



# **BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA**

---

Facultad de Ciencias Químicas

Departamento de Farmacia

Título de tesis:

## **ESTUDIO QSAR DE LOS DERIVADOS DE 3-FORMIL CUMARINA DIRIGIDOS A TIMIDINA FOSFORILASA HUMANA EN SOBRE EXPRESIÓN**

Octubre 2020

Tesis presentada para obtener el grado de:

LICENCIATURA EN QUÍMICO FARMACOBIOLOGO.

PRESENTA:

Juan Carlos García Hernández

DIRECTOR DE TESIS:

D.C. Thomas Scior

CODIRECTOR DE TESIS:

D.C José Manuel Pérez Aguilar

ASESORES DE TESIS:

D.C. María de Guadalupe Muñoz Arenas

D.C. Teresa Soledad Cid Pérez

D.C. Julián Alejandro Yunes Rojas

## Abreviaturas

**TP:** Timidina Fosforilasa

**SAR:** Relación estructura actividad

**QSAR:** Relación cuantitativa estructura-actividad

**RLM:** Regresión lineal múltiple

**RLS:** Regresión lineal simple

**VD:** Variable dependiente

**VID:** Variable independiente

$\alpha$ : Intercepto.

$\beta$ : Pendiente.

**IC<sub>50</sub>:** Constante inhibitoria 50

**K<sub>i</sub>:** Constante de inhibición.

**EC<sub>50</sub>:** Concentración efectiva 50

**OMS:** Organización Mundial de la Salud

**PDB:** Protein Data Bank

**BCI<sub>4</sub>:** Bond Information Content index

**TPSA(Tot):** Topological polar surface area

**IC<sub>5</sub>:** Information Content index

**Mor03i:** signal 03 / weighted by ionization potential

**B10[O-CI]:** Presence/absence of O - Cl at topological distance 10

**Pi:** Fosfato inorgánico

## Agradecimientos

En el tiempo de mi trayectoria como estudiante doy gracias al maestro Juvenal Cruz Vega quien con el conocimiento interdisciplinario de las lenguas clásicas me introdujo a la cultura clásica, de la cual nace mi amor por el conocimiento ya que la escuela de humanidades clásicas es una fuente de sabiduría.

Durante mi formación universitaria tuve varios profesores a quienes les agradezco cada uno de mis aprendizajes. En especial al Dr. Thomas Scior quien con sus clases abrió la curiosidad en mí de aprender más sobre los métodos *in silico*. Siendo esta la manera en la cual termine realizando mi tesis en el laboratorio de simulaciones computacionales. También le debo mi agradecimiento a mis compañeras Laura Flores y Kenia Salvador quienes me han guiado en la realización de este trabajo, así como a Diana Brenda Ruíz por su apoyo y amistad.

A la D.C Guadalupe Muñoz Arenas, D.C Teresa Soledad Cid Pérez, D.C Julián Alejandro Yunes Rojas, gracias por sus comentarios y correcciones para mejorar este trabajo, así como por su disposición para cada una de las presentaciones online. Por último, a mi codirector de tesis D.C José Manuel Pérez Aguilar a quien agradezco sus comentarios y observaciones de mi trabajo para mejorarlo.

## **Dedicatoria**

Este trabajo está dedicado principalmente a mis padres que han puesto siempre su confianza en mí y porque a ellos les debo mis primeros aprendizajes. También a mi hermana Nalleli quien de una u otra manera me ha dado grandes enseñanzas.

Por último, a mis amigos Luis, Heber, Mario y Aldo quienes me han apoyado en los buenos y malos momentos de mi vida.

# Índice

<b>Abreviaturas</b>	<b>2</b>
<b>1. Introducción</b>	<b>5</b>
<b>2. Antecedentes</b>	<b>24</b>
<b>3. Hipotesis</b>	<b>31</b>
<b>4. Justificación</b>	<b>32</b>
<b>5. Objetivos</b>	<b>33</b>
<b>6. Diagrama de trabajo</b>	<b>34</b>
<b>7. Metodología</b>	<b>35</b>
<b>8. Resultados</b>	<b>36</b>
<b>9. Conclusión</b>	<b>59</b>
<b>10. Bibliografía</b>	<b>60</b>
<b>11. Anexos</b>	<b>63</b>

# 1. Introducción

## 1.2 El Cáncer

En la enfermedad del Cáncer se implica el desarrollo de sucesivas alteraciones genéticas y epigenéticas que permiten a las células escapar de los controles homeostáticos que en condiciones normales suprimen la proliferación inapropiada, así como también inhiben la supervivencia de las células que se multiplican de manera aberrante fuera de sus nichos celulares. Los cánceres en la mayoría de los casos surgen de las células epiteliales y se manifiestan como carcinomas en órganos como el pulmón, la piel, el seno, el hígado y el páncreas. Por el contrario, los sarcomas provienen de los tejidos llamados mesenquimales, que se producen en los fibroblastos, miocitos, adipocitos y osteoblastos. Por otro lado, también los tumores no epiteliales pueden desarrollarse en células del sistema nervioso como (por ejemplo; gliomas, neuroblastomas y meduloblastomas) y tejidos hematopoyéticos (leucemia y linfoma).<sup>1</sup>

En el mundo, principalmente en los países económicamente desarrollados el cáncer es la principal causa de muerte y la segunda causa de muerte en los países en desarrollo.<sup>2</sup> Se esperan más de 21 millones de nuevos casos de cáncer y 13 millones de muertes para el año 2030, esto según el informe de la Organización Mundial de la Salud (OMS).<sup>3</sup> Entre otros estudios se estimó que una de cada cinco personas antes de los 75 años sufrirá cáncer durante su vida.<sup>4</sup>

En México fallecen debido a esta enfermedad 14 de cada 100 mexicano siendo el cáncer la tercera causa de muerte. Por otro lado, la expectativa de vida de quienes la padecen es de alrededor de 63 años. De acuerdo con el director del INCAN en México aún hay retos por cumplir. Entre ellos, implementar el Plan Nacional de Prevención y Control del Cáncer, así como mejorar los programas de prevención, diagnóstico oportuno, tratamiento e incrementar el acceso a terapias innovadoras.<sup>5</sup> Para ello hoy en día una de las terapias de gran interés por su importancia en la progresión del cáncer es evitar la angiogénesis por medio de la inhibición de la

enzima Timidina Fosforilasa (TP), ya que esta enzima es la responsable de la formación de nuevos vasos sanguíneos que son esencial para el crecimiento tumoral. A pesar del éxito clínico de los agentes anti angiogénicos como el bevacizumab y otros inhibidores, muchos cánceres desarrollan resistencia a las terapias anti angiogénicas disponibles actualmente.<sup>6</sup> En consecuencia, el descubrimiento de un nuevo agente anticancerígeno con actividad prometedora y un alto índice terapéutico el cual indica que la administración del fármaco no causara efectos de toxicidad al paciente es por ello la necesidad urgente de una terapia segura.<sup>7</sup>

### **1.3 Timidina Fosforilasa**

La timidina fosforilasa (TP) es una enzima involucrada en el metabolismo de nucleósidos. En 1953 se describió por primera vez cuando se descubrió que Timidina Fosforilasa cataliza la conversión reversible de timidina a timina y 2-desoxi- $\alpha$ -D-ribosa-1-fosfato (dRib-1-P). De manera adicional, TP también posee actividad desoxirribosil transferasa acción por la cual el grupo desoxirribosil se transfiere a otra base de pirimidina, lo que resulta en la formación de un nuevo nucleósido de pirimidina, en la figura uno se ilustra la reacción catalizada por timidina fosforilaza.<sup>8</sup>

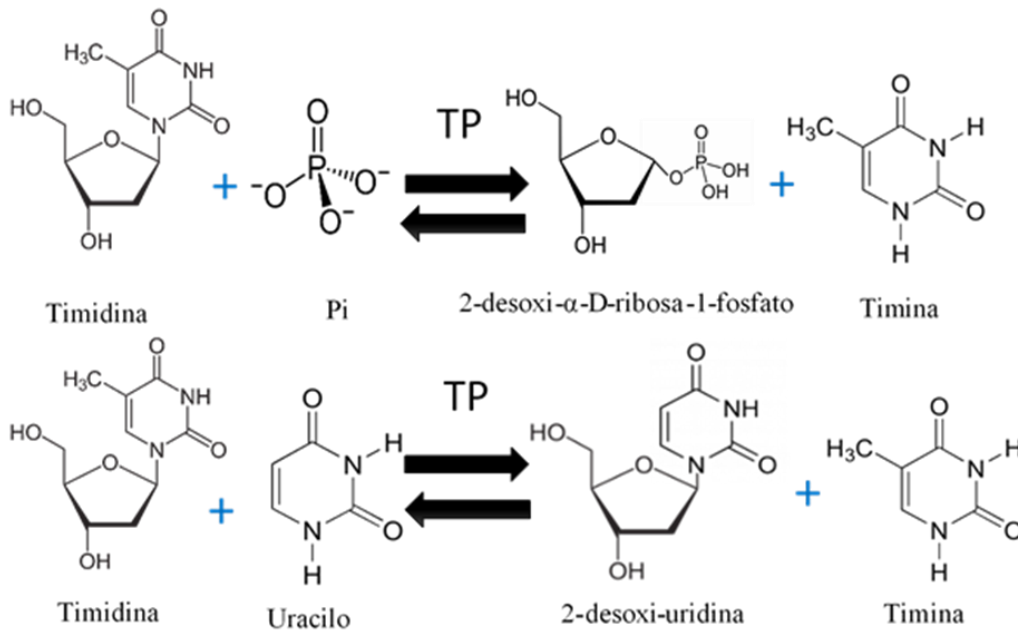


Figura 1: Reacción catalizada por timidina fosforilasa

Se pueden encontrar altos niveles de TP en varias células incluyendo; macrófagos, células del estroma, células gliales y algunos epitelios. Se sabe que TP se expresa tanto en el núcleo donde modula el conjunto de nucleósidos de pirimidina para la síntesis de ADN como en el citoplasma, donde también muestra sus funciones enzimáticas.<sup>9</sup>

### 1.3.1 Rol de la timidina fosforilasa en la angiogénesis tumoral

En los tumores sólidos, las alteraciones promueven típicamente la formación de un conjunto de células proliferativas relativamente benignas (hiperplasias) a una masa de células con morfología anormal, apariencia citológica y organización celular. Después de que un tumor se expande, el núcleo del tumor pierde acceso al oxígeno y a los nutrientes, lo que a menudo conduce al crecimiento de nuevos vasos sanguíneos (angiogénesis), que restauran el acceso a los nutrientes y al

oxígeno. Posteriormente, las células tumorales pueden desarrollar la capacidad de invadir el tejido más allá de sus límites normales, ingresar a la circulación y sembrar nuevos tumores en otros lugares (metástasis), la característica definitoria de la malignidad.<sup>1</sup>

La Timidina Fosforilasa a menudo se sobre expresa en los tumores y se informa que una alta expresión de TP en estos sitios tumorales se ha relacionado con una alta densidad de microvasos, lo cual causa la inducción de metástasis siendo este un mal pronóstico para el paciente. Timidina Fosforilasa juega un papel clave en la angiogénesis pues la 2-desoxi-D-ribosa (D-dRib) es un agente quimio atrayente endotelial que actúa promoviendo tanto la formación de nuevos vasos sanguíneos, como la quimiotaxis de las células endoteliales. La angiogénesis es la formación de nuevos vasos sanguíneos a partir de la vasculatura ya existente, dicho proceso es esencial para el crecimiento tumoral y la progresión del cáncer. La TP Mejora la angiogénesis dentro del tumor por dos mecanismos; el primero es estimulando la migración de células endoteliales y el segundo es por medio de estimulación para la liberación de factores angiogénicos de las células malignas y las células del estroma en el microambiente tumoral. En otros estudios se ha demostrado que el bloqueo de la función enzimática de TP elimina su función angiogénica.<sup>6</sup>

## **1.4 Inhibidores de Timidina fosforilasa**

En 1971, el biólogo y oncólogo Judah Folkman, postuló que el crecimiento tumoral es dependiente de la angiogénesis y que el desarrollo del tumor y la metástasis podrían ser abolidos bloqueando el suministro de sangre del tumor.<sup>10</sup>

Aunque se sabe que las bases de pirimidina libres liberadas durante la reacción de fosforilación, exhiben una inhibición relativamente débil de la enzima. La introducción de un grupo amino en la posición 6 de la pirimidina ha aumentado la actividad inhibitoria hasta por nueve veces.<sup>11</sup> Por ello, hoy en día se busca

novedosos inhibidores, dando como resultado la propuesta de la cumarina y sus derivados que están siendo objeto de estudio de relación estructura actividad (SAR) para encontrar un inhibidor eficiente de la timidina fosforilasa.

### **1.4.1 Inhibidores derivados de cumarina**

Las cumarinas son naturalmente provenientes de los metabolitos secundarios de las semillas, raíces y hojas de muchas especies de plantas, y son encontradas especialmente en altas concentraciones en el *haba tonka*, el cual es un árbol de legumbres.<sup>12</sup>

Los compuestos cumarínicos muestran una gran variedad de acciones bioquímicas y farmacológicas. En este contexto, se ha encontrado que, ciertos miembros de este grupo de compuestos pueden afectar significativamente la función de varios sistemas celulares de mamíferos.<sup>13</sup>

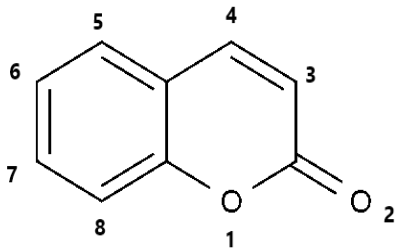
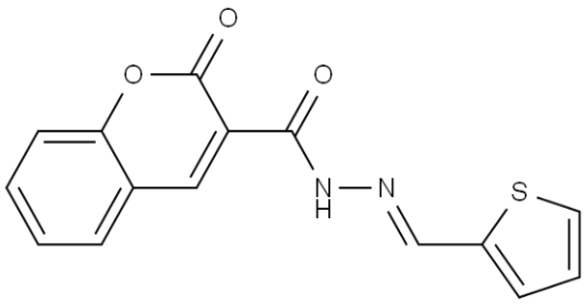
Algunas cumarinas muestran propiedades citostáticas (inhibidoras del crecimiento) mientras que otras exhiben actividades citotóxicas.<sup>14</sup> Aun mas se ha encontrado que los derivados de la cumarina tienen numerosas aplicaciones terapéuticas, entre las que destaca su posible uso en terapia antitumoral. Por ejemplo, la cumarina y su metabolito activo, la 7-hidroxycumarina, demostraron actividad citostática al inhibir crecimiento en líneas celulares de cáncer humano, como A549 (pulmón), ACHN (renal), H727 (pulmón), MCF-7 (seno) y HL -60 (leucemia), además se ha reportado que dichos compuestos demuestran actividad contra el cáncer de próstata, melanoma maligno y carcinoma metastásico de células renales en ensayos clínicos.

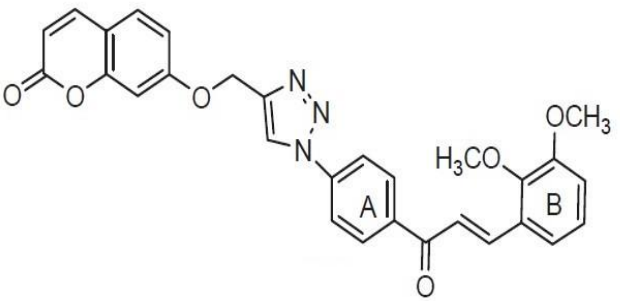
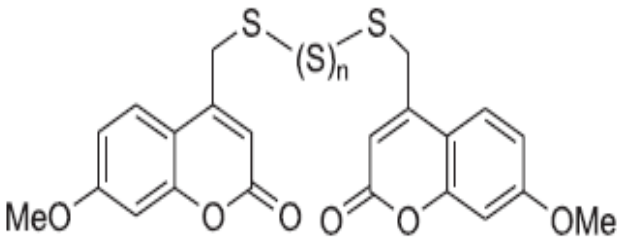
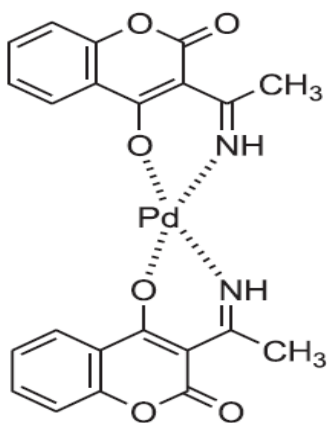
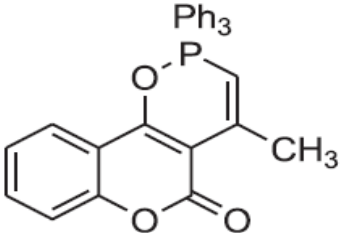
<sup>12</sup>

Por los numerosos efectos benéficos que tienen las cumarinas, estos compuestos bioactivos siguen siendo una de las clases de moléculas más versátiles para el diseño y descubrimiento de nuevos fármacos contra el cáncer.

Los agentes anticancerígenos basados en cumarina, según sea el tipo y la posición en la que se encuentra el sustituyente del farmacóforo principal en su estructura central, se han clasificado de la siguiente manera como se muestra en la tabla 1:

*Tabla 1: Clasificación de las cumarinas con ejemplos representativos de cada tipo de derivado.*

Clasificación de derivados cumarínicos	Ejemplo
Cumarinas sustituidas	
Cumarinas 3-sustituidas	

<p>Híbridos de cumarina</p>	
<p>Biscumarinas</p>	
<p>Complejos metálicos de cumarina</p>	
<p>Cumarinas fusionadas</p>	

Es importante destacar los reportes sobre las cumarinas con fracción hidrazida en la posición C-3, los cuales en han demostrado que el grupo hidrazida-hidrazona ( $-\text{CO}-\text{NH}-\text{NCH}-$ ) tiene un papel importante dentro de las moléculas que actúan como agentes antitumorales.<sup>15</sup>

En un estudio tipo SAR (structure-activity relationship) realizado en el año 2014 por Tamer y colaboradores demostraron que cuando el heteroarilo presente es un grupo pirrol o tiofeno, los compuestos mostraban una mayor actividad contra las líneas celulares HepG2 (células de cáncer hepático) y CCRF (células linfoblastoides T) y cuando el heterociclo contenía un grupo furano, los compuestos demostraron una buena actividad contra células PANC1 (células cancerígenas de páncreas). Por otra parte, también se observó que la fracción de hidrazida en la porción del  $-\text{CO}-\text{NH}-\text{NCH}-$ , especialmente el oxígeno presente tiene una buena actividad contra células PANC1 las cuales son provenientes de cáncer pancreático, mientras que la presencia del átomo de azufre del grupo presente tiofeno o el átomo de nitrógeno en el grupo pirrol, tienen actividad contra las células cancerígenas de hígado (Hepg2) así como contra las células linfoblastoides T (CCRF). En la figura 2 se representa con una molécula la información descrita anteriormente observando un derivado cumarínico con fracción hidrazida y un grupo tiofeno. Debido al mecanismo de angiogénesis y mecanismos anti apoptóticos de este tipo de compuestos han motivado el desarrollo de nuevas terapias para tratar más eficazmente el cáncer.<sup>16</sup>

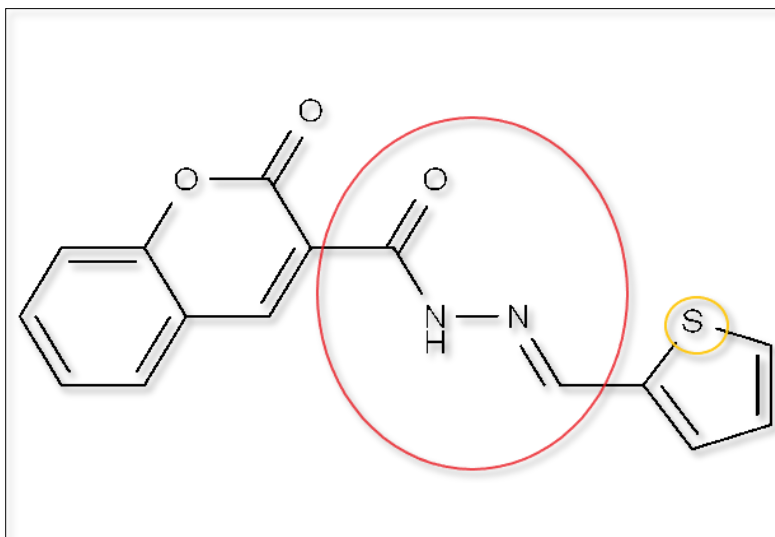


Figura 2: *Cumarina con la fracción hidrazida encerrada en un círculo rojo y el tiofeno, se encuentra encerrado en un círculo amarillo.*

## 1.5 Descubrimiento y desarrollo de nuevos fármacos

### 1.5.1 Métodos *in Silico*

El nombre *in silico* hace referencia al material del cual están fabricados los componentes de un ordenador. Con la invención de la computadora se ha logrado recrear varios procesos naturales, complejos con la ayuda de esta tecnología, se puede entender mejor los procesos fisiológicos en los cuales se busca una respuesta para así poder obtener mejores resultados en el desarrollo de nuevos tratamientos.<sup>17</sup>

## 1.5.2 Modelado molecular

Hoy en día el diseño de fármacos asistidos por computadora está siendo usados con mayor frecuencia, tanto en la investigación de sus detalles funcionales como en su desarrollo. En el descubrimiento de nuevos fármacos se realiza un proceso por el cual se identifican las nuevas moléculas terapéuticamente potenciales, todo esto haciéndolo posible con el uso combinado de modelos computacionales, experimentales, traslacionales y clínicos. Con frecuencia, aunque no necesariamente en el diseño de medicamentos se basa en el uso de técnicas de modelado por computadora y enfoques bioinformáticos. Por otra parte, las moléculas pequeñas, los productos biofarmacéuticos y con mayor importancia las terapias con anticuerpos son una clase de fármacos realizados con métodos computacionales para mejorar la afinidad, selectividad y estabilidad de estas terapias basadas en proteínas. El acoplamiento molecular, denominado en inglés *molecular Docking*, es uno de estos métodos mencionados El cual consiste en la unión de ligandos potenciales que van dirigidos a un blanco farmacológico molecular para investigar su posible efecto sobre dicha proteína.<sup>17</sup> Otra de las aplicaciones del método computacional es el buscar y proponer nuevos cambios a las estructuras químicas para incrementar su afinidad a su blanco molecular (también denominado diana química ) y al mismo tiempo potencializar su efecto terapéutico.<sup>18</sup>

### 1.5.3 Etapas del desarrollo de fármacos

Durante el desarrollo de fármacos en las pruebas *in vitro*, posibles candidatos pueden mostrar actividad, sin embargo, es común que en pruebas siguientes fallen ya que tiene limitadas propiedades farmacocinéticas y en algunos casos llegan a ser tóxicos. Además, se estima que un aproximado de cada 9,000 compuestos activos biológicamente, solo una llega a tener uso clínico ya que, aunque actúe en su diana molecular, llega a afectar otros procesos fisiológicos que evitan su uso seguro en humanos.<sup>19</sup> El descubrir y desarrollar un fármaco suele ser un procedimiento largo y complejo que normalmente transcurre de diez a quince años, desde que inicia la investigación hasta el lanzamiento del medicamento en el mercado una vez obtenida la aprobación regulatoria. Los años transcurridos incluyen los ensayos clínicos.<sup>20, 21</sup> Como ya se ha mencionado, el periodo de tiempo de obtención de un medicamento en el mercado podría en principio reducirse significativamente con la ayuda de métodos computacionales.

Las diversas etapas por las que transcurre el desarrollo de un fármaco se describen a continuación:

#### I. Etapa de descubrimiento y optimización.

En esta etapa se inicia la investigación básica en la cual se identifica la diana molecular, así como de los compuestos líderes que tienen una posible acción contra el objetivo molecular. Éste proceso se puede llevar a cabo utilizando una diversidad de métodos computacionales los cuales tienen como objetivo:

- a. El diseño e identificación de nuevos compuestos
- b. La selección de las moléculas candidatas.

- c. Optimizar las estructuras líderes para mejorar sus propiedades y disminuir los efectos adversos.

Todo lo anterior tiene como principio el entender la relación entre la estructura y la actividad biológica de los compuestos.<sup>19</sup> Si la optimización es exitosa los compuestos líderes pueden continuar el proceso de desarrollo del fármaco.<sup>22</sup>

## II. Etapa de desarrollo

Durante esta acción puede haber varias razones por las cuales se puede interrumpir el desarrollo de un fármaco, como se muestra en la siguiente grafica de pastel:

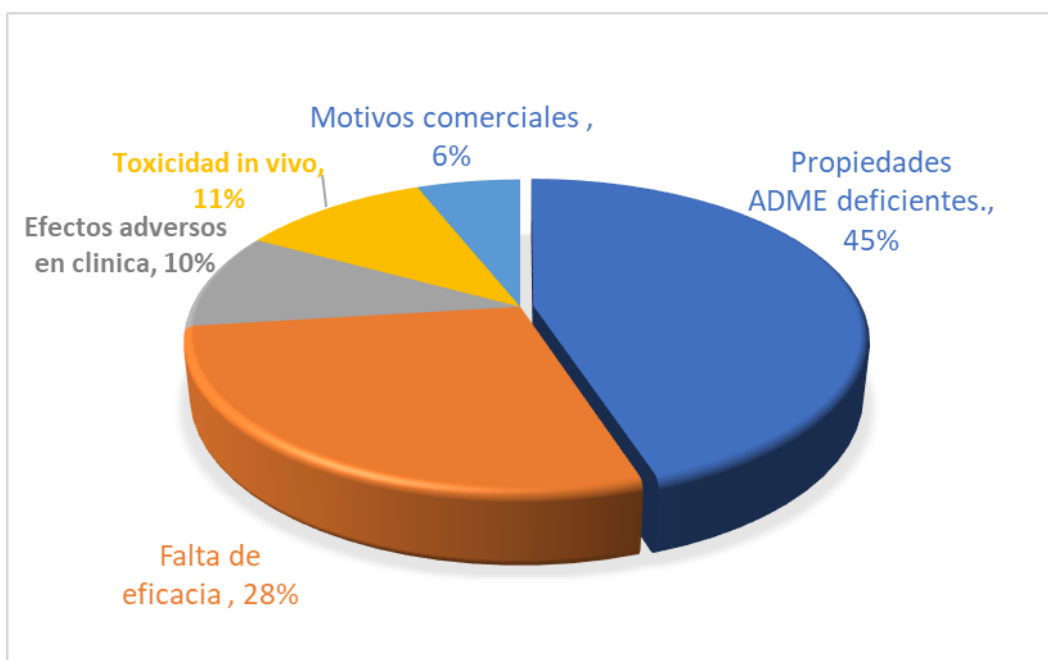


Figura3. Gráfico tomado y modificado de (Patel, 2006), que indica el porcentaje de los motivos por los cuales se interrumpe el desarrollo de un fármaco.<sup>23</sup>

El desarrollo preclínico es una etapa que se inicia en el laboratorio con un ensayo *in vitro* o utilizando animales de laboratorio (*in vivo*) y sirve para demostrar que los efectos adversos del fármaco son pocos o nulos. Las pruebas mencionadas son un

paso previo al ensayo con humanos y son un requisito exigido por las autoridades regulatorias. Una vez aprobado este paso, el siguiente será la fase clínica.<sup>24</sup>

## **Etapas de fases clínicas**

### **Fase I**

Se administra por vez primera el fármaco en humanos, en esta fase se hace un grupo reducido de voluntarios sanos, al menos 100. Por lo general los voluntarios son jóvenes adultos del género masculino. El principal objetivo de esta fase es detectar signos de toxicidad. Esta fase suele durar entre 6 meses a un año.

### **Fase II**

En esta fase el fármaco se administra a un grupo casi homogéneo de entre 100 a 200 personas, las cuales se dividen en dos grupos. A unas se le administra el fármaco a prueba y a otras se les administra un placebo o bien el mejor fármaco para tratar dicha patología. Esto con el objetivo de verificar la eficacia del fármaco. Esta fase dura aproximadamente 2 a 3 años.

### **Fase III**

Ya en este punto, los individuos de estudio suelen ser cientos o miles con características heterogéneas para así tener una gran diversidad biológica. En esta fase se pueden detectar indicios de toxicidad no detectados en los estudios previos. La duración de este ensayo suele tener una duración de aproximadamente 3 y 5 años.

### **Fase IV**

Una vez que ya se ha autorizado el fármaco se hace un seguimiento de este después de su comercialización, esto se le conoce como estudios de farmacovigilancia, y su objetivo es el identificar efectos adversos que pueden surgir en cierta porción de la población.<sup>24</sup>

## 1.6 QSAR (Relación cuantitativa estructura-actividad)

La técnica de relación cuantitativa de la estructura-actividad o QSAR por sus siglas en inglés (Quantitative Structure – Activity Relationship), es una herramienta dirigida a el área del desarrollo de nuevos fármacos y frecuentemente utilizada, tanto en industrias farmacéuticas, como en institutos de investigación públicos y privados.

Esta disciplina usa la metodología de relacionar numéricamente las estructuras químicas de varios compuestos con su actividad biológica. La herramienta QSAR entonces nos sirve para conocer la relación entre la actividad biológica y las propiedades moleculares de los ligantes, es decir, las propiedades fisicoquímicas de un compuesto. Dichas propiedades se expresan mediante parámetros conocidos como descriptores moleculares. En dicho análisis se intenta expresar la actividad biológica como una combinación lineal de diferentes descriptores, postulando de esta manera la forma de una relación lineal entre la actividad y las propiedades moleculares relevantes. Los coeficientes en las ecuaciones aportan contribuciones para predecir la actividad. Por lo regular, una serie de análogos con actividad conocida se usa como un conjunto de datos para la construcción de modelos. Los modelos QSAR resultantes se evalúan utilizando técnicas de validación cruzada y los modelos con buen rendimiento se utilizan para predecir modificaciones estructurales que podrían aumentar el efecto de la molécula. La modificación estructural de la molécula tiene como finalidad el optimizar su actividad farmacológica, esto con el fin de obtener fármacos más selectivos y con una menor toxicidad y por ende con una mejor farmacocinética, así como también con mejor estabilidad en las formulaciones farmacéuticas.<sup>25</sup>

En QSAR hay diferentes métodos computacionales entre los cuales se encuentran los datos de dos dimensiones 2D-QSAR la construcción de modelos se basa en la representación 2D de moléculas y descriptores 2D. Sin embargo, este tipo de estudios son insensibles a los arreglos moleculares en el espacio. Por otro lado, en el 3D-QSAR, se predicen las conformaciones bioactivas de los compuestos de prueba y se generan campos estéricos y electrostáticos dependientes de la

conformación de estos compuestos y se sondean en representaciones de cuadrícula para sitios de interacción favorables. En este caso tales descriptores también deben ser 3D o también se pueden usar en combinación con descriptores 1D o 2D para así tener modelos QSAR.<sup>26, 27</sup>

Una vez mencionado todo lo anterior, el QSAR tiene la finalidad de hacer una predicción teórica de la actividad biológica que permite el diseño teórico de nuevos fármacos totalmente nuevos a desarrollar en un futuro.<sup>28</sup>

Por ello el modelado QSAR reúne la relación de las estructuras químicas para el desarrollo de una relación matemática de respuesta o atributos químicos cuantitativos que definen las características de nuestras moléculas analizadas. El nombre del estudio depende de la naturaleza de la respuesta teniendo lugar tres clases de estudios principales mencionados a continuación: QSPR (estructura-propiedad), QSAR (estructura- actividad), QSTR (estructura-toxicidad).

Para iniciar el estudio QSAR se necesita las moléculas relacionadas estructuralmente que tengan la misma farmacodinamia (mismo mecanismo de acción) y que tengan en común la misma diana molecular. Por otra parte se necesitan datos de respuesta biológicas de cada uno de los ligantes, tales como: EC50 (concentración efectiva en 50% de la población), IC50 (concentración de inhibición de 50% de la población), LD50 (la dosis requerida para matar a la mitad de la población total), por último se necesitan las propiedades fisicoquímicas también llamados descriptores moleculares, las cuales son calculadas utilizando el software apropiado y usando los modelos de las estructuras químicas diseñadas previamente.<sup>28, 29</sup>

## Descriptores moleculares

Con frecuencia los descriptores se dividen en descriptores 1D, 2D o 3D, dependiendo de la dimensionalidad de la representación molecular a partir de la que se calculan. De acuerdo con esto, los descriptores 1D se calculan a partir de fórmulas de composición molecular e incluyen propiedades generales como el peso molecular. Por otra parte, los descriptores bidimensionales incluyen: índices de conectividad molecular, descriptores de carga etc. Finalmente, los descriptores tridimensionales (3D) capturan propiedades o características que dependen de la conformación molecular, como son: el momento dipolar, la superficie accesible al solvente o los arreglos específicos del farmacóforo. Por último, la contribución de los diferentes tipos de descriptores puede combinarse en una formulación bastante compleja de descriptores que combinan topología molecular y propiedades de átomos electrónicos o descriptores que mapean características tales como caracteres hidrófobos y polares en representaciones de superficie molecular.<sup>26</sup>

En la tabla número 2 se enlistan algunos ejemplos de descriptores fisicoquímicos.

*Tabla 2: Ejemplos del tipo de descriptores*

Categoría de descriptores	Ejemplos
Propiedades físicas	Peso molecular, log P(o/w)
Conteo de átomos y enlaces	Número de átomos de Nitrógeno Número de átomos aromáticos Número de enlaces rotables
Características del farmacóforo	Número de enlaces de hidrógenos aceptores Total de áreas de superficie de van der waal de átomos básicos.

Descriptores relacionados a la carga	Total de cargas parciales positivas Momento dipolar de cargas parciales
--------------------------------------	--

Tendiendo una vez todo este conjunto de descriptores moleculares que son propiedades fisicoquímicas de las moléculas análogas y la actividad biológica se hace el uso de la bioinformática, así como de la estadística para evaluar el modelo QSAR y predecir la actividad biológica con los descriptores elegidos.

## 1.7 Regresión lineal múltiple en R software para el análisis estadístico para el análisis de los modelos QSAR

La regresión lineal múltiple es una herramienta estadística que permite establecer la relación entre una variable dependiente y un conjunto de variables independientes ( $X_1, X_2, \dots$ ). En esta forma de regresión a diferencia de la regresión lineal simple, se permite hacer un análisis de situaciones más reales (ya que los fenómenos en este caso biológicos son altamente complejos), por lo cual al tener un conjunto de varias variables permite explicar de una mejor manera dicha relación.<sup>30, 31</sup>

A continuación, se escribe la ecuación lineal uno (1):

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e \quad (1)$$

Donde:

Y: Es la variable dependiente a predecir

$\alpha$  y  $\beta$  son parámetros desconocidos a estimar

e: es el error, la diferencia entre el valor observado y el estimado por el modelo.

La regresión lineal múltiple (MLR) es favorecida por su simplicidad y facilidad de interpretación ya que el modelo asume una relación lineal entre la propiedad del compuesto denotado con la letra Y y su vector de características, denotado con una X, que generalmente son los descriptores moleculares. Por lo tanto, con el modelo ajustado se puede predecir la propiedad de un compuesto desconocido.

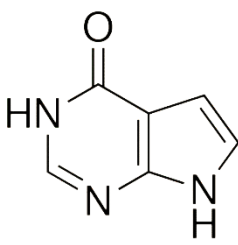
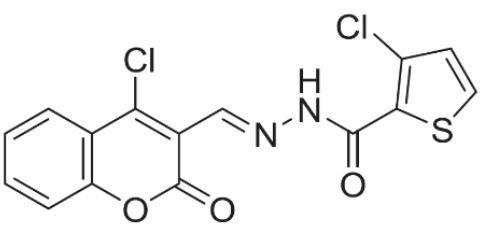
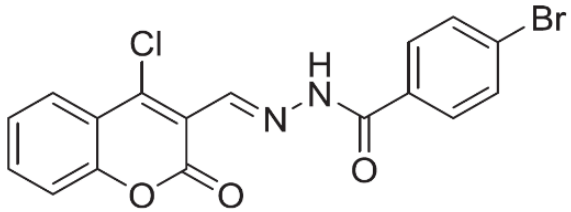
El tamaño de los coeficientes puede revelar el grado de influencia de los descriptores moleculares correspondientes en la propiedad objetivo. Además, un coeficiente positivo sugiere que el descriptor molecular correspondiente contribuye positivamente a la propiedad objetivo, mientras que un coeficiente negativo sugiere una contribución negativa. Sin embargo, estas interpretaciones pueden no ser precisas ya que los descriptores colineales tienen el potencial de influir en los coeficientes, de modo que se pueden asignar valores erróneos. Por lo tanto, los descriptores moleculares en el modelo deben ser independientes entre sí y el número de instancias para la construcción del modelo debe ser al menos cinco veces el número de descriptores utilizados.<sup>32</sup>

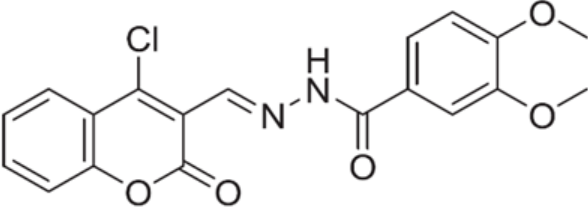
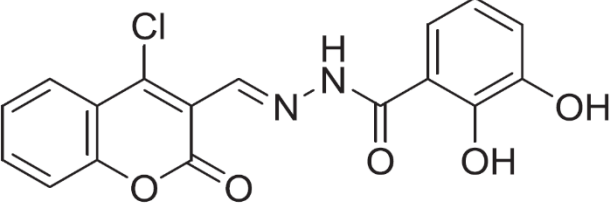
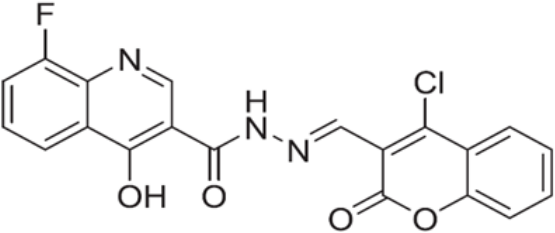
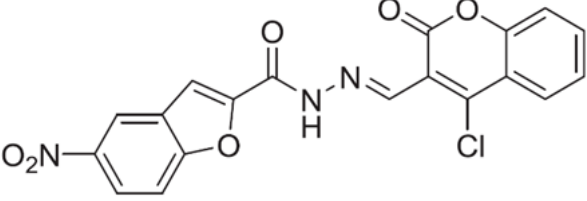
## 2. Antecedentes

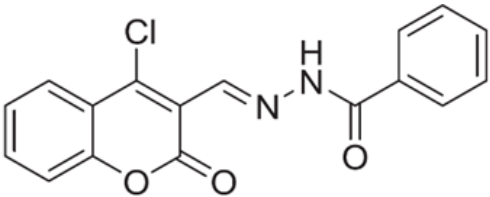
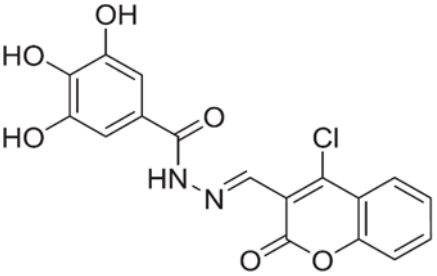
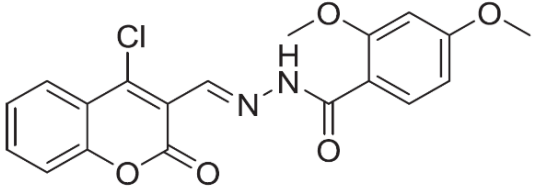
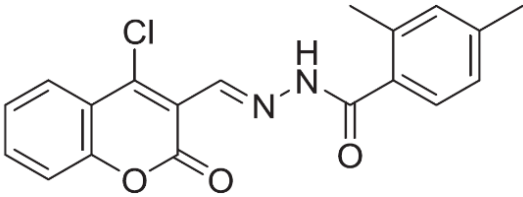
En 2018 Muhammad y colaboradores proponen 17 moléculas derivadas de la 3-formilcumarina como agentes inhibidores de la proteína timidina fosforilasa. El artículo mencionado lleva como título: *Synthesis, molecular docking study and thymidine phosphorylase inhibitory activity of 3-formylcoumarin derivatives*. Traducido al español como: Síntesis, estudio de acoplamiento molecular y actividad inhibitoria de timidina fosforilasa de los derivados de 3-Formil cumarina. En dicho artículo se menciona que las cumarinas y sus derivados son atractivos para la química orgánica como para la medicina química ya que tienen varios usos que van desde aditivos en perfumes, alimentos o como antiinflamatorios, anticonceptivos, antimicrobianos y antitumorales.<sup>33</sup>

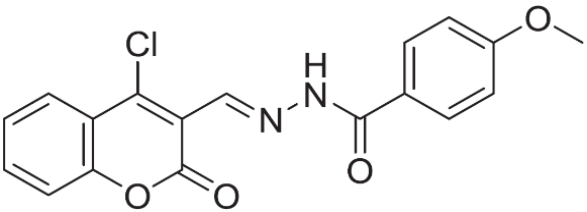
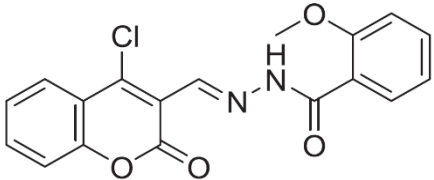
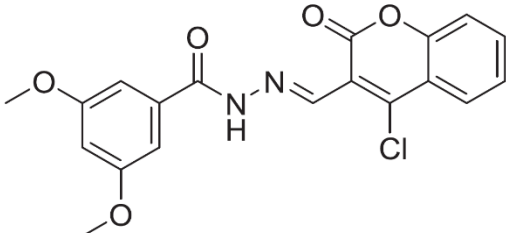
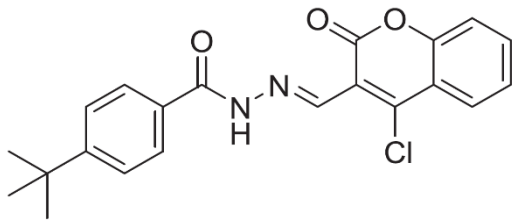
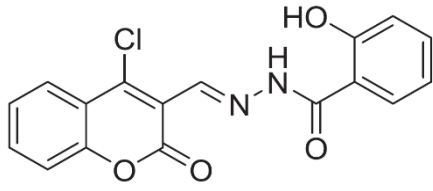
En dicho estudio se investigaron 17 derivados de 3-formilcumarina cumarina, los cuales fueron sintetizados y caracterizados por técnicas espectroscópicas tales como: HNMR y EI-MS. Posteriormente fueron probadas experimentalmente *in vitro* las 17 moléculas derivadas de cumarina. En dicho análisis se determinó su constante de inhibición 50 ( $IC_{50}$ ) teniendo como estándar la molécula 7-Deazaxanthine, la cual tiene una  $IC_{50}$  de  $38.68 \pm 1.20 \mu\text{M}$ . Este procedimiento se llevó a cabo usando usando la técnica espectrofotométrica en la cual a una longitud de onda de 290nm en un vial se midió la absorbancia de la mezcla de un total de 200 $\mu\text{L}$  los cuales contenían 145  $\mu\text{L}$  de fosfato de potasio como buffer pH (7.4), 30  $\mu\text{L}$  de enzima (Timidina fosforilasa humana y de *E.coli*) estando a una concentración de 0.05 y 0.002 U respectivamente. También se agregó 5  $\mu\text{L}$  y 20  $\mu\text{L}$  del sustrato. La lectura fue tomada 10, 20 y 30 minutos Los experimentos se realizaron por triplicado. Después de todo lo anterior se obtuvo los valores de  $IC_{50}$  para cada derivado, véase la tabla 3.

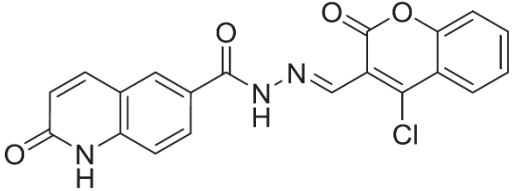
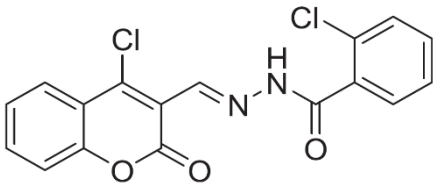
Tabla 3: 17 derivados de 3-Formilcumarina con su respectiva  $IC_{50}$  así como su compuesto de referencia el inhibidor 7-Deazaxanthine.

No	Molécula 2D	$IC_{50}$
7-Deazaxanthine		$38.68 \pm 1.20 \mu\text{M}$
1		$13.40 \pm 0.1 \mu\text{M}$
2		$29.30 \pm 0.60 \mu\text{M}$

3		23.50 ± 0.20 μM
4		3.50 ± 0.1 μM
5		1.20 ± 0.01 μM
6		4.50 ± 0.01 μM

7		29.30 ± 0.80 μM
8		1.30 ± 0.10 μM
9		24.40 ± 0.50 μM
10		31.20 ± 0.90 μM

11		34.50 ± 0.1 μM
12		38.30 ± 0.90 μM
13		41.60 ± 1.10 μM
14		53.50 ± 1.20 μM
15		13.20 ± 0.20 μM

16		$0.90 \pm 0.01 \mu\text{M}$
17		$16.30 \pm 0.20 \mu\text{M}$

En la explicación de los resultados se mencionan que el análogo 3 con  $IC_{50}$  de  $23.50 \pm 0.20 \mu\text{M}$ , el análogo 9 con un valor  $IC_{50}$   $24.40 \pm 0.50 \mu\text{M}$  y el análogo 13 con un valor  $IC_{50}$   $41.60 \pm 1.10 \mu\text{M}$  tienen grupos metoxi en el anillo de fenilo. De manera interesante, ellos proponen que debido a que la posición de los grupos metoxi en cada análogo es distinta, estos compuestos presentan tales variaciones de los valores de  $IC_{50}$ . Aunque el modelo 13 no tiene un mejor  $IC_{50}$  en comparación al estándar, el análogo 3 y 9 tienen un buen valor de inhibición en comparación al mismo.

Al comparar también el análogo 11, el cual tiene un valor  $IC_{50}$   $34.50 \pm 0.1 \mu\text{M}$  con el análogo 12, el cual tiene un valor  $IC_{50}$   $38.30 \pm 0.90 \mu\text{M}$ , ambos análogos presentan un grupo metoxi, sin embargo, la disposición del grupo metoxi es diferente en ellos. Este hecho sugiere que la posición de los sustituyentes influye en gran medida en los potenciales inhibitorios. Mientras que, entre los derivados halogenados, como el análogo 17 sustituido con cloro, el cual tiene un valor  $IC_{50}$   $16.30 \pm 0.20 \mu\text{M}$  mostró un mayor potencial inhibitorio en comparación con el análogo 2 que tiene un bromo mostrando un valor  $IC_{50}$   $29.30 \pm 0.60 \mu\text{M}$ . Este hecho puede ser explicado por la mayor electronegatividad del grupo cloro.

El mejor derivado cumarínico fue el 16, con un valor de  $IC_{50}$  de  $0.90 \pm 0.01 \mu M$ , lo cual podría explicarse por la presencia del grupo quinolona en el anillo fenólico. Por lo tanto, de acuerdo al docking molecular (acoplamiento molecular) realizado en citado estudio, el derivado número 16 ocupa el sitio de unión de fosfato con un total de cuatro interacciones de enlace de hidrógeno. Por otro lado, el oxígeno en el carbonilo (C=O) del anillo de cumarina forma un enlace de hidrógeno con el átomo de nitrógeno del residuo de Glicina 118 (Gly 118). Además, a una distancia de 2,87 Å, también se encontró que el grupo amino (NH) en el grupo hidrazona interactúa con el átomo de oxígeno de la cadena principal (O) del residuo hidrofóbico de Leucina 117 (distancia de 2,88 Å). Esto hace que la enzima no pueda unirse a los ribonucleótidos de pirimidina. Otra interacción importante en la posible explicación del mecanismo de acción de tal derivado es la interacción del oxígeno del grupo carbonilo de la quinolinona el cual puede formar enlaces de hidrógeno con aminoácidos importantes tales como la Glicina 114 y la Ser 113. Finalmente, de acuerdo con la investigación de Muhammad y colaboradores esta interacción es importante para disminuir las propiedades nucleofílicas del fosfato, evitando así que tenga lugar cualquier fosforolisis.<sup>33, 34</sup>

### **3. Hipótesis**

El estudio cuantitativo de la relación estructura-actividad (QSAR), por medio de una regresión lineal múltiple predecirá la relación de la actividad biológica de los derivados de 3-Formil cumarina con sus descriptores moleculares.

## 4. Justificación

Hay múltiples investigaciones de los derivados cumarínicos siendo objeto reciente de estudio las cumarinas sustituidas en su carbono número 3 como inhibidores de la enzima timidina fosforilasa. Hoy en día se sabe que esta enzima está implicada en el crecimiento tumoral por su efecto angiogénico y quimiotáctico por lo que Muhammad y colaboradores han propuesto en 2018 un estudio con 17 derivados de 3-formil cumarina que solo cuenta con un estudio tipo SAR. Dicho estudio, sólo evalúa la relación que hay entre la estructura y su actividad inhibitoria la cual está dada por la constante de inhibición 50 que determinaron experimentalmente, por lo que se pretende tomar esta única variable dependiente y complementar el estudio con variables independientes calculadas a partir de las moléculas modeladas.

## **5. Objetivo General**

Generar un modelo QSAR que represente cuantitativamente la relación de la actividad biológica y los descriptores moleculares de las 17 moléculas derivadas de 3-Formil cumarina.

### **5.1 Objetivos particulares**

1. Considerar el ligante del PDB 4EAD para el modelado y superposición de las moléculas derivadas de cumarina.
2. Calcular los descriptores moleculares por medio de programas computacionales especializados.
3. Analizar los modelos propuestos mediante una regresión lineal múltiple.
4. Determinar el mejor modelo QSAR para los derivados de 3- Formil Cumarina inhibidores de la enzima timidina fosforilasa.

## 6. Diagrama de trabajo

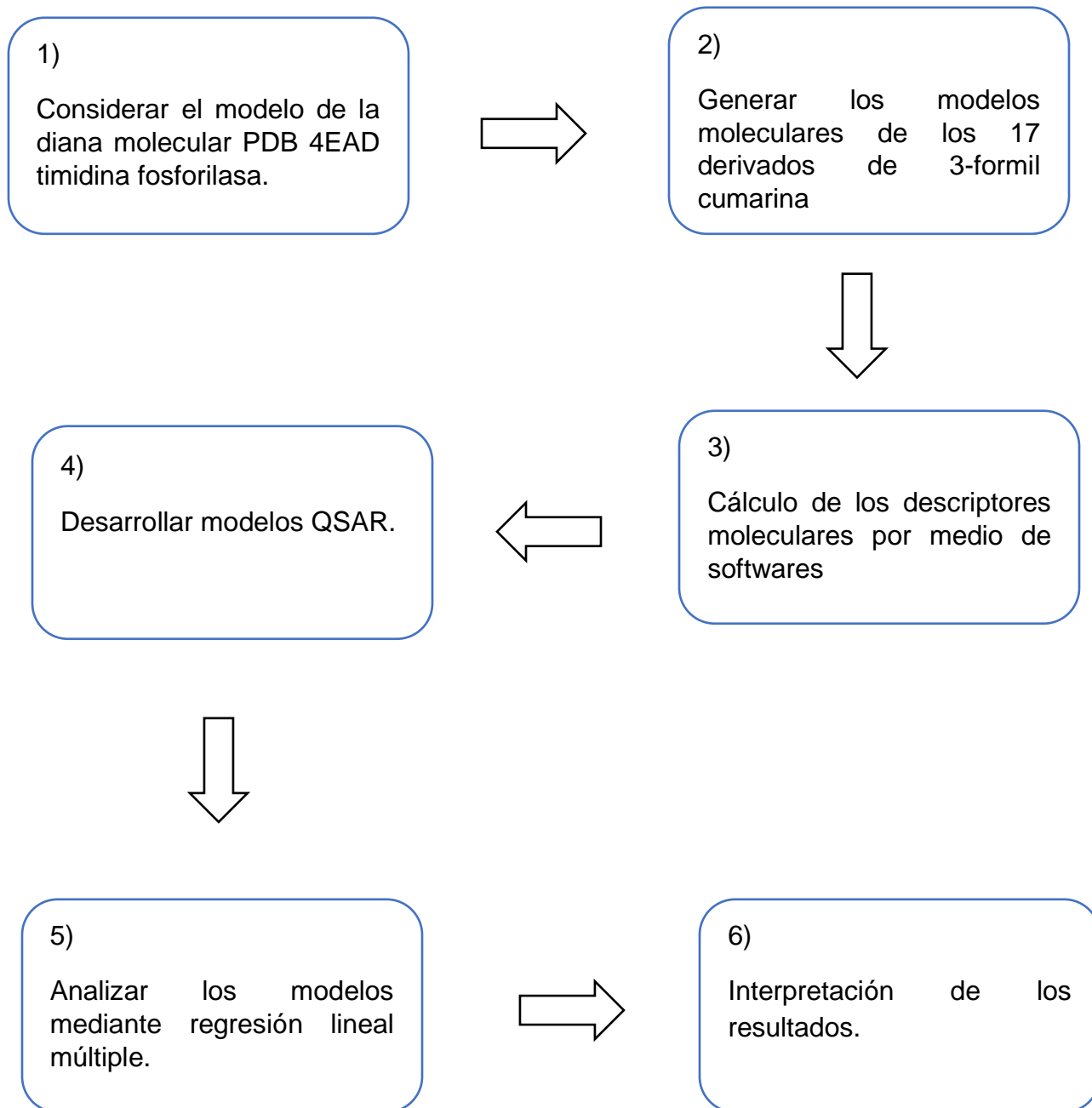


Figura 4: Diagrama de trabajo

## **7. Metodología**

### **7.1 Considerar el modelo de la diana molecular, timidina fosforilasa**

Del Protein Data Bank se tomó de referencia la estructura depositada 4EAD.pdb, la cual es un cristal de la proteína timidina fosforilasa proveniente de *E.coli*. Dicha estructura está cocrystalizada con un ligante inhibidor el cual servirá para la modelación y optimización de las moléculas a desarrollar.

### **7.2 Generación de los modelos moleculares**

Con la ayuda del programa Vega ZZ se generó las estructuras de las 17 moléculas derivadas de cumarina, siendo estas optimizadas con el campo de fuerza Tripos y con cargas de tipo gasteiger.

### **7.3 Cálculo de los descriptores o parámetros fisicoquímicos**

Mediante el uso de varios programas computacionales especializados como Vega ZZ, SwissADME, E-Dragon y Alva Desc 1.0., se obtuvieron los descriptores fisicoquímicos que servirán para poder construir la ecuación lineal de regresión múltiple. Éstos descriptores serán seleccionados utilizando métodos teóricos y estadísticos para representar apropiadamente al grupo de ligantes de estudio.

### **7.4 Desarrollar los modelos QSAR**

Una vez elegidos los descriptores moleculares se realiza nuestro modelo QSAR, utilizando un análisis estadístico.

### **7.5 Análisis del modelo QSAR por regresión lineal múltiple**

El modelo obtenido se analizó haciendo uso del programa de libre acceso R. Gracias a su amplia paquetería, dicha herramienta computacional estadística permite realizar análisis lineales y no lineales, así como también la posibilidad de representar a través de distintos tipos de gráficos.

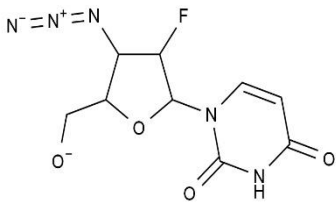
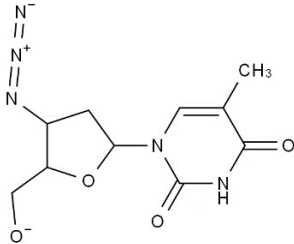
## 8. Resultados y discusión de resultados

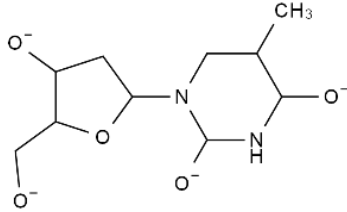
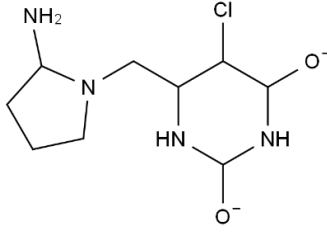
### 8.1 Consideración del PDB 4EAD y construcción de las moléculas análogas inhibidoras de la proteína timidina fosforilasa

Se buscó el PDB 4EAD, el cual contiene la estructura de rayo X de la enzima timidina fosforilasa, teniendo en su sitio activo un ligante inhibidor. También se encontraron cinco estructuras adicionales de la proteína de las cuales solo 3 tienen una molécula de ligante inhibidor en el sitio activo.

En la tabla 4 se muestra los PDBs buscados y utilizados.

Tabla 4: Proteínas cristalizadas de la proteína timidina fosforilasa con su respectivo ligante.

AÑO	PDB	TÍTULO TRADUCIDO	ORGANISMO	RESOLUCIÓN	LIGANDO	Bibl.
2013	4EAD	Timidina fosforilasa de <i>E. coli</i> con 3'-azido-2'-fluoro-didesoxiuridina	<i>Escherichia coli.</i>	1.5 Å		35
2014	4LHM	Timidina fosforilasa de <i>E. coli</i> con 3'-azido-3'-desoxitimidina	<i>Escherichia coli.</i>	1.52 Å		36

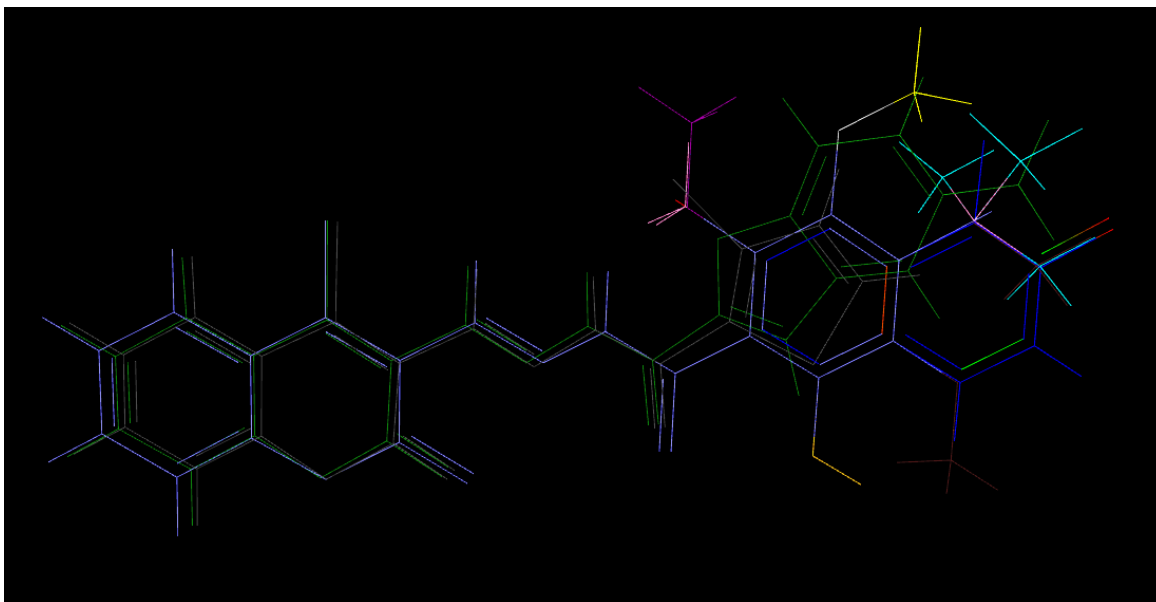
1999	1AZY	Los estudios estructurales y teóricos sugieren que el movimiento de dominio produce una conformación activa de la timidina fosforilasa	<i>Escherichia coli.</i>	3 Å	No hay	37
1999	1BRW	La estructura de cristal de la fosforilasa de nucleosidos de pirimidina en una conformación cerrada	<i>Geobacillus stearothermophilus</i>	2.1 Å	No hay	38
2016	4YEK	Estructura de rayos X de la timidina fosforilasa de <i>Salmonella typhimurium</i> en complejo con timidina	<i>Samlmonella typhimurium.</i>	2.55 Å		39
2004	1UOU	Estructura cristalina de la timidina fosforilasa humana en complejo con un inhibidor de molécula pequeña.	<i>Homo sapiens</i>	2.11 Å		40

El ligante del PDB 4EAD (3'-azido-2'-fluoro-didesoxiuridina) se utilizó como plantilla para construir los 17 moléculas derivados de 3-formil cumarina y al mismo tiempo ajustarlas al sitio activo.

## 8.2 Construcción de los 17 modelos moleculares

Las moléculas fueron construidas en el software Vega ZZ y se guardaron en formato SMILE. Posteriormente, con el fin de que las moléculas estuvieran colocadas en el espacio del sitio activo de la enzima TP, se ajustaron las moléculas con el ligante del PDB 4EAD mostrado en la tabla 4.

De esta forma se logró obtener la superposición de las moléculas las cuales se pueden observar en la figura 5.

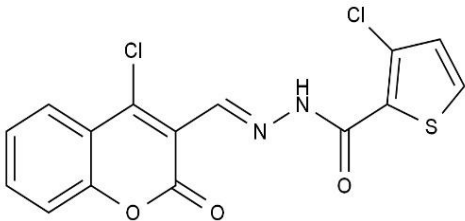
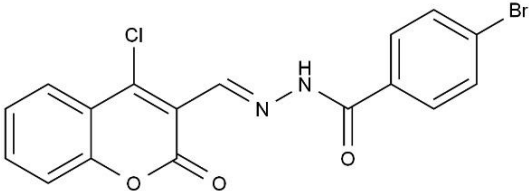


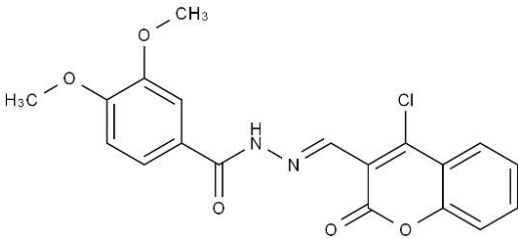
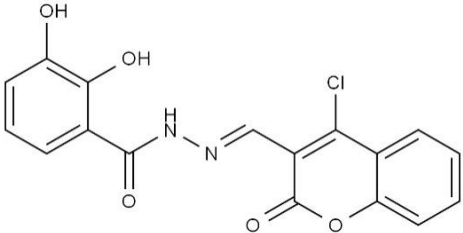
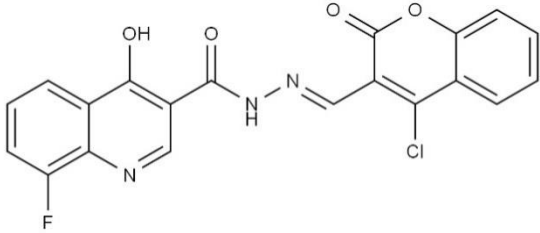
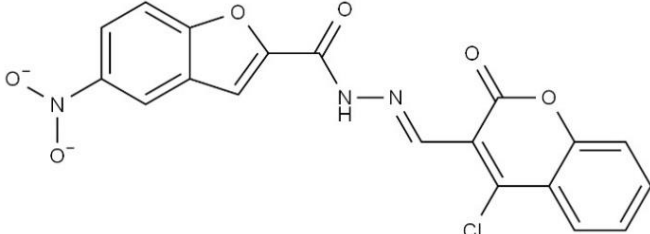
*Figura 5: 17 moléculas derivadas de 3-formilcumarina en superposición.*

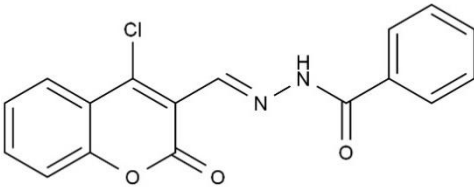
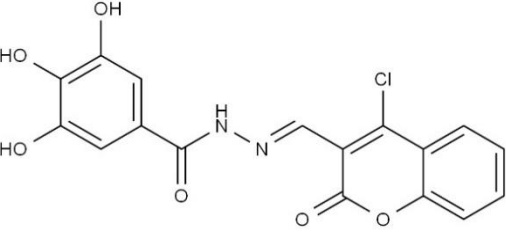
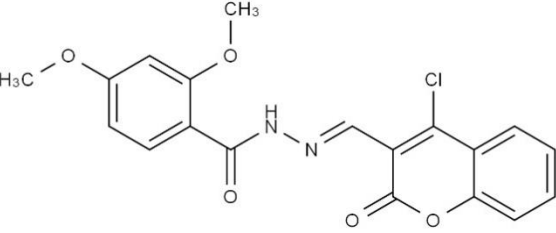
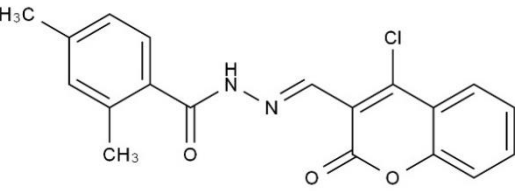
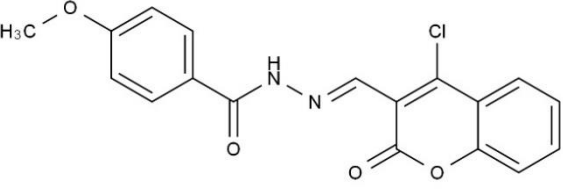
## 8.3 Cálculo de los descriptores moleculares mediante programas especializados

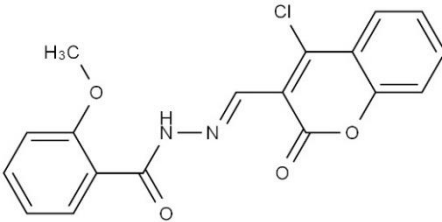
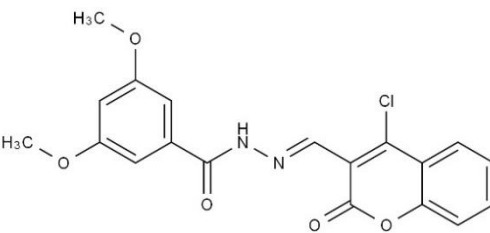
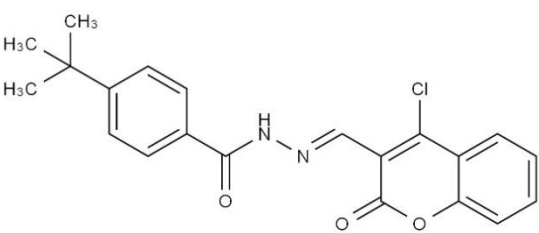
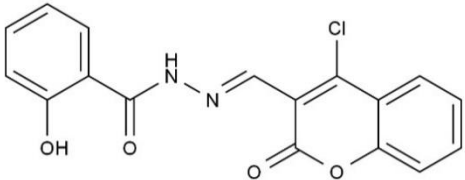
En la tabla 5 se presentan los 17 derivados de cumarina con sus respectivas estructuras en formato SMILES y en la tabla 6 los mismos formatos de las moléculas catiónicas como fueron obtenidas de VegaZZ e introducidas en diferentes software para calcular los descriptores fisicoquímicos.

Tabla 5: 17 moléculas derivadas de cumarina con sus respectivas estructuras moleculares catiónicas en el formato SMILES.

ID	Molécula 2D
1	 <chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1SC=CC=1Cl</chem>
2	 <chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1C=CC(Br)=CC=1</chem>

3	 <p>Chemical structure showing a benzene ring with two methoxy groups (H<sub>3</sub>C-O) and a carbonyl group (C=O) attached to a hydrazide linkage (-NH-N=) which is further connected to a chlorinated benzofuran derivative.</p> <chem>COC1=CC(=CC=C1OC)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
4	 <p>Chemical structure showing a benzene ring with two hydroxyl groups (OH) and a carbonyl group (C=O) attached to a hydrazide linkage (-NH-N=) which is further connected to a chlorinated benzofuran derivative.</p> <chem>OC1C(=CC=CC=1O)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
5	 <p>Chemical structure showing a quinoline ring system with a fluorine atom (F) and a hydroxyl group (OH) on the quinoline ring, and a carbonyl group (C=O) attached to a hydrazide linkage (-NH-N=) which is further connected to a chlorinated benzofuran derivative.</p> <chem>OC1C2=CC=CC(F)=C2N=CC=1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
6	 <p>Chemical structure showing a furan ring system with a nitro group (NO<sub>2</sub>) and a carbonyl group (C=O) attached to a hydrazide linkage (-NH-N=) which is further connected to a chlorinated benzofuran derivative.</p> <chem>[O]N([O])C1C=C2C=C(OC2=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>

7	 <p>Chemical structure of 2-chloro-3-(benzylidenehydrazono)chromone, showing a chromone ring with a chlorine atom at position 2 and a benzylidenehydrazono group at position 3.</p> <chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1C=CC=CC=1</chem>
8	 <p>Chemical structure of 3-(3,4,5-trihydroxybenzylidenehydrazono)-2-chlorochromone, showing a chromone ring with a chlorine atom at position 2 and a 3-(3,4,5-trihydroxybenzylidenehydrazono) group at position 3.</p> <chem>OC1C=C(C=C(O)C=1O)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
9	 <p>Chemical structure of 3-(3,4-dimethoxybenzylidenehydrazono)-2-chlorochromone, showing a chromone ring with a chlorine atom at position 2 and a 3-(3,4-dimethoxybenzylidenehydrazono) group at position 3.</p> <chem>COC1C=C(OC)C(=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
10	 <p>Chemical structure of 3-(3,4-dimethylbenzylidenehydrazono)-2-chlorochromone, showing a chromone ring with a chlorine atom at position 2 and a 3-(3,4-dimethylbenzylidenehydrazono) group at position 3.</p> <chem>CC1=CC(C)=CC=C1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
11	 <p>Chemical structure of 3-(4-methoxybenzylidenehydrazono)-2-chlorochromone, showing a chromone ring with a chlorine atom at position 2 and a 3-(4-methoxybenzylidenehydrazono) group at position 3.</p> <chem>COC1C=CC(=CC=1)C(=O)NN=CC1C(=O)OC2=CC=CC=C2C1Cl</chem>

12	 <p style="text-align: center;"><chem>COC1=CC=CC=C1C(=O)NN=CC1C(=O)OC2=CC=CC=C2C=1Cl</chem></p>
13	 <p style="text-align: center;"><chem>COC1C=C(C=C(C=1)OC)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem></p>
14	 <p style="text-align: center;"><chem>CC(C)(C)C1C=CC(=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem></p>
15	 <p style="text-align: center;"><chem>OC1=CC=CC=C1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem></p>

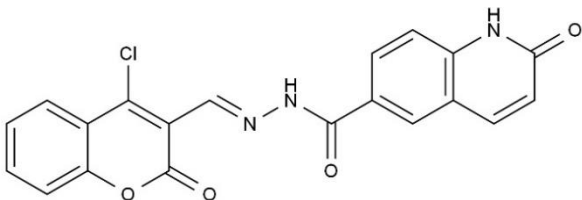
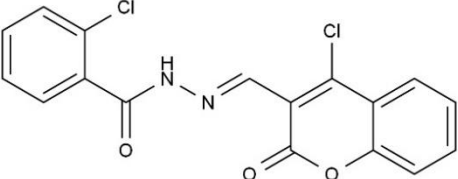
16	 <chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1=CC2C=CC(=O)NC=2C=C1</chem>
17	 <chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1=CC=CC=C1Cl</chem>

Tabla 6: Sonrisa catiónica de las 17 moléculas derivadas de cumarina

ID	SMILE
1	<chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1SC=CC=1Cl</chem>
2	<chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1C=CC(Br)=CC=1</chem>
3	<chem>COC1=CC(=CC=C1OC)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
4	<chem>OC1C(=CC=CC=1O)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
5	<chem>OC1C2=CC=CC(F)=C2N=CC=1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
6	<chem>[O-]N([O])C1C=C2C=C(OC2=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
7	<chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1C=CC=CC=1</chem>
8	<chem>OC1C=C(C=C(O)C=1O)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
9	<chem>COC1C=C(OC)C(=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
10	<chem>CC1=CC(C)=CC=C1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
11	<chem>COC1C=CC(=CC=1)C(=O)NN=CC1C(=O)OC2=CC=CC=C2C=1Cl</chem>
12	<chem>COC1=CC=CC=C1C(=O)NN=CC1C(=O)OC2=CC=CC=C2C=1Cl</chem>

13	<chem>COC1C=C(C=C(C=1)OC)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
14	<chem>CC(C)(C)C1C=CC(=CC=1)C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
15	<chem>OC1=CC=CC=C1C(=O)NN=CC1=C(Cl)C2=CC=CC=C2OC1=O</chem>
16	<chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1=CC2C=CC(=O)NC=2C=C1</chem>
17	<chem>O=C(NN=CC1=C(Cl)C2=CC=CC=C2OC1=O)C1=CC=CC=C1Cl</chem>

En el programa Vega ZZ se obtuvieron 11 descriptores, mientras que en el programa Swis ADME se obtuvieron 23. Además, utilizando el programa gratuito de E-Dragon se consideraron 21 mientras que en el programa AlvaDesc 1.0 también se tomaron 21 descriptores. La cantidad tomada de E-Dragon y AlvaDesc 1.0 fueron la misma ya que resultaron ser los mismos datos. Véase anexos para ver la lista detallada de todos los descriptores obtenidos para cada software utilizado.

La tabla 7 muestra los descriptores seleccionados para construir los modelos QSAR.

*Tabla7: Lista de descriptores seleccionados.*

Descriptor	Description
IC5	Information Content index (neighborhood symmetry of 5-order)
R4p	R autocorrelation of lag 4 / weighted by polarizability
R3s	R autocorrelation of lag 3 / weighted by I-state
CATS3D_05_AL	CATS3D Acceptor-Lipophilic BIN 05 (5.000 - 6.000 Å)
CATS3D_06_AL	CATS3D Acceptor-Lipophilic BIN 06 (6.000 - 7.000 Å)
CATS3D_08_DL	CATS3D Donor-Lipophilic BIN 08 (8.000 - 9.000 Å)
SPH	Spherosity
VE1_RG	coefficient sum of the last eigenvector (absolute values) from reciprocal squared geometrical matrix
Mor03i	signal 03 / weighted by ionization potential
Mor20s	signal 20 / weighted by I-state
F06[N-O]	Frequency of N - O at topological distance 6
TDB10p	3D Topological distance based descriptors - lag 10 weighted by polarizability
B10[O-Cl]	Presence/absence of O - Cl at topological distance 10
BIC4	Bond Information Content index (neighborhood symmetry of 4-order)

TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions
SpMax3_Bh(s)	largest eigenvalue n. 3 of Burden matrix weighted by I-state

Durante la elección de los descriptores se hace uso tanto de las herramientas estadísticas como de diversos criterios teóricos, es por ello que de todos los datos obtenidos se seleccionan y con ellos se construyen los modelos QSAR. En la tabla 8 se muestran las variables independientes tomadas para cada modelo siendo los modelos más grandes aquellos con un máximo de 4 descriptores y los más pequeños aquellos que solo tienen 2 variables independientes.

*Tabla 8: Lista de los modelos formados con nombre de los descriptores y número de variables en cada uno de los 9 modelos.*

No.	Número de variables independientes	Variables
1	4	IC5, R4p, R3s, CATS3D_05_AL
2	4	IC5, R4p, R3s, CATS3D_06_AL
3	4	IC5, SPH, VE1_RG, CATS3D_05_AL
4	3	IC5, Mor03i, F06[N-O]
5	3	IC5, TDB10p, CATS3D_08_DL
6	3	IC5, Mor03i, B10[O-Cl]
7	2	BIC4, Mor20s
8	2	BIC4, TPSA(Tot)
9	2	BIC4, SpMax3_Bh(s)

De todos los modelos anteriores se eligió el número 8, siendo los siguientes descriptores aquellos que conforman nuestro modelo QSAR a utilizar en el presente trabajo. En el subtítulo 8.4 “Análisis del modelo QSAR”, se habla sobre la razón por la cual se escogió dicho modelo.

## 8.4 Análisis de los modelos QSAR

Tanto los datos estadísticos, como los gráficos fueron obtenidos con el software R. A continuación, se presenta en la tabla 9 los coeficientes de determinación de los 9 modelos obtenidos.

*Tabla 9: Listado de modelos QSAR obtenidos con su R<sup>2</sup> resultante de su análisis de Regresión lineal múltiple (RLM)*

Modelo	Numero de descriptores	R <sup>2</sup>	R <sup>2</sup> -ajustada
Modelo1	4	0.9203	0.8938
Modelo2	4	0.9181	0.8907
Modelo3	4	0.8354	0.7805
Modelo4	3	0.8275	0.7877
Modelo5	3	0.6649	0.5875
Modelo6	3	0.8142	0.7713
Modelo7	2	0.6794	0.6336
Modelo8	2	0.8883	0.8724
Modelo9	2	0.8525	0.8314

El modelo número uno es el mejor de acuerdo del coeficiente de determinación. Dicho modelo tiene 4 descriptores los cuales se mostraron en la tabla 8 sin embargo al realizar los análisis estadísticos de cada uno, se consideró más adecuado el modelo número 8, debido a su buena predicción de la actividad biológica, así como por la aportación de información fisicoquímica de la contribución en la actividad inhibitoria de las moléculas. Aclarado esta parte a continuación se procede a desarrollar los resultados del análisis estadístico del modelo 8. En el apartado de anexos puede encontrarse una tabla con el número 34 y cuadro comparativo con gráficos de la validación de cada ecuación de regresión lineal múltiple.

La ecuación número 2 representa nuestro modelo QSAR elegido:

$$DV = \alpha + \beta_1 \text{BIC4} + \beta_2 \text{TPSA(Tot)} \quad (2)$$

Los datos de la información fisicoquímica que aportan los descriptores del modelo elegido (número 8) se muestran a continuación en la tabla número 10:

Tabla 10: Descriptores utilizados para formar nuestros modelos QSAR.

Descriptor molecular/IDV	Información	Fuente
BIC4	Bond Information Content index (neighborhood symmetry of 4-order)	Information indices
TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	Molecular properties

El descriptor BIC4 se tomó en cuenta porque describe de manera topológica el carácter del enlace. El carácter de enlace está estrechamente relacionado con la capacidad de los átomos unidos para intercambiar electrones, y dicha capacidad está comúnmente bien representada por la electronegatividad de los átomos unidos.<sup>41</sup>

TPSA(Tot) se ha elegido ya que muestra la correlación con el transporte molecular pasivo a través de las membranas, lo que permite predecir la absorción intestinal humana, la permeabilidad de las monocapas celulares y la penetración de la barrera hematoencefálica. Este descriptor se obtiene de un software que lo determina mediante la suma de las contribuciones superficiales tabuladas de los tipos de átomos polares, siendo estos los fragmentos polares con nitrógeno y oxígeno más fragmentos "ligeramente polares" que contienen fósforo y azufre.<sup>41, 42</sup>

Esta información que da cada uno sobre la actividad inhibitoria de la molécula se puede evaluar con métodos estadísticos como se mencionó en el apartado anterior y de esta forma justificar y explicar cómo de manera cuantitativa influyen nuestras variables independientes sobre los valores de la variable dependiente (IC50).

### 8.5 Análisis estadístico preliminar de los descriptores contenidos en el modelo elegido.

Se evaluó la constante de inhibición 50 de las 17 moléculas análogas mediante un histograma, para evaluar la frecuencia de los datos. La figura 6 muestra que los valores de IC50 tienen mayor frecuencia en una concentración de 0 a 10  $\mu\text{M}$ .

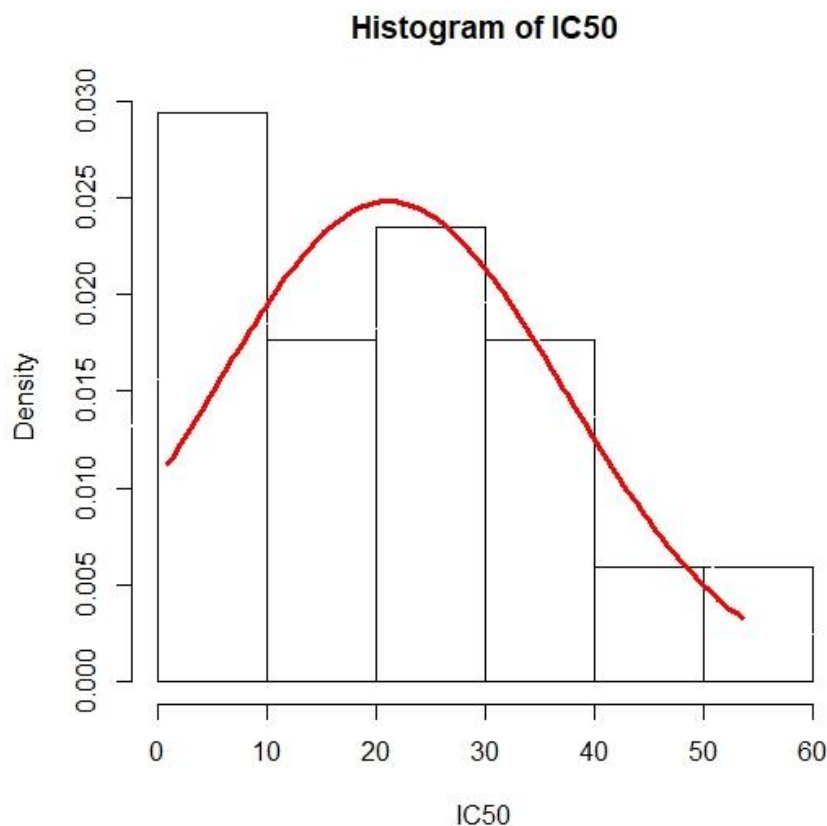


Figura 6: Histograma de la constante de inhibición 50

Por otra parte, en un estudio preliminar se busca analizar con una matriz de correlación la relación que hay entre la constante de inhibición 50 la cual es nuestra variable dependiente (VD) y los descriptores moleculares que son nuestras variables independientes (VID). Se conoce que una buena correlación tiene valores de 1 a -1, siendo así valores cercanos a 0 indicadores de una nula relación entre las variables. Conociendo que evalúa nuestra matriz de correlación a continuación se muestra la figura 7 obtenida de software R.

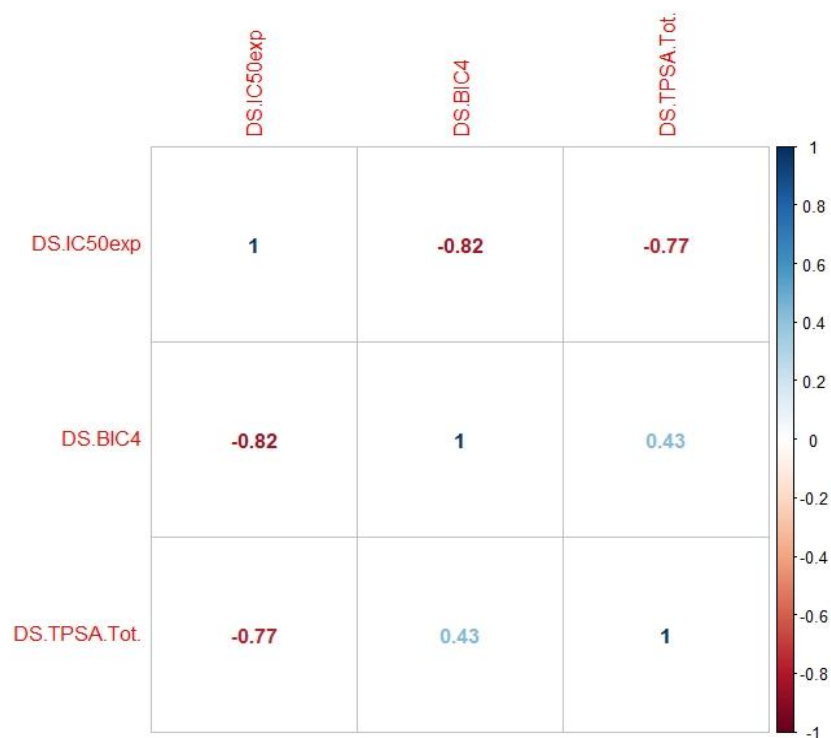
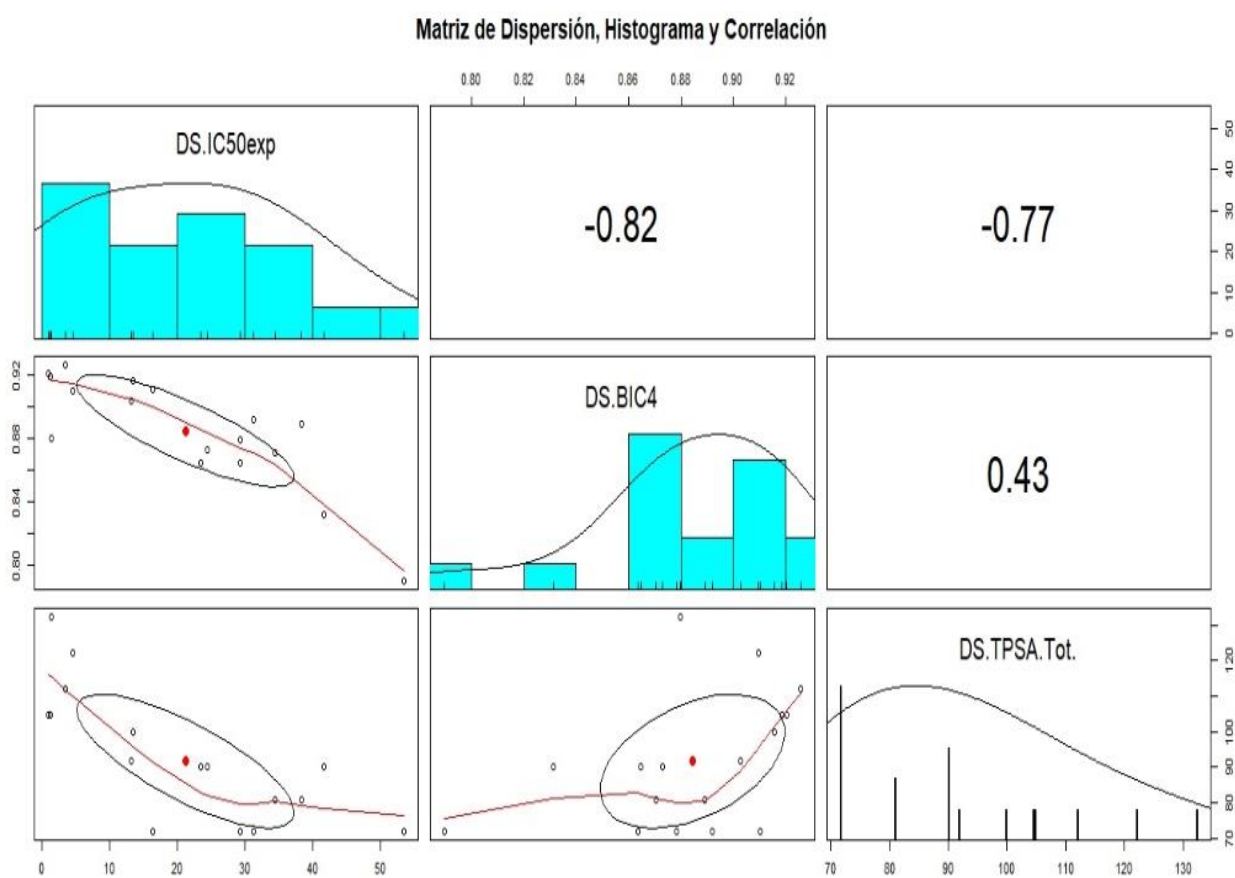


Figura 7: Matriz de correlación en base a colores que en rojo fuerte indican valores cercanos a -1 y en azul fuerte valores de 1 en los tonos de color menos intensos son aquellos que guardan poca correlación es decir cercanos a 0.

El gráfico de matriz de correlación nos da la relación que hay entre cada una de nuestras variables independientes con la variable dependiente, esto indicada por valores cercanos a -1 y 1 con colores rojo fuerte para relación con valor negativo y azul rey para relación con numero positivo, observando nula correlación con el valor 0 o colores rojos o azules más claros. Con esto se sabe que nuestros descriptores tienen relación con IC50, esto es porque BIC4 tiene un valor de -0.82 y TPSA de -0.77, por otra parte, en el presente estudio también se busca que entre variables independientes no tengan relación ya que nos podría estar indicando que ambos descriptores dan la misma información molecular lo cual no se necesita. Siendo no el caso de nuestros dos descriptores ya que su correlación es de 0.43 la cual indica baja relación entre ellos. Desde este primer análisis podemos observar que ambos descriptores son adecuados para nuestro estudio QSAR.

En otro gráfico mostrado en la figura 8 se realizó una matriz de dispersión, histograma y correlación. En esta figura se muestran tres tipos de gráficos que muestran la relación de nuestras variables independientes con nuestra variable dependiente vista desde tres tipos de gráficos en uno solo. Los coeficientes de correlación como se mencionó anteriormente muestran una buena relación de BIC4 y TPSA(Tot) con IC50. El histograma muestra la frecuencia de cada En IC50 hay una mayor frecuencia en una concentración de 0 a 10  $\mu\text{M}$ , en BIC4 se observa una mayor frecuencia está en datos con valores de 0.85 a 0.87 y en TPSA(Tot) la frecuencia está en un rango de 71 a 72. Los gráficos de dispersión muestran la relación que hay entre la IC50 contra cada descriptor, siendo cada punto un derivado de 3-formil cumarina.

*Figura 8: Matriz de dispersión, histograma y correlación obtenida de software R.*



## 8.6 Construcción de los modelos de regresión lineal simple

En esta parte se presentan los modelos de regresión lineal simple (SLR), para cada descriptor contra la actividad biológica (VD). La figura 9 y 10 se puede observar la gráfica de regresión lineal simple obtenida para ambas ecuaciones.

En el gráfico de regresión lineal simple de IC50exp vs BIC4 se observa que la constante de inhibición 50 la cual se encuentra en el eje de las ordenadas disminuye cuando el valor de la variable independiente aumenta, por lo tanto, cuando el valor del bonding information content (BIC) aumenta, la actividad inhibitoria de la molécula disminuye por lo tanto habrá una mejor inhibición de la enzima TP y de igual manera el descriptor TPSA(Tot) se ve una tendencia negativa por lo tanto cuando aumenta el valor de la variable independiente como resultado el valor de IC50exp disminuye sucediendo el mismo comportamiento que con BIC4.

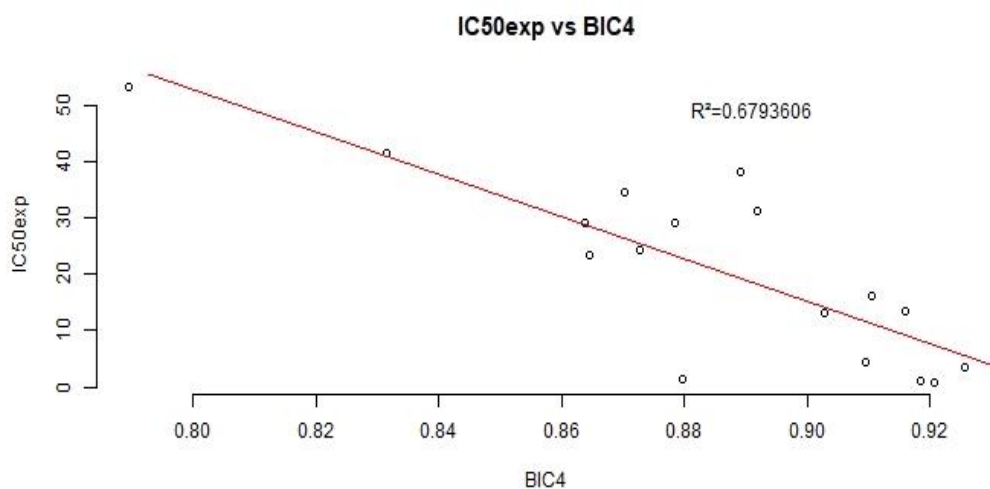
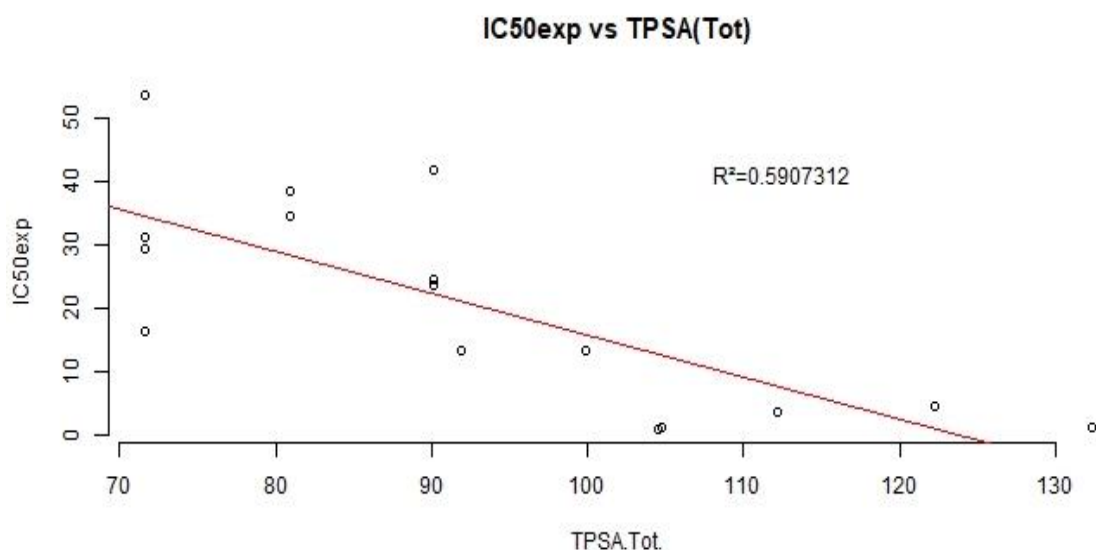


Figura 9: Regresión lineal simple (SLR) con un coeficiente de determinación  $R^2=0.6793606$  obtenida de R software.



*Figura 10: Regresión lineal simple (SLR) con un coeficiente de determinación  $R^2=0.6793606$  obtenida de R software.*

La información anterior nos da noción de cómo estas variables pueden influir en la actividad inhibitoria de cada una de las moléculas análogas de 3-Formil cumarina.

## **8.7 Análisis estadístico del modelo mediante regresión lineal múltiple (MLR)**

Una vez analizando la importancia de los descriptores para nuestro análisis QSAR se procede al análisis de regresión lineal múltiple para así saber cómo los descriptores pueden predecir la actividad biológica de acuerdo con su importancia, por lo tanto, con este análisis se obtendrán los valores del intercepto y las pendientes

A continuación, se muestran los resultados resumidos del análisis del modelo QSAR presentados en la tabla 11.

Tabla 11: Resumen de los resultados obtenidos del modelo 1 de regresión lineal múltiple (RLM)

<b>Min</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max</b>
-6.368	-3.812	-2.177	3.47	13.737
	<b>Estimate</b>	<b>std.Error</b>	<b>t Value</b>	<b>pr(&gt; t )</b>
<b>(Intercept)</b>	304.73773	37.25656	8.179	1.06E-06
<b>BIC4</b>	-275.73141	45.13938	-6.108	2.70E-05
<b>TPSA(Tot)</b>	-0.43316	0.08462	-5.119	0.000156
<b>Residual std error</b>	5.736 on 14 degrees of freedom			
<b>R2</b>	<b>R2-ajustada</b>	<b>F-static</b>	<b>p-value</b>	
0.8883	0.8724	55.69	2.165E-07	

Se puede observar una  $R^2=0.8883$  y una  $R$ -ajustada = 0.8724, la cual es un valor bueno. Con este análisis podemos ver que es un modelo aceptable ya que para parámetros biológicos un  $R^2=0.7$  es considerado aceptable.

A continuación, en la figura 11 se muestran los gráficos que validan nuestro modelo de regresión lineal múltiple, el cual se explica detalladamente.

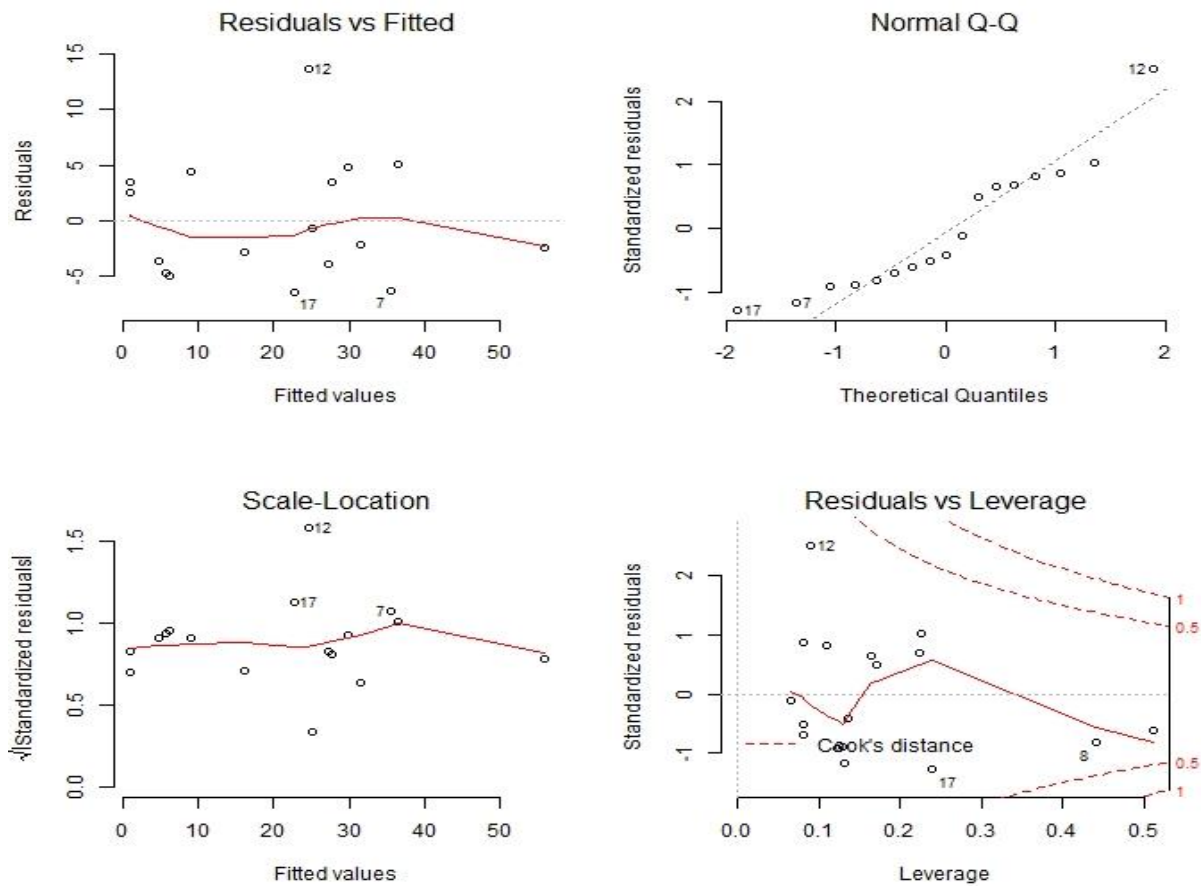


Figura 11: Análisis estadístico del modelo de Regresión lineal múltiple obtenido de R software.

En la figura 11 se muestra el gráfico de Residuales vs valores ajustados (Residuals vs fitted values) en el cual se puede observar si nuestros datos tienen patrones lineales o no lineales. En nuestro caso podemos observar que los residuales guardan una buena distribución ya que la línea roja se mantiene en el centro y no forma una parábola.

Por otra parte, en el gráfico de cuantiles Q-Q, el cual nos indica los cuartiles o percentiles de una distribución, se observa una distribución normal ya que esta función de R toma los datos ordenándolos en forma ascendente y los compara con datos calculados de una distribución normal (theoretical quantiles). Una vez sabido esto vemos que nuestros datos tienen una distribución normal, ya que los puntos

caen cerca de la línea recta y no observamos a los puntos formando colas curvadas o puntos muy dispersos.

El gráfico de escala locación (Scale-Location) nos muestra la distribución de los residuos. Observando una buena distribución de nuestros datos, ya que la línea roja es horizontal (no diagonal, ni mucho menos inclinada) y nuestros puntos (datos) están dispersos aleatoriamente sin alejarse mucho de la línea a lo largo del eje X.

Por último, el gráfico de residuales vs apalancamiento (Residuals vs Leverage) nos indica que algunos valores atípicos pueden estar o no influyendo en nuestros resultados de la regresión, esto se puede observar por las líneas de *cook's distance*, teniendo que los datos extremos que estén cerca de las líneas largas están influyendo en nuestra regresión lineal y si removemos esos datos puede afectar el resultado de nuestro estudio.

Una vez explicado lo anterior, en el gráfico de Residuals vs Leverage , el dato número 12 es bastante influyente en nuestro análisis y omitir dicho valor podría alterar nuestro coeficiente de determinación ( $R^2$ ). Es por ello que nuestros datos se deben mantener intactos y no manipularlos para obtener resultados favorables, los cuales no serían verídicos ni tampoco representativos.

## **8.8 Predicción de la actividad biológica a partir de nuestra ecuación del modelo obtenido**

Por último, se calculó la constante de inhibición 50, utilizando la ecuación 4 obtenida a partir de regresión lineal múltiple que se escribe a continuación, los resultados se muestran en la tabla 12:

$$IC50 = 304.73773 + (-275.73141) BIC4 + (-0.43316) TPSA(Tot) \quad (4)$$

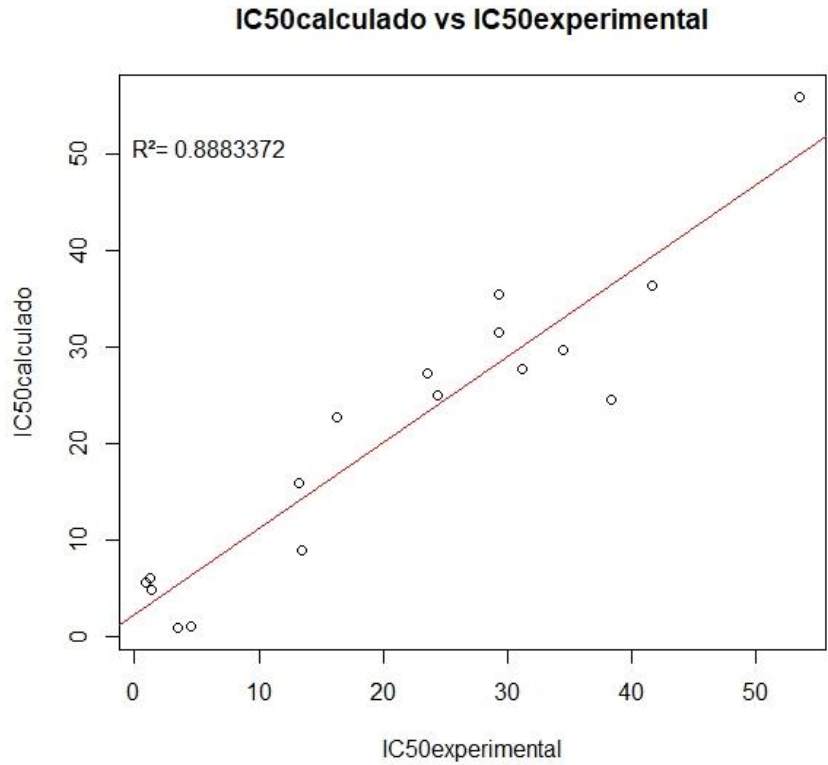
*Tabla 12: Resultados del cálculo de la IC50 a partir del modelo 8 elegido como más representativo.*

ID	Nombre	IC50exp	IC50calculada	Resta de la predicción del modelo.
1	114_L01_Fit16	13.4	8.93182	-4.46818
2	112_L02_Fit16	29.3	31.47711	2.17711
3	110_L03_Fit16	23.5	27.31199	3.81199
4	108_L04_Fit16	3.5	0.92663	-2.57337
5	106_L05_Fit16	1.2	6.10973	4.90973
6	104_L06_Fit16	4.5	1.02980	-3.47020
7	102_L07_Fit16	29.3	35.52186	6.22186
8	100_L08_Fit16	1.3	4.84754	3.54754
9	098_L09_Fit16	24.4	25.03044	0.63044
10	096_L10_Fit16	31.2	27.77017	-3.42983
11	094_L11_Fit16	34.5	29.70530	-4.79470
12	092_L12_FitL16	38.3	24.56265	-13.73735
13	090_L13_FitL16	41.6	36.43819	-5.16181
14	088_L14_FitL16	53.5	55.94658	2.44658
15	086_L15_FitL16	13.2	15.98440	2.78440
16	082_L16_VegaFit	0.9	5.63986	4.73986
17	084_L17_FitL16	16.3	22.66785	6.36785

Tabla 13: Tabla con los valores de IC50calculado vs IC50experimental.

IC50calculado	IC50experimental
8.93182	13.4
31.47711	29.3
27.31199	23.5
0.92663	3.5
6.10973	1.2
1.0298	4.5
35.52186	29.3
4.84754	1.3
25.03044	24.4
27.77017	31.2
29.7053	34.5
24.56265	38.3
36.43819	41.6
55.94658	53.5
15.9844	13.2
5.63986	0.9
22.66785	16.3

Teniendo los datos de IC50experimental se realizó un análisis de regresión lineal simple con los datos de la tabla 13 obteniendo el grafico observado en la figura 12 en el cual se puede observar tanto los datos del análisis como el de determinación de 0.8883372. Aquí no solamente podemos ver la linealidad de nuestros datos, sino que también la buena distribución de los mismos.



*Figura 12: Grafico obtenido de R software del análisis de regresión lineal simple entre IC50calculada VS IC50experimental*

Por último, en la sección de anexos puede verse una tabla con los valores de IC50 calculados con cada modelo y con ellos gráficos que validan cada regresión lineal múltiple.

## 9. Conclusión

Los modelos QSAR obtenidos y analizados estadísticamente mediante el uso del programa especializado R, fueron los más representativos. A partir del modelo 8 se determinó que los descriptores moleculares BIC4 (bond index content) y TPSA(Tot) son cuantitativamente los mejores para relacionar la estructura de los ligantes con su correspondiente actividad inhibitoria 50. Prediciendo su actividad y dando una noción de cómo un estudio cuantitativo puede mejorar la visualización de la relación que existe entre la estructura de los ligantes y su respectiva interacción de su diana molecular. Corroborando la importancia de los estudios QSAR como una herramienta computacional importante para el desarrollo de nuevos fármacos. Por lo tanto, utilizando estos hallazgos, se puede prever la realización de un estudio más profundo de las 17 moléculas derivadas de cumarina, en las cuales sea posible optimizar su estructura o ver a detalle su interacción con la enzima timidina fosforilasa. De esta manera tener un entendimiento del posible mecanismo de acción por el cual estos compuestos derivados de las cumarinas funcionan como buenos inhibidores de la proteína timidina fosforilasa y de porqué serían una buena alternativa terapéutica para el tratamiento de tumores localizados en distintas partes del cuerpo humano.

## Bibliografía

1. Richard Sever, J. S. (2015). Signal Transduction in Cancer. *Cold Spring Harb Perspect Med*, 1-21.
2. Organization, W. H. (2008). *The Global Burden of Disease*. Geneva: World Health Organization.
3. Kim, S.-K. (2015). *Handbook of Anticancer*. Switzerland: Springer International Publishing Switzerland.
4. Jacques Ferlay, H.-R. S. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, 2893-2917.
5. Salud, S. d. (8 de Agosto de 2017). *Gobierno de Mexico* . Obtenido de Secretaria de Salud: <https://www.gob.mx/salud/prensa/318-cancer-tercera-causa-de-muerte-en-mexico>
6. Yasir Y. Elamin, s. R. (2016). Thymidine Phosphorylase in Cancer; Enemy or Friend? *Cancer Microenvironment*, 33-43.
7. Sabrina Riedl, D. Z. (2011). Membrane-active host defense peptides – Challenges and perspectives for the development of novel anticancer drugs. *Chemistry and physics of lipids*, 766–781.
8. Annelies Bronckaers, F. G. (2009). The Dual Role of Thymidine Phosphorylase in Cancer Development and Chemotherapy. *Medicinal Research Reviews*, Vol. 29, No. 6, 903--953.
9. Harris, D. S. (1995). Platelet-derived endothelial cell growth factor/thymidine phosphorylase expression in normal tissues: An immunohistochemical study. *The Journal of Pathology* , 183-190.
10. Judah Folkman, M. (1971). Tumor Angiogenesis: Therapeutic Implications. *The New England Journal of Medicine* , 1182-1186.

11. Razzell, W. E. (1964). Substrate specificity and induction of thymidine phosphorylase in *Escherichia coli*. *the Journal of Biological chemistry*, 1789–1793.
12. Musiliyu A. Musa, J. S. (2008). A Review of Coumarin Derivatives in Pharmacotherapy of Breast. *Curr Med Chem*, 2664–2679.
13. I. Kostova, I. S. (2006). Structure-Activity Relationships of Synthetic Coumarins as HIV-1 Inhibitors. *Bioinorganic Chemistry and Applications*, 1-9.
14. M. E. Marshall, J. L. (1994). An updated review of the clinical development of coumarin (1,2-benzopyrone) and 7-hydroxycoumarin. *Journal of Cancer Research Clinical Oncology*, 39-42.
15. Saeed Emami, S. D. (2015). Current developments of coumarin-based anti-cancer agents in medicinal chemistry. *European Journal of Medicinal Chemistry*, 611-630.
16. Tamer Nasr, S. B. (2014). Anticancer activity of new coumarin substituted hydrazide-hydrazone derivatives. *European Journal of Medicinal Chemistry*, 539-548.
17. Thomas Scior, E. M. (2014). Los modelos in Silico, una herramienta para el conocimiento farmacológico. *Elementos: Ciencia y Cultura*, 45-48.
18. José L. Medina Franco, E. F. (2015). Avances en el diseño de fármacos asistido por computadora. *Educación Química*, 180-186.
19. Fernanda Saldívar Gonzáles, F. D. (2017). Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación química*, 51-58.
20. Shu-Feng Zhou, W.-Z. Z. (2017). Drug Design and Discovery: Principles and Applications. *molecules*, 279.
21. Novartis. (s.f.). *Novartis*. Obtenido de Novartis de España : <https://www.novartis.es/innovacion/proceso-de-desarrollo-de-farmacos>

22. Cancerques. (06 de 6 de 2020). *Cancerquest*. Obtenido de Cancerques: <https://www.cancerquest.org/es/para-los-pacientes/descubrimiento-y-desarrollo-de-medicamentos>
23. Patel, N. (2006). Predictive ADME and Toxicology Strategies . *Business insights Ltd* .
24. Mercedes de Miguel, C. G. (2012). *Guía de desarrollo preclínicos*. España: Fundacion para el desarrollo de la investigación en genómica y proteómica .
25. Avedaño, C. (2001). *Introducción a la química farmacéutica*. Aravaca Madrid: McGRAW-HILL/INTERAMERICANA DE ESPAÑA.
26. Gad, S. C. (2005). *Drug discovery Handbook*. New Jersey: Wiley & Sons.
27. T. Scior, J. M.-F.-T.-M. (2009). How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review. *Current Medicinal Chemistry*, 4297-4313.
28. Thomas Scior, J. L. (2012). ¿Qué sabe Ud. acerca de...QSAR? What do you know about QSAR? *Rev Mex Cienc Farm* 43 (2) 2012, 82-84.
29. Kunal Roy, S. K. (2015). *A primer on QSAR/QSPR Modeling Fundamental Concepts* . Springer.
30. Maria José Rodríguez Jaume, R. M. (2001). *Estadística informático: Casos y ejemplos con el SPSS*. España: Publicaciones Universidades alicante.
31. Rodrigo, J. A. (01 de Julio de 2016). *Rpubs by RStudio*. Obtenido de Rpubs by RStudio: [https://rpubs.com/Joaquin\\_AR/226291](https://rpubs.com/Joaquin_AR/226291)
32. Matthias Dehmer, D. B. (2012). *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. Weinheim : WILEY.
33. Muhammad Taha, S. A. (2018). Synthesis, molecular docking study and thymidine phosphorylase inhibitory activity of 3-formylcoumarin derivatives. *Bioorganic Chemistry*, 17–23.

34. Hayat Ullah, F. R. (2018). Synthesis, Molecular docking study and in vitro Thymidine Phosphorylase Inhibitory Potential of Oxadiazole Derivatives. *Bioorganic Chemistry*, 58-67.
35. V.I. Timofeev, Y. A. (2013). Three- Dimensional Structure of Thymidine Phosphorylase from E. coli in complex with 3'- Azido-2'-Fluoro-2',3'- Dideoxyuridine. *Crystallography Reports*, 842–853.
36. Vladimir Timofeev, Y. A. (2014). Acta Crystallographica Section D Biological Crystallography, 1155–1165.
37. Matthew J. Pugmire, W. J. (1998). Structural and Theoretical Studies Suggest Domain Movement Produces an Active Conformation of Thymidine Phosphorylase. *JB*, 285±299.
38. Matthew J. Pugmire, S. E. (1998). The crystal structure of pyrimidine nucleoside phosphorylase in a closed conformation. *Structure*, 1467-1479.
39. Vladislav V. Balaev, A. A. (2016). Structural investigation of the thymidine phosphorylase from Salmonella typhimurium in the unliganded state and its complexes with thymidine and uridine. *Structural Biology Communications* , 224–233.
40. Richard A. Norman, S. T. (2004). Crystal Structure of Human Thymidine Phosphorylase in Complex with a Small Molecule Inhibitor. *Structure* , 75–84.
41. Roberto Todeschini, V. C. (2000). *Handbook of Molecular Descriptors*. Weinheim: WILEY-VCH.
42. Consonni, R. T. (2009). *topological lipophilicity potential*. WILEY-VCH.

## Anexos

Tabla 14: Descriptores calculados en Vega ZZ (1/2)

ID	Nombre	IC50	Dipole	Surf Area Å <sup>2</sup>	PSA Å <sup>2</sup>	Volume Å <sup>3</sup>	Ovality
1	114_L01_Fit16	13.4	5.9835	546.1	143.1	275.1	1.5706
2	112_L02_Fit16	29.3	6.4460	619.6	113.3	310.8	1.6727
3	110_L03_Fit16	23.5	5.1059	676.9	135.8	347.0	1.7323
4	108_L04_Fit16	3.5	7.3577	614.6	199.3	311.3	1.6704
5	106_L05_Fit16	1.2	2.9622	670.9	169.4	351.3	1.7104
6	104_L06_Fit16	4.5	5.1896	639.3	217.7	318.0	1.6870
7	102_L07_Fit16	29.3	6.5055	587.0	113.3	293.6	1.6394
8	100_L08_Fit16	1.3	4.6164	635.6	271.7	321.6	1.7114
9	098_L09_Fit16	24.4	5.2472	672.60	133.20	346.90	1.7338
10	096_L10_Fit16	31.2	6.6299	635.2	100.1	325.9	1.6816
11	094_L11_Fit16	34.5	4.8845	636.1	134.4	321.1	1.6895
12	092_L12_FitL16	38.3	7.0057	623.6	112.0	319.4	1.6865
13	090_L13_FitL16	41.6	4.8596	688.4	156.9	348.3	1.7545
14	088_L14_FitL16	53.5	6.5414	684.1	113.3	361.6	1.7585
15	086_L15_FitL16	13.2	6.4152	598.1	147.5	301.8	1.6517
16	082_L16_VegaFit	0.9	6.5076	671.5	194.2	347.1	1.7163
17	084_L17_FitL16	16.3	5.5861	595.6	102.8	304.9	1.6391

Tabla 15: Descriptores calculados en Vega ZZ (2/2)

ID	Nombre	Log P Broto	Lipole Broto	Virtual logP	Davies' HLB	Griffin's HLB	PSA-based HLB
1	114_L01_Fit16	1.7810	1.5495	1.7810	7.37	5.45	5.94
2	112_L02_Fit16	1.6790	2.0284	1.6790	6.43	4.00	3.80
3	110_L03_Fit16	1.0540	0.8786	1.0540	8.55	4.49	5.64
4	108_L04_Fit16	0.0220	2.3546	0.0220	10.70	6.34	6.19
5	106_L05_Fit16	0.2940	1.0713	0.2940	8.80	5.23	5.26
6	104_L06_Fit16	2.4040	2.1825	2.4040	7.72	6.31	6.75
7	102_L07_Fit16	0.7720	1.4202	0.7720	5.95	4.24	4.71
8	100_L08_Fit16	-0.3530	2.8183	-0.3530	13.08	7.57	6.83
9	098_L09_Fit16	1.0540	1.3417	1.0540	8.55	4.41	5.64
10	096_L10_Fit16	1.6020	1.7558	1.6020	5.95	3.59	4.34
11	094_L11_Fit16	0.9130	1.1843	0.9130	7.25	4.46	5.21
12	092_L12_FitL16	0.9130	1.5568	0.9130	7.25	4.21	5.21
13	090_L13_FitL16	1.0540	1.1060	1.0540	8.55	4.63	5.64
14	088_L14_FitL16	2.6330	3.0788	2.6330	5.00	3.44	4.02
15	086_L15_FitL16	0.3970	1.4579	0.3970	8.32	5.20	5.49
16	082_L16_VegaFit	-0.3950	3.0180	-0.3950	5.95	5.41	5.49
17	084_L17_FitL16	1.4290	1.6463	1.4290	6.43	3.91	4.26

Tabla 16: Cálculo de descriptores en SwissADME parte (1/5)

ID	Nombre	IC50	Heavy Atoms	Rotatable Bonds	H-bond Acceptors	H-bond Donors
1	114_L01_Fit16	13.4	23	4	4	1
2	112_L02_Fit16	29.3	24	4	4	1
3	110_L03_Fit16	23.5	27	6	6	1
4	108_L04_Fit16	3.5	25	4	6	3
5	106_L05_Fit16	1.2	29	4	7	2
6	104_L06_Fit16	4.5	29	5	7	1
7	102_L07_Fit16	29.3	23	4	4	1
8	100_L08_Fit16	1.3	26	4	7	4
9	098_L09_Fit16	24.4	27	6	6	1
10	096_L10_Fit16	31.2	25	4	4	1
11	094_L11_Fit16	34.5	25	5	5	1
12	092_L12_FitL16	38.3	25	5	5	1
13	090_L13_FitL16	41.6	27	6	6	1
14	088_L14_FitL16	53.5	27	5	4	1
15	086_L15_FitL16	13.2	24	4	5	2
16	082_L16_VegaFit	0.9	28	4	5	2
17	084_L17_FitL16	16.3	24	4	4	1

Tabla 17: Cálculo de descriptores en SwissADME parte (2/5)

ID	Nombre	MR	TPSA	iLOGP	XLOGP3	WLOGP	MLOGP
1	114_L01_Fit16	91.55	99.91	2.42	4.02	3.93	2.93
2	112_L02_Fit16	96.36	71.67	2.04	4.07	3.97	3.48
3	110_L03_Fit16	101.65	90.13	2.17	3.32	3.23	2.25
4	108_L04_Fit16	92.71	112.13	1.55	3.22	2.62	1.79
5	106_L05_Fit16	105.94	104.79	2.10	3.94	4.02	2.68
6	104_L06_Fit16	107.65	134.17	2.05	3.25	3.76	2.64
7	102_L07_Fit16	88.66	71.67	2.24	3.38	3.21	2.87
8	100_L08_Fit16	94.73	132.36	1.14	2.31	2.33	1.27
9	098_L09_Fit16	101.65	90.13	2.02	3.32	3.23	2.25
10	096_L10_Fit16	98.59	71.67	2.21	4.10	3.83	3.33
11	094_L11_Fit16	95.15	80.90	1.93	3.35	3.22	2.55
12	092_L12_FitL16	95.15	80.90	1.66	3.35	3.22	2.55
13	090_L13_FitL16	101.65	90.13	2.28	3.32	3.23	2.25
14	088_L14_FitL16	107.93	71.67	3.00	5.05	4.51	3.77
15	086_L15_FitL16	90.69	91.90	1.67	3.57	2.92	2.32
16	082_L16_VegaFit	106.79	104.53	2.21	2.79	3.05	2.77
17	084_L17_FitL16	93.67	71.67	2.42	4.00	3.86	3.37

Tabla 18: Cálculo de descriptores en SwissADME parte (3/5)

ID	Nombre	Silicos-IT LogP	Consensus LogP	ESOL LogS	ESOL Sol-mg	ESOL Sol-mol
1	114_L01_Fit16	5.38	3.74	-4.87	4.98e-0.4	1.51e-06
2	112_L02_Fit16	4.78	3.67	-5.15	2.88e-03	7.11e-06
3	110_L03_Fit16	4.22	3.04	-4.37	1.64e-0.2	4.24e-05
4	108_L04_Fit16	3.14	2.46	-4.30	1.79e-02	4.98e-05
5	106_L05_Fit16	4.51	3.45	-5.12	3.11e-03	7.56e-06
6	104_L06_Fit16	2.21	2.78	-4.60	1.05e-02	2.54e-05
7	102_L07_Fit16	4.10	3.16	-4.25	1.85e-02	5.68e-05
8	100_L08_Fit16	2.67	1.94	-3.81	5.80e-02	1.55e-04
9	098_L09_Fit16	4.22	3.01	-4.37	1.64e-02	4.24e-05
10	096_L10_Fit16	5.13	3.72	-4.83	5.22e-03	1.47e-05
11	094_L11_Fit16	4.15	3.04	-4.31	1.76e-02	4.94e-05
12	092_L12_FitL16	4.15	2.99	-4.31	1.76e-02	4.94e-05
13	090_L13_FitL16	4.22	3.06	-4.89	4.99e-05	1.29e-05
14	088_L14_FitL16	5.46	4.36	-5.50	1.20e-03	3.14e-06
15	086_L15_FitL16	3.62	2.82	-4.44	1.23e-02	3.60e-05
16	082_L16_VegaFit	4.71	3.11	-4.30	1.96e-02	4.97e-05
17	084_L17_FitL16	4.74	3.68	-4.83	5.36e-03	1.48e-05

Tabla 19: Cálculo de descriptores en SwissADME parte (4/5)

ID	Nombre	Ali-LogS	Ali-Sol mg	Ali-Sol mol	Silicos-IT-LogSw
1	114_L01_Fit16	-5.82	5.55e-04	1.51e-06	-6.67
2	112_L02_Fit16	-5.28	2.13e-03	5.25e-06	-7.61
3	110_L03_Fit16	-4.89	4.99e-03	1.29e-05	-7.03
4	108_L04_Fit16	-5.25	2.03e-03	5.66e-06	-5.64
5	106_L05_Fit16	-5.84	5.95e-04	1.44e-06	-7.76
6	104_L06_Fit16	-5.74	7.47e-04	1.81e-06	-5.82
7	102_L07_Fit16	-4.56	8.92e-03	2.73e-05	-6.81
8	100_L08_Fit16	-4.73	7.01e-03	1.87e-05	-5.06
9	098_L09_Fit16	-4.89	4.99e-03	1.29e-05	-7.03
10	096_L10_Fit16	-5.31	1.73e-03	4.89e-08	-7.57
11	094_L11_Fit16	-4.73	6.70e-03	1.88e-05	-6.92
12	092_L12_FitL16	-4.73	6.70e-03	1.88e-05	-6.92
13	090_L13_FitL16	-7.03	3.64e-05	9.42e-08	-7.03
14	088_L14_FitL16	-6.30	1.93e-04	5.05e-07	-7.98
15	086_L15_FitL16	-5.19	2.23e-03	6.52e-06	-6.23
16	082_L16_VegaFit	-4.64	8.99e-03	2.28e-5	-7.98
17	084_L17_FitL16	-5.21	2.24e-03	6.21e-06	-7.41

Tabla 20: Cálculo de descriptores en SwissADME parte (5/5)

ID	Nombre	Silicos-IT Sol-mg	Silicos-IT Sol-mol	log Kp-cm	Syn Access
1	114_L01_Fit16	7.81e-05	2.13e-070	-5.69	3.32
2	112_L02_Fit16	1.00e-05	2.47e-08	-5.88	3.19
3	110_L03_Fit16	3.64e-05	9.42e-08	-6.30	3.35
4	108_L04_Fit16	8.16e-04	2.28e-06	-6.20	3.19
5	106_L05_Fit16	7.21e-06	1.75e-08	-6.01	3.16
6	104_L06_Fit16	6.17e-04	1.50e-06	-6.50	3.50
7	102_L07_Fit16	5.03e-05	1.54e-07	-5.89	3.19
8	100_L08_Fit16	3.29e-03	8.79e-06	-6.95	3.22
9	098_L09_Fit16	3.64e-05	9.42e-08	-6.30	3.37
10	096_L10_Fit16	9.55e-06	2.59e-08	-5.55	3.38
11	094_L11_Fit16	4.20e-05	1.20e-07	-6.10	3.18
12	092_L12_FitL16	4.28e05	1.20e-07	-6.10	3.21
13	090_L13_FitL16	3.64e-05	9.42e-08	-6.30	3.37
14	088_L14_FitL16	3.97e-06	1.04e-08	-5.05	3.56
15	086_L15_FitL16	2.03e-04	5.92e-07	-5.86	3.15
16	082_L16_VegaFit	4.16e-06	1.06e-08	-6.72	3.25
17	084_L17_FitL16	1.41e-05	3.92e-08	-5.66	3.22

Tabla 21: Cálculo de descriptores en AlvaDesc 1.0 parte (1/4)

ID	Nombre	IC50	IC5	R4p	R3s	CATS3D _05_AL	CATS3D _06_AL
1	114_L01_Fit16	13.4	4.95419631	0.33161988	4.27909397	6	4
2	112_L02_Fit16	29.3	4.85216872	0.3379083	3.71870126	8	6
3	110_L03_Fit16	23.5	5.02303707	0.38663108	3.34527712	8	5
4	108_L04_Fit16	3.5	5.169925	0.31430355	4.75422947	8	5
5	106_L05_Fit16	1.2	5.28540222	0.35005002	5.279094	7	4
6	104_L06_Fit16	4.5	5.23412017	0.38369892	4.95975961	9	8
7	102_L07_Fit16	29.3	4.85216872	0.3093756	3.5399986	9	7
8	100_L08_Fit16	1.3	4.9391831	0.31711284	5.88695099	6	5
9	098_L09_Fit16	24.4	5.16589421	0.38987477	3.24026046	9	6
10	096_L10_Fit16	31.2	5.08418372	0.45472699	3.77803658	9	8
11	094_L11_Fit16	34.5	4.91227258	0.39691378	3.68488685	8	6
12	092_L12_FitL16	38.3	5.1227989	0.36507681	3.31675829	10	7
13	090_L13_FitL16	41.6	4.78494183	0.34562666	2.88350252	7	6
14	088_L14_FitL16	53.5	4.62607908	0.61726137	4.64935698	9	7
15	086_L15_FitL16	13.2	5.12928302	0.31582836	3.87567039	10	7
16	082_L16_VegaFit	0.9	5.3219281	0.36281331	4.20087819	10	8
17	084_L17_FitL16	16.3	5.08746284	0.33837873	3.80386202	9	7

Tabla 22: Cálculo de descriptores en AlvaDesc 1.0 parte (2/4)

ID	Nombre	CATS3D _09_AL	CATS3D_10_A L	CATS3D _08_DL	CATS3D _11_DL	BIC4
1	114_L01_Fit16	3	1	2	0	0.91585104
2	112_L02_Fit16	3	5	1	0	0.87844923
3	110_L03_Fit16	3	5	1	0	0.86455522
4	108_L04_Fit16	3	5	1	1	0.92568661
5	106_L05_Fit16	7	6	3	1	0.91841973
6	104_L06_Fit16	6	4	2	0	0.90950869
7	102_L07_Fit16	3	4	1	0	0.86378005
8	100_L08_Fit16	4	4	2	2	0.87968626
9	098_L09_Fit16	4	7	1	0	0.87282976
10	096_L10_Fit16	3	5	1	0	0.89189325
11	094_L11_Fit16	3	5	1	0	0.87037522
12	092_L12_FitL16	4	6	1	0	0.88902615
13	090_L13_FitL16	4	3	1	0	0.83145706
14	088_L14_FitL16	4	5	2	0	0.78970537
15	086_L15_FitL16	4	6	2	1	0.90285662
16	082_L16_VegaFit	5	5	2	1	0.92053226
17	084_L17_FitL16	3	4	1	0	0.91039792

Tabla 23: Cálculo de descriptores en AlvaDesc 1.0 parte (3/4)

ID	Nombre	TPSA(Tot)	SpMax3_Bh(s)	J_Dz(p)	VE1RG	SPH
1	114_L01_Fit16	99.91	4.961021351	1.348187	2.35186201	1
2	112_L02_Fit16	71.67	4.958472488	1.325037	3.54480985	1
3	110_L03_Fit16	90.13	4.964963425	1.267049	3.05033112	0.99393603
4	108_L04_Fit16	112.13	6.528300595	1.318362	3.83798737	1
5	106_L05_Fit16	104.79	7.428338089	1.1595	3.92816051	1
6	104_L06_Fit16	122.19	7.428471841	1.075002	2.46705883	1
7	102_L07_Fit16	71.67	4.958500968	1.325401	4.03564405	1
8	100_L08_Fit16	132.36	6.562056727	1.309774	3.28750612	1
9	098_L09_Fit16	90.13	4.963691951	1.277768	3.14210658	0.99336572
10	096_L10_Fit16	71.67	4.958584622	1.33976	2.69426041	0.99243229
11	094_L11_Fit16	80.9	4.960086658	1.269716	2.26746827	0.99647184
12	092_L12_FitL16	80.9	4.960148607	1.297053	2.39228892	0.99565555
13	090_L13_FitL16	90.13	4.963800392	1.277261	2.90884761	0.99370272
14	088_L14_FitL16	71.67	4.958861742	1.320938	3.59848532	0.96826397
15	086_L15_FitL16	91.9	6.468470641	1.321152	3.80030662	1
16	082_L16_VegaFit	104.53	7.427508257	1.141586	3.60524171	1
17	084_L17_FitL16	71.67	4.96352134	1.339529	3.61562884	1

Tabla 24: Cálculo de descriptores en AlvaDesc 1.0 parte (4/4)

ID	Nombre	Mor20s	Mor03i	F06 [N-O]	TDB10p	B10 [O-CI]	TDB04m
1	114_L01_Fit16	0.95836763	-3.86898378	0	7.13201935	0	4.613161
2	112_L02_Fit16	0.27427199	-5.34913954	0	7.19473795	0	4.13933
3	110_L03_Fit16	-0.69847901	-7.01287833	2	6.0813889	1	3.357281
4	108_L04_Fit16	0.97833341	-5.87521999	1	5.9990738	1	3.628827
5	106_L05_Fit16	0.82341539	-7.06335035	0	7.02549089	0	4.096194
6	104_L06_Fit16	-1.03769608	-4.9301897	0	6.57475306	0	3.494051
7	102_L07_Fit16	0.12790416	-5.11113704	0	6.27593671	0	3.562987
8	100_L08_Fit16	0.5412859	-4.72348665	3	6.10298543	1	3.587647
9	098_L09_Fit16	-0.71195894	-7.43205414	1	5.89825926	0	3.459545
10	096_L10_Fit16	0.44745172	-6.84848212	0	5.4530031	0	2.939304
11	094_L11_Fit16	-0.22026827	-6.59020624	1	6.24441069	0	3.348088
12	092_L12_FitL16	-0.24778822	-6.7174195	0	5.88530665	0	3.553696
13	090_L13_FitL16	-0.55693062	-6.99292396	2	5.95173403	1	3.41959
14	088_L14_FitL16	0.10191378	-8.0924042	0	6.29284977	0	2.263471
15	086_L15_FitL16	-0.30925516	-5.54515423	0	5.91206952	0	3.636518
16	082_L16_VegaFit	-0.69956611	-6.86203111	1	6.87319327	0	3.35396
17	084_L17_FitL16	-0.02313581	-5.57610791	0	6.70063251	0	4.211925

Tabla 25: Cálculo de descriptores en E-Dragon parte (1/4)

ID	Nombre	IC50	IC5	R4p	R3s	CATS3D _05_AL	CATS3D _06_AL
1	114_L01_Fit16	13.4	4.95419631	0.33161988	4.27909397	6	4
2	112_L02_Fit16	29.3	4.85216872	0.3379083	3.71870126	8	6
3	110_L03_Fit16	23.5	5.02303707	0.38663108	3.34527712	8	5
4	108_L04_Fit16	3.5	5.169925	0.31430355	4.75422947	8	5
5	106_L05_Fit16	1.2	5.28540222	0.35005002	5.279094	7	4
6	104_L06_Fit16	4.5	5.23412017	0.38369892	4.95975961	9	8
7	102_L07_Fit16	29.3	4.85216872	0.3093756	3.5399986	9	7
8	100_L08_Fit16	1.3	4.9391831	0.31711284	5.88695099	6	5
9	098_L09_Fit16	24.4	5.16589421	0.38987477	3.24026046	9	6
10	096_L10_Fit16	31.2	5.08418372	0.45472699	3.77803658	9	8
11	094_L11_Fit16	34.5	4.91227258	0.39691378	3.68488685	8	6
12	092_L12_FitL16	38.3	5.1227989	0.36507681	3.31675829	10	7
13	090_L13_FitL16	41.6	4.78494183	0.34562666	2.88350252	7	6
14	088_L14_FitL16	53.5	4.62607908	0.61726137	4.64935698	9	7
15	086_L15_FitL16	13.2	5.12928302	0.31582836	3.87567039	10	7
16	082_L16_VegaFit	0.9	5.3219281	0.36281331	4.20087819	10	8
17	084_L17_FitL16	16.3	5.08746284	0.33837873	3.80386202	9	7

Tabla 26: Cálculo de descriptores en E-Dragon parte (2/4)

ID	Nombre	CATS3D _09_AL	CATS3D_10_A L	CATS3D _08_DL	CATS3D _11_DL	BIC4
1	114_L01_Fit16	3	1	2	0	0.91585104
2	112_L02_Fit16	3	5	1	0	0.87844923
3	110_L03_Fit16	3	5	1	0	0.86455522
4	108_L04_Fit16	3	5	1	1	0.92568661
5	106_L05_Fit16	7	6	3	1	0.91841973
6	104_L06_Fit16	6	4	2	0	0.90950869
7	102_L07_Fit16	3	4	1	0	0.86378005
8	100_L08_Fit16	4	4	2	2	0.87968626
9	098_L09_Fit16	4	7	1	0	0.87282976
10	096_L10_Fit16	3	5	1	0	0.89189325
11	094_L11_Fit16	3	5	1	0	0.87037522
12	092_L12_FitL16	4	6	1	0	0.88902615
13	090_L13_FitL16	4	3	1	0	0.83145706
14	088_L14_FitL16	4	5	2	0	0.78970537
15	086_L15_FitL16	4	6	2	1	0.90285662
16	082_L16_VegaFit	5	5	2	1	0.92053226
17	084_L17_FitL16	3	4	1	0	0.91039792

Tabla 27: Cálculo de descriptores en E-Dragon parte (3/4)

ID	Nombre	TPSA(Tot)	SpMax3_Bh(s)	J_Dz(p)	VE1RG	SPH
1	114_L01_Fit16	99.91	4.961021351	1.348187	2.35186201	1
2	112_L02_Fit16	71.67	4.958472488	1.325037	3.54480985	1
3	110_L03_Fit16	90.13	4.964963425	1.267049	3.05033112	0.99393603
4	108_L04_Fit16	112.13	6.528300595	1.318362	3.83798737	1
5	106_L05_Fit16	104.79	7.428338089	1.1595	3.92816051	1
6	104_L06_Fit16	122.19	7.428471841	1.075002	2.46705883	1
7	102_L07_Fit16	71.67	4.958500968	1.325401	4.03564405	1
8	100_L08_Fit16	132.36	6.562056727	1.309774	3.28750612	1
9	098_L09_Fit16	90.13	4.963691951	1.277768	3.14210658	0.99336572
10	096_L10_Fit16	71.67	4.958584622	1.33976	2.69426041	0.99243229
11	094_L11_Fit16	80.9	4.960086658	1.269716	2.26746827	0.99647184
12	092_L12_FitL16	80.9	4.960148607	1.297053	2.39228892	0.99565555
13	090_L13_FitL16	90.13	4.963800392	1.277261	2.90884761	0.99370272
14	088_L14_FitL16	71.67	4.958861742	1.320938	3.59848532	0.96826397
15	086_L15_FitL16	91.9	6.468470641	1.321152	3.80030662	1
16	082_L16_VegaFit	104.53	7.427508257	1.141586	3.60524171	1
17	084_L17_FitL16	71.67	4.96352134	1.339529	3.61562884	1

Tabla 28: Cálculo de descriptores en E-Dragon parte (4/4)

ID	Nombre	Mor20s	Mor03i	F06 [N-O]	TDB10p	B10 [O-Cl]	TDB04m
1	114_L01_Fit16	0.95836763	-3.86898378	0	7.13201935	0	4.613161
2	112_L02_Fit16	0.27427199	-5.34913954	0	7.19473795	0	4.13933
3	110_L03_Fit16	-0.69847901	-7.01287833	2	6.0813889	1	3.357281
4	108_L04_Fit16	0.97833341	-5.87521999	1	5.9990738	1	3.628827
5	106_L05_Fit16	0.82341539	-7.06335035	0	7.02549089	0	4.096194
6	104_L06_Fit16	-1.03769608	-4.9301897	0	6.57475306	0	3.494051
7	102_L07_Fit16	0.12790416	-5.11113704	0	6.27593671	0	3.562987
8	100_L08_Fit16	0.5412859	-4.72348665	3	6.10298543	1	3.587647
9	098_L09_Fit16	-0.71195894	-7.43205414	1	5.89825926	0	3.459545
10	096_L10_Fit16	0.44745172	-6.84848212	0	5.4530031	0	2.939304
11	094_L11_Fit16	-0.22026827	-6.59020624	1	6.24441069	0	3.348088
12	092_L12_FitL16	-0.24778822	-6.7174195	0	5.88530665	0	3.553696
13	090_L13_FitL16	-0.55693062	-6.99292396	2	5.95173403	1	3.41959
14	088_L14_FitL16	0.10191378	-8.0924042	0	6.29284977	0	2.263471
15	086_L15_FitL16	-0.30925516	-5.54515423	0	5.91206952	0	3.636518
16	082_L16_VegaFit	-0.69956611	-6.86203111	1	6.87319327	0	3.35396
17	084_L17_FitL16	-0.02313581	-5.57610791	0	6.70063251	0	4.211925

Tabla 29: Modelos QSAR generados

Modelo	Tamaño	R2	Ecuación
1	4	0.8938	DV = 240.2298 + (-43.3127)IC5 + 84.7136R4p + (-9.8641)R3s + 0.9055 CATS3D_05_AL
2	4	0.8907	DV = 231.9315+ (-40.4472)IC5 + 89.7641R4p + (-10.4208)R3s + 0.2939 CATS3D_05_AL
3	4	0.7805	DV = 752.176 + (-61.805)IC5 + (-442.471) SPH + (-7.411)VE1_RG + 5.33 CATS3D_05_AL
4	3	0.7877	DV = 305.446+ ( -64.255)IC5 + (-6.853)Mor03i+ (-4.796)F06[N-O]
5	3	0.5875	DV = 350.959 + ( -55.898)IC5 + (-6.299)TDB10p + (-5.875)CATS3D_08_DL
6	3	0.7713	DV = 309.187+ ( -64.811)IC5 + (-6.551)Mor03i + (-9.296)B10[O-Cl]
7	2	0.6336	DV = 354.1527+ (-376.4974 )BIC4 + (0.2098)Mor20s
8	2	0.8724	DV = 304.73773+ (-275.73141) BIC4 + (-0.43316) TPSA(Tot)
9	2	0.8314	DV = 286.728+ (-250.419)BIC4 + (-7.773)SpMax3_Bh(s)

Tabla 30: Comparación de la IC50 experimental y la IC50 calculada (1/2)

<b>Molécula</b>	<b>Nombre</b>	<b>IC50 experimental</b>	<b>IC50 Calculado modelo 1</b>	<b>IC50 Calculado modelo 2</b>	<b>IC50 Calculado modelo 3</b>	<b>IC50 Calculado modelo 4</b>
1	114_L01_Fit16	13.4	16.96648	17.89971	18.0612477	13.62826
2	112_L02_Fit16	29.3	29.25706	29.01845	26.1861262	30.32755
3	110_L03_Fit16	23.5	29.66726	30.07834	21.9733210	21.15801
4	108_L04_Fit16	3.5	3.27998	2.96231	4.37446094	8.71935
5	106_L05_Fit16	1.2	-4.77625	-4.26308	-8.7608817	14.23762
6	104_L06_Fit16	4.5	5.25638	5.33492	15.8968300	2.91420
7	102_L07_Fit16	29.3	29.50819	28.61337	27.8785539	28.69652
8	100_L08_Fit16	1.3	0.52774	0.74348	12.0550809	6.06084
9	098_L09_Fit16	24.4	25.69592	25.97959	18.0462353	19.64733
10	096_L10_Fit16	31.2	29.42421	30.08970	26.8283552	25.69442
11	094_L11_Fit16	34.5	31.98591	32.23637	33.4988939	30.17461
12	092_L12_FitL16	38.3	25.61268	24.99345	30.5834538	22.31503
13	090_L13_FitL16	41.6	40.15567	41.13386	32.5105662	36.32007
14	088_L14_FitL16	53.5	54.44004	53.83475	59.1340805	63.65454
15	086_L15_FitL16	13.2	15.64666	14.48613	17.8255908	13.86486

16	082_L16_VegaFit	0.9	8.07506	7.81671	7.36478780	5.71501
17	084_L17_FitL16	16.3	19.17115	18.95015	16.4489337	16.76414

*Tabla 31: Comparacion de la IC50 experimental y la IC50 calculada (2/2)*

<b>Molécula</b>	<b>Nombre</b>	<b>IC50 experimental</b>	<b>IC50 Calculado modelo 5</b>	<b>IC50 Calculado modelo 6</b>	<b>IC50 Calculado modelo 7</b>	<b>IC50 Calculado modelo 8</b>	<b>IC50 Calculado modelo 9</b>
1	114_L01_Fit16	13.4	17.35474	13.44630	9.53823	8.93182	18.81948
2	112_L02_Fit16	29.3	28.53782	29.75531	23.47639	31.47711	28.20542
3	110_L03_Fit16	23.5	25.99961	20.28431	28.50337	27.31199	31.63429
4	108_L04_Fit16	3.5	18.30737	3.31156	5.83935	0.92663	4.17400
5	106_L05_Fit16	1.2	-6.36298	12.90680	8.54281	6.10973	-1.00222
6	104_L06_Fit16	4.5	5.21778	2.25611	11.50734	1.02980	1.22823
7	102_L07_Fit16	29.3	34.32535	28.19615	28.96859	35.52186	31.87864
8	100_L08_Fit16	1.3	24.67584	10.72117	23.06667	4.84754	15.43098
9	098_L09_Fit16	24.4	19.16771	23.06762	25.38520	25.03044	29.57207
10	096_L10_Fit16	31.2	26.53983	24.54038	18.45108	27.77017	24.83791
11	094_L11_Fit16	34.5	31.16424	33.99014	26.41248	29.70530	30.21475
12	092_L12_FitL16	38.3	21.65824	21.17910	19.38468	24.56265	25.54373

13	090_L13_FitL16	41.6	40.12535	35.58478	40.99444	36.43819	39.93173
14	088_L14_FitL16	53.5	40.98177	62.37953	56.85206	55.94658	50.42554
15	086_L15_FitL16	13.2	15.25221	13.07934	14.16465	15.98440	10.35613
16	082_L16_VegaFit	0.9	-1.57038	9.22068	7.42793	5.63986	-1.52479
17	084_L17_FitL16	16.3	18.49772	15.99253	11.38540	22.66785	20.16561

Tabla 32: Diferencia de IC50 experimental y la IC50 calculada (1/2)

Molécula	Nombre	Diferencia del IC50exp Vs IC50 Calculado modelo 1	Diferencia del IC50exp Vs IC50 Calculado modelo 2	Diferencia del IC50exp Vs IC50 Calculado modelo 3	Diferencia del IC50exp Vs IC50 Calculado modelo 4	Diferencia del IC50exp Vs IC50 Calculado modelo 5
1	114_L01_Fit16	3.56648	4.49971	-4.6612477	0.22826	3.95474
2	112_L02_Fit16	-0.04294	-0.28155	3.1138738	1.02755	-0.76218
3	110_L03_Fit16	6.16726	6.57834	1.52667891	-2.34199	2.49961
4	108_L04_Fit16	-0.22002	-0.53769	-0.8744609	5.21935	14.80737
5	106_L05_Fit16	-5.97625	-5.46308	9.96088170	13.03762	-7.56298
6	104_L06_Fit16	0.75638	0.83492	-11.396830	-1.58580	0.71778
7	102_L07_Fit16	0.20819	-0.68663	1.42144602	-0.60348	5.02535
8	100_L08_Fit16	-0.77226	-0.55652	-10.755080	4.76084	23.37584
9	098_L09_Fit16	1.29592	1.57959	6.35376467	-4.75267	-5.23229
10	096_L10_Fit16	-1.77579	-1.11030	4.37164472	-5.50558	-4.66017
11	094_L11_Fit16	-2.51409	-2.26363	1.00110605	-4.32539	-3.33576
12	092_L12_FitL16	-12.68732	-13.30655	7.71654619	-15.98497	-16.64176
13	090_L13_FitL16	-1.44433	-0.46614	9.08943373	-5.27993	-1.47465

14	088_L14_FitL16	0.94004	0.33475	-5.6340805	10.15454	-12.51823
15	086_L15_FitL16	2.44666	1.28613	-4.6255907	0.66486	2.05221
16	082_L16_VegaFit	7.17506	6.91671	-6.4647878	4.81501	-2.47038
17	084_L17_FitL16	2.87115	2.65015	-0.1489337	0.46414	2.19772

*Tabla 33: Diferencia de IC50 experimental y la IC50 calculada (2/2)*

<b>Molécula</b>	<b>Nombre</b>	<b>Diferencia del IC50exp Vs IC50 Calculado modelo 6</b>	<b>Diferencia del IC50exp Vs IC50 Calculado modelo 7</b>	<b>Diferencia del IC50exp Vs IC50 Calculado modelo 8</b>	<b>Diferencia del IC50exp Vs IC50 Calculado modelo 9</b>
1	114_L01_Fit16	0.04630	-3.86177	-4.46818	5.41948
2	112_L02_Fit16	0.45531	-5.82361	2.17711	-1.09458
3	110_L03_Fit16	-3.21569	5.00337	3.81199	8.13429
4	108_L04_Fit16	-0.18844	2.33935	-2.57337	0.67400
5	106_L05_Fit16	11.70680	7.34281	4.90973	-2.20222
6	104_L06_Fit16	-2.24389	7.00734	-3.47020	-3.27177
7	102_L07_Fit16	-1.10385	-0.33141	6.22186	2.57864
8	100_L08_Fit16	9.42117	21.76667	3.54754	14.13098

9	098_L09_Fit16	-1.33238	0.98520	0.63044	5.17207
10	096_L10_Fit16	-6.65962	-12.74892	-3.42983	-6.36209
11	094_L11_Fit16	-0.50986	-8.08752	-4.79470	-4.28525
12	092_L12_FitL16	-17.12090	-18.91532	-13.73735	-12.75627
13	090_L13_FitL16	-6.01522	-0.60556	-5.16181	-1.66827
14	088_L14_FitL16	8.87953	3.35206	2.44658	-3.07446
15	086_L15_FitL16	-0.12066	0.96465	2.78440	-2.84387
16	082_L16_VegaFit	8.32068	6.52793	4.73986	-2.42479
17	084_L17_FitL16	-0.30747	-4.91460	6.36785	3.86561

Tabla 34: Gráficos que validan cada ecuación de regresión lineal múltiple de los 9 modelos QSAR obtenidos (1/3)

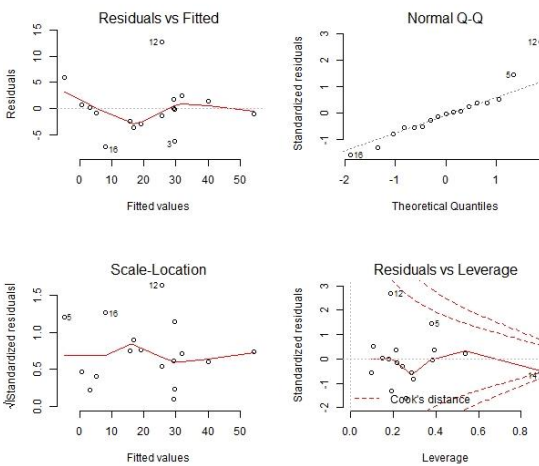
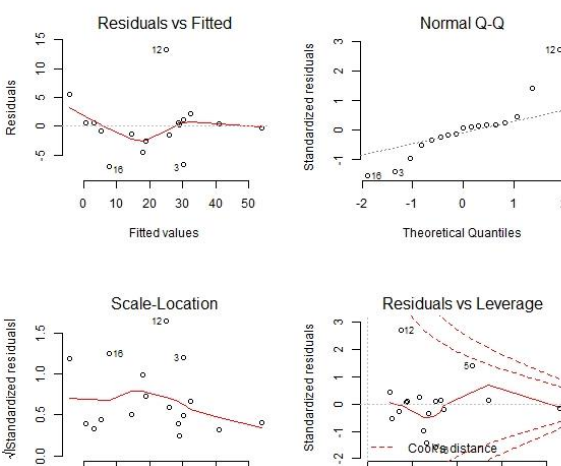
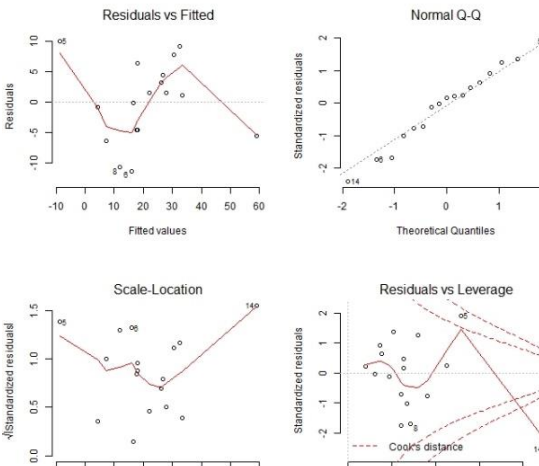
Modelo 1	Modelo 2	Modelo 3
$DV = 240.2298 + (-43.3127)IC5 + 84.7136R4p + (-9.8641)R3s + 0.9055 CATS3D\_05\_AL$	$DV = 231.9315 + (-40.4472)IC5 + 89.7641R4p + (-10.4208)R3s + 0.2939 CATS3D\_05\_AL$	$DV = 752.176 + (-61.805)IC5 + (-442.471) SPH + (-7.411)VE1\_RG + 5.33 CATS3D\_05\_AL$
 <p>Residuals vs Fitted</p> <p>Normal Q-Q</p> <p>Scale-Location</p> <p>Residuals vs Leverage</p>	 <p>Residuals vs Fitted</p> <p>Normal Q-Q</p> <p>Scale-Location</p> <p>Residuals vs Leverage</p>	 <p>Residuals vs Fitted</p> <p>Normal Q-Q</p> <p>Scale-Location</p> <p>Residuals vs Leverage</p>

Tabla 35: Gráficos que validan cada ecuación de regresión lineal múltiple de los 9 modelos QSAR obtenidos (2/3)

Modelo 4	Modelo 5	Modelo 6
$DV = 305.446 + (-64.255)IC5 + (-6.853)Mor03i + (-4.796)F06[N-O]$	$DV = 350.959 + (-55.898)IC5 + (-6.299)TDB10p + (-5.875)CATS3D\_08\_DL$	$DV = 309.187 + (-64.811)IC5 + (-6.551)Mor03i + (-9.296)B10[O-C]$
<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with outliers 012 and 140.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, suggesting constant variance.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Cook's distance contours are shown. Outliers 012 and 140 are identified.</p>	<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with outliers 120 and 140.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, suggesting constant variance.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Cook's distance contours are shown. Outliers 120 and 140 are identified.</p>	<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with outliers 120 and 140.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, suggesting constant variance.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Cook's distance contours are shown. Outliers 120 and 140 are identified.</p>

Tabla 35: Gráficos que validan cada ecuación de regresión lineal múltiple de los 9 modelos QSAR obtenidos (3/3)

Modelo 7	Modelo 8	Modelo 9
$DV = 354.1527 + (-376.4974)BIC4 + (0.2098)Mor20s$	$DV = 304.73773 + (-275.73141)BIC4 + (-0.43316)TPSA(Tot)$	$DV = 286.728 + (-250.419)BIC4 + (-7.773)SpMax3\_Bh(s)$
<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with a smoothing line. Points 012 and 010 are labeled.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality. Point 120 is labeled.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, indicating constant variance.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Points are mostly within Cook's distance boundaries. Points 012, 010, 08, and 14 are labeled.</p>	<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with a smoothing line. Points 012, 00, 17, and 7 are labeled.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality. Points 017 and 07 are labeled.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, indicating constant variance. Point 012 is labeled.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Points are mostly within Cook's distance boundaries. Points 012, 017, and 70 are labeled.</p>	<p>Residuals vs Fitted: Plot of Residuals vs Fitted values. Points are scattered around zero, with a smoothing line. Points 120, 08, and 30 are labeled.</p> <p>Normal Q-Q: Plot of Standardized residuals vs Theoretical Quantiles. Points follow a straight line, indicating normality. Point 08 is labeled.</p> <p>Scale-Location: Plot of <math>\sqrt{ \text{Standardized residual} }</math> vs Fitted values. Shows a relatively flat line, indicating constant variance. Points 08 and 120 are labeled.</p> <p>Residuals vs Leverage: Plot of Standardized residuals vs Leverage. Points are mostly within Cook's distance boundaries. Points 012, 08, 120, 30, and 140 are labeled.</p>