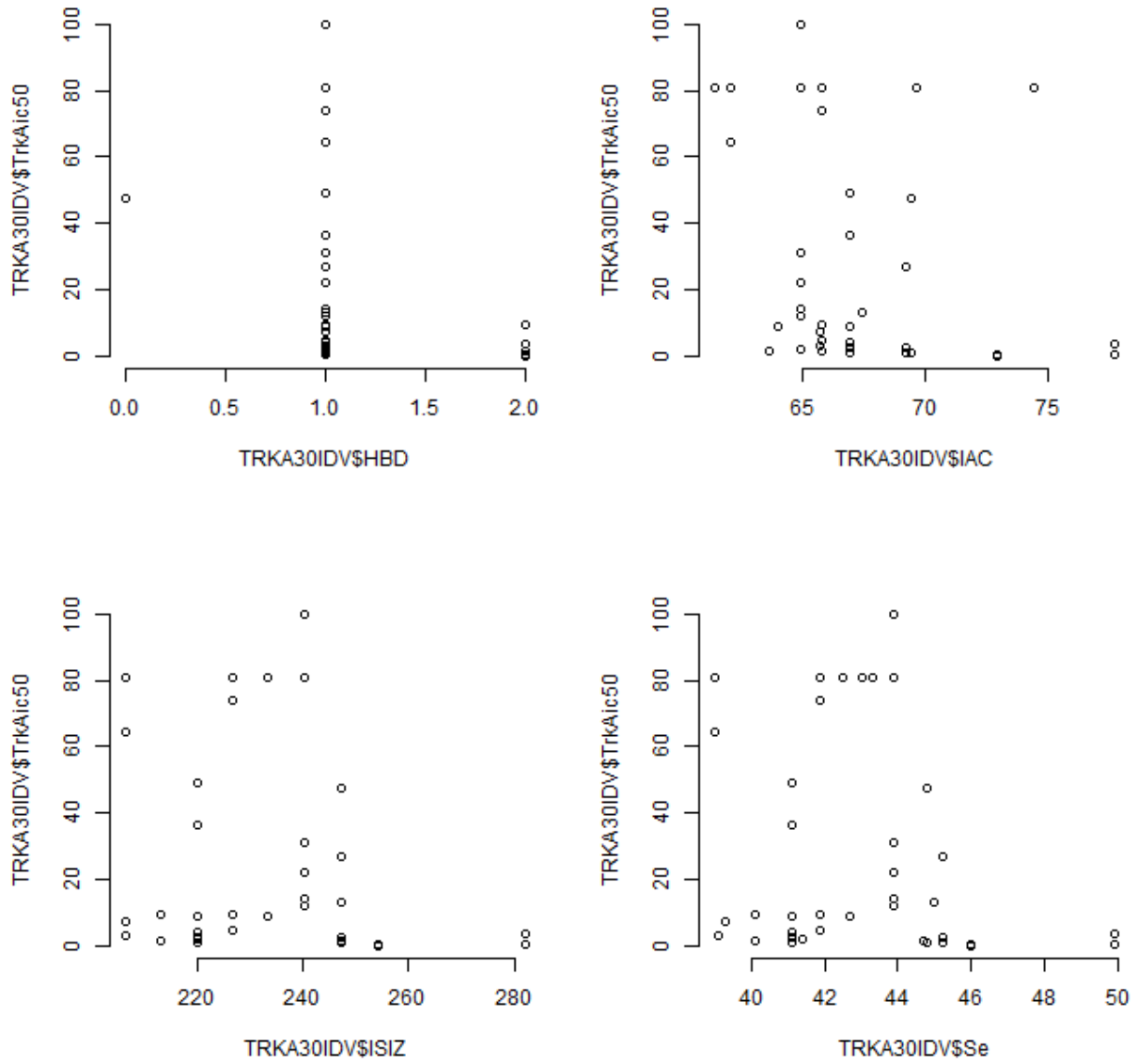


Figura 3. Representación gráfica de la correlación entre IDV vs DV

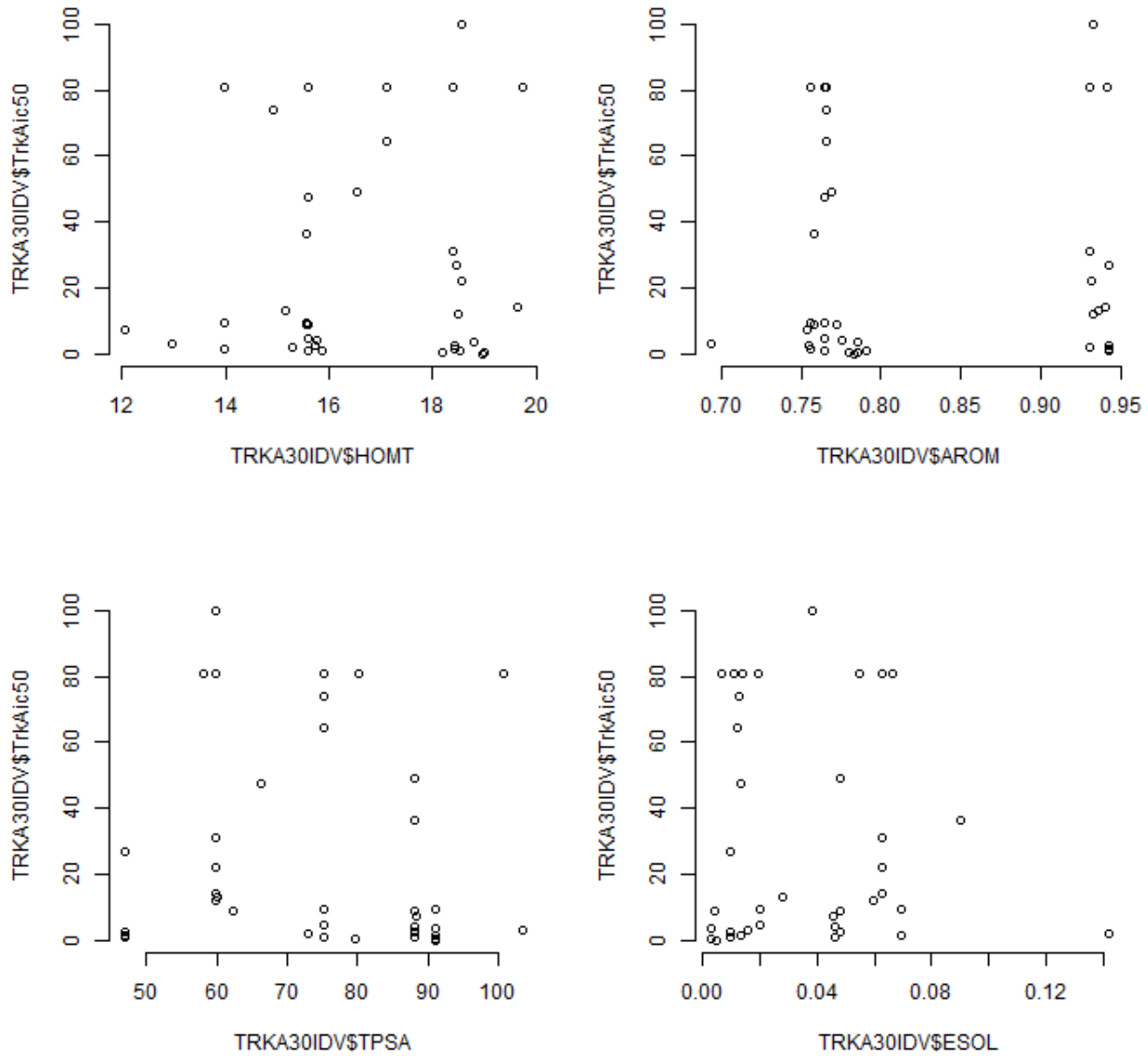


Figura 3. Representación gráfica de la correlación entre IDVs vs DV
 }

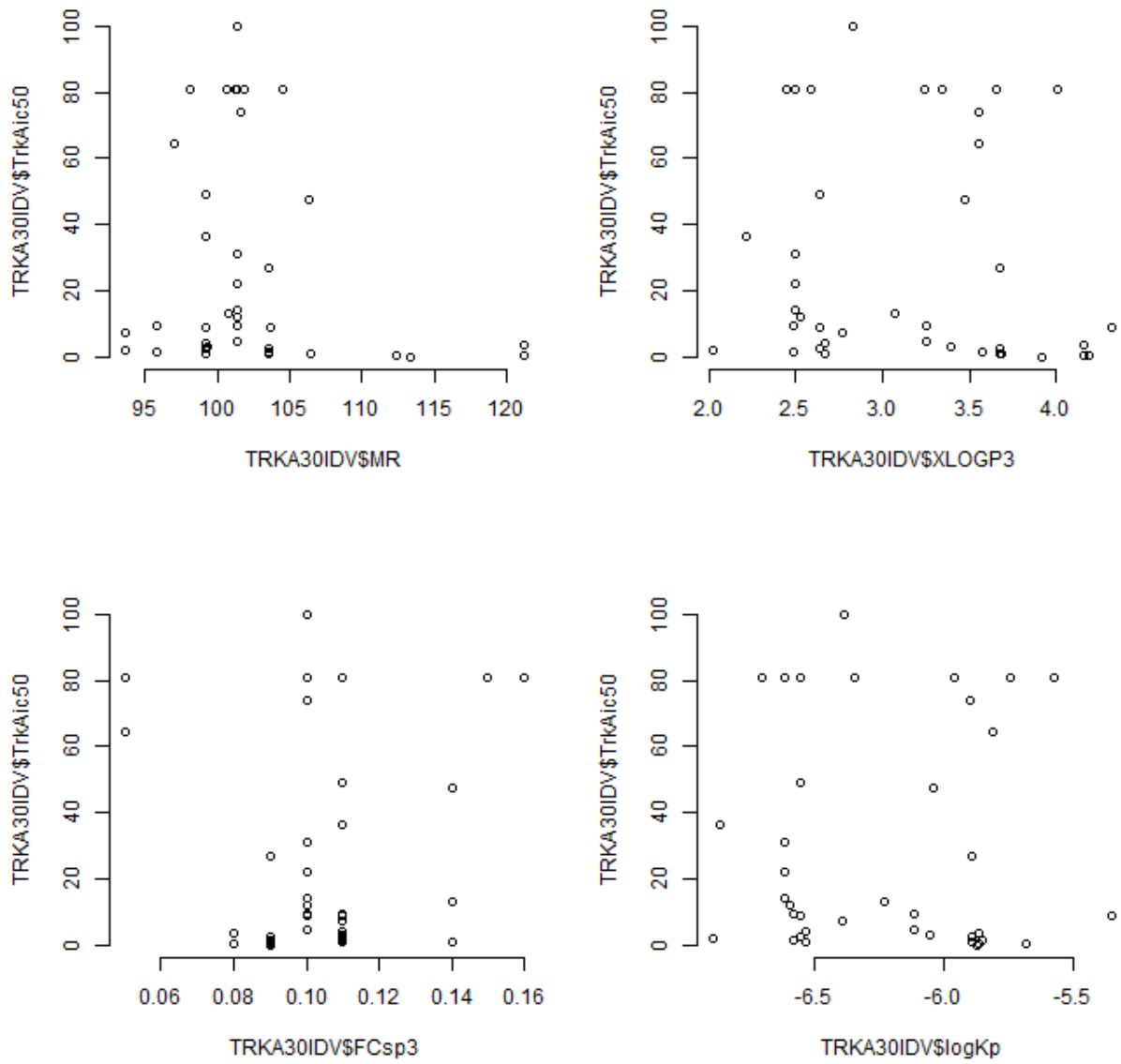
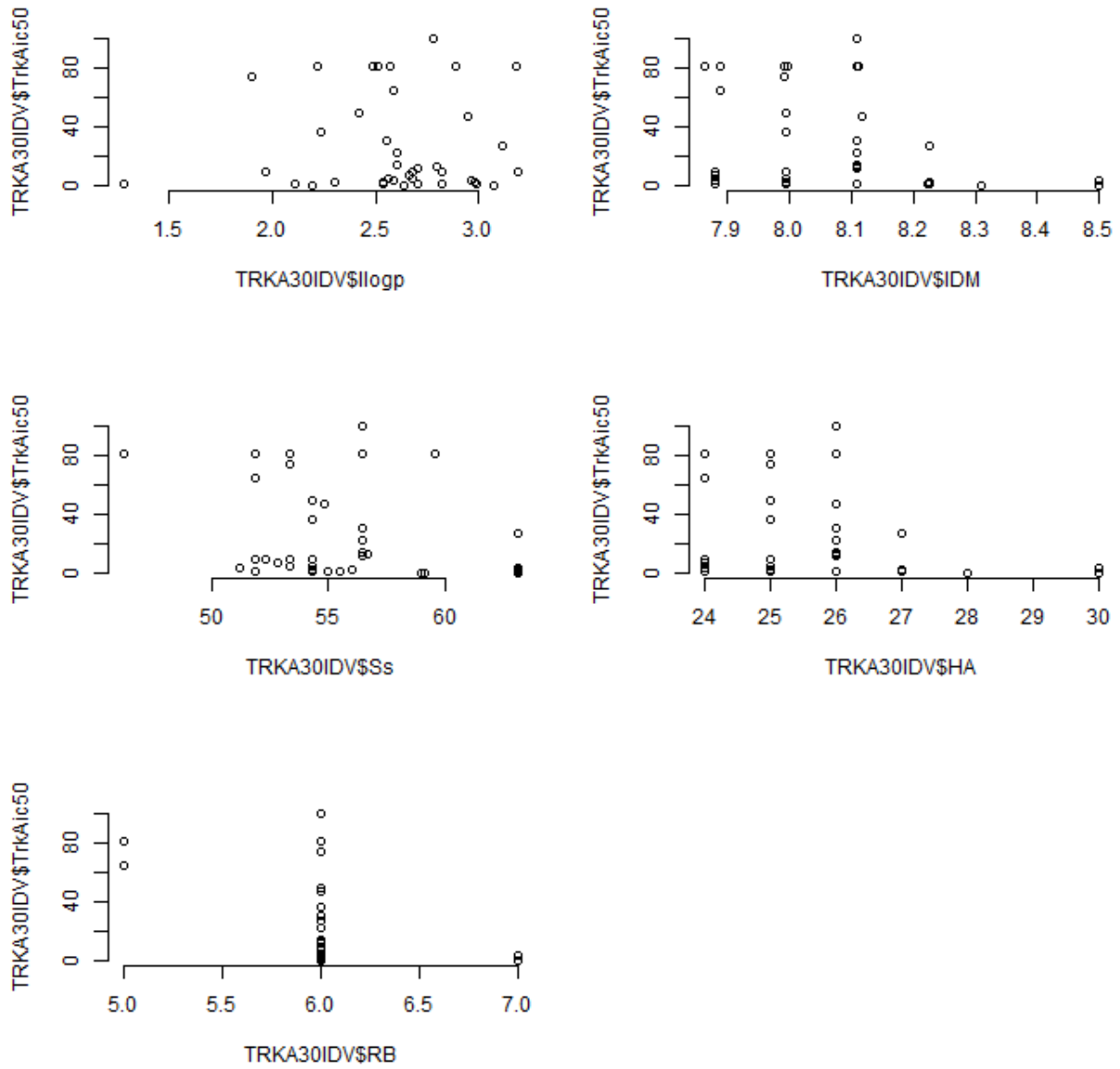


Figura 3. Representación gráfica de la correlación entre IDV vs DV



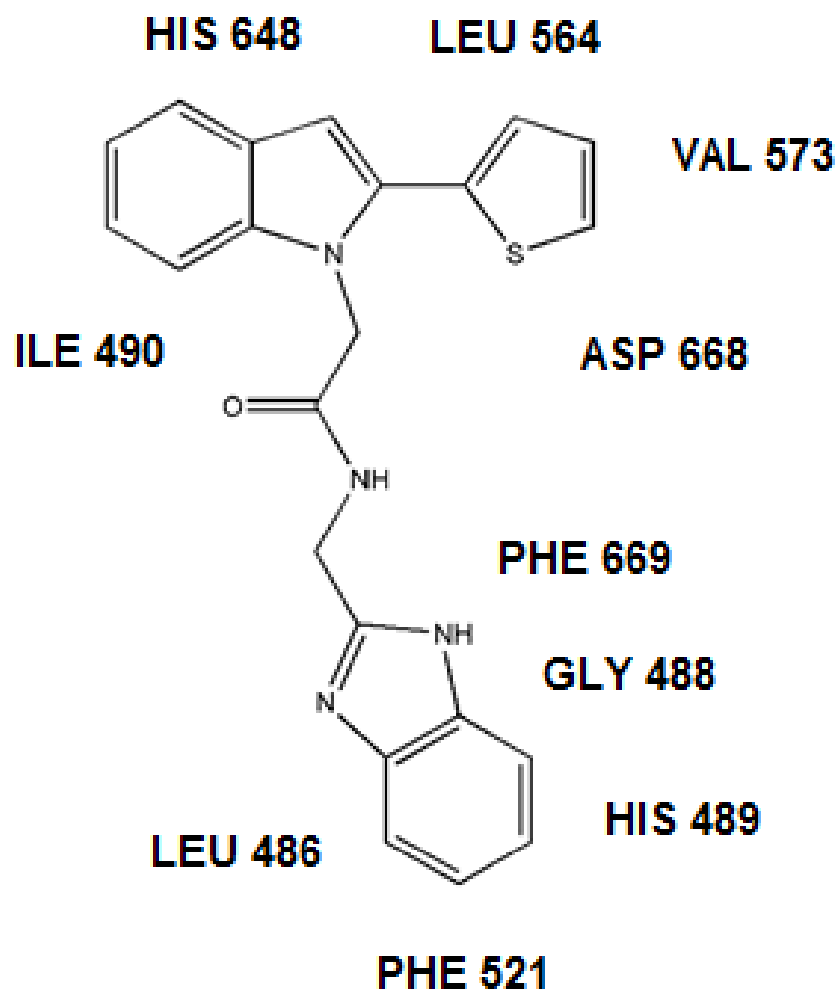


Figura S4: Representación de la molécula D29 y los aminoácidos con los que interacciona

Figura S5: Cálculos para la selección de IDVs por el método de *stepwise* mixto en el programa R

```

R Console (64-bit)
Archivo Editar Misc Paquetes Ventanas Ayuda

F-statistic: 1.551 on 20 and 19 DF, p-value: 0.172

> step(object = ModelR, direction = "both", trace = 1)
Start: AIC=282.26
TrkAic50 ~ LBE + MW + VOL + Sv + HOMT + AROM + TPSA + ESOL +
  AliSol + AMR + Pol + HBD + IAC + ISIZ + Se + XLOGP3 + logPC +
  MR + FCsp3 + logKp

      Df Sum of Sq  RSS   AIC
- AMR   1     5.30 16250 280.28
- LBE   1    58.38 16303 280.41
- Pol   1   264.71 16509 280.91
- logPC 1   415.65 16660 281.27
- VOL   1   477.79 16722 281.42
- ISIZ  1   487.80 16732 281.45
- HBD   1   757.69 17002 282.09
<none>          16244 282.26
- Se    1  1025.03 17269 282.71
- IAC   1  1062.70 17307 282.80
- AliSol 1  1083.01 17327 282.85
- TPSA  1  1195.87 17440 283.11
- FCsp3 1  1208.21 17453 283.13
- ESOL  1  1296.85 17541 283.34
- HOMT  1  1494.81 17739 283.79
- MR    1  1530.40 17775 283.87
- AROM  1  1578.12 17823 283.97
- XLOGP3 1  2102.95 18347 285.13
- MW    1  2124.66 18369 285.18
- logKp 1  2128.40 18373 285.19
- Sv    1  2204.39 18449 285.36

Step: AIC=280.28
TrkAic50 ~ LBE + MW + VOL + Sv + HOMT + AROM + TPSA + ESOL +
  AliSol + Pol + HBD + IAC + ISIZ + Se + XLOGP3 + logPC + MR +
  FCsp3 + logKp

      Df Sum of Sq  RSS   AIC
- LBE   1     56.61 16306 278.42
- Pol   1    309.26 16559 279.03
- logPC 1    416.38 16666 279.29
- VOL   1    505.93 16756 279.50
- ISIZ  1    702.48 16952 279.97
<none>          16250 280.28
- HBD   1    858.69 17108 280.34
- AliSol 1  1111.54 17361 280.93
- FCsp3 1  1387.16 17637 281.56

```

	Df	Sum of Sq	RSS	AIC
- LBE	1	56.61	16306	278.42
- Pol	1	309.26	16559	279.03
- logPC	1	416.38	16666	279.29
- VOL	1	505.93	16756	279.50
- ISIZ	1	702.48	16952	279.97
<none>			16250	280.28
- HBD	1	858.69	17108	280.34
- AliSol	1	1111.54	17361	280.93
- FCsp3	1	1387.16	17637	281.56
- ESOL	1	1439.54	17689	281.67
- IAC	1	1448.74	17698	281.69
- Se	1	1465.58	17715	281.73
- MR	1	1646.71	17896	282.14
- TPSA	1	1677.15	17927	282.21
- HOMT	1	1685.95	17936	282.23
+ AMR	1	5.30	16244	282.26
- AROM	1	1955.18	18205	282.82
- XLOGP3	1	2282.36	18532	283.54
- logKp	1	2310.50	18560	283.60
- MW	1	2365.80	18616	283.71
- Sv	1	2618.27	18868	284.25

Step: AIC=278.42

TrkAic50 ~ MW + VOL + Sv + HOMT + AROM + TPSA + ESOL + AliSol +
 Pol + HBD + IAC + ISIZ + Se + XLOGP3 + logPC + MR + FCsp3 +
 logKp

	Df	Sum of Sq	RSS	AIC
- Pol	1	311.8	16618	277.18
- VOL	1	536.5	16843	277.71
- logPC	1	598.9	16905	277.86
- ISIZ	1	704.4	17011	278.11
<none>			16306	278.42
- HBD	1	1189.4	17496	279.23
- AliSol	1	1431.9	17738	279.78
+ LBE	1	56.6	16250	280.28
- MR	1	1694.8	18001	280.37
+ AMR	1	3.5	16303	280.41
- Se	1	1735.2	18042	280.46
- ESOL	1	1784.2	18091	280.57
- HOMT	1	1837.8	18144	280.69
- TPSA	1	2180.9	18487	281.44
- IAC	1	2333.9	18640	281.77
- MW	1	2404.6	18711	281.92

<

Archivo Editar Misc Paquetes Ventanas Ayuda

	Df	Sum of Sq	RSS	AIC
- VOL	1	574.0	17192	276.53
<none>			16618	277.18
- logPC	1	1010.8	17629	277.54
+ Pol	1	311.8	16306	278.42
- ISIZ	1	1463.9	18082	278.55
+ LBE	1	59.1	16559	279.03
+ AMR	1	44.3	16574	279.07
- AliSol	1	2048.5	18667	279.82
- HBD	1	2409.0	19027	280.59
- TPSA	1	2429.8	19048	280.63
- IAC	1	2434.6	19053	280.64
- MR	1	2684.5	19303	281.17
- Se	1	2857.8	19476	281.52
- ESOL	1	2877.2	19495	281.56
- XLOGP3	1	2904.8	19523	281.62
- logKp	1	2960.6	19579	281.73
- MW	1	3009.7	19628	281.83
- AROM	1	3237.4	19856	282.29
- FCsp3	1	4011.4	20630	283.82
- HOMT	1	4069.2	20687	283.94
- Sv	1	7308.1	23926	289.75

Step: AIC=276.53

TrkAic50 ~ MW + Sv + HOMT + AROM + TPSA + ESOL + AliSol + HBD +
IAC + ISIZ + Se + XLOGP3 + logPC + MR + FCsp3 + logKp

	Df	Sum of Sq	RSS	AIC
<none>			17192	276.53
- logPC	1	932.9	18125	276.65
- ISIZ	1	1015.6	18208	276.83
+ VOL	1	574.0	16618	277.18
+ Pol	1	349.3	16843	277.71
- AliSol	1	1651.9	18844	278.20
+ AMR	1	107.3	17085	278.28
+ LBE	1	91.7	17100	278.32
- IAC	1	1861.7	19054	278.65
- HBD	1	1895.4	19088	278.72
- TPSA	1	1903.5	19096	278.73
- Se	1	2287.7	19480	279.53
- ESOL	1	2334.2	19526	279.62
- XLOGP3	1	2350.0	19542	279.66
- logKp	1	2409.2	19601	279.78
- MW	1	2448.3	19640	279.86
- MR	1	2784.0	19976	280.54
- AROM	1	2991.3	20183	280.95

```

- logKp 1 2960.6 19579 281.73
- MW 1 3009.7 19628 281.83
- AROM 1 3237.4 19856 282.29
- FCsp3 1 4011.4 20630 283.82
- HOMT 1 4069.2 20687 283.94
- Sv 1 7308.1 23926 289.75

```

Step: AIC=276.53

```

TrkAic50 ~ MW + Sv + HOMT + AROM + TPFA + ESOL + AliSol + HBD +
IAC + ISIZ + Se + XLOGP3 + logPC + MR + FCsp3 + logKp

```

	Df	Sum of Sq	RSS	AIC
<none>			17192	276.53
- logPC	1	932.9	18125	276.65
- ISIZ	1	1015.6	18208	276.83
+ VOL	1	574.0	16618	277.18
+ Pol	1	349.3	16843	277.71
- AliSol	1	1651.9	18844	278.20
+ AMR	1	107.3	17085	278.28
+ LBE	1	91.7	17100	278.32
- IAC	1	1861.7	19054	278.65
- HBD	1	1895.4	19088	278.72
- TPFA	1	1903.5	19096	278.73
- Se	1	2287.7	19480	279.53
- ESOL	1	2334.2	19526	279.62
- XLOGP3	1	2350.0	19542	279.66
- logKp	1	2409.2	19601	279.78
- MW	1	2448.3	19640	279.86
- MR	1	2784.0	19976	280.54
- AROM	1	2991.3	20183	280.95
- HOMT	1	3526.7	20719	282.00
- FCsp3	1	4648.2	21840	284.11
- Sv	1	6917.5	24110	288.06

Call:

```

lm(formula = TrkAic50 ~ MW + Sv + HOMT + AROM + TPFA + ESOL +
AliSol + HBD + IAC + ISIZ + Se + XLOGP3 + logPC + MR + FCsp3 +
logKp, data = TRKA30IDV)

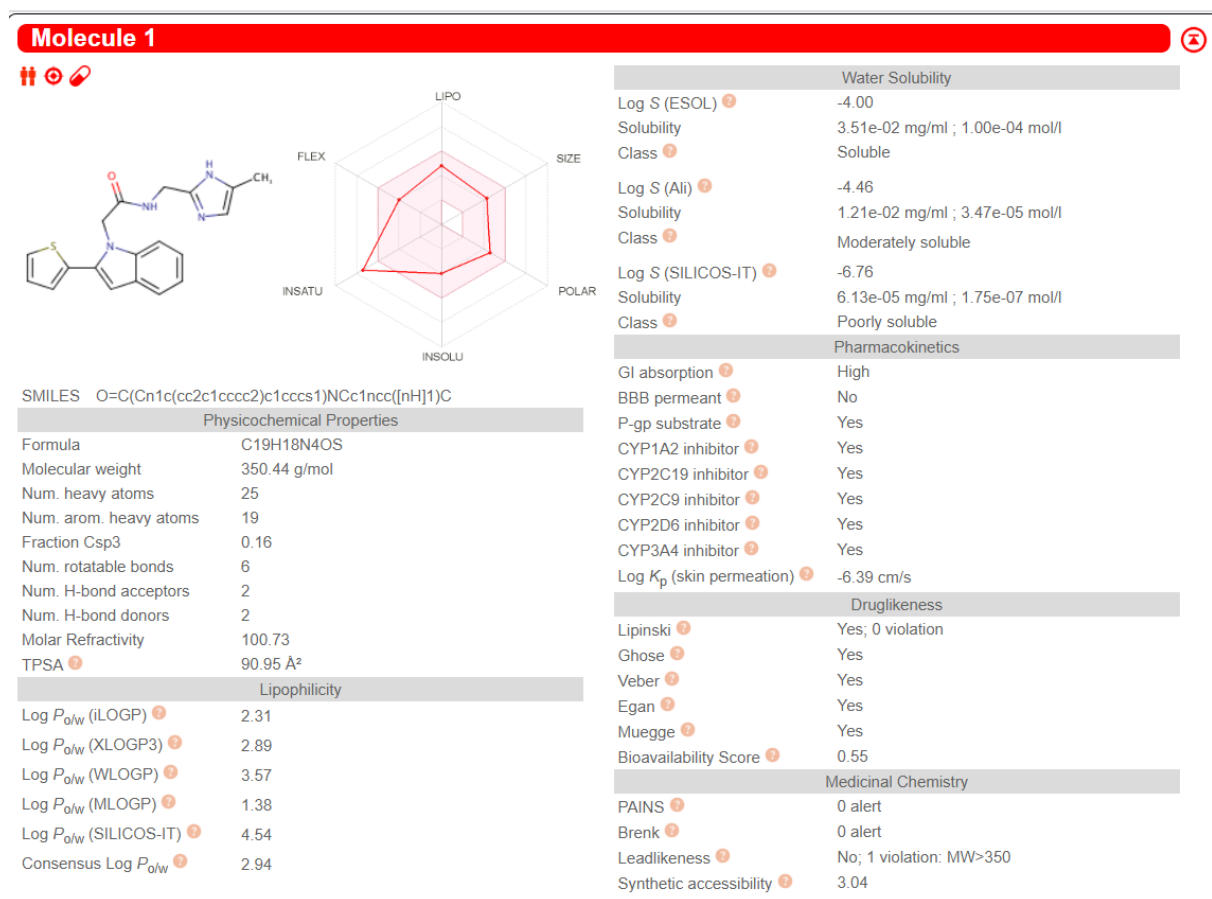
```

Coefficients:

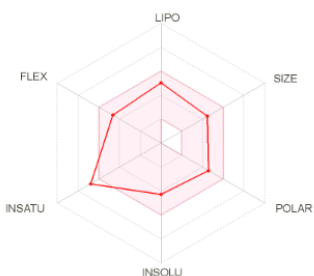
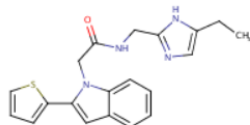
(Intercept)	MW	Sv	HOMT	AROM	TPFA	ESOL	AliSol	HBD	IAC	ISIZ	Se	XLOGP3	logPC
-41429.248	-40.891	-416.237	19.781	-1446.841	3.265	-1830.070	678.891	-38.490	-39.131	-41.629	553.886	4464.630	17.406
	MR	FCsp3	logKp										
	35.930	-2164.229	-6392.984										

> |

Figura S6: Resultados de Swiss ADME de las nuevas moléculas.



Molecule 2



SMILES CCc1cnc([nH]1)CNC(=O)Cn1c(cc2c1cccc2)c1cccs1

Physicochemical Properties

Formula	C ₂₀ H ₂₀ N ₄ O ₂ S
Molecular weight	364.46 g/mol
Num. heavy atoms	26
Num. arom. heavy atoms	19
Fraction Csp ³	0.20
Num. rotatable bonds	7
Num. H-bond acceptors	2
Num. H-bond donors	2
Molar Refractivity	105.53
TPSA	90.95 Å ²

Lipophilicity

Log <i>P</i> _{ow} (iLOGP)	1.67
Log <i>P</i> _{ow} (XLOGP3)	3.33
Log <i>P</i> _{ow} (WLOGP)	3.82
Log <i>P</i> _{ow} (MLOGP)	1.61
Log <i>P</i> _{ow} (SILICOS-IT)	4.93
Consensus Log <i>P</i> _{ow}	3.07

Water Solubility

Log S (ESOL)	-4.28
Solubility	1.93e-02 mg/ml ; 5.29e-05 mol/l
Class	Moderately soluble
Log S (Ali)	-4.92
Solubility	4.42e-03 mg/ml ; 1.21e-05 mol/l
Class	Moderately soluble
Log S (SILICOS-IT)	-7.15
Solubility	2.57e-05 mg/ml ; 7.06e-08 mol/l
Class	Poorly soluble

Pharmacokinetics

GI absorption	High
BBB permeant	No
P-gp substrate	Yes
CYP1A2 inhibitor	Yes
CYP2C19 inhibitor	Yes
CYP2C9 inhibitor	Yes
CYP2D6 inhibitor	Yes
CYP3A4 inhibitor	Yes
Log <i>K</i> _p (skin permeation)	-6.16 cm/s

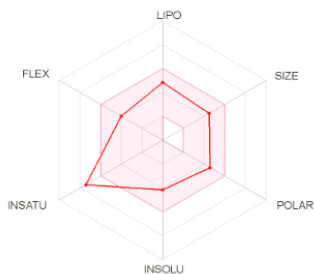
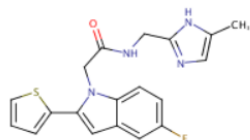
Druglikeness

Lipinski	Yes; 0 violation
Ghose	Yes
Veber	Yes
Egan	Yes
Muegge	Yes
Bioavailability Score	0.55

Medicinal Chemistry

PAINS	0 alert
Brenk	0 alert
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	3.19

Molecule 3



SMILES O=C(Cn1c(cc2c1ccc(c2)F)c1cccs1)NCc1ncc([nH]1)C

Physicochemical Properties

Formula	C ₁₉ H ₁₇ FN ₄ O ₂ S
Molecular weight	368.43 g/mol
Num. heavy atoms	26
Num. arom. heavy atoms	19
Fraction Csp ³	0.16
Num. rotatable bonds	6
Num. H-bond acceptors	3
Num. H-bond donors	2
Molar Refractivity	100.69
TPSA	90.95 Å ²

Lipophilicity

Log <i>P</i> _{0/w} (iLOGP)	2.61
Log <i>P</i> _{0/w} (XLOGP3)	2.99
Log <i>P</i> _{0/w} (WLOGP)	4.13
Log <i>P</i> _{0/w} (MLOGP)	1.76
Log <i>P</i> _{0/w} (SILICOS-IT)	4.96
Consensus Log <i>P</i> _{0/w}	3.29

Water Solubility

Log S (ESOL)	-4.15
Solubility	2.59e-02 mg/ml ; 7.04e-05 mol/l
Class	Moderately soluble
Log S (Ali)	-4.56
Solubility	1.01e-02 mg/ml ; 2.73e-05 mol/l
Class	Moderately soluble
Log S (SILICOS-IT)	-7.02
Solubility	3.49e-05 mg/ml ; 9.48e-08 mol/l
Class	Poorly soluble

Pharmacokinetics

GI absorption	High
BBB permeant	No
P-gp substrate	Yes
CYP1A2 inhibitor	Yes
CYP2C19 inhibitor	Yes
CYP2C9 inhibitor	Yes
CYP2D6 inhibitor	Yes
CYP3A4 inhibitor	Yes
Log <i>K</i> _p (skin permeation)	-6.42 cm/s

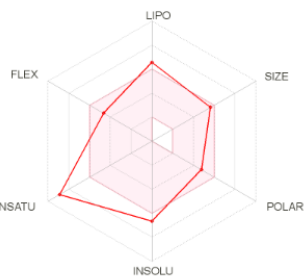
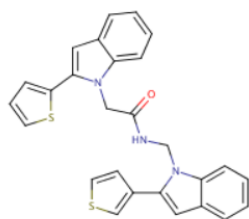
Druglikeness

Lipinski	Yes; 0 violation
Ghose	Yes
Veber	Yes
Egan	Yes
Muegge	Yes
Bioavailability Score	0.55

Medicinal Chemistry

PAINS	0 alert
Brenk	0 alert
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	3.08

Molecule 4



SMILES O=C(Cn1c(cc2c1cccc2)c1cccs1)NCn1c(cc2c1cccc2)c1ccsc1

Physicochemical Properties

Formula	C27H21N3OS2
Molecular weight	467.61 g/mol
Num. heavy atoms	33
Num. arom. heavy atoms	28
Fraction Csp3	0.07
Num. rotatable bonds	7
Num. H-bond acceptors	1
Num. H-bond donors	1
Molar Refractivity	138.72
TPSA	95.44 Å²

Lipophilicity

Log P_{ow} (iLOGP)	3.34
Log P_{ow} (XLOGP3)	5.97
Log P_{ow} (WLOGP)	6.68
Log P_{ow} (MLOGP)	3.84
Log P_{ow} (SILICOS-IT)	6.74
Consensus Log P_{ow}	5.31

Water Solubility

Log S (ESOL)	-6.67
Solubility	1.01e-04 mg/ml ; 2.16e-07 mol/l
Class	Poorly soluble
Log S (Alii)	-7.75
Solubility	8.31e-06 mg/ml ; 1.78e-08 mol/l
Class	Poorly soluble
Log S (SILICOS-IT)	-9.28
Solubility	2.48e-07 mg/ml ; 5.31e-10 mol/l
Class	Poorly soluble

Pharmacokinetics

GI absorption	Low
BBB permeant	No
P-gp substrate	Yes
CYP1A2 inhibitor	Yes
CYP2C19 inhibitor	Yes
CYP2C9 inhibitor	Yes
CYP2D6 inhibitor	No
CYP3A4 inhibitor	Yes
Log K_p (skin permeation)	-4.91 cm/s

Druglikeness

Lipinski	Yes; 0 violation
Ghose	No; 2 violations: WLOGP>5.6, MR>130
Weber	Yes
Egan	No; 1 violation: WLOGP>5.88
Muegge	No; 1 violation: XLOGP3>5
Bioavailability Score	0.55

Medicinal Chemistry

PAINS	0 alert
Brenk	0 alert
Leadlikeness	No; 2 violations: MW>350, XLOGP3>3.5
Synthetic accessibility	3.69