

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE CIENCIAS

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA



**ESTUDIO SAR/QSAR DE COMPUESTOS CON POTENCIAL
APLICACIÓN ANTINEOPLÁSICA**

T E S I S

**que para cubrir parcialmente los requisitos para obtener el grado de MAESTRO
presenta**

JOSÉ ANTONIO RODRIGUEZ ARENAS

**DIRECTOR DE TESIS:
Dr. JUAN CARLOS GARCÍA RAMOS**

ENSENADA, B. C.

DICIEMBRE DE 2022

112

UNIVERSIDAD AUTONOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

**ESTUDIO SAR/QSAR DE COMPUESTOS CON POTENCIAL APLICACIÓN
ANTINEOPLÁSICA**

T E S I S

que para cubrir parcialmente los requisitos para obtener el grado de MAESTRO presenta

JOSÉ ANTONIO RODRÍGUEZ ARENAS

APROBADA POR EL SIGUIENTE COMITÉ:



Dr. Juan Carlos García Ramos
Director



Dra. Yanis Toledano Magaña
Sinodal



Dra. Juana Claudia Leyva Aguilera
Sinodal



Dr. Francisco Casillas Figueroa
Sinodal

123456

RESUMEN de la Tesis de José Antonio Rodríguez Arenas, presentada como requisito parcial para la obtención del grado de MAESTRO. Enseñada, Baja California, México. Diciembre de 2022.

ESTUDIO SAR/QSAR DE COMPUESTOS CON POTENCIAL APLICACIÓN ANTINEOPLÁSICA.

Resumen aprobado por:

Dr. Juan Carlos García Ramos
Director de tesis

El cáncer es causado por la pérdida del control en el crecimiento celular y un funcionamiento anormal de células. En pacientes de sexo femenino, los tipos de cáncer de mama y cervicouterino son de los más agresivos, más comunes y que causan la mayor cantidad de muertes por cáncer a nivel mundial. Es de interés global el desarrollar tratamientos cada vez más efectivos y con menos efectos adversos. De esta forma, el estudio de compuestos con cualidades que favorecen la biocompatibilidad con células sanas y que han demostrado capacidad antiproliferativa frente a células cancerígenas en ambientes controlados (líneas tumorales), tales como los compuestos basados en azoles, son de particular interés. El estudio de nuevos compuestos mediante modelos matemáticos obtenidos por técnicas *in silico*, tales como los análisis cuantitativos de correlación estructura-actividad (QSAR), han facilitado el proponer compuestos con mejores características antitumorales. En el presente trabajo se realizó un análisis QSAR del comportamiento de 40 compuestos probados en las líneas tumorales de cáncer de mama MCF-7 y cervicouterino HeLa reportados en la literatura por 4 trabajos independientes. La actividad biológica se reporta como la concentración inhibitoria media (IC_{50} , por sus siglas en inglés, *inhibitory concentration*). Se generaron un total de 28 diferentes modelos QSAR con excelente capacidad descriptiva y predictiva de la actividad antiproliferativa frente a cada línea tumoral y relacionados con características químicas simples. Estos modelos fueron utilizados para analizar un total de 280 compuestos derivados de las 4 estructuras principales. Se generaron descripciones de comportamiento químico que concuerda y amplía las conjeturas individuales de los trabajos independientes. Se identificó el alcance de dichos modelos QSAR y se propuso compuestos con modificaciones teóricas que, de acuerdo con los modelos QSAR generados con la mayor información posible, presenten una excelente capacidad antitumoral, superando en su mayoría a las IC_{50} experimentales de sus respectivos farmacóforos.

Palabras clave: actividad antiproliferativa, análisis QSAR, compuestos basados en azoles.

ABSTRACT of the thesis, presented by José Antonio Rodríguez Arenas, in order to obtain the MASTER'S DEGREE. Ensenada, Baja California, México. December 2022.

ESTUDIO SAR/QSAR DE COMPUESTOS CON POTENCIAL
APLICACIÓN ANTINEOPLÁSICA.

Approved by:



Dr. Juan Carlos García Ramos
Thesis Advisor

Cancer is caused by loss of cell growth control and abnormal cell function. In female patients, breast and cervical cancers are among the most aggressive, most common, and cause the highest numbers of cancer deaths worldwide. It is of global interest to develop increasingly effective treatment measures with fewer adverse effects. Thus, the study of compounds with qualities that favors healthy cells biocompatibility and demonstrated antiproliferative capacity against cancer cells in controlled environments (tumor cell lines), such as azole-based compounds, are of particular interest. The study of new compounds through mathematical models obtainable by *in silico* techniques such as quantitative structure-activity relationship (QSAR) has facilitated proposing compounds with better antitumor characteristics. In this work, a QSAR analysis of the behavior of 40 compounds evaluated in human cancer cell lines of breast MCF-7 and cervix HeLa and reported in the literature by 4 independent works was carried out. The biological activity was reported as the mean inhibitory concentration (IC_{50}). A total of 28 different QSAR models were generated with excellent descriptive and predictive capability of the antiproliferative activity against each tumor cell line and related to simple chemical characteristics. These models were used to analyze 280 compounds derived from the 4 leading structures. Descriptions of chemical behavior have been generated that match and extend individual conjectures from independent work. The scope of these QSAR models was identified and compounds with theoretical modifications that present excellent antitumor capacity were proposed, according to the QSAR models generated with the greatest possible information, and mostly exceeding the experimental IC_{50} of their respective pharmacophores.

Keywords: antiproliferative activity, QSAR analysis, azole-based compounds.

DEDICATORIA

“Vuelen tan alto, como sus alas se lo permitan”

A mi familia. Los amo.

A todos los que me han acompañado en este tiempo. Siempre estarán en mi corazón

AGRADECIMIENTOS

A mi director, Juan Carlos, y a mi codirectora de tesis, Yanis. Por su guía. Por su apoyo.

Por su conocimiento. Por su amistad.

A mi familia. Porque a pesar de la distancia, siempre han estado conmigo.

A Marylin, por estar conmigo en este periodo de mi vida.

A CONACyT, por su apoyo como becario.

Contenido

índice de Figuras	V
Índice de Tablas	VII
Índice de Material Suplementario	VIII
Cáncer a nivel mundial.	1
Búsqueda de alternativas para tratar y combatir el cáncer.	2
Uso de líneas tumorales	3
Evaluación de moléculas mediante técnicas <i>in vitro</i>	4
Análisis QSAR como alternativa en la búsqueda de compuestos con mejor actividad antiproliferativa.....	5
Regresión lineal.....	6
Aplicación de modelos QSAR	7
Descriptores usados en análisis QSAR.	8
Resultados esperados	13
Objetivo.....	16
Hipótesis.....	16
Metodología	17
Búsqueda bibliográfica	17
Primera búsqueda bibliográfica.....	17
Segunda búsqueda bibliográfica	18
Generación de descriptores	19
Moléculas evaluadas experimentalmente.....	19
Moléculas teóricas.....	19
Preanálisis de datos	20
Selección de descriptores	20
Actividad antiproliferativa experimental	21

Etiquetado	21
Generación de Modelos QSAR.....	21
Modelos QSAR AB.	23
Modelos QSAR ABCD.	24
Predicción de actividad antiproliferativa de moléculas teóricas y significado químico de descriptores	25
Resultados	26
<i>Modelos QSAR AB complejos</i>	26
<i>Modelos QSAR AB simples</i>	30
MODELOS QSAR ABCD.....	32
Predicción de moléculas teóricas basadas en los farmacóforos originales	34
DISCUSION	40
PLANTEAMIENTOS BASADOS EN OBSERVACIONES DE AUTORES ORIGINALES	40
MODELO QSAR. Relación con descriptores simples.....	46
MODELOS AB.....	47
Relación entre <i>modelos QSAR AB</i> vs <i>modelos QSAR ABCD</i>	48
MODELO ABCD. Modelo robusto.	53
<i>Predicción de actividad biológica mediante los modelos QSAR ABCD</i>	53
Significado de los <i>modelos QSAR ABCD</i> y su implicación química	54
DISCUSION BASADA EN ESTRUCTURA	58
Efecto de sustituyentes en posiciones específicas frente a MCF-7	58
Efecto de sustituyentes en posiciones específicas frente a HeLa	64
Contraste con autores originales	68
Otros estudios teóricos de los compuestos reportados	70
ALCANCE DE MODELOS QSAR y PRPUESTA DE NUEVAS MOLÉCULAS	71

PREDICCIÓN DE NUEVOS COMPUESTOS. ALCANCE DE LOS MODELOS PROPUESTOS.....	71
RECOMENDACIONES PARA LA BUSQUEDA DE NUEVOS FARMACÓFOROS.	73
Conclusiones	76
Referencias.....	78
Material Suplementario	84

índice de Figuras

Figura 1. Principales actividades de las distintas fases en el desarrollo de nuevos compuestos. Los detalles varían en función del tipo de fármaco que se esté desarrollando. Adaptada de [12].....	2
Figura 2. Esquema general de la metodología comúnmente utilizada durante la evaluación de actividad antiproliferativa de mu. Las imágenes marcadas con *, de Autor desconocido está bajo licencia CC BY-SA, ^a Link, ^b Extraída de Link, ^c Link, licencia CC BY-NC.....	5
<i>Figura 3 Algunos ejemplos de ajustes de regresión lineal que muestran diferentes grados de ajuste.</i>	7
Figura 4. Representación en gráfica de radar de diferentes espacios formados a partir de distintos eigenvalores.	10
Figura 5. Ejemplificación de cálculo de descriptores, considerando una separación de 0, 1, 2 y 3 átomos y tomando como ponderación la masa atómica (m) de cada átomo.....	12
Figura 6. Estructura de a) 1,3-oxazol y b) 1,2-oxazol.....	13
Figura 7 Compuestos de estudio. Compuestos azólicos provenientes de 4 fuentes distintas que cumplieron los criterios de inclusión del presente trabajo.	15
Figura 8. Estructura principal de moléculas de interés mostrando las posiciones sustituibles identificadas con la etiqueta Rn.....	20
Figura 9. Modelos QSAR AB 1.1 complejos de la actividad biológica de los compuestos frente a las líneas tumorales HeLa y MCF-7. IC50 representada en forma -Log(1/IC50).	29
Figura 10. Modelo QSAR de la actividad biológica en MCF-7 generado a partir de los 4 grupos de compuestos.....	32
Figura 11 Modelo QSAR de la actividad biológica en HeLa generado a partir de los 4 grupos de compuestos.....	33
Figura 12 Cambio relativo de IC50 para cada posición y sustituyente evaluado respecto a cada farmacóforo de cada grupo de compuestos, representado mediante un mapa de calor. Colores más azules y oscuros representan cambios más drásticos asociados con una mejor actividad antiproliferativa.	34
Figura 13. Representación de interacciones contempladas para el cálculo de descriptores.....	43
Figura 14. Valor de los descriptores de compuestos de los grupos A y B, usados en la generación de los modelos QSAR AB simples, agrupadas por su actividad biológica cualitativa en HeLa. Cada color representa una categoría cualitativa de actividad biológica (ningún compuesto del grupo A o B presentó una actividad cualitativa ‘mala’ (IC ₅₀ >150).	47

Figura 15. Comparativa de predicción de modelo QSAR AB 1.1 vs modelo QSAR ABCD, para a) HeLa y b) MCF-7. Cada color representa a un grupo de compuestos distinto.	49
Figura 16. Descriptores efectivos para distinguir actividad biológica del grupo A y B, actualizado con información de los valores de los descriptores de los grupos C y D. Cada color representa una de las 5 categorías de actividad biológica.	50
Figura 17. Relación entre compuestos activos del grupo D y el descriptor sumHBs propuesto.	51
Figura 18. Cambio en los valores de descriptores relacionados con el número de caminos moleculares posibles desde 2 hasta 10 átomos de distancia en función del sustituyente empleado, cada color representa un sustituyente diferente. Se muestra información de.....	57
Figura 19 Cambios relativos de IC50 frente a MCF-7 de compuestos de cada grupo con respecto a su molécula base correspondiente (mostrado como el único compuesto con sustituyente H), generados por cada sustituyente evaluado. Valores negativos indican una mejor	60
Figura 20 Cambios relativos de IC50 frente a HeLa de compuestos de cada grupo con respecto a su molécula base correspondiente (mostrado como el único compuesto con sustituyente H), generados por cada sustituyente evaluado. Valores negativos indican una mejor	66
Figura 21 Representación visual de modificaciones de IC50 de los farmacóforos de los grupos A, B, C y D con cada uno de los sustituyentes en posiciones específicas. Valores más bajos representan una mejora en la potencia de los compuestos y valores negati	72

Índice de Tablas

Tabla 1 IC50 experimental reportada para compuestos de los grupos A, B, C y D. IC50 presentada en μM . Se homogeneizaron los valores en función de la IC50 del fármaco de referencia cisplatino...	19
Tabla 2 Significado de descriptores usados en las distintas aproximaciones usadas en la generación de modelos QSAR.....	22
Tabla 3 Ecuación y estadísticas de los Modelos QSAR AB obtenidos mediante esta aproximación.	27
Tabla 4. Valores de IC50 teóricos y experimentales de los compuestos usados en el entrenamiento y validación de los modelos QSAR AB 1.1.	29
Tabla 5 Ecuaciones y descripción estadística de modelos QSAR simples.	30
Tabla 6. Mejor IC50 teórica posible obtenida por cada sustituyente para los diferentes grupos. HeLa y MCF-7 QSAR AB.	36
Tabla 7. Mejor IC50 teórica posible obtenida por cada sustituyente para los diferentes grupos. HeLa y MCF-7. QSAR ABCD	37
Tabla 8. Comparación de valor en descriptores de los mejores representantes del grupo B respecto a sus variantes del grupo C. y la IC50 obtenida.	46
Tabla 9. Descriptores cuya presencia o aumento de valor se relaciona con un cambio positivo o negativo en la actividad biológica	51
Tabla 10 Propuesta de evaluación de moléculas teóricas basadas en los grupos A, B, C y D.	72
Tabla 11 Modelos QSAR de actividad biológica frente a líneas tumoral HeLa y MCF-7, obtenidos por diferentes aproximaciones.....	75

Índice de Material Suplementario

Material Suplementario 1. IC ₅₀ experimental extraída directamente de los trabajos originales de los compuestos evaluados. En el último renglón se muestra la IC ₅₀ de su respectivo fármaco de referencia Cisplatino. MS0	84
Material Suplementario 2. Actividad antiproliferativa calculada a partir de los modelos QSAR AB de información compleja (1.1-1.4) para el conjunto de entrenamiento	84
Material Suplementario 3. Actividad antiproliferativa calculada a partir de los modelos QSAR AB de información compleja (1.1-1.4) para el conjunto de validación	84
Material Suplementario 4. IC ₅₀ teórica frente a HeLa del conjunto de entrenamiento usado en los modelos QSAR simples	85
Material Suplementario 5. IC ₅₀ teórica frente a HeLa del conjunto de validación usado en los modelos QSAR simples	85
Material Suplementario 6. IC ₅₀ teórica frente a MCF-7 del conjunto de entrenamiento usado en los modelos QSAR simples	85
Material Suplementario 7. IC ₅₀ teórica frente a MCF-7 del conjunto de validación usado en los modelos QSAR simples	86
Material Suplementario 8. Modelos QSAR ABCD. IC ₅₀ set de entrenamiento. Valores representados como IC ₅₀ (μM)	86
Material Suplementario 9. Modelos QSAR ABCD. Set de validación.	87
Material Suplementario 10. Identificador de compuestos teóricos y experimentales.	87
Material Suplementario 11. Actividad antiproliferativa (IC ₅₀ [μM]) predicha por los mejores modelos QSAR AB y ABCD.....	88
Material Suplementario 12. Descriptores relacionados con los modelos QSAR AB complejos y su actividad biológica cualitativa. Se observa que modelos QSAR AB catalogan como activos a los compuestos que presenten bajos valores en la mayoría de los descriptores mostrados.	91
Material Suplementario 13. Valores de los descriptores usados en los modelos QSAR AB 1.1 y ABCD para HeLa y MCF-7. Redondeado a 5 decimales. MS2.....	91
Material Suplementario 14. Descriptores de características presentes principalmente en el a) grupo A y b) grupo D. Los descriptores son asociados a características propias de compuestos con buena o mala actividad biológica.....	92
Material Suplementario 15. Cambio en los valores de descriptores relacionados con el número de caminos moleculares posibles desde 2 hasta 10 átomos de distancia en función del sustituyente (izq.)	

y la posición (der.) empleada. A mayor número de caminos posibles, más elementos son considerados para el cálculo de cada descriptor y, por tanto, mayor será valor del descriptor. Los valores de los descriptores aquí mostrados corresponden a los compuestos del grupo B y C.	93
Material Suplementario 16. Descriptores característicos de cada grupo usado en la generación de los modelos QSAR ABCD complejos . a) Cada color representa un grupo de compuestos; b) agrupados por tipo de actividad biológica cualitativa para HeLa y c) MCF-7	93
Material Suplementario 17. Valor del descriptor minHBa redondeado a 2 dígitos, para todas las moléculas basadas en los farmacóforos de los grupos A-D. Los valores correspondientes a moléculas sustituidas con Cl son las únicas con valores debajo de 1.	95

Capítulo 1.

Cáncer a nivel mundial.

El cáncer es un padecimiento cada vez más común que puede llegar a ser tan agresivo como para causar la muerte de aquel que lo padezca. El cáncer es causado por la pérdida del control en el crecimiento celular, causando un crecimiento y funcionamiento anormal de células [1]. En 2020, más de 10 millones de personas murieron a causa de algún tipo de cáncer a nivel mundial [2], y se estima que para 2050 el cáncer sea la principal causa de muerte a nivel mundial [3]. Existen diversos tipos de cáncer puesto que pueden verse afectados varios tipos de tejidos distintos al mismo tiempo, e incluso a múltiples órganos a la vez. Así mismo, factores como la dieta, los patrones de actividad física y composición corporal [3], síndrome metabólico [4], factores genéticos y hereditarios [5], pueden ser asociados a su la ocurrencia de diversos tipos de cáncer. El conjunto del tipo de células, órganos afectados y los diversos factores asociados dificultan la búsqueda de alternativas efectivas para el tratamiento de personas con cáncer.

En pacientes de sexo femenino, dos tipos de cáncer de mayor ocurrencia son el cáncer cervicouterino y el de mama. El cáncer de mama se ha convertido en uno de los tipos de cáncer más común, con alrededor de 2.3 millones de casos reportados para 2022 [5]. y se ha visto que en mujeres de origen mexicano la ocurrencia de este tipo de cáncer parece incrementar más rápido de lo esperado [6]. Por otro lado, el cáncer cervicouterino, tipo de cáncer de alta mortalidad común en mujeres, muestra una alta incidencia en el mundo cuya mayor cantidad de víctimas ocurren en países menos desarrollados, de las cuales el 15% de muertes ocurren en Latinoamérica y, particularmente en México, solo el 15% de los casos son diagnosticados [7] derivado de la falta de acceso para la

detección temprana [8]. En este sentido, la búsqueda de mejores tratamientos para combatir diferentes variedades de estos tipos de cáncer son líneas de estudio de importancia nacional y global para mejorar la calidad de vida en pacientes de sexo femenino.

Búsqueda de alternativas para tratar y combatir el cáncer.

El tratamiento y combate contra el cáncer en pacientes es de interés y colaboración global [3], [9]. El tratamiento se realiza comúnmente mediante la administración en dosis controladas de un compuesto químico denominado *fármaco*, el cual es definido como aquella sustancia que sirve para curar, prevenir, aliviar, reducir tratar y/o controlar [10] los efectos de alguna enfermedad o dolor físico. Se conoce como *fármacos de primera elección* a los compuestos que han demostrado efectividad y que han sido aprobados para su uso por organizaciones regulatorias, tales como la FDA, y que son utilizados como la primera alternativa para el combate contra un tipo de cáncer en específico.

En general, la búsqueda de nuevos compuestos con aplicación biomédica parte de su estudio bajo condiciones controladas de interacción con los blancos biológicos de interés. Existen varias etapas *in vitro* e *in vivo* en la búsqueda de nuevas alternativas de compuestos para combatir el cáncer [11]. Dicha búsqueda se divide en fases que enfocan los esfuerzos en la selección de los compuestos candidatos que vayan presentando las mejores cualidades de efectividad, seguridad y menor severidad de efectos adversos. La Figura 1 muestra un esquema general de cómo se dividen las fases de estudio para el desarrollo de nuevos fármacos.

Fases de un proyecto "típico", destinado a producir un fármaco comercializable que cubra una necesidad clínica específica							
Fases	Descubrimiento del Fármaco	Desarrollo preclínico		Desarrollo Clínico		Aprobación por las autoridades sanitarias	Fase IV
Actividades realizadas	Selección de la diana Búsqueda de la molécula de partida Optimización de la molécula de partida Evaluación de características farmacológicas	Farmacocinética a corto plazo Toxicología a gran escala Formulación	Fase I Farmacocinética, tolerabilidad, efectos secundarios en voluntarios sanos	Fase II Ensayos a pequeña escala en pacientes para valorar eficacia y dosis Estudios toxicológicos a largo plazo	Fase III Ensayos clínicos controlados a gran escala	Se remiten los datos completos y son revisados por las autoridades sanitarias	Vigilancia tras la comercialización
Tiempo promedio	2 – 5 años	1 – 5 años		5 – 7 años		1 – 2 años	
Compuestos evaluados promedio	100 proyectos	20 compuestos	10	5	2	1, 2	1
Estado de fármacos	Fármaco candidato						
	Compuesto en desarrollo						
						Remisión a la comisión de registro sanitario Fármaco a probado para su comercialización	

Figura 1. Principales actividades de las distintas fases en el desarrollo de nuevos compuestos. Los detalles varían en función del tipo de fármaco que se esté desarrollando. Adaptada de [12]

Durante las primeras fases de estudio, se generan (sintetizan) y evalúan diversos tipos de *compuestos* con la finalidad de identificar sus potenciales aplicaciones como fármacos frente a una o más enfermedades, tratando de identificar y promover propiedades particulares que los vuelvan efectivos para una aplicación específica. Al conjunto de propiedades necesarias para asegurar óptimas interacciones con un blanco biológico específico y así desencadenar (o bloquear) su respuesta biológica se le conoce como *farmacóforo* [13]. Dicho farmacóforo es modificado estructuralmente añadiendo diferentes tipos y cantidad de sustituyentes para generar así nuevos compuestos a probar frente a un blanco específico. A la evaluación de series de compuestos frente a un blanco biológico específico se le conoce como abordaje químico. Como resultado del abordaje químico empleado se seleccionan compuestos con modificaciones que resultaron potencial la actividad deseada y descarta al resto para futuros ensayos.

Para la evaluación de compuestos durante los ensayos preclínicos, se utiliza un tipo particular de células denominadas *líneas celulares*. Se conoce como *línea celular* a aquellas células immortalizadas que son distribuidas globalmente y que son aceptadas como estándar en la realización de pruebas biológicas. Una línea celular ideal es definida como un cultivo puro genética, epigenética y fenotípicamente [14]. La primera línea celular fue establecida en 1951 y desde un principio resultaron de gran valor para la realización de pruebas biológicas [14]. El uso de líneas celulares como modelos en las primeras fases de estudio de fármacos es una práctica bien establecida debido a ser modelos simples cuyos resultados obtenidos pueden ser extrapolados hacia resultados *in vivo*. Su uso es parte importante de las pruebas *in vitro* a realizar para el estudio de nuevos fármacos.

Uso de líneas tumorales

Las células extraídas de algún tipo de cáncer son conocidas como líneas tumorales, existiendo diversas variedades para un mismo tipo de cáncer derivado del tipo de tejido y cáncer del cual hayan sido extraídas. Estas líneas tumorales son usadas en estudios por servir como modelo de tipos de cáncer muy específicos, a la par que permiten una comparación directa de resultados entre distintos laboratorios que hayan realizado estudios con el mismo tipo de línea tumoral.

Dos líneas tumorales ampliamente utilizadas como modelos para cáncer de mama y cáncer cervicouterino son las líneas tumorales MCF-7 y HeLa, respectivamente. La línea tumoral MCF-7 fue establecida de un paciente humano con derrame pleural en la “Michigan Cancer Foundation” en 1973. Esta línea celular tiene por características ser no invasiva, presentar receptores de estrógeno funcionales (ER+) y dependencia de estrógeno para su crecimiento, razones por las que esta línea tumoral es usada como modelo para representar la etapa temprana del cáncer de mama, principalmente para terapias antiestrógeno. [11]. Por otro lado, la línea tumoral de cáncer cervical humano HeLa fue la primer línea tumoral humana establecida obtenida de una biopsia de Henrietta Lacks, de 30 años de edad [15], establecida en 1951 y fue la primera línea tumoral creada [16], es tan agresiva e invasiva que se ha registrado que esta línea tumoral ha llegado a invadir otras líneas tumorales [14], [16].

Evaluación de moléculas mediante técnicas *in vitro*.

Comúnmente, cuando se quiere evaluar que tan efectivo es un compuesto particular para inhibir el crecimiento o proliferación de un tipo de célula específica, se evalúa la actividad antiproliferativa del compuesto, medida cualitativa o cuantitativamente, frente a cultivos celulares controlados de dichas células, usando generalmente líneas celulares. Este tipo de estudios permiten identificar las potenciales aplicaciones biomédicas que un compuesto pueda tener como, por ejemplo, agente anticancerígeno.

Se da uso a las líneas celulares como modelos *in vitro* estandarizados para evaluar el desempeño de la actividad antiproliferativa de compuestos. Este tipo de estudios de carácter cuantitativo, se hacen normalmente en función del tiempo de exposición y la concentración aplicada. A la concentración requerida para disminuir la población celular a un 50% de su total, se le conoce como concentración inhibitoria media (IC₅₀) y sirve de referencia para medir la capacidad de un compuesto específico para disminuir o evitar la proliferación celular. Por esta razón, determinar la IC₅₀ de este frente a múltiples líneas celulares o, por otro lado, evaluar múltiples compuestos frente a una o más líneas tumorales es un estudio común. Estas mediciones se realizan para un tiempo de exposición específico, siendo comúnmente medidas a 24, 48 o 72 horas. Las características específicas de la metodología a usar dependerán del tipo de cultivo celular empleado, y de las características de dicho cultivo tales como si las

células son adherentes o no, o si tienen una alta capacidad de dispersión [16]. La Figura 2 muestra un esquema general de este tipo de metodologías.

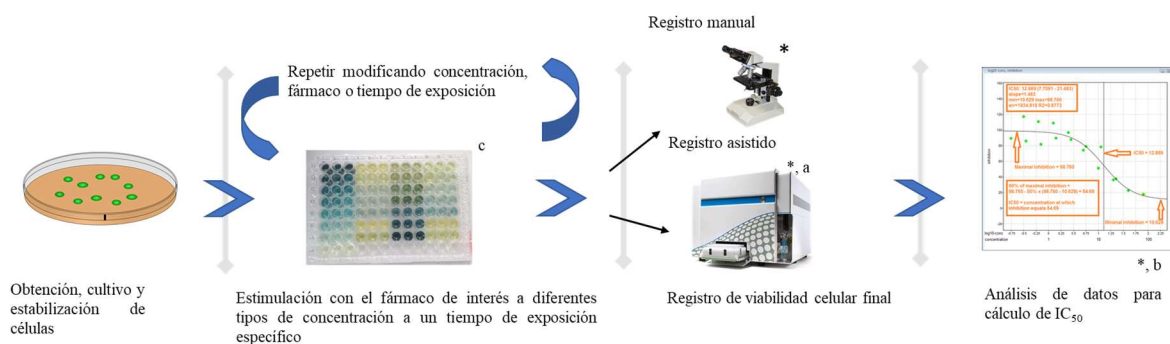


Figura 2. Esquema general de la metodología comúnmente utilizada durante la evaluación de actividad antiproliferativa de mu. Las imágenes marcadas con *, de Autor desconocido está bajo licencia [CC BY-SA](#), ^a [Link](#), ^b Extraída de [Link](#), ^c [Link](#), licencia [CC BY-NC](#)

Históricamente la búsqueda de nuevos compuestos se ha realizado por métodos experimentales basados en prueba y error de los potenciales candidatos [17], principalmente en las fases previas al comienzo de ensayos en seres vivos, con la intención de clasificar a los compuestos en función de factores tales como la *potencia* (la cantidad necesaria del compuesto para observar el efecto deseado), o el blanco celular afectado. Previo al comienzo de la experimentación en seres vivos se realizan ensayos *in vitro*, a fin de conocer las características que los compuestos puedan tener y los efectos que puedan causar en sistemas vivos. Derivado de esto, la necesidad de obtener/sintetizar previamente los compuestos a evaluar genera un problema de eficiencia en la distribución de recursos. No obstante, dada de la capacidad de análisis que sistemas de cómputo actuales logran realizar, el uso de estudios *in silico* (a computadora) previos al comienzo de ensayos experimentales se ha vuelto más común con la finalidad de aumentar la probabilidad de encontrar compuestos con buen desempeño de actividad antiproliferativa [18], [19].

Análisis QSAR como alternativa en la búsqueda de compuestos con mejor actividad antiproliferativa

Se han propuesto métodos que mejoren las posibilidades de encontrar compuestos efectivos como agentes antitumorales al mismo tiempo que contribuyan a reducir la inversión de tiempos y gastos requeridos. Esto debido a que si bien, el abordaje químico usado en la

evaluación de potenciales fármacos permitiría aumentar las posibilidades de encontrar compuestos con IC_{50} 's más altas, la búsqueda de compuestos con mejores IC_{50} se vuelve cuestión de prueba y error al tener que sintetizar y evaluar cada uno de los compuestos para poder discriminar a los compuestos de interés. Uno de estos métodos que ayudan en la búsqueda de compuestos es el análisis cuantitativo de correlación estructura- actividad (modelos QSAR, por sus siglas en inglés).

Los modelos QSAR son, en esencia, protocolos basados en conocimiento estadístico y matemático aplicado a la predicción o clasificación de datos biológicos relacionados con un grupo particular de moléculas de estudio [18]. La metodología básica consiste en realizar aproximaciones de la actividad biológica de interés de las moléculas de estudio, calculable a partir de un modelo matemático basado en un ajuste lineal o no lineal, simple o múltiple, y usando propiedades cuantitativas del conjunto de moléculas de interés como variables del modelo matemático usado. De tal forma que, con esa información, se genere una ecuación que se ajuste a la actividad biológica de interés, obteniendo así una descripción matemática del comportamiento de dichas moléculas.

En esta sección se discutirán los aspectos básicos sobre el fundamento matemático de la regresión lineal múltiple, la cual será usada para generar nuestros modelos QSAR. Así mismo, introduciremos el concepto de descriptor molecular y abordaremos particularmente el significado matemático de 3 grupos principales de descriptores y las interpretaciones químicas comúnmente asociadas a ellos.

Regresión lineal

La calidad de los modelos QSAR es medida principalmente por la calidad del ajuste al reproducir con la mayor similitud los valores reales de la actividad biológica en cuestión. Normalmente el ajuste es medido por el coeficiente de correlación (R^2). A mayor coeficiente en una escala de $[-1,1]$ implica mejor representación de los datos por el modelo empleado. La figura 3 muestra una representación de diferentes ajustes de regresión lineal a distintos conjuntos de datos.

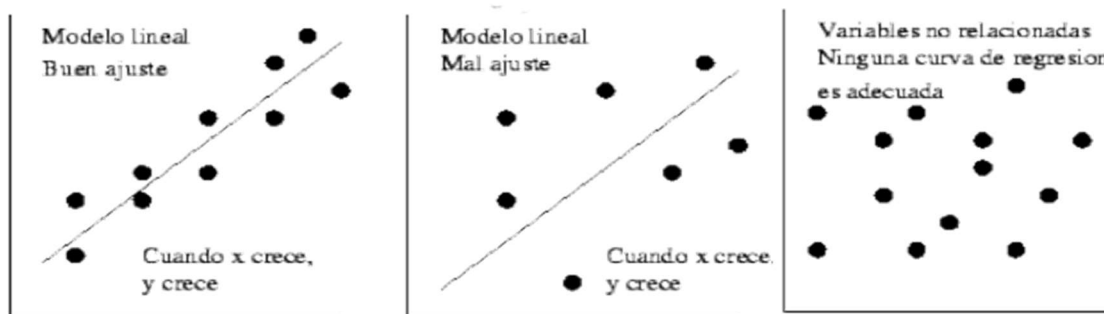


Figura 3 Algunos ejemplos de ajustes de regresión lineal que muestran diferentes grados de ajuste.

Desde una perspectiva matemática, el uso de menos variables involucradas en la generación del ajuste de regresión lineal implica mayor simplicidad y facilidad para describir el comportamiento de las observaciones independientes bajo una misma ecuación. Así mismo, mientras menor sea la colinealidad entre dos o más variables para generar el ajuste, indica que la información que se esté usando está describiendo dos o más comportamientos distintos. En este sentido, mientras más cercana a 0 sea la colinealidad entre variables, más información distinta se estará usando. Un criterio de colinealidad aplicada a modelos QSAR es utilizar información con una colinealidad menor a 0.7 ($R < 0.7$)[20].

Aplicación de modelos QSAR

Un modelo QSAR, generado a partir de un ajuste de regresión lineal entre información del comportamiento químico de un conjunto de moléculas e información de una característica biológica, puede ser usado principalmente para 2 aplicaciones: descripción de comportamiento, y predicción de comportamiento. La descripción del comportamiento se basa en dar a los descriptores que forman parte de la ecuación final el sentido químico más simple posible para explicar el comportamiento biológico estudiado. Mientras que la predicción del comportamiento se basa en evaluar el desempeño que el modelo QSAR presente para predecir la actividad de interés de compuestos desconocidos por el modelo, es decir, que no fueron usados durante la generación del modelo. De esta forma, se pueden identificar características y cualidades presentes en las moléculas como benéficas o perjudiciales para una actividad dada y así formular conjeturas que sirvan para explicar la

razón del comportamiento biológico de cada molécula estudiada y ser usadas para la búsqueda de nuevos compuestos sin necesidad de haber sido probados experimentalmente.

Un modelo QSAR puede ser usado para generar predicciones del comportamiento de moléculas teóricas similares a las moléculas con las que se realizó el modelo matemático. Comúnmente, se realizan pequeñas modificaciones a una estructura base añadiendo un sustituyente simple, -OH, -NO, -NO₂, CH₃, o algún halógeno son comúnmente usados. Se determina la actividad específica que se quiere estudiar y se obtiene un modelo de correlación. Con este modelo se calculan los valores de los descriptores y la actividad biológica teórica de la nueva molécula. De este modo, se puede estudiar la actividad biológica de nuevos compuestos sin la necesidad de evaluar experimentalmente a todas las nuevas moléculas candidatas. Ayudando así a enfocar futuros esfuerzos experimentales en el estudio de aquellas moléculas que presenten las mejores características teóricas, aumentando la probabilidad de obtener compuestos más potentes en menor tiempo.

Para poder considerar al modelo QSAR con capacidad predictiva se toma como criterio de la confiabilidad de la predicción tanto a los ajustes entre el modelo QSAR y el conjunto de entrenamiento (R^2), como entre el modelo QSAR y el conjunto de validación (R^2_{test}). El conjunto de entrenamiento es usado para la generación del modelo QSAR, un buen ajuste R^2 implica mayor capacidad descriptiva de los compuestos conocidos por el modelo. Mientras que el conjunto de validación son un conjunto de datos del que se conozca su respectiva respuesta biológica pero que no haya sido usado como parte del conjunto de entrenamiento. Esta validación debe constar de al menos el 30% del total de datos conocidos para poder validar la capacidad predictiva del modelo. Se dice que un modelo QSAR tiene capacidad predictiva cuando el ajuste R^2_{test} es mayor a 0.6 ($R^2_{\text{test}} > 0.6$) [20]. La confiabilidad y precisión de la predicción mejora mientras más cercanos sean los valores de R^2 y R^2_{test} a 1. Para que la predicción sea válida, los compuestos a predecir deberán tener comportamiento químico similar al de los conjuntos de entrenamiento y de validación.

Descriptores usados en análisis QSAR.

En el área de los análisis QSAR, se le conoce como descriptor molecular, o simplemente descriptor, a una característica o comportamiento específico medible de una

molécula [21]. Los descriptores son clasificados en función de que tan simple o compleja sea el cálculo del descriptor como 1D, 2D, 3D y 4D. Llegando a ser desde simples conteos de un tipo de átomos presentes en la estructura molecular (1D), conteo de átomos en una posición específica o el número de ramificaciones (2D); hasta contemplar información sobre su disposición espacial (3D) o su interacción con otras moléculas (4D). Se han generado aplicaciones que permiten calcular descriptores ya reportados en la literatura, facilitando su obtención. Software de acceso libre tales como PAdel Descriptor (<http://www.yapcwsoft.com/dd/padeldescriptor/>), en su versión más reciente, permite el cálculo de 1444 descriptores 1D y 2D. A pesar de existir una basta cantidad de descriptores ya establecidos, debido a su complejidad, la experiencia previa del autor, y/o el objetivo particular de cada estudio, pocos descriptores son usados en la construcción de cada modelo [22].

El trabajo aquí presentado partió de la contemplación de los 1444 descriptores de información 1D y 2D calculables por el software PAdel descriptor, abordando principalmente 3 grupos de descriptores a lo largo del manuscrito. Estos 3 grupos de descriptores de mayor interés, de acuerdo con el sistema de clasificación del software implementado, forman parte las familias de descriptores de autocorrelación (descriptores *AATSC#** y *MATS#**), eigenvalores modificados de burden (descriptores *SpMin#_Bh** y *SpMax#_Bh**) y de la clase de tipo de estado atómico electrotopológico (descriptores *minHB**, *minwHB** y *maxHB**). Familias de descriptores que serán abordadas a continuación.

Eigenvalores modificados de burden

Uso común para describir similitud y diferencia entre moléculas con distintas características. A fin de representar una aplicabilidad de los eigenvalores modificados de burden, la Figura 4/FCV muestra una ejemplificación del comportamiento de 4 compuestos imaginarios distintos para los primeros 8 eigenvalores.



Figura 4. Representación en gráfica de radar de diferentes espacios formados a partir de distintos eigenvalores.

Considérese que valores cercanos entre diferentes moléculas para un mismo eigenvalor o más al mismo tiempo, estaría representando que uno o varios comportamientos son similares entre dichas moléculas. Así mismo, valores más alejados entre diferentes moléculas para 1 o más eigenvalores estaría indicando que los compuestos se comportan distinto. Bajo esta premisa, se pueden considerar los siguientes tres casos:

Mismo comportamiento en múltiples eigenvalores al mismo tiempo, hablaría de que la similitud sea constante en diferentes características de dos o más moléculas no necesariamente basadas en la misma estructura base. La idea principal es que, mientras más eigenvalores se consideren al mismo tiempo para decidir si el comportamiento es similar o no, estaría dando mayor robustez. Bajo esta premisa, se podría declarar, por ejemplo, que el compuesto W y X presentan comportamiento similar ya que presentan valores similares en todos los eigenvalores, mientras que serían distintos al comportamiento del compuesto Z (Figura 4).

Mismo comportamiento en un espacio dimensional (representado por el valor del eigenvalor), pero muy distinto en otros, lo que podría significar que hay un aspecto en donde dos o más moléculas son similares, pero si no se considera el comportamiento en el resto de las dimensiones, podrían generarse declaraciones invalidas como que, en la Figura3, el **compuesto 4** se comporta igual que el **compuesto 1** y el **compuesto 2** si se considera únicamente su similitud en el eigenvalor 4.

Diferente comportamiento en todos los eigenvalores. Este sería uno de los comportamientos más fáciles de identificar bajo este ejemplo analizando el comportamiento en todos los eigenvalores de interés al mismo tiempo para una misma molécula, y comparándolo con los valores de otros compuestos.

Tipo de estado atómico electrotopológico

Miden energía eletrotopológica máxima y mínima entre aceptores y donadores de enlaces de hidrógeno con características particulares, es decir, la energía asociada a la deformación de la densidad electrónica de una molécula al encontrarse cerca de otra. Los descriptores enfocados en medir la energía mínima requerida para generar un tipo particular de enlaces hidrógeno estaría hablando de la mayor facilidad que la molécula tiene para formar su primer puente de hidrógeno. Por otro lado, los descriptores enfocados en medir la energía máxima requerida para generar un tipo particular de enlaces de hidrógeno estarían hablando de la máxima dificultad que un compuesto tiene para formar su último puente de hidrógeno.

Descriptores de autocorrelación

Los descriptores de autocorrelación miden cómo una característica es consistente entre átomos cercanos y si es qué dicha característica cambia en átomos más lejanos [23]. Para este trabajo son de particular interés los descriptores de tipo AATSC y MATS.

Los descriptores de tipo *ATS*, *AATS* y *AATSC* miden la interacción entre algún átomo en particular con cualquier otro separado por un número específico de átomos entre ellos [24]. El cálculo de los primeros tipos de descriptores de esta familia, los *ATS* y *AATS*, depende de 2 factores principales: el número de enlaces de separación entre dos o más átomos y la ponderación de interés. Ambas etiquetas consisten en un número y una letra en el nombre del descriptor, justo después del tipo de descriptor (por ejemplo, *ATS1p*, *ATSC2c*, *AATS3v*, *AATSC6e*, ...). La ponderación hace referencia a alguna propiedad específica de cada átomo, y considera la electronegatividad (*e*), el volumen de van Der Waals (*v*) o la carga (*c*), entre otras, que los átomos tengan. Mientras que el número de enlaces de separación entre átomos hace referencia a que tan cerca o lejos se encuentran los átomos a considerar en el cálculo.

En este sentido, mientras más grande sea el número de enlaces de separación indicado por el descriptor, el número de combinaciones totales con las que un átomo interaccione

dentro de la misma molécula será mayor o menor dependiendo de su estructura. De este modo, moléculas más lineales tenderán a tener más contribuciones mientras menor sea el número de enlaces de separación considerado por el descriptor usado. Por otro lado, moléculas con más bifurcaciones podrían tener más coeficientes en función de la posibilidad de seguir diferentes caminos enlazados dentro de la estructura molecular, ya que es más probable que un camino seguido se bifurque en dos o más caminos mientras mayor sea el número de enlaces de separación considerada. La Figura 5 muestra una ejemplificación del cálculo de los descriptores ATS y AATS.

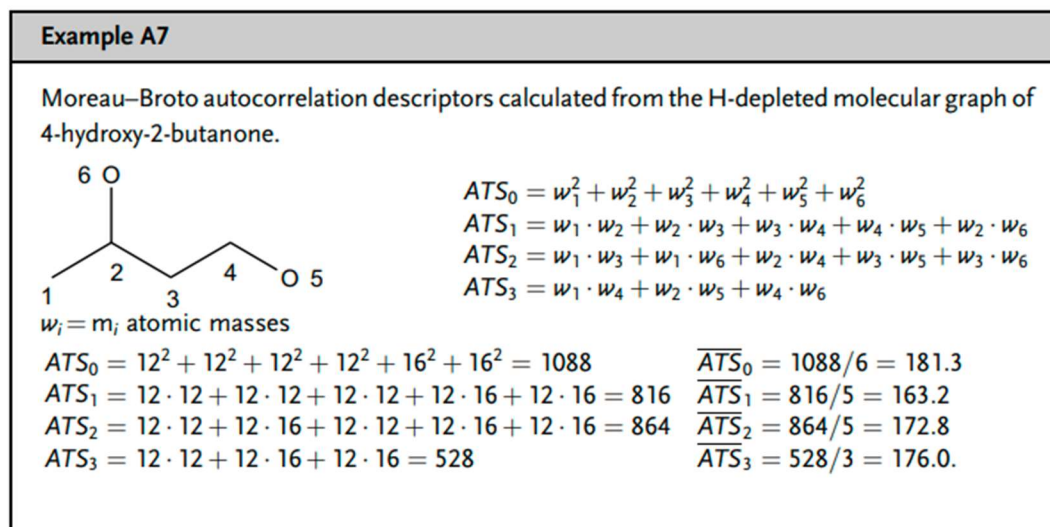


Figura 5. Ejemplificación de cálculo de descriptores, considerando una separación de 0, 1, 2 y 3 átomos y tomando como ponderación la masa atómica (m) de cada átomo.

Los descriptores de tipo AATSC siguen el mismo principio que los descriptores ATS y AATS, con la diferencia que presentan una modificación a fin de evitar correlaciones indeseadas con otros descriptores [25]. Esta modificación es conocida como centrado de datos (en inglés: centering) y es la razón de la letra C al final de la abreviación. Dicho “centrado” considera la distribución poblacional y de posición interna propios de la configuración específica de la molécula. Esto implica que en la desviación estándar de los valores de la ponderación considerada (e, p, v, ...), afectará directamente al valor final calculado para dichos descriptores. Así mismo, implica indirectamente que la contribución que un único átomo pueda generar será mayor o menor en función de qué tanta desviación estándar presente la molécula en cuestión, la posición elegida y la cantidad de ramificaciones

finales. Por esta razón, se podría usar esta familia de descriptores para identificar al mismo tiempo, la mejor posición donde colocar un sustituyente que favorezca o perjudique una característica específica de una ponderación de interés a fin de mejorar la descripción de un fenómeno específico, por ejemplo, la actividad biológica final.

Por otro lado, los descriptores de tipo *MATS* son descriptores que miden la autocorrelación espacial general de su conjunto de datos, es decir, cómo un objeto es similar a otros que lo rodean y por lo general, estos coeficientes presentan valores en el intervalo [-1,1] [21]. Pueden ser aplicados para medir la diferencia de características específicas a diferentes distancias. Estos descriptores pueden medir características de electronegatividad, polarizabilidad, u otras, a diferentes distancias topológicas de manera similar al que los descriptores de la *familia ATS* miden la interacción a distintos enlaces de distancia.

Resultados esperados

Se ha observado que ciertos compuestos heterociclos presentan actividad biológica que puede ser usada para tratar diversos padecimientos. Basado en evidencia, ha quedado establecido en la literatura que compuestos heterocíclicos formados por azoles presentan excelente actividad biológica para aplicaciones en la industria médica y veterinaria [1], [26], Los azoles son estructuras químicas cíclicas de 5 miembros donde al menos uno de los miembros es un átomo N. Dos de las estructuras de interés en este trabajo son los anillos de 5 miembros 1,3-oxazol y el 1,2-oxazol (ver Figura 6). Estos anillos están presentes en varios productos naturales y sintéticos y varios compuestos basados en estas estructuras son usados como fármacos. Compuestos basados en azoles han sido objeto de estudio en múltiples aplicaciones médicas, incluida la actividad antitumoral.

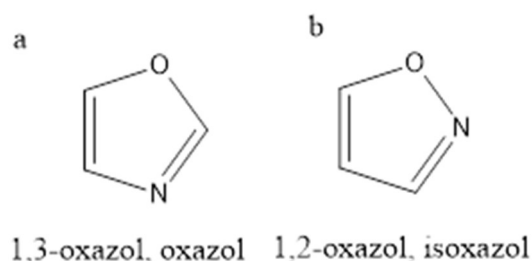


Figura 6. Estructura de a) 1,3-oxazol y b) 1,2-oxazol

Los compuestos basados en azoles han sido identificados como excelentes agentes antitumorales frente a diversas líneas tumorales. Se han generado compuestos con excelente actividad anticancerígena *in vitro* a bajas concentraciones, generalmente entre 10^{-3} a 10^{-5} M [1], [27]. Diversos estudios han demostrado la capacidad antiproliferativa frente a diversas líneas tumorales [28], incluidas cáncer de mama [29]–[32] y cervicouterino [33]. Particularmente frente a las líneas tumorales MCF-7 y HeLa, se ha identificado que algunos compuestos pueden llegar a presentar una excelente actividad biológica en la misma ventana de concentración, sin embargo, de acuerdo con nuestra búsqueda, solo unos cuantos trabajos han probado su efectividad en ambas líneas tumorales al mismo tiempo [34]–[37].

Un estudio teórico como el análisis QSAR podría ayudar a identificar cualidades que vuelven a los compuestos basados en azoles efectivos contra ambas líneas tumorales, cáncer de mama MCF-7 y cáncer cervicouterino HeLa [38]. De tal forma que, a partir de compuestos que demostraron buena actividad antiproliferativa en ambas líneas tumorales se pueda proponer modificaciones que favorezcan las propiedades necesarias para potenciar la actividad antiproliferativa de los farmacóforos, así mismo, el identificar dichas características benéficas permitiría que, en un futuro y bajo las limitantes del modelo QSAR obtenible, se enfoquen esfuerzos en identificar compuestos que hayan sido evaluados únicamente frente a una línea tumoral que potencialmente presenten actividad antiproliferativa contra la segunda línea tumoral.

Cuatro grupos de compuestos, reportados por trabajos individuales fueron identificados con actividad antitumoral frente a las líneas MCF-7 y HeLa. Los compuestos se muestran en la Figura 7. Para su estudio, los compuestos fueron separados en 4 grupos distintos en función de la estructura principal identificada como farmacóforo. El grupo A está conformado por los compuestos **1-12**, el grupo B por los compuestos **13-22**, el grupo C por los compuestos **23-32** y el grupo D por los compuestos **33-40**.

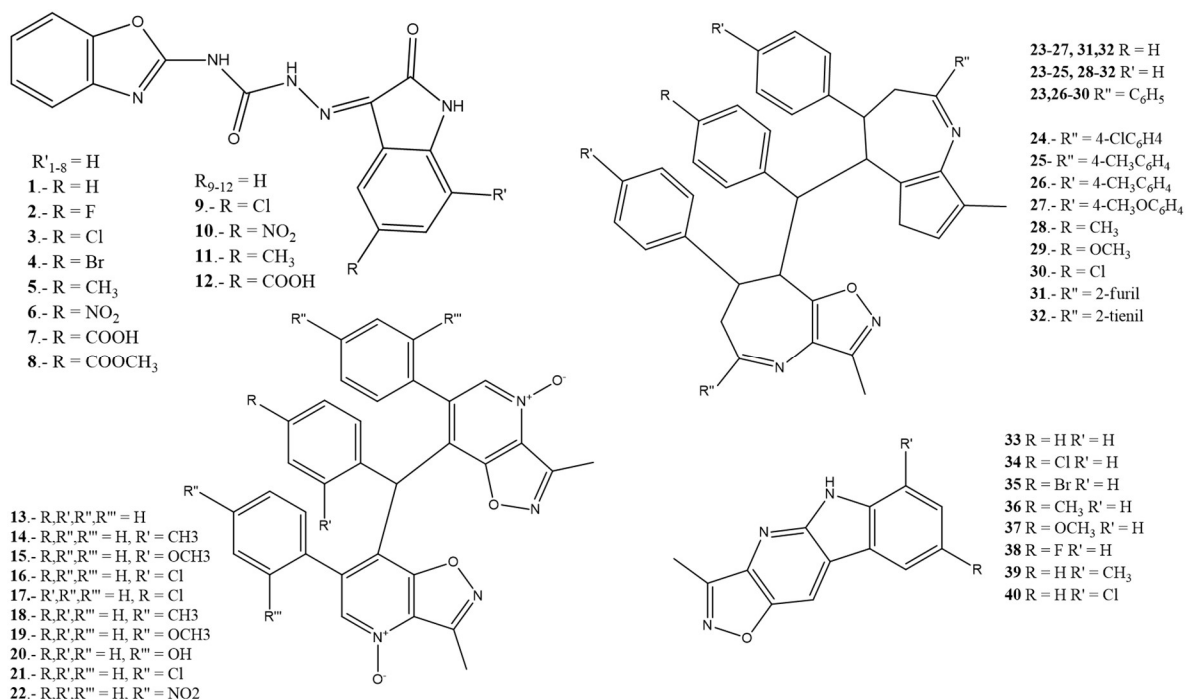


Figura 7 Compuestos de estudio. Compuestos azólicos provenientes de 4 fuentes distintas que cumplieron los criterios de inclusión del presente trabajo.

En el presente trabajo se realizó un análisis QSAR del comportamiento de **40** compuestos reportados por 4 trabajos independientes. Cuya actividad biológica, reportada como IC_{50} , fue probada, al menos, frente a las líneas tumorales de cáncer de mama MCF-7 y cáncer cervicouterino HeLa. A partir de dicha información, se generaron múltiples modelos QSAR con diferentes capacidades descriptivas y predictivas. El modelo QSAR fue usado para estudiar un total de **280 compuestos** derivados de las 4 estructuras principales de los compuestos experimentalmente estudiados. De este modo, se identificaron potenciales nuevos compuestos con un excelente comportamiento biológico teórico frente a las líneas tumorales HeLa y MCF-7, y se generó una descripción del comportamiento químico que unifica las conjeturas individuales de los trabajos independientes.

Objetivo

Objetivo General:

Analizar la actividad biológica de compuestos heterocíclicos con actividad anticancerígena efectiva frente a cáncer cervicouterino y cáncer de mama mediante el uso de modelos cualitativos/cuantitativos de análisis de correlación estructura actividad (SAR/QSAR).

Objetivos específicos:

Identificar y clasificar derivados de azoles con actividad antiproliferativa frente a las líneas tumorales humanas MCF-7 y HeLa mediante revisión de la literatura de los últimos 10 años.

Generar bases de datos de compuestos que cumplan con los parámetros de inclusión.

Calcular los descriptores moleculares para las moléculas seleccionadas

Generar los modelos de correlación cuantitativa estructura-actividad (QSAR) que describan y predigan la actividad de derivados azoles en las líneas tumorales MCF-7 y HeLa.

Identificar las moléculas teóricas que presentan la mejor actividad antiproliferativa y proponer aquellas que tendrían que ser sintetizadas.

Hipótesis

La generación de modelos SAR/QSAR permitirá describir de forma cuantitativa el comportamiento de compuestos azólicos con actividad antiproliferativa frente a líneas celulares de cáncer cervicouterino y de mama, e identificar potenciales compuestos teóricos con mejor actividad antiproliferativa

Capítulo 2

Metodología

La metodología seguida en este trabajo consta de 5 pasos principales: 1) Búsqueda bibliográfica, 2) generación de descriptores, 3) preanálisis de datos, 4) generación de modelos QSAR, 5) Predicción de actividad biológica. Los pasos se describen a continuación.

Búsqueda bibliográfica

Primera búsqueda bibliográfica

Para la búsqueda bibliográfica se usó el motor de búsqueda ScienceDirect y Taylor&Francis. Se siguieron los siguientes criterios:

- i) Espacio temporal de 10 años atrás a partir del comienzo de nuestro estudio;
- ii) Que se haya evaluado compuestos basados en azoles;
- iii) Que los artículos hayan realizado evaluación de actividad biológica en alguna línea celular cancerígena humana;
- iv) Que se haya reportado la actividad biológica en forma de IC_{50} ;
- v) Que su ensayo haya sido probado exponiendo las células tumorales al compuesto evaluado por 48 horas;
- vi) Que hayan evaluado un fármaco de referencia en las mismas condiciones experimentales (fármaco de primera elección o algún otro compuesto con

probada actividad antiproliferativa en las líneas tumorales humanas seleccionadas).

A partir de esta primera búsqueda bibliográfica. Se seleccionaron los trabajos que hayan evaluado el mismo fármaco de referencia. Conservando un total de 2 artículos de investigación reportados cuyo fármaco de referencia usado fue cisplatino.

Segunda búsqueda bibliográfica

Se realizó una búsqueda de trabajos posteriores que hayan citado a los artículos identificados. Para esta búsqueda, fue de principal interés identificar trabajos que abordaran alguno de los siguientes puntos: 1) nuevos compuestos sintetizados que cumplan los criterios de inclusión del presente trabajo; 2) trabajos que hayan abordado estudios biológicos experimentales donde se propusiese algún mecanismo de acción o; 3) que abordasen directamente algún análisis de correlación estructura-actividad cualitativo (SAR) o cuantitativo (QSAR). De esta segunda búsqueda, se identificaron 2 artículos más en donde se describen evaluaciones antitumorales de compuestos basados en azoles y que cumplieran los criterios de inclusión. Así mismo, se identificó sólo 1 trabajo que abordó un estudio SAR y ningún análisis QSAR.

Los artículos evaluaron un total de 40 compuestos basados en azoles los cuales fueron agrupados en 4 grupos: el grupo A fue conformado por las 12 moléculas probadas por [34]; el grupo B por las 10 moléculas evaluadas por [37]; el grupo C por las 10 moléculas probadas por [35]; y el grupo D por las 8 moléculas evaluadas por [36].

La Tabla 1 muestra los valores de IC_{50} de los compuestos reportados por los cuatro trabajos discutidos. Presentando los valores de IC_{50} ya reparametrizados respecto al promedio de la IC_{50} del fármaco de referencia, cisplatino, reportado por cada trabajo. Los valores extraídos directamente de los trabajos originales se muestran en el Material Suplementario 1.

Tabla 1 IC50 experimental reportada para compuestos de los grupos A, B, C y D. IC50 presentada en μM . Se homogeneizaron los valores en función de la IC50 del fármaco de referencia cisplatino.

Molécula, grupo A	IC ₅₀ HeLa	IC ₅₀ MCF-7	Molécula, grupo B	IC ₅₀ HeLa	IC ₅₀ MCF-7	Molécula, grupo C	IC ₅₀ HeLa	IC ₅₀ MCF-7	Molécula, grupo D	IC ₅₀ HeLa	IC ₅₀ MCF-7
1	40.04	43.25	13	52.84	62.30	23	100.42	105.28	33	273.46	209.88
2	16.89	21.99	14	112.32	115.65	24	80.01	84.88	34	214.21	166.92
3	15.23	19.83	15	114.62	106.20	25	77.40	82.08	35	226.64	175.27
4	14.25	18.55	16	137.19	132.42	26	74.87	73.56	36	93.13	69.54
5	19.16	25.04	17	122.80	136.47	27	61.32	75.62	37	213.79	161.00
6	26.63	32.20	18	73.80	82.30	28	55.01	62.75	38	234.63	186.59
7	21.16	25.59	19	79.63	80.86	29	92.31	83.90	39	111.55	83.21
8	29.07	36.88	20	57.24	61.92	30	77.83	85.86	40	241.95	185.43
9	23.29	27.46	21	115.01	137.05	31	49.23	56.50			
10	21.23	26.62	22	126.72	127.73	32	40.95	36.19			
11	18.82	23.25									
12	16.62	20.80									

Generación de descriptores

Moléculas evaluadas experimentalmente

Los descriptores fueron calculados mediante el software libre PAdel descriptor [24]. El software fue alimentado con información de las moléculas en el formato “.mol” generada por el software Biovia Draw for Academics [39], obtenido a partir del redibujado de las moléculas. Para el cálculo de descriptores se usó la siguiente configuración: “Standardized < Remove salt, Detect aromaticity, standarize nitro groups> y Advanced <Log, Max. threads - 1, Max. waiting Jobs -1, Max. running time per molecule -1, Retain molecules order>”. Se calcularon todos los descriptores catalogados como descriptores de información 1D y 2D, obteniendo un total de 1444 descriptores para cada uno de los compuestos, esta información fue almacenada en un archivo “.csv” para su posterior uso.

Moléculas teóricas

Se propusieron un total de 280 moléculas teóricas de potencial interés basadas en las estructuras principales de moléculas de cada grupo con una única sustitución en una de las posiciones específicas mostradas en la Figura 8. Se evaluaron 8 posiciones para moléculas basadas en la estructura principal del grupo A, 6 posiciones para el grupo B, 9 para el grupo C y 5 para el grupo D. Se analizaron 10 sustituyentes distintos para cada posición: -H, -Cl, -Br, -F, -I, -CH₃, -OCH₃, -COOH, -OH, -NO₂.

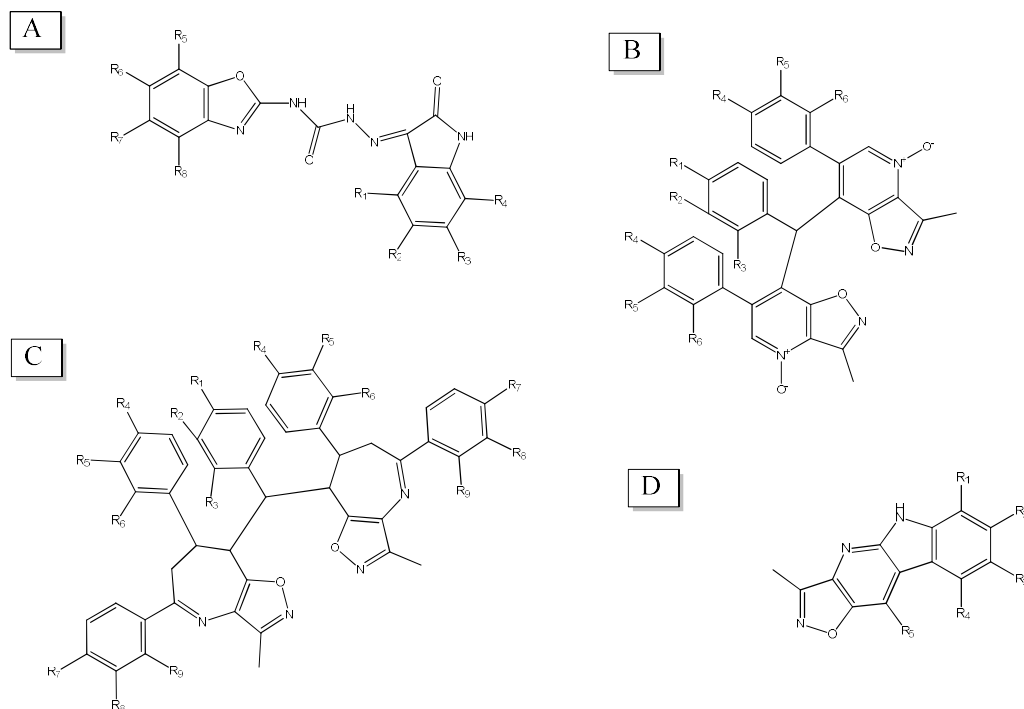


Figura 8. Estructura principal de moléculas de interés mostrando las posiciones sustituibles identificadas con la etiqueta R_n .

Para la generación de descriptores de las moléculas teóricas se siguió la misma dinámica que para la generación de descriptores de moléculas evaluadas experimentalmente. Los descriptores obtenidos fueron almacenados en un nuevo archivo “.csv”.

Preanálisis de datos

Selección de descriptores

Se realizó un prefiltrado de la información de los descriptores de las moléculas experimentales descartando a aquellos descriptores que:

- a) fueran descriptores con información nula, es decir, aquellos cuyos valores medidos para cada molécula sean iguales a 0,
- b) aquellos descriptores que presentaran una varianza de datos <0.01 o
- c) descriptores que presentaran menos del 25% de valores distintos a 0.

Entre el 45-50% de los descriptores fueron eliminadas a causa de estos criterios. Por último, debido a que se busca que los modelos [20] cuenten con capacidad predictiva, se eliminaron los descriptores conformados por conteos (descriptores 1D) directamente

relacionados con un único tipo de los sustituyentes empleados, eliminando así 12 descriptores más.

Actividad antiproliferativa experimental

Se evaluó si la actividad antiproliferativa cuantificada por los trabajos originales presentaban en conjunto una distribución normal considerando diferentes combinaciones de agrupaciones. Para mejorar la distribución de los datos y disminuir el intervalo de variación, la actividad antiproliferativa (IC_{50}) fue transformada a notación logarítmica. Para ser consistentes con la literatura del área de modelos QSAR, se reportó esta actividad de la forma $-\text{Log}(1/IC_{50})$. Para reducir la variación en las mediciones de la IC_{50} en cada trabajo original, se reparametrizaron todos los valores de IC_{50} en función del promedio del valor de la IC_{50} reportada para el cisplatino por cada uno de los artículos originales. La actividad biológica de los compuestos experimentales fue capturada y almacenada de forma independiente para cada línea tumoral en 2 archivos “.csv” para su futuro análisis. Por último, para corroborar que la actividad antiproliferativa era estadísticamente similar en ambas líneas tumorales, se analizó la homocedasticidad entre las IC_{50} reportadas por medio de la Prueba de Bartlett.

Etiquetado

Usando el mismo orden de las moléculas en los archivos de la actividad biológica y de los archivos de descriptores. Se generaron archivos independientes para las moléculas evaluadas experimentalmente y otro para las moléculas teóricas, con información sobre el grupo al que cada molécula pertenecía, el tipo de sustituyente usado y, la posición del sustituyente en su respectiva molécula.

Generación de Modelos QSAR

En cada análisis realizado, para la generación de los modelos QSAR se generaron dos conjuntos a partir de la información biológica experimental reportada. Conservando $\approx 70\%$ de los datos para la generación del modelo (*conjunto de entrenamiento*) y el restante se reservó para realizar la validación externa (*conjunto de validación*) de los modelos QSAR. Se procuró que el conjunto de entrenamiento fuese conformado a partes iguales de moléculas de cada grupo empleado. Las moléculas teóricas fueron agrupadas en un tercer conjunto denominado *conjunto de predicción*.

Los modelos QSAR fueron generados a partir de un modelo de regresión lineal múltiple. Diferentes enfoques fueron usados para la selección de modelos QSAR. Un primer enfoque priorizó la selección de modelos basados en descriptores con el mejor ajuste a los datos experimentales conocidos para el **conjunto de entrenamiento** y el **conjunto de validación**. Mientras que un segundo enfoque contempló el significado químico de los descriptores empleados en la generación de los modelos QSAR. Así mismo, se realizaron diferentes análisis QSAR contemplando 2 o más grupos de compuestos a la vez. Cada análisis QSAR fue repetido para la actividad biológica de los compuestos frente a cada línea tumoral. Como referencia, la Tabla 2 muestra la descripción de los descriptores usados a lo largo del presente trabajo.

Tabla 2 Significado de descriptores usados en las distintas aproximaciones usadas en la generación de modelos QSAR.

Número	Descriptor	Significado
1	SpMin5_Bhp	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative polarizabilities
2	SpMin5_Bhv	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative van der Waals volumes
3	SpMin5_Bhe	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative Sanderson electronegativities
4	SpMin5_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative first ionization potential
5	minHBa	Minimum E-States for (strong) Hydrogen Bond acceptors. Electrotopological State Atom Type descriptor.
6	AATSC3v	Average centered Broto-Moreau autocorrelation - lag 3 / weighted by van der Waals volumes
7	ATSC6e	Centered Broto-Moreau autocorrelation - lag 6 / weighted by Sanderson electronegativities
8	MATS6c	Moran autocorrelation - lag 6 / weighted by charges
9	MATS7e	Moran autocorrelation - lag 7 / weighted by Sanderson electronegativities
10	Si	Sum of first first ionization potentials (scaled on carbon atom)
11	apol	Sum of the atomic polarizabilities (including implicit hydrogens)
12	McGowan_volume	McGowan characteristic volume
13	Zagreb	Sum of the squares of atom degree over all heavy atoms i
14	WPOL	Weiner polarity number
15	Kier1	First kappa shape index
16	Kier2	Second kappa shape index
17	MDEN-22	Molecular distance edge between all secondary nitrogen
18	LipinskyFailures	Number failures of the Lipinski's Rule Of 5
19	MLogP	Mannhold LogP
20	MWC2(3, 4, ..., 10)	Molecular walk count of order 2 (3, 4, ..., 10) (ln(1+x))
21	RingCountDescriptor	Number of rings whit specific characteristics

Modelos QSAR AB.

Los **modelos QSAR AB**, fueron realizados contemplando únicamente la información disponible para los grupos A y B por ser los grupos con los compuestos de mejor actividad biológica (IC_{50} experimentales más bajas). Se realizaron dos aproximaciones que contemplaron como criterio de mayor importancia a las características estadísticas de los modelos o la simplicidad de los descriptores usados.

Modelos AB. Selección modelos QSAR de mejor característica estadística

Para la identificación de los mejores modelos matemáticos se programó un algoritmo que calculase toda posible combinación entre dos descriptores para la generación de un ajuste de regresión lineal hacia la IC_{50} de los compuestos [20], [40]. Para ello, dos criterios principales fueron contemplados para la elección de los mejores modelos. El primer criterio fue usado para asegurar que los descriptores usados en los modelos no fuesen colineales entre ellos (es decir, que no estuvieran correlacionados), lo cual se mide estadísticamente por el coeficiente de correlación de Pearson, limitando a un valor no mayor a 0.7 ($R < 0.7$). El segundo criterio contempla un buen ajuste entre los valores calculados por los modelos de regresión lineal y sus correspondientes valores experimentales. Se usó el coeficiente de determinación ajustado (R^2) para este criterio y un límite de 0.85 ($R^2 > 0.85$). Este criterio establece un límite inferior del ajuste mínimo aceptable para contemplar a los modelos QSAR semifinalistas.

Con la intención de corroborar su capacidad predictiva, se realizó la validación predictiva de cada modelo calculando el coeficiente de determinación entre los valores calculados para las moléculas del set de validación y sus correspondientes valores experimentales (R^2_{test}). Si bien existen trabajos publicados que prescinden de este ajuste [41], en el presente trabajo se consideraron válidos aquellos modelos cuyo $R^2_{test} > 0.6$ de acuerdo con el límite referido por [20], para los conjuntos de validación externa, límite que es considerado como el valor mínimo para asegurar una buena predicción de nuevos datos.

Por último, se enlistaron en orden descendente a las combinaciones de modelos QSAR en función de sus correspondientes valores de R^2 . Limitándonos a los primeros 50 mejores modelos, se realizó lo siguiente:

- Se analizó la prevalencia de descriptores presentes en la mayoría de los modelos QSAR obtenidos. Identificando a descriptores de la familia <Burden Matrix eigenvalues> y al descriptor minHBa como los principales descriptores presentes en los modelos con mejores estadísticas.
- Los descriptores de la familia <SpMXX_BhX> presentan modificación en el número y ponderación usada para su cálculo. Contemplando esto, se analizó la prevalencia del número y ponderación más común entre los modelos cuyo segundo descriptor fuese minHBa.
- Se repitió todo el proceso para evaluar los modelos QSAR obtenidos para la actividad biológica frente a la segunda línea tumoral.

Se seleccionaron 4 modelos QSAR finalistas de acuerdo con las siguientes condicionantes. En primer lugar, se seleccionaron a los primeros dos mejores modelos QSAR por presentar la misma combinación de descriptores para describir la actividad biológica tanto en HeLa como en MCF-7. En segundo lugar, se seleccionaron las combinaciones cuyo número y ponderación fuesen los más comunes.

Modelos AB. Selección de modelos QSAR de explicación simple

Se identificaron los descriptores que estaban linealmente correlacionados con los cuatro descriptores de la familia SpMin5_BhX que conformaron los modelos QSAR de mejor característica estadística. A partir de ello, se seleccionaron arbitrariamente aquellos 10 descriptores cuya descripción química fuera lo más simple posible. Posteriormente, se identificaron las combinaciones de estos descriptores con el descriptor minHBa que generaron modelos QSAR con $R^2 > 0.9$ y $R^2_{\text{test}} > 0.6$.

Modelos QSAR ABCD.

Un último análisis se llevó a cabo contemplando la información de los 4 grupos de compuestos encontrados que cumplieron los criterios de selección. Se generaron sus respectivos conjuntos de entrenamiento y validación siguiendo las mismas pautas que las utilizadas para la generación de los **modelos QSAR AB**. Así mismo, se utilizaron los mismos criterios de colinealidad entre descriptores ($R < 0.7$) y correlación con el modelo final ($R^2 > 0.85$) usados para las aproximaciones anteriores. Para la elección final de modelos QSAR son de interés aquellos con los mejores valores estadísticos. Finalmente, se analizó el

significado matemático de los descriptores presentes en los *modelos QSAR ABCD* finalistas, con la intención de dar un significado químico lo más simple posible.

Predicción de actividad antiproliferativa de moléculas teóricas y significado químico de descriptores

Usando los modelos QSAR finalistas de cada aproximación, se calculó el valor predicho de las moléculas teóricas de los respectivos *conjuntos de predicción*. Para ello se usó la información de los correspondientes descriptores de cada uno de dichos modelos QSAR. Por otro lado, se analizó el significado de cada descriptor identificado como parte de estos modelos QSAR finalistas con la intención de generar una explicación descriptiva basada en los resultados de los modelos QSAR obtenidos. Y en conjunto, se analizaron los cambios en los valores de dichos descriptores a partir de los cambios de sustituyente y su posición en cada molécula generada. De esta forma, se esboza una nueva teoría del funcionamiento de los compuestos evaluados y se plantean posibles nuevas modificaciones a dichos compuestos.

Capítulo 3

Resultados

Se identificaron un total de 4 trabajos que cumplieron los criterios establecidos para la búsqueda bibliográfica. Los compuestos que cumplieron los criterios de inclusión para este trabajo fueron agrupados en 4 grupos: grupo A [34], grupo B [37], grupo C [35] y grupo D [36]. Cada grupo está conformado por compuestos derivados de una misma estructura principal, todos conteniendo al menos una unidad azol en su estructura. Las moléculas originales se muestran en la Figura 7 y sus respectivas IC_{50} se enlistan en la Tabla 1. La Figura 8 muestra las estructuras principales y sus respectivas sustituciones planteadas para cada grupo de moléculas planteadas en el presente trabajo (ver sección de introducción). Los resultados de esta sección abordan los modelos QSAR obtenidos mediante las distintas aproximaciones probadas. La primera aproximación fue estudiar los compuestos del grupo A y del grupo B por ser los grupos de compuestos con mejor actividad biológica general. Mientras que la segunda aproximación tuvo por finalidad evaluar la actividad biológica de los cuatro grupos bajo un mismo modelo QSAR con la intención de ser enfocado para la búsqueda de nuevas estructuras principales.

Modelos QSAR AB complejos

De acuerdo con la aproximación seguida para los *modelos QSAR AB* descrita en la metodología, se identificaron 4 modelos QSAR (1.1 – 1.4) con las mejores estadísticas posibles para el modelado de actividad biológica de los compuestos de interés frente a las líneas tumorales de cáncer cervicouterino (HeLa) y de mama (MCF-7). La Tabla 3 muestra estos 4 modelos matemáticos usados para describir la capacidad antiproliferativa de los compuestos de los grupos A y B.

Tabla 3 Ecuación y estadísticas de los Modelos QSAR AB obtenidos mediante esta aproximación.

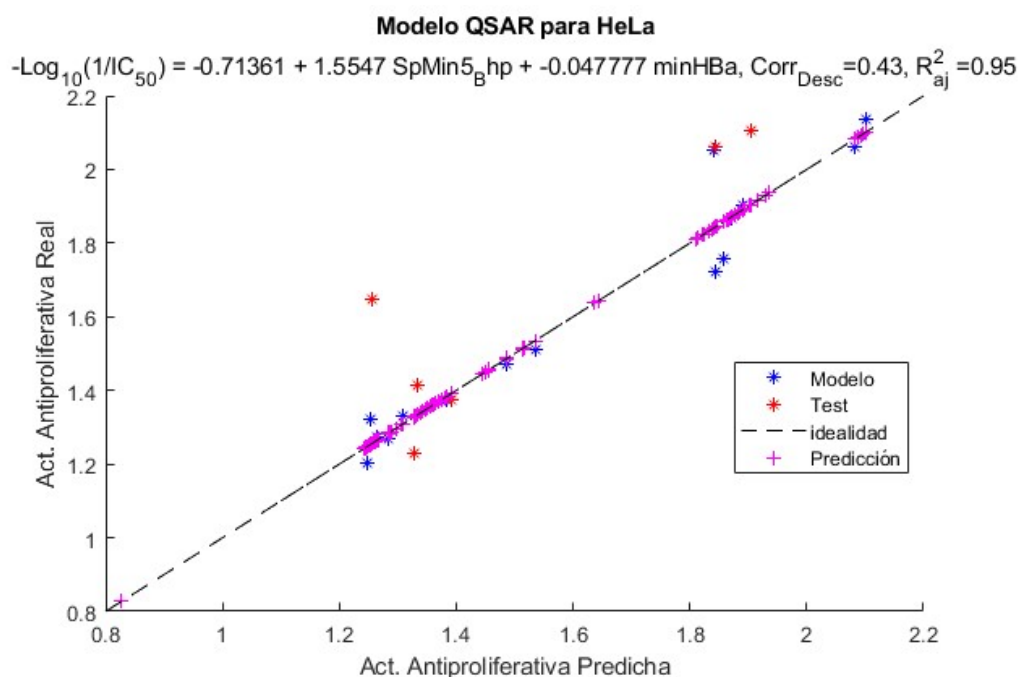
ID	
1.1 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = -0.714 + 1.555 SpMin5_{Bhp} - 0.048 minHBa$ $R_1 = 0.428, \quad R^2 = 0.954, \quad R^2_{aj} = 0.947, \quad Fstat = 135.751,$ $p - valor = 1.94572 E^{-9}, RMSE = 0.078,$ $R^2_{test} = 0.778 R^2_{aj_{test}} = 0.722$
1.1 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = -0.285 + 1.328 SpMin5_{Bhv} - 0.047 minHBa$ $R_1 = 0.427, \quad R^2 = 0.962, R^2_{aj} = 0.956, \quad Fstat = 162.973,$ $p - valor = 6.234 E^{-10}, RMSE = 0.068$ $R^2_{test} = 0.817 R^2_{aj_{test}} = 0.771$
1.2	$HeLa - Log\left(\frac{1}{IC_{50}}\right) = -0.404 + 1.388 SpMin5_{Bhv} - 0.047 minHBa$ $R_1 = 0.427, \quad R^2 = 0.952, \quad R^2_{aj} = 0.945, \quad Fstat = 128.812,$ $p - valor = 2.69309 E^{-9}, RMSE = 0.080$ $R^2_{test} = 0.788 R^2_{aj_{test}} = 0.735$
	$MCF-7 - Log\left(\frac{1}{IC_{50}}\right) = -0.578 + 1.485 SpMin5_{Bhp} - 0.047 minHBa$ $R_1 = 0.428, \quad R^2 = 0.961, R^2_{aj} = 0.955, \quad Fstat = 160.185,$ $p - valor = 6.944 E^{-10}, RMSE = 0.068$ $R^2_{test} = 0.807 R^2_{aj_{test}} = 0.759$
1.3	$HeLa - Log\left(\frac{1}{IC_{50}}\right) = 0.316 + 1.060 SpMin5_{Bhe} - 0.047 minHBa$ $R_1 = 0.426, \quad R^2 = 0.941, R^2_{aj} = 0.932, \quad Fstat = 103.879,$ $p - valor = 1.01201 E^{-8}, RMSE = 0.088$ $R^2_{test} = 0.771 R^2_{aj_{test}} = 0.714$
	$MCF - 7 - Log\left(\frac{1}{IC_{50}}\right) = 0.403 + 1.015 SpMin5_{Bhe} - 0.047 minHBa$ $R_1 = 0.426, \quad R^2 = 0.952, R^2_{aj} = 0.944, \quad Fstat = 128.314,$ $p - valor = 2.758 E^{-9}, RMSE = 0.076$ $R^2_{test} = 0.798 R^2_{aj_{test}} = 0.747$
1.4	$HeLa - Log\left(\frac{1}{IC_{50}}\right) = -0.578 + 1.485 SpMin5_{Bhi} - 0.047 minHBa$ $R_1 = 0.425, \quad R^2 = 0.940, R^2_{aj} = 0.930, \quad Fstat = 101.171,$ $p - valor = 1.18933 E^{-8}, RMSE = 0.089$ $R^2_{test} = 0.753 R^2_{aj_{test}} = 0.691$
	$MCF - 7 - Log\left(\frac{1}{IC_{50}}\right) = 0.450 + 1.000 SpMin5_{Bhi} - 0.046 minHBa$ $R_1 = 0.425, \quad R^2 = 0.950, R^2_{aj} = 0.942, \quad Fstat = 123.805,$ $p - valor = 3.441 E^{-9}, RMSE = 0.077$ $R^2_{test} = 0.78 R^2_{aj_{test}} = 0.725$

Cabe destacar que los modelos QSAR de la tabla anterior están compuestos de la siguiente forma: una variante de la *familia de descriptores Burden* con el coeficiente de mayor contribución en combinación con el descriptor *minHBa* con el coeficiente de menor contribución. Las variantes de la *familia de descriptores Burden* siempre muestra una contribución positiva, mientras que el descriptor *minHBa* siempre muestra una contribución

negativa. Esto indica que, para obtener compuestos con una actividad biológica más potente posible, es necesario que la molécula presente bajos valores en las variantes de la **familia de descriptores Burden** al mismo tiempo que se obtenga valores altos para el descriptor minHBa (Ver Material Suplementario 13).

De acuerdo con estas ecuaciones, las propiedades químicas principales para una buena actividad biológica están relacionadas con el volumen de van Der Waals, la polarizabilidad, la electronegatividad y con el potencial de ionización de las moléculas (ponderaciones v, p, e, i). Estas propiedades son independientes de la línea tumoral estudiada.

La Figura 9 muestra los gráficos de los **modelos QSAR AB 1.1**, los cuales presentan el mejor ajuste a los datos experimentales de la actividad biológica de los compuestos del grupo A y B frente a las líneas tumorales HeLa y MCF-7, respectivamente. Los **modelos QSAR AB 1.1** muestran un excelente ajuste a los datos del **conjunto de entrenamiento**, superior a 0.95 en ambos casos. Al mismo tiempo, presentan un buen ajuste para los compuestos pertenecientes al **conjunto de validación**, siendo en ambos casos superior a 0.7. Lo anterior habla sobre una buena capacidad para predecir la actividad biológica de compuestos teóricos.



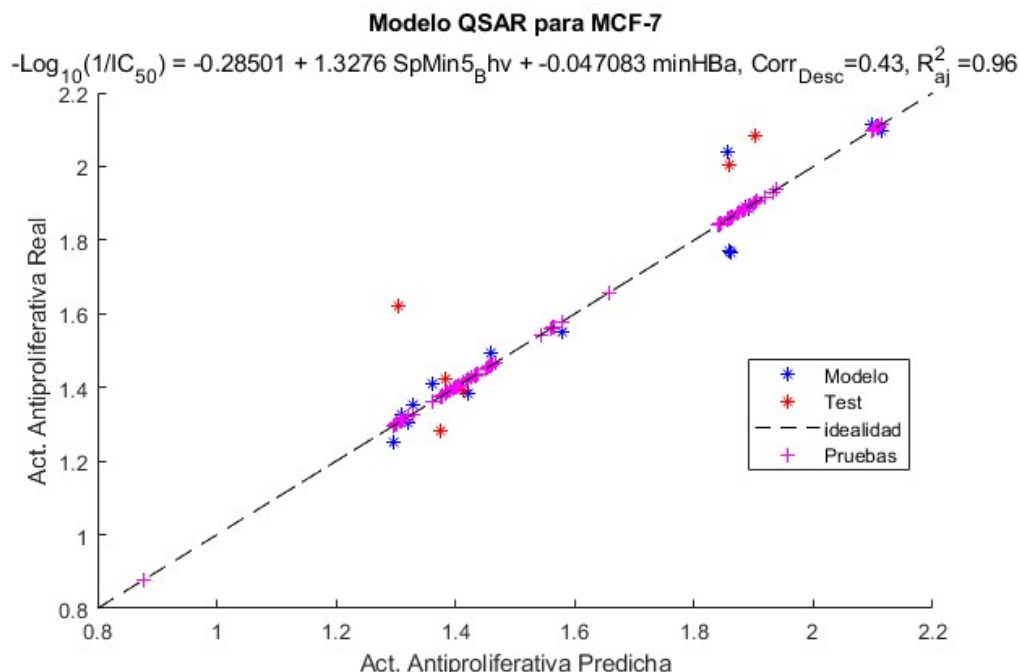


Figura 9. Modelos QSAR AB 1.1 complejos de la actividad biológica de los compuestos frente a las líneas tumorales HeLa y MCF-7. IC₅₀ representada en forma $-\text{Log}(1/\text{IC}_{50})$.

La Tabla 4 R# muestra un resumen de las características de los **modelos QSAR AB 1.1**. En el Material Suplementario 2 y 3 se muestra los valores de IC₅₀ frente a MCF-7 y HeLa, calculados para compuestos del conjunto de entrenamiento y de validación evaluados por los **modelos QSAR AB (1.1-1.4)**. muestra una comparación en escala lineal de los valores de IC₅₀ experimentales vs los predichos por estos modelos QSAR.

Tabla 4. Valores de IC₅₀ teóricos y experimentales de los compuestos usados en el entrenamiento y validación de los modelos QSAR AB 1.1.

Modelo	Conjunto de entrenamiento		Conjunto de Validación		Compuestos más activos		
	D.A.M.* (%) grupo A**)	D.A.M.* (%) ambos grupos)	D.A.M.* (%) grupo A**)	D.A.M.* (%) ambos grupos)	Mejor (IC ₅₀ [μM])	2do mejor (IC ₅₀ [μM])	3er mejor (IC ₅₀ [μM])
AB 1.1 HeLa	6.82	10.36	15.79	30.57	4 (15.69)	11 (15.94)	2 (16.36)
AB 1.1 MCF-7	7.33	9.45	12.58	25.20	4 (20.49)	2 (21.07)	12 (21.72)

*D.A.M. = Diferencia absoluta media, calculado como el promedio de la diferencia absoluta entre valor predicho y experimental, dividida por el valor experimental de la actividad antiproliferativa
 ** valor calculado excluyendo al compuesto **1** por ser un compuesto no sustituido.

Los modelos presentan una discrepancia promedio respecto a los valores experimentales de 10.36% para el conjunto de entrenamiento mientras que para el conjunto de validación asciende al 30.57%. Presenta particular problema con la predicción de los compuestos **14** del conjunto de entrenamiento, y **1**, **15**, **22**, del conjunto de validación para ambas líneas celulares (ver Material

Suplementario 4 y 5). Son estos compuestos, sin sustituir y sustituidos con -OCH₃, -CH₃ y -NO₂ los que presentan mayor problema al predecir su IC₅₀, no obstante, el modelo se mantiene con un ajuste $R^2 > 0.90$ debido a la precisión con la que se predice la actividad antiproliferativa del resto de compuestos evaluados. Como se discutirá más adelante, la discrepancia generada por el compuesto 1 se atribuye a la ausencia de los sustituyentes estudiados en el farmacóforo, mientras que las diferencias en los compuestos 14, 15 y 22 resultan sobreestimaciones que pueden ser mejor resueltas mediante los *modelos QSAR ABCD*.

Modelos QSAR AB simples

Se encontró un total de 280 descriptores correlacionados con los 4 descriptores de la *familia de descriptores Burden* que conforman los *modelos QSAR AB complejos*. Se identificaron alrededor de 100 descriptores con información que permite dar una explicación lo más sencilla posible al comportamiento de las moléculas estudiadas. A partir de esta información, se generaron modelos QSAR 1.2-1.10 a partir de alguno de los 9 descriptores que mostraron una buena relación directa entre sus valores y la actividad biológica, y se seleccionaron aquellos modelos QSAR que presentaran las mejores características estadísticas. La Tabla 5 muestra los modelos QSAR simples identificados a partir del análisis previo.

Tabla 5 Ecuaciones y descripción estadística de modelos QSAR simples.

ID	Modelo QSAR
1.5 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.563 - 0.063 * minHBa + 0.003 * Si$ $R1 = 0.498; R^2 = 0.939; R_{aj}^2 = 0.929 Fstat = 99.623; p - valor = 1.3067e - 08; RMSE = 0.09$ $R_{test}^2 = 0.768 R_{aj_{test}}^2 = 0.71$
1.5 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.644 - 0.062 * minHBa + 0.003 * Si$ $R1 = 0.498; R^2 = 0.941; R_{aj}^2 = 0.931 Fstat = 102.757; p - valor = 1.0815E^{-0}; RMSE = 0.084$ $R_{test}^2 = 0.795 R_{aj_{test}}^2 = 0.743$
1.6 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.628 - 0.051 * minHBa + 0.018 * apol$ $R1 = 0.448; R^2 = 0.932; R_{aj}^2 = 0.921 Fstat = 88.642; p - value = 2.6577e - 08; RMSE = 0.095$ $R_{test}^2 = 0.758 R_{aj_{test}}^2 = 0.697$
1.6 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.704 - 0.051 * minHBa + 0.017 * apol$ $R1 = 0.448; R^2 = 0.937; R_{aj}^2 = 0.928 Fstat = 97.274; p - value = 1.5114e - 08; RMSE = 0.086$ $R_{test}^2 = 0.784 R_{aj_{test}}^2 = 0.73$
1.7 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.521 + 0.395 * McGowan_{volume} - 0.047 * minHBa$ $R1 = 0.428; R^2 = 0.93; R_{aj}^2 = 0.919 Fstat = 86.198; p - value = 3.1475e - 08; RMSE = 0.096$ $R_{test}^2 = 0.749 R_{aj_{test}}^2 = 0.687$

1.7 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.602 + 0.377 * McGowan_{Volume} - 0.046 * minHBa$ $R1 = 0.428; R^2 = 0.936; R^2_{aj} = 0.926 Fstat = 94.462; p - value = 1.8069e - 08; RMSE = 0.088$ $R^2_{test} = 0.78 R^2aj_{test} = 0.725$
1.8 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.477 - 0.047 * minHBa + 0.007 * Zagreb$ $R1 = 0.431; R^2 = 0.93; R^2_{aj} = 0.919 Fstat = 85.754; p - value = 3.2474e - 08; RMSE = 0.096$ $R^2_{test} = 0.751 R^2aj_{test} = 0.689$
1.8 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.561 - 0.047 * minHBa + 0.007 * Zagreb$ $R1 = 0.431; R^2 = 0.934; R^2_{aj} = 0.924 Fstat = 92.146; p - value = 2.101e - 08; RMSE = 0.089$ $R^2_{test} = 0.787 R^2aj_{test} = 0.734$
1.9 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.79 - 0.049 * minHBa + 0.017 * WPOL$ $R1 = 0.436; R^2 = 0.929; R^2_{aj} = 0.918 Fstat = 85.121; p - value = 3.3959e - 08; RMSE = 0.097$ $R^2_{test} = 0.76 R^2aj_{test} = 0.7$
1.9 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.86 - 0.048 * minHBa + 0.016 * WPOL$ $R1 = 0.436; R^2 = 0.931; R^2_{aj} = 0.921 Fstat = 88.309; p - value = 2.719e - 08; RMSE = 0.09$ $R^2_{test} = 0.793 R^2aj_{test} = 0.742$
1.10 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.337 + 0.057 * Kier1 - 0.048 * minHBa$ $R1 = 0.434; R^2 = 0.926; R^2_{aj} = 0.915 Fstat = 81.503; p - value = 4.4126e - 08; RMSE = 0.099$ $R^2_{test} = 0.729 R^2aj_{test} = 0.661$
1.10 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.428 + 0.055 * Kier1 - 0.047 * minHBa$ $R1 = 0.434; R^2 = 0.928; R^2_{aj} = 0.917 Fstat = 83.844; p - value = 3.7204e - 08; RMSE = 0.093$ $R^2_{test} = 0.768 R^2aj_{test} = 0.71$
1.11 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 0.274 + 0.14 * Kier2 - 0.049 * minHBa$ $R1 = 0.439; R^2 = 0.924; R^2_{aj} = 0.913 Fstat = 79.317; p - value = 5.1965e - 08; RMSE = 0.1$ $R^2_{test} = 0.754 R^2aj_{test} = 0.693$
1.11 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 0.37 + 0.134 * Kier2 - 0.048 * minHBa$ $R1 = 0.439; R^2 = 0.924; R^2_{aj} = 0.913 Fstat = 79.58; p - value = 5.0942e - 08; RMSE = 0.095$ $R^2_{test} = 0.791 R^2aj_{test} = 0.739$
1.12 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 2.156 - 0.232 * MDEN - 22 - 0.046 * minHBa$ $R1 = -0.426; R^2 = 0.916; R^2_{aj} = 0.903 Fstat = 70.68; p - value = 1.0355e - 07; RMSE = 0.105$ $R^2_{test} = 0.795 R^2aj_{test} = 0.744$
1.12 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 2.164 - 0.222 * MDEN - 22 - 0.045 * minHBa$ $R1 = -0.426; R^2 = 0.924; R^2_{aj} = 0.912 Fstat = 78.746; p - value = 5.4271e - 08; RMSE = 0.095$ $R^2_{test} = 0.819 R^2aj_{test} = 0.774$
1.13 HeLa	$-Log\left(\frac{1}{IC_{50}}\right) = 1.445 + 0.682 * LipinskiFailures - 0.046 * minHBa$ $R1 = 0.426; R^2 = 0.916; R^2_{aj} = 0.903 Fstat = 70.68; p - value = 1.0355e - 07; RMSE = 0.105$ $R^2_{test} = 0.795 R^2aj_{test} = 0.744$
1.13 MCF-7	$-Log\left(\frac{1}{IC_{50}}\right) = 1.484 + 0.652 * LipinskiFailures - 0.045 * minHBa$ $R1 = 0.426; R^2 = 0.924; R^2_{aj} = 0.912 Fstat = 78.746; p - value = 5.4271e - 08; RMSE = 0.095$ $R^2_{test} = 0.819 R^2aj_{test} = 0.774$

Estos modelos QSAR están ligados directamente a características químicas más simples, por lo que, la descripción del comportamiento de estos compuestos mediante alguno de estos modelos QSAR permite una descripción más simple a costa de una precisión más baja respecto a los *modelos QSAR AB 1.1-1.4*. Por esta razón, los *modelos QSAR AB 1.5-1.13* pueden ser usados como modelos descriptivos. En el Material Suplementario 4-7 se muestra los valores predichos para compuestos del grupo de entrenamiento y validación frente a HeLa y MCF-7.

MODELOS QSAR ABCD.

La Figura 10 muestra el modelo QSAR para la actividad antiproliferativa en MCF-7 de los grupos A, B, C y D evaluados bajo un mismo modelo QSAR.

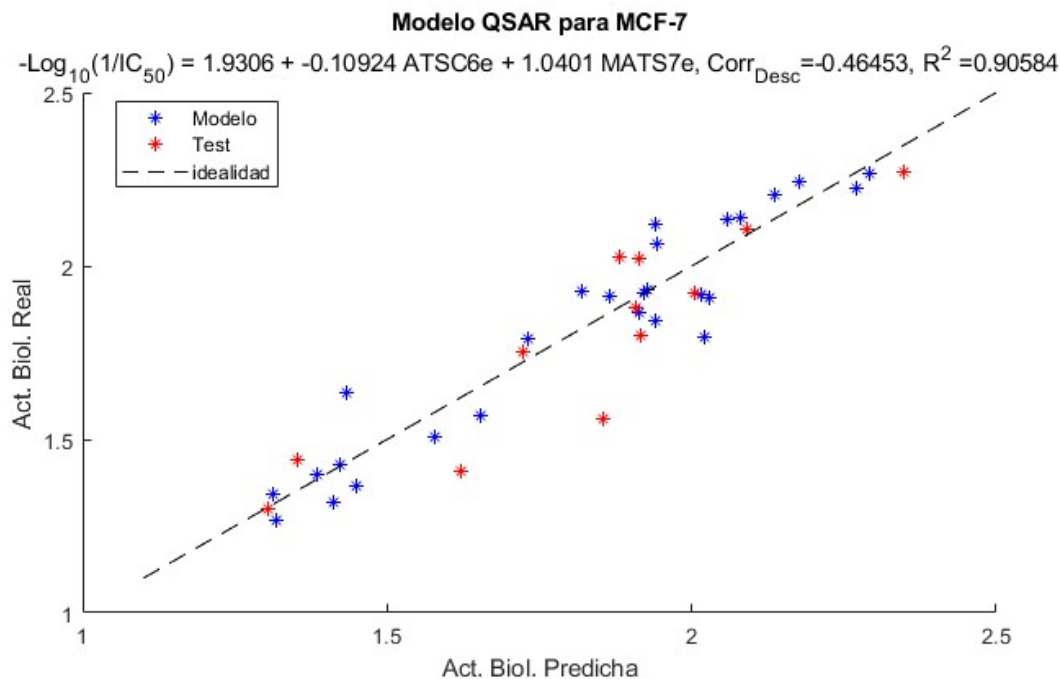


Figura 10. Modelo QSAR de la actividad biológica en MCF-7 generado a partir de los 4 grupos de compuestos.

El **modelo QSAR ABCD MCF-7** (Figura 10) fue generado inicialmente con el conjunto de entrenamiento original. Sin embargo, al evaluar la capacidad predictiva con dicho conjunto de validación (ver Material Suplementario 8-9), se identificó como caso aislado que la predicción del compuesto base del grupo D (compuesto **33**) sale de la tendencia de los valores predichos por el modelo, razón por la que el ajuste del conjunto de validación fuese tan bajo ($R^2_{\text{test}} = 0.46$). El compuesto **33** puede aislarse de futuros cálculos

al considerarse con comportamiento anómalo. Al descartarlo del cálculo de R^2_{test} , se obtuvo un valor de $R^2_{\text{test}} = 0.852$. Lo anterior indica que, para el resto de las moléculas de dicho conjunto, la predicción supera el valor mínimo para considerar una capacidad predictiva por parte del modelo. La predicción de los valores de IC_{50} con este modelo es de más del 80% de precisión. El descarte del compuesto **33** del conjunto de validación no elimina la validez matemática para la predicción, puesto que el nuevo conjunto con 12 compuestos aún mantiene al menos el 30% de los datos considerados (27 del conjunto de entrenamiento + 12 del conjunto de validación) para esta aproximación

En lo que respecta a la actividad antiproliferativa predicha por esta aproximación para la línea tumoral HeLa, el modelo QSAR respectivo se muestra en la Figura 11.

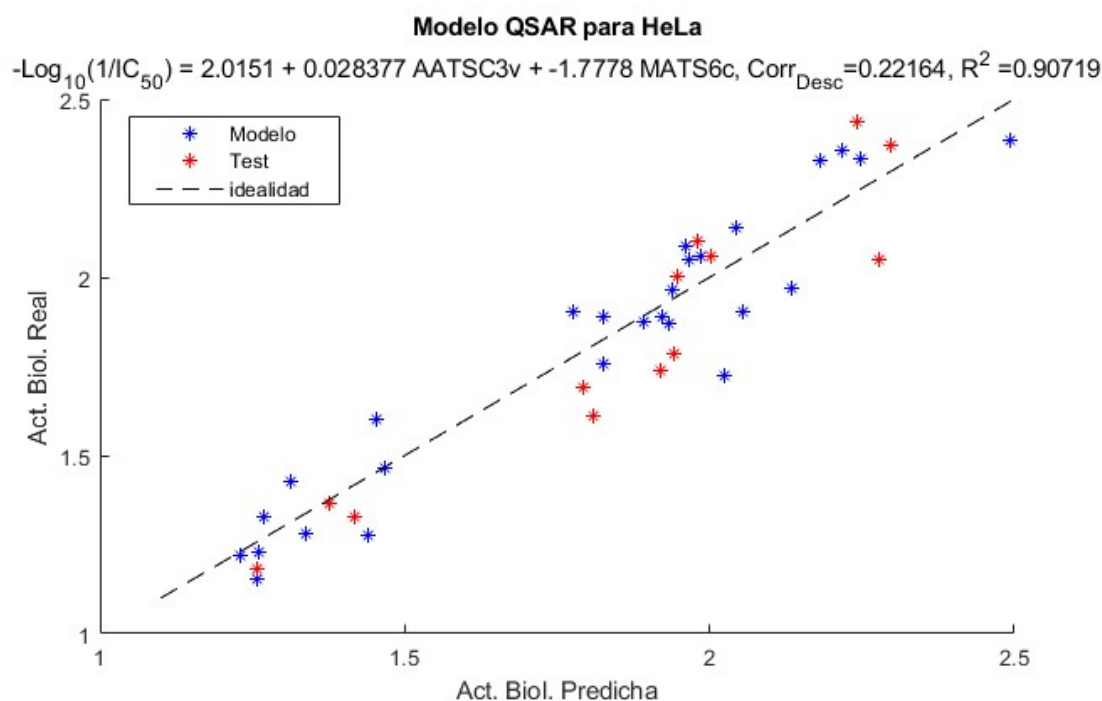


Figura 11 Modelo QSAR de la actividad biológica en HeLa generado a partir de los 4 grupos de compuestos.

El **modelo QSAR ABCD HeLa** modela el comportamiento de los compuestos pertenecientes a los 4 grupos estudiados. El modelo emplea dos variantes de los mismos descriptores usados en el modelo análogo para MCF-7, *AATSC3v* y *MATS6c*, por lo que, matemáticamente hablando, la IC_{50} puede ser modelada por los mismos principios que los usados para discutir el comportamiento frente a MCF-7. A diferencia de su versión para

modelar el comportamiento en MCF-7, no presentó problema para describir el comportamiento del compuesto **33** ni ningún otro como lo observado en su versión para MCF-7. El modelo presentó un ajuste para el conjunto de validación de $R^2_{\text{test}} = 0.748$. De este modo, tanto para MCF-7 como para HeLa, los **modelos QSAR ABCD** tienen capacidad predictiva.

Predicción de moléculas teóricas basadas en los farmacóforos originales

La capacidad predictiva de los principales modelos QSAR generados fue validada al determinar el coeficiente de correlación entre la actividad biológica real y la predicha para los respectivos *conjuntos de validación* de cada modelo (resumidos posteriormente en la Tabla 11). Habiendo superado el valor mínimo de R^2_{test} los **modelos QSAR AB 1.1 y ABCD** para ambas líneas celulares, se procedió a realizar la predicción de compuestos basados en los farmacóforos de los 4 grupos de compuestos.

Se usaron como referencia a los **modelos QSAR ABCD** por ser los más robustos en cuestión de mayor número de compuestos usados para generar los modelos y mejor predicción para sus respectivos sets de validación. La Figura 12 muestra una representación en escala de colores al cambio provocado por cada sustituyente en cada posición específica.

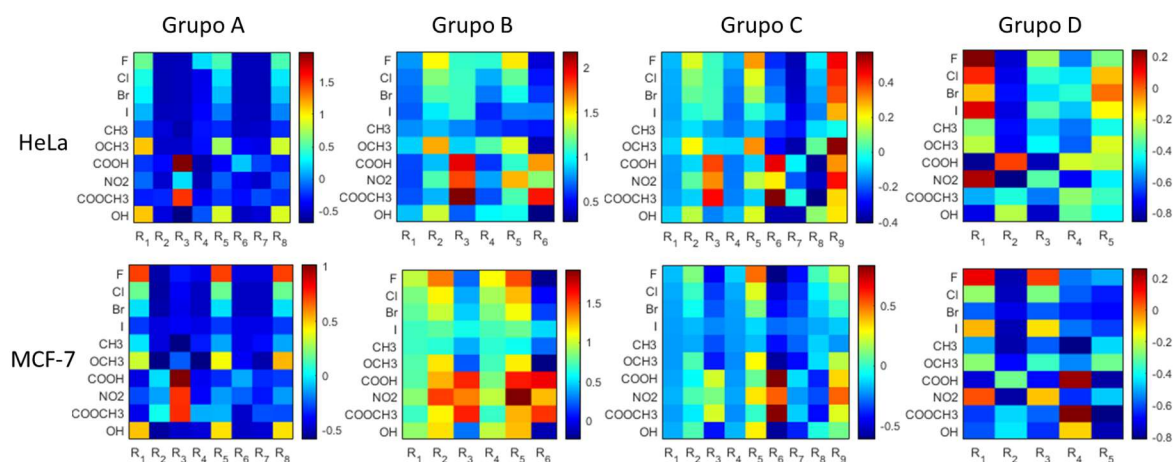


Figura 12 Cambio relativo de IC50 para cada posición y sustituyente evaluado respecto a cada farmacóforo de cada grupo de compuestos, representado mediante un mapa de calor. Colores más azules y oscuros representan cambios más drásticos asociados con una mejor actividad antiproliferativa.

Con base en los **modelos QSAR ABCD**, se identifica que el comportamiento general de los sustituyentes resulta muy similar entre varios de los sustituyentes. Además, no

discriminando entre las características particulares de cada sustituyente, las posiciones efectivas tanto en MCF-7 como en HeLa son:

- las posiciones R₂, R₃, R₆ y R₇ para el grupo A
- R₁, R₆ para el grupo B
- R₆ y R₇ para el grupo C y,
- R₂ para el grupo D.

Cabe destacar que algunas posiciones favorecen a conjuntos de sustituyentes específicos. Esto es observable al comparar los cambios de IC₅₀ observadas en ambas líneas celulares con sustituyentes en posiciones tales como R₃ de los grupos A, B y C. Los sustituyentes halógenos muestran cambios que favorecen la actividad antiproliferativa respecto a lo mostrado para sustituyentes como COOH, NO₂, y COOCH₃. La diferencia de comportamiento en función de la posición y el sustituyente se abordará a profundidad en diferentes secciones del Capítulo 4.

En términos generales, los sustituyentes halógenos parecen mostrar una mejor actividad antiproliferativa respecto al resto de sustituyentes, en varias posiciones estudiadas (R₂, R₃, R₆ y R₇ del grupo A, R₆ del B, R₇ del C y R₂ del D), concordando con diversos estudios relacionados con azoles [42]. Sin embargo, se encontraron posiciones donde sustituyentes no halógenos resultan efectivos para mejorar la actividad antiproliferativa, principalmente en posiciones R₁, R₂ y R₄ del grupo A, R₃ y R₆ del B, R₆₋₈ del C y, R₂ y R₅ del D. Estos resultados amplían las suposiciones generadas por los autores en sus trabajos originales, ya que dicha actividad no se ve favorecida principal o únicamente en función de la electronegatividad del sustituyente, sino que depende de la posición en la que se encuentre y las cualidades asociadas a dicha posición en términos del volumen de van Der Waals, cargas, y electronegatividad, principalmente. Esto se discutirá a fondo en el capítulo 4.

Algunos de los compuestos teóricos presentan mejor actividad teórica que la mejor actividad antiproliferativa reportada por sus respectivos trabajos originales. Unos cuantos de estos compuestos presentan mejor actividad biológica que el compuesto más activo evaluado experimentalmente. Las tablas 6 y 7 muestran las mejores IC₅₀ teóricas obtenidas para cada sustituyente de cada uno de los 4 farmacóforos calculada por los *modelos QSAR AB o ABCD*.

Tabla 6. Mejor IC₅₀ teórica posible obtenida por cada sustituyente para los diferentes grupos. HeLa y MCF-7 QSAR AB.

		Mejor IC ₅₀ de cada sustituyente para todas las moléculas teóricas.									
Línea tumoral	Grupo	F	Cl	Br	I	CH ₃	OCH ₃	COOH	OH	NO ₂	COOCH ₃
HeLa	A	(R ₄) 5.71	(R _{2,7}) 18.87	(R ₈) 15.64	(R ₈) 15.49	(R ₄) 15.94	(R ₄) 17.34	(R ₄) 17.11	(R ₆) 16.06	(R ₄) 21.90	(R ₁) 30.05
	B	(R ₁) 69.67	(R ₄) 120.31	(R ₄) 65.96	(R ₄) 63.97	(R ₃) 68.99	(R ₃) 69.54	(R ₃) 69.58	(R ₁) 69.48	(R ₁) 69.64	(R ₃) 74.97
MCF-7	A	(R ₄) 7.74	(R ₂) 24.54	(R ₈) 20.48	(R ₈) 20.40	(R ₄) 22.04	(R ₄) 28.10	(R ₄) 21.72	(R ₆) 20.98	(R ₄) 23.79	(R ₄) 36.44
	B	(R ₁) 76.59	(R ₄) 132.11	(R ₄) 73.88	(R ₄) 73.23	(R ₃) 75.84	(R ₃) 76.45	(R ₁) 76.44	(R ₁) 76.38	(R ₁) 76.56	(R ₃) 82.34

En negritas. Mejores modificaciones para cada grupo.

De acuerdo con los resultados de los **modelos QSAR AB** mostrados en la tabla anterior, se identifica que:

- Las IC₅₀ más bajas son en general, las mismas posiciones para HeLa y MCF-7.
- Independientemente de la línea tumoral, las mejores sustituciones con -Br y -I son obtenidas en una misma posición tanto para el grupo A (R₈) como para el grupo B (R₃). Del mismo modo, las mejores sustituciones con -CH₃, -OCH₃ y -COOH son obtenidas en las mismas posiciones tanto para el grupo A (R₄) como para el grupo B (R₄). En ambos casos obteniendo una IC₅₀ similar
- Para el grupo A, las posiciones más activas son principalmente en las posiciones más cercanas al N de cada anillo (R₄ y R₈). Respecto a compuestos del grupo B, son posiciones -para en los anillos laterales (R₄) y posiciones -para y -orto en el anillo central (R₁ y R₃).
- Las predicciones de actividad biológica para cada sustituyente varían por lo general ± 2 μ M entre posiciones ocupadas en un mismo anillo (ver Material Suplementario 2 y 3).
- La sustitución con -F, específicamente en R₄ del grupo A, genera una excelente disminución de IC₅₀ bajo este modelo.

Algunas de estas observaciones se mantienen al evaluar la predicción de los compuestos mediante los **modelos QSAR ABCD**. Bajo los **modelos QSAR AB**, las mejores posiciones a sustituir para cada sustituyente varían principalmente en función de si el sustituyente es electronegativo o electropositivo, ya que fueron identificadas mismas posiciones para sustituyentes halógenos, por ejemplo. Los **modelos QSAR AB** presentan un mejor ajuste R^2 con su conjunto de entrenamiento por lo que sus predicciones serían más confiables que los **modelos QSAR ABCD**, pero únicamente para sustituyentes y posiciones evaluadas en los trabajos originales de los grupos A y B. Los **modelos QSAR ABCD** generalizan el comportamiento de cada sustituyente para cada posición específica, dado que fueron entrenados con más información para cada farmacóforo, la predicción de actividad biológica para posiciones o sustituyentes no evaluadas experimentalmente sería más confiable que la mostrada por los **modelos QSAR AB**. La validez predictiva de los modelos será abordada a fondo en la sección de discusión. La tabla 7 muestra las mejores predicciones para cada sustituyente obtenidas por los **modelos QSAR ABCD**. En el Material Suplementario 10 y 11 se muestra el etiquetado y la predicción de IC_{50} obtenida para todos los compuestos evaluados del conjunto de predicción mediante los **modelos QSAR AB 1.1 y ABCD**.

Tabla 7. Mejor IC_{50} teórica posible obtenida por cada sustituyente para los diferentes grupos. HeLa y MCF-7. QSAR ABCD

A

B

C

D

Línea tumoral	Grupo	Mejor IC ₅₀ provocada por cada sustituyente presente en las moléculas teóricas propuestas.									
	F	Cl	Br	I	CH ₃	OCH ₃	COOH	OH	NO ₂	COOCH ₃	

HeLa	A	(R ₂) 18.234	(R ₂) 18.065	(R ₂) 18.090	(R ₂) 18.702	(R ₃) 17.396	(R ₃) 20.257	(R ₄) 16.999	(R₃) 12.746	(R ₄) 18.517	(R ₄) 24.703
	B	(R ₆) 75.309	(R ₆) 81.194	(R ₆) 84.472	(R ₄) 83.193	(R ₆) 81.246	(R ₆) 72.074	(R ₄) 84.599	(R₆) 66.997	(R ₁) 86.116	(R ₄) 87.367
	C	(R ₇) 64.426	(R ₇) 65.487	(R ₇) 66.375	(R ₇) 69.807	(R ₇) 73.575	(R ₇) 63.522	(R ₈) 61.270	(R ₇) 63.888	(R ₈) 65.605	(R₈) 59.425
	D	(R ₂) 62.284	(R ₂) 67.976	(R ₂) 70.736	(R ₂) 79.716	(R ₂) 79.187	(R ₂) 74.318	(R ₁) 45.193	(R₂) 39.447	(R ₃) 113.315	(R ₃) 60.135
MCF-7	A	(R ₂) 20.57	(R ₂) 20.13	(R ₂) 20.84	(R ₂) 20.47	(R ₃) 18.11	(R₄) 18.09	(R ₁) 24.65	(R ₂) 20.08	(R ₄) 26.47	(R ₆) 23.15
	B	(R ₆) 49.45	(R ₆) 63.63	(R ₆) 72.61	(R ₆) 93.48	(R ₆) 78.19	(R₆) 47.78	(R ₁) 109.4	(R ₆) 53.93	(R ₁) 116.27	(R ₄) 105.58
	C	(R₆) 40.60	(R ₆) 53.00	(R ₆) 60.35	(R ₆) 75.34	(R ₆) 60.97	(R ₆) 45.32	(R ₈) 64.02	(R ₆) 44.52	(R ₈) 70.80	(R ₈) 60.60
	D	(R ₂) 48.77	(R ₂) 49.62	(R ₂) 51.98	(R ₂) 61.42	(R ₄) 39.71	(R ₂) 67.28	(R ₅) 46.39	(R ₂) 46.50	(R₅) 38.76	(R ₅) 49.85

En negritas. Mejores modificaciones para cada grupo.

De acuerdo con los resultados mostrados en la tabla anterior se identifica que:

- Salvo una excepción, los sustituyentes -CH₃ y -OCH₃, presentan su mejor desempeño en las mismas posiciones de cada farmacóforo (R₃ del grupo A, R₆ del B, R₇ del C y R₂ del D) frente a ambas líneas celulares
- Salvo una única excepción, las mejores posiciones para los sustituyentes -F, -Cl, -Br, -I, y -OH son exactamente las mismas para los 4 farmacóforos, tanto frente a HeLa como en MCF-7,
- La sustitución en R₄ con -OCH₃ (18.09 μ M) y R₃ con -OH (12.75 μ M) en el farmacóforo del grupo A, fueron las mejores actividades biológicas teóricas frente a MCF-7 y HeLa, respectivamente. Cabe destacar que estas combinaciones de posición y sustituyente no fueron evaluadas experimentalmente por el trabajo original.
- Los mejores compuestos teóricos superan la mejor actividad biológica experimental cuantificada tanto para MCF-7 (18.55 μ M) como en HeLa (14.25 μ M).
- La mejor actividad teórica del grupo B y C en MCF-7 supera a los mejores valores experimentales de sus respectivos grupos en la misma línea tumoral.
- Las IC₅₀ teóricas obtenidas por las mejores sustituciones para el grupo D se encuentran en un intervalo medio de actividad (45 μ M < IC₅₀ < 150 μ M), resultan ser entre 3 a 6 veces más potentes que la media de las IC₅₀ evaluadas experimentalmente.
- Para el grupo D, la posición R₂ sustituida con -OH, -F, -Cl y -Br resulta más benéfica que las posiciones evaluadas experimentalmente (R₁ y R₃) frente a ambas líneas tumorales. Es en R₂ donde gran parte de los sustituyentes resultan superar por mucho el valor de la mejor actividad biológica experimental de su grupo.

- Contemplando que las predicciones de los modelos QSAR AB son muy similares para sustituciones en un mismo anillo, se da preferencia a la actividad biológica predicha por los modelos QSAR ABCD. La razón de esto se discutirá en la sección de Discusión.

Por último, la relación asociada directamente entre una buena actividad antiproliferativa y el carácter electropositivo o electronegativo identificada con los *modelos QSAR AB* no prevalece al añadir información de sustituyentes evaluados por los grupos C y D, sino más bien se puede asociar al aprovechamiento de zonas en la molécula que propician interacciones favorables para sustituyentes con características específicas. Esta propuesta, generada a partir de la información de los descriptores presentes en los *modelos QSAR ABCD* logra unir las observaciones y sugerencias de comportamiento que los autores originales plantearon en sus respectivos trabajos. La validez y robustez de esta nueva propuesta será discutida a fondo a lo largo de la siguiente sección.

Capítulo 4.

DISCUSION

PLANTEAMIENTOS BASADOS EN OBSERVACIONES DE AUTORES ORIGINALES

SELECCIÓN DE ANILLOS MÁS BENÉFICOS

Una pregunta que responder común entre los 4 grupos de compuestos está relacionada con identificar si existe una diferencia sustancial respecto a en qué anillos benceno de cada estructura base se debería realizar preferentemente las sustituciones. Las observaciones experimentales realizadas en los 4 trabajos dejan ver claras diferencias de actividad biológica provocadas por un mismo sustituyente dispuesto en anillos distintos. Además, su efectividad cambia en función de la posición para-, meta- u orto- del sustituyente en un mismo anillo. Sin embargo, dado que no en todos los casos las mismas posiciones fueron evaluadas en diferentes anillos, no queda claro si dicha diferencia de actividad biológica se debe a la elección de posiciones no favorables para dichos sustituyentes o si dicho cambio se debe a haber sustituido anillos menos activos para la actividad biológica. De acuerdo con nuestros modelos, podemos dar una razón al por qué unos anillos son más activos que otros, y por qué una posición específica de un anillo resulta más favorable a la actividad biológica que las otras posiciones en el mismo anillo.

Un anillo benéfico para la actividad biológica está relacionado con qué tan bien interaccione con los átomos a su alrededor. Considerando en principio la información

proporcionada por los descriptores de la familia ATS: AATSC3v y ATSC6e. Un anillo aromático contribuirá más mientras más posibilite a sus sustituyentes interactuar con átomos a 3 o 6 enlaces de distancia. La interacción con más átomos implica más contribuciones al cálculo de ambos descriptores. Un anillo aromático unido a una distribución estructural que facilite obtener valores más negativos de AATSC3v y, a su vez, más positivos para el descriptor ATSC6e será lo ideal. Estructuralmente, esto se puede traducir a dos características principales: priorizar las interacciones de átomos con mayor diferencia de radio atómico a 3 enlaces de distancia ya que así beneficiaría la actividad biológica debido a generar valores de AATSC3v más negativos. Del mismo modo, priorizar interacciones de sustituyentes electronegativos interaccionando con átomos distintos a C o H, o átomos electropositivos con átomos de tipo C o H a 6 enlaces de distancia, permite obtener valores positivos de ATSC6e.

Algo a considerar es que las posiciones benéficas para un descriptor no serán necesariamente las mismas para el segundo. De acuerdo con el modelo, los descriptores MATS6c y MATSC6c tienen una mayor contribución en la descripción de la actividad antiproliferativa de los derivados azólicos frente a HeLa y MCF-7, mientras que ATSC6e presenta una mayor contribución que AATSC3v. Sin embargo, los descriptores **MATS** presentan valores generalmente un orden de magnitud más bajo que los presentados por los descriptores de tipo **ATS**. Con esto en mente, una óptima selección de un anillo benéfico y posición a sustituir sería considerar la contribución de cada descriptor en función de la magnitud de los valores involucrados, mismos que cuantificaran el cambio a presentarse.

Gudipati [34] identifica con base en sus observaciones que para los compuestos del grupo A, sustituciones en el anillo lejano al oxazol (anillo A) son mejores que las del anillo unido al oxazol (anillo B). Su suposición se basa en que los átomos de H del anillo B se encuentran más protegidos que los del otro anillo debido a su cercanía con el O del oxazol. Este comportamiento puede ser descrito por los *modelos QSAR ABCD* basado en la facilidad para interactuar con N y/o O a 3 (AATSC3v), 6 (MATS6e y ATSC6e) o 7 (MATS7e) átomos de distancia, ya que se identifica el mismo comportamiento experimentalmente observado para las posiciones evaluadas. No obstante, el modelo QSAR amplía esta suposición al identificar posiciones que resultan generar actividades antiproliferativas similares en ambos anillos principalmente para halógenos en

posiciones R₆ y R₇, dependen del tipo de sustituyente y posición empleada y que serán abordadas a mayor profundidad en una siguiente sección.

En el caso de las estructuras del grupo B, se nota un claro favorecimiento de los sustituyentes halógenos, -OCH₃ y -OH por sustituciones realizadas en los anillos laterales, principalmente en R₆, obteniendo IC₅₀ hasta 13 y 29 µM más bajas que el promedio de las mejores IC₅₀ posibles para HeLa y MCF-7 respectivamente (ver Tabla 7: QSAR_ABCD). Mientras que los mejores representantes de los sustituyentes -COOH y -NO₂ son favorecidos por posiciones en el anillo central, principalmente en R₁ (ver Tabla 7: QSAR ABCD). Si bien, los autores no reportan una propuesta para explicar este comportamiento, los *modelos QSAR ABCD* generados pueden proporcionarla.

El hecho que -COOCH₃, -COOH y -NO₂ resulten mejorar su comportamiento general en el anillo central se debe a que las interacciones a 3, 6 o 7 átomos de distancia que pueden tener sustituciones en el anillo central (R₁-R₃) son principalmente entre C y H o directamente entre O y C, lo cual contribuye a generar valores más altos de los descriptores MATS7e y MATS6c y más bajos de ATSC6e y AATSC3v, comparado con los que se obtienen en otras posiciones (R₄-R₆), particularmente para estos sustituyentes. Algo similar ocurre con los halógenos que interaccionan en las pociones orto- de los anillos laterales (R₆), aunque aquí la contribución positiva a la actividad antiproliferativa se presenta a mayores interacciones con átomos N comparado con el resto de las posiciones, e interaccionan con un mínimo número de átomos C o H a 3 o 6 enlaces de distancia en dicha posición. La Figura 13 muestra un esquema de dichas interacciones tomando como ejemplo sustituyentes -COOH y -F en las posiciones R₁ y R₆.

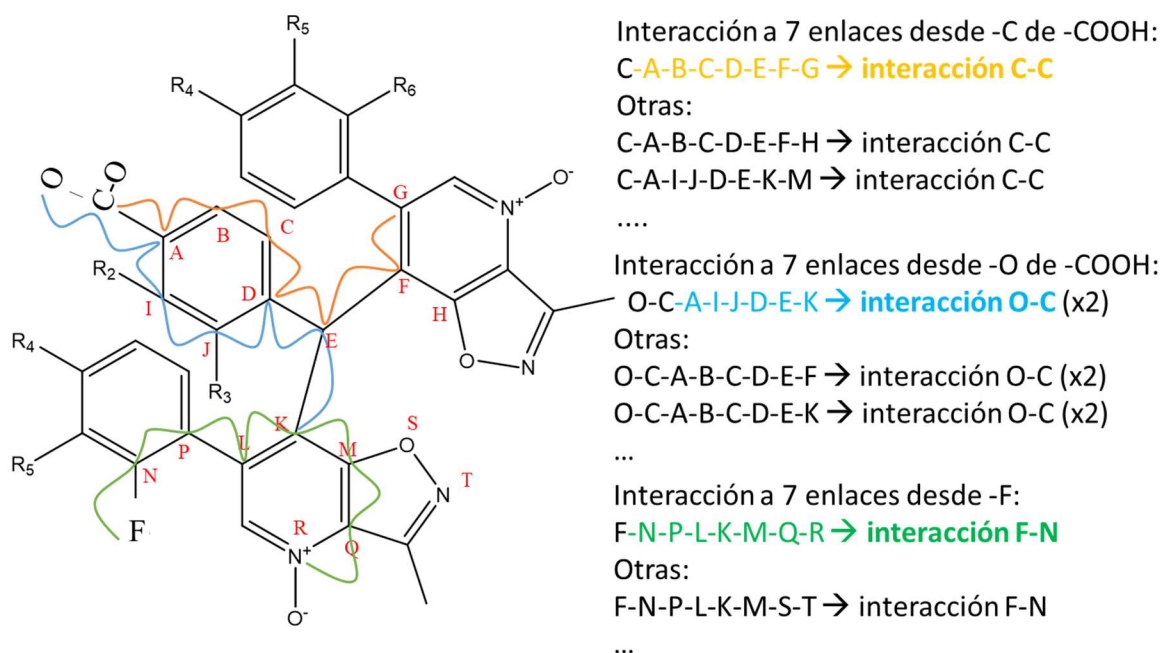


Figura 13. Representación de interacciones contempladas para el cálculo de descriptores

MATS, ATSC y AATSC del sustituyente -COOH en R_1 y -F en R_6 del grupo B. Cabe señalar que, para halógenos en posición R_4 , posición asociada a menor actividad antiproliferativa sustituida por halógenos, existiría menor número de interacciones a 6 o 7 enlaces de distancia con N u O, las cuales se asocian a contribuciones positivas a la actividad antiproliferativa.

Para las estructuras del grupo C, al igual que en el grupo B, la sustitución en R_6 (correspondiente a la posición orto- en los anillos aromáticos laterales) resultó ser una posición que promueve una excelente actividad antiproliferativa para los mismos sustituyentes que en caso del grupo B ($IC_{50}^{-F}=40.6 \mu M$, $IC_{50}^{-OH}=44.52 \mu M$, $IC_{50}^{-OCH_3}=45.32 \mu M$, $IC_{50}^{-CH_3}=60.97$) (ver Tabla 7) Por otro lado, los anillos aromáticos presentes en la estructura principal del grupo C que evalúan las posiciones R_7 - R_9 resultan ser más benéficos para sustituciones -COOCH₃, -COOH y -NO₂ que el resto de los anillos estudiados presentes en ambos grupos B y C. En este caso, se atribuye dicha mejora al hecho que estos anillos permiten una interacción del O con un mayor número de átomos C o H, principalmente, a 6 (MATSC6e y ATSC6e) y N-O a 7 (MATS7e) enlaces de distancia (principalmente la posición R7) son características benéficas de acuerdo con lo descrito por el *modelo QSAR ABCD* obtenido.

Las estructuras del grupo D cuentan con un único anillo benceno. En este caso, la particularidad buscada es identificar la posición a sustituir y no el elegir un anillo aromático a sustituir. Por lo que esta estructura se discutirá principalmente en la siguiente sección. No obstante, cabe destacar que, la posición R₅, correspondiente a una posición intermedia en la estructura no asociada al anillo aromático mejoró la actividad antiproliferativa teórica de los sustituyentes COOCH₃, COOH y NO₂, en la misma proporción que las mejores sustituciones en el único anillo aromático (R₁-R₄) de este grupo.

DISTRIBUCIÓN ESTRUCTURAL DE LA MOLÉCULA Y TIPO DE SUSTITUYENTES

El identificar teóricamente si existen posiciones que al agregar diferentes grupos proporcionen una mejor actividad antiproliferativa que los compuestos evaluados experimentalmente e identificar si existe una relación directa entre una buena actividad antiproliferativa y la naturaleza electroattractora o electrodonadora del sustituyente son dos puntos indispensables que discutir. En esta sección se plantea la relevancia del **modelo QSAR ABCD** para dar sentido descriptivo al comportamiento observado basado en el significado de los descriptores que conforman dicho modelo.

En primer lugar, se identifica que la distribución estructural y el tipo de sustituyente elegido en cada molécula son los principales factores que definirán la actividad antiproliferativa final, debido a la naturaleza y valores de los descriptores empleados para construir los **modelos QSAR AB** y **ABCD**. De acuerdo con los **modelos QSAR ABCD**, la mejora de la actividad antitumoral depende principalmente de las características de los vecindarios cercanos y lejanos respecto a una sustitución específica, en términos de la información medida por los descriptores AATSC3v, ATSC6e, MATS6c y MATS7e.

El descriptor AATSC3v mide el contraste entre el volumen de van Der Waals de átomos que interaccionen con cada átomo del sustituyente en cuestión. Considerando que el coeficiente del descriptor AATSC3v es positivo (**ABCD HeLa**), sustituyentes que generen un cambio durante el “centrado” más bajo (realizado durante el cálculo del descriptor entre todas las posibles interacciones en una molécula) generará valores del descriptor AATSC3v más bajos o negativos, lo que implica que se beneficiaría una de las características de compuestos más activos. Para conseguir un valor más bajo para este descriptor debe existir

un mayor contraste entre el volumen de van Der Waals de átomos que interaccionen con cada átomo del sustituyente. Así mismo, este efecto se podría conseguir con un mayor número de interacciones posibles para los átomos que conforman el sustituyente en cuestión, lo que hablaría de que una estructura lo suficientemente ramificada en posiciones que favorezcan la interacción con este sustituyente potencialmente presentaría una mejor actividad antiproliferativa.

El descriptor ATSC6e (***ABCD MCF-7***) cuantifica la electronegatividad (e) de los átomos interactuantes a 6 enlaces de distancia. Usando el mismo análisis que para el descriptor anterior, se identifica que el coeficiente es negativo y que átomos con electronegatividades más grandes generarán una contribución tal que aumente los valores del descriptor ATSC6e. Al igual que en el caso anterior, esta contribución será mayor mientras mayor sea el número de átomos con los que interaccione, es decir, mientras más ramificaciones existan en la molécula. Para este descriptor, una buena combinación entre número de interacciones con átomos con alta electronegatividad aumentaría el valor del descriptor ATSC6e, provocando que disminuya el valor predicho de la actividad biológica para el respectivo compuesto, consiguiendo compuestos más activos. Esta podría ser una de las razones por las cuales, átomos halógenos resultan generalmente más efectivos en la mayoría de las posiciones.

Esto implica que, dependiendo de la conformación estructural del compuesto, las interacciones no benéficas que se agregue podrían contribuir lo suficiente como para “*opacar*” la contribución aportada por su interacción con los sustituyentes de interés aplicable para los descriptores de las ***familias ATS y MATS***. Un ejemplo de este efecto de “opacado” puede ser observado en los valores de los descriptores de los grupos B y C usados en el ***modelo QSAR ABCD MCF-7*** (ver Tabla 8). Los compuestos de estos grupos difieren únicamente por la adición de anillos aromáticos a los extremos de la molécula (ver Figura 4) por lo que se puede hacer una comparación directa de las posiciones R₁₋₆. El efecto de *opacado* se observa en compuestos del grupo B respecto a sus variantes del grupo C ya que, debido a la falta de dichos anillos en su estructura, los valores de los descriptores de interés tienden a desfavorecer la tendencia deseada para una buena actividad antiproliferativa (valores positivos de ATSC6e y negativos de MATS7e) frente a MCF-7. Estos descriptores podrían

servir para identificar la cercanía o lejanía de zonas electrodonadoras, sustituyentes ricos en electrones y áreas de enlaces de hidrógenos, las cuales son características favorables a la actividad antiproliferativa [31]

Tabla 8. Comparación de valor en descriptores de los mejores representantes del grupo B respecto a sus variantes del grupo C. y la IC50 obtenida.

Sustituyente	Posición	ATSC6e (IC ₅₀)		MATS7e (IC ₅₀)		MCF-7 (IC ₅₀)		Diferencia (IC ₅₀ C – B)
		(B)	(C)	(B)	(C)	(B)	(C)	
-F	R ₆	1.64	2.98	-0.05	0.003	49.4	40.6	-8.8
-Cl	R ₆	0.50	1.91	-0.07	0.002	63.6	53.0	-10.6
-Br	R ₆	-0.09	1.36	-0.08	-0.001	72.6	60.3	-12.3
-I	R ₄	-1.72	-0.14	-0.08	-0.027	107.4	82.7	-24.7
-I	R ₆	-1.15	0.40	-0.08	-0.010	93.5	75.3	-18.1
-CH ₃	R ₆	-1.01	0.82	-0.14	-0.054	78.2	61.0	-17.2
-OCH ₃	R ₆	1.83	2.79	-0.05	0.029	47.8	45.3	-2.5
-COOH	R ₁	-1.90	-0.29	-0.10	-0.034	109.4	84.4	-25.0
-COOH	R ₄	-1.88	-0.29	-0.07	-0.042	114.4	82.9	-31.4
-NO ₂	R ₁	-2.19	-0.44	-0.10	-0.040	116.3	86.4	-29.9
-COOCH ₃	R ₄	-1.35	-0.10	-0.05	-0.033	105.6	80.9	-24.7
-OH	R ₆	1.04	2.51	-0.08	-0.008	53.9	44.5	-9.4
Contribución al <i>modelo</i> <i>QSAR ABCD MCF-7</i>		-0.10924		1.0401				

Con esta perspectiva, se habla entonces que la actividad biológica no depende específicamente de si el sustituyente usado es electropositivo o electronegativo sino, más bien, que depende de un buen balance y aprovechamiento de las ramificaciones estructurales, el número y tipo de átomos presentes en la estructura molecular, y el conjunto de características dominantes en zonas específicas de una molécula que interaccionen con sustituyentes que puedan aportar coeficientes favorecedores reflejados en los valores de los descriptores AATSC3v, ATSC6e, MATS6c y MATS7e.

MODELO QSAR. Relación con descriptores simples.

El primer acercamiento involucró el análisis de los grupos A y B únicamente con la intención de identificar características químicas simples y como resultado inicial se generaron los *modelos QSAR AB simples* relacionadas con los descriptores empleados por los *modelos QSAR AB complejos*.

En esta sección se estudió la relación entre los **modelos QSAR AB 1.1 y ABCD** y descriptores que pudieran ser ligados a un comportamiento químico más simple. Para estudiarlo se analizaron 331 descriptores que fuesen fácilmente relacionables con un aspecto químico y cuyo cálculo fuese simple. Se agrupó en 5 categorías el nivel de actividad antiproliferativa predicha por todos los modelos QSAR: *Excelente* ($IC_{50} < 20 \mu M$), *Muy Buena* ($20 \mu M \leq IC_{50} < 40 \mu M$), *Buena* ($40 \mu M \leq IC_{50} < 75 \mu M$), *Regular* ($75 \mu M \leq IC_{50} < 150 \mu M$) y *Mala* ($IC_{50} \geq 150 \mu M$). Se analizó cada descriptor a fin de identificar tendencias de comportamiento lineal (valores más altos o bajos) que pudieran ligarse a una actividad antiproliferativa predicha por estos modelos QSAR.

MODELOS AB

La Figura 14 muestra una representación visual del comportamiento de los compuestos en función de los descriptores usados para construir los **modelos QSAR AB simples**. Dado que el valor de los descriptores es independiente de la actividad biológica y que, varios compuestos tienen valores predichos similares de actividad antiproliferativa para HeLa y MCF-7, comparten la misma etiqueta de *actividad biológica cualitativa*. Se muestra una única gráfica que representa el comportamiento general estudiado en ambas líneas tumorales, de los compuestos de los grupos A y B.

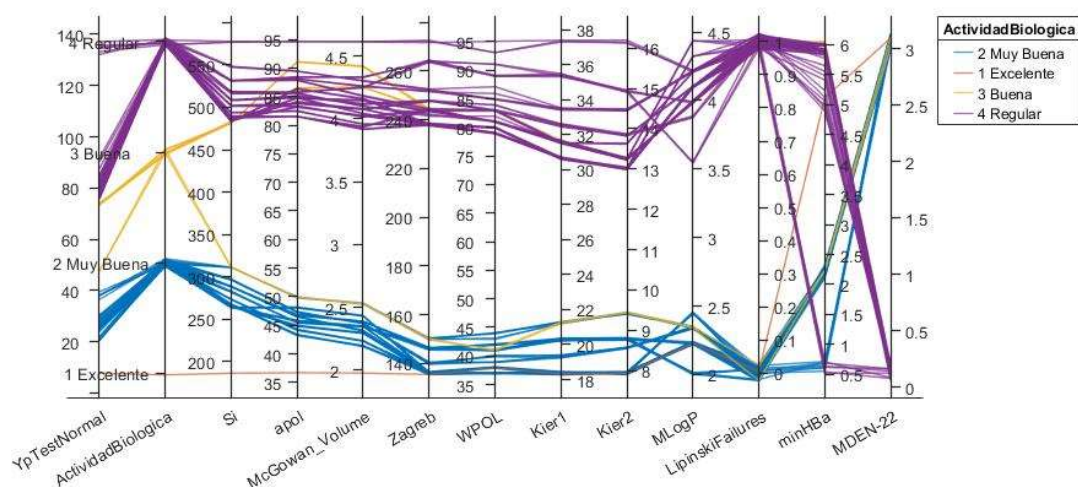


Figura 14. Valor de los descriptores de compuestos de los grupos A y B, usados en la generación de los modelos QSAR AB simples, agrupadas por su actividad biológica cualitativa en HeLa. Cada color representa una categoría cualitativa de actividad biológica (ningún compuesto del grupo A o B presentó una actividad cualitativa 'mala' ($IC_{50} > 150$)).

Valores más bajos de los 9 primeros descriptores mostrados en la Figura 8 U y la presencia de la característica descrita por el descriptor *MDEN-22* (distancia molecular entre los átomos de nitrógeno secundario) son asociados a una buena actividad biológica cualitativa. Se observa una clara diferencia entre los compuestos de los grupos A y B respecto al valor de los descriptores mostrados en la Figura 14. Si bien estos descriptores podrían ser usados como una referente general, se requiere validar su generalidad contemplando las características de los grupos C y D.

Relación entre *modelos QSAR AB* vs *modelos QSAR ABCD*

Una versión extendida de la Figura 14 se muestra en el Material Suplementario 12 donde se aprecia que, debido a que algunos compuestos del grupo D presentan un comportamiento similar a los del grupo A en algunos descriptores, varios de estos compuestos son erróneamente predichos y catalogados con un comportamiento excelente. Ésta falla predictiva está directamente relacionada con el hecho de no haber contado con información de los grupos C y D en el conjunto de entrenamiento usado para estos modelos QSAR y posiblemente con un bajo número de datos experimentales usados tanto en el conjunto de entrenamiento como en el de validación. Identificando así que la predicción mediante los *modelos QSAR AB* está limitada a compuestos con ligeras modificaciones de los farmacóforos de los grupos A o B ya que, en caso contrario, la predicción de actividad biológica no será tan confiable.

Ambos acercamientos de análisis QSAR mantuvieron un buen ajuste general frente al *conjunto de entrenamiento* sin perder su capacidad predictiva validada por el *conjunto de validación*. Debido a que tanto los *modelos QSAR AB 1.1* como los *modelos QSAR ABCD* presentan excelentes características predictivas ($R^2_{\text{test}} > 0.85$), ambos tipos de modelos QSAR fueron usados para predecir la actividad biológica de los compuestos teóricos del conjunto de predicción. La Figura 15 PEP muestra la comparativa entre las IC_{50} calculadas por los modelos QSAR para cada compuesto teórico.

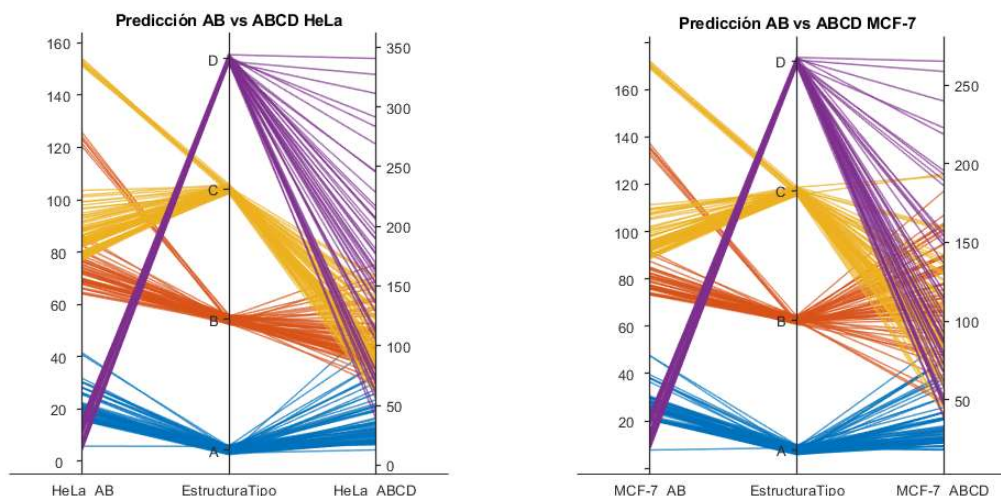


Figura 15. Comparativa de predicción de modelo QSAR AB 1.1 vs modelo QSAR ABCD, para a) HeLa y b) MCF-7. Cada color representa a un grupo de compuestos distinto.

Se observa que la actividad biológica de los compuestos teóricos basados en los farmacóforos de los grupos A y B presentan una distribución similar en la predicción por parte de **modelos QSAR AB** y **ABCD**. No obstante, la actividad biológica predicha entre los grupos C y D resulta distinta, siendo más severa para compuestos basados en el farmacóforo del grupo D que en el de C. Así mismo, la distribución de actividad biológica predicha es ligeramente más alta en los modelos QSAR ABCD. En ambas versiones de modelos QSAR las actividades biológicas predichas por el mismo modelo frente a HeLa y MCF-7 se encuentran en los mismos intervalos, con excepción de las moléculas del grupo D.

Al incorporar a los compuestos del grupo C y D en este análisis se pudo estudiar la prevalencia de los descriptores identificados en la aproximación inicial (Figura 14). Se abordó la aproximación de los **modelos QSAR ABCD** a fin de evaluar si los **modelos QSAR AB 1.1**, con una alta capacidad predictiva de acuerdo con el conjunto de validación usado, lograban predecir el comportamiento de compuestos desconocidos. La Figura 16 muestra los valores de los descriptores de los 4 grupos estudiados agrupados por su actividad biológica cualitativa.

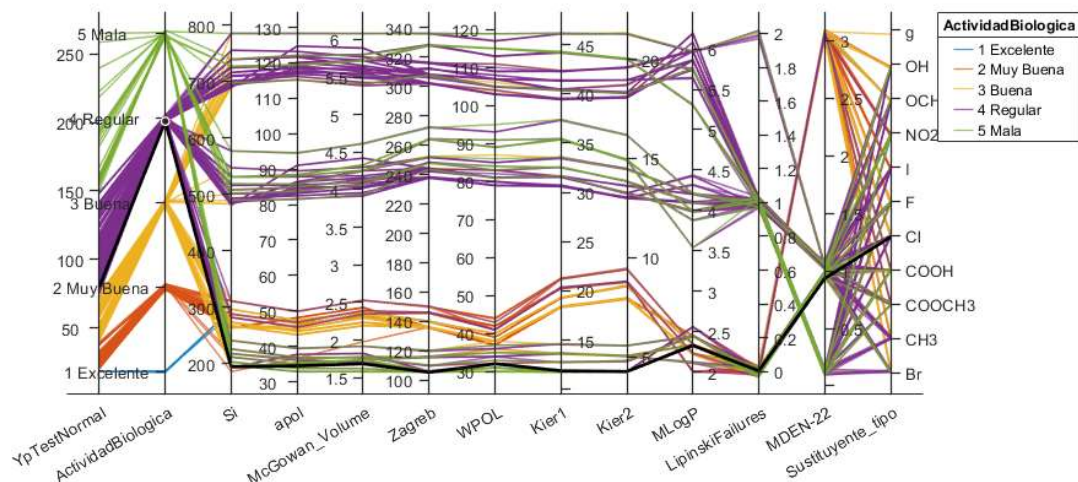


Figura 16. Descriptores efectivos para distinguir actividad biológica del grupo A y B, actualizado con información de los valores de los descriptores de los grupos C y D. Cada color representa una de las 5 categorías de actividad biológica.

Se comprueba que altos niveles de MDEN-22 es una característica general que contribuye a la actividad biológica. Así mismo, valores bajos en varios descriptores útiles para el grupo A (ver Material Suplementario 13 y 16) podrían ser ligados a una buena actividad biológica, sin embargo, compuestos del grupo D con valores igualmente bajos que los encontrados en el grupo A, presentan una actividad biológica *mala* ($IC_{50} \geq 150 \mu M$). De este modo, los compuestos del grupo D no sigue el comportamiento de actividad marcado por los grupos A y B durante el análisis de los **modelos QSAR AB**. Por otro lado, el grupo C presenta valores similares a los del grupo B, siendo esto algo esperado, ya que el farmacóforo del grupo C presenta mucha similitud con el farmacóforo del grupo B. Con base en esto, se plantean dos posibilidades: la existencia de descriptores simples distintos a los ya identificados que puedan ser ligados a una buena actividad biológica o, la existencia de características únicas en el grupo D que los vuelva menos activos.

El grupo D presenta en varios descriptores simples valores similares a los del grupo A, pero no muestran una actividad biológica similar en alguna línea tumoral, razón por la cual la predicción por parte del **modelo QSAR AB 1.1** falle en predecir la actividad biológica del grupo D (ver Figura 15 y Material Suplementario 12). Para corroborar la posibilidad de que exista una característica que, de estar presente en la molécula, se inviertan o desproporcionen sus características, o inhiba su capacidad antitumoral, se buscaron características en común entre compuestos activos del grupo D. No se halló una característica

única medida por un descriptor simple, no obstante, se encontró una característica particular en común para compuestos activos del grupo D. Se identificó que la suma de 5 descriptores independientes: *SHBint4*, *SHBint6*, *nHBint6*, *SHBint8* y *nHBint8*, en uno solo (denominado *sumHBs*), puede servir como un discriminante de compuestos con mejor actividad antiproliferativa de este conjunto (ver Figura 17). Logrando identificar a 9 de los mejores compuestos teóricamente activos independientemente del sustituyente, posición sustituida o de la línea tumoral.

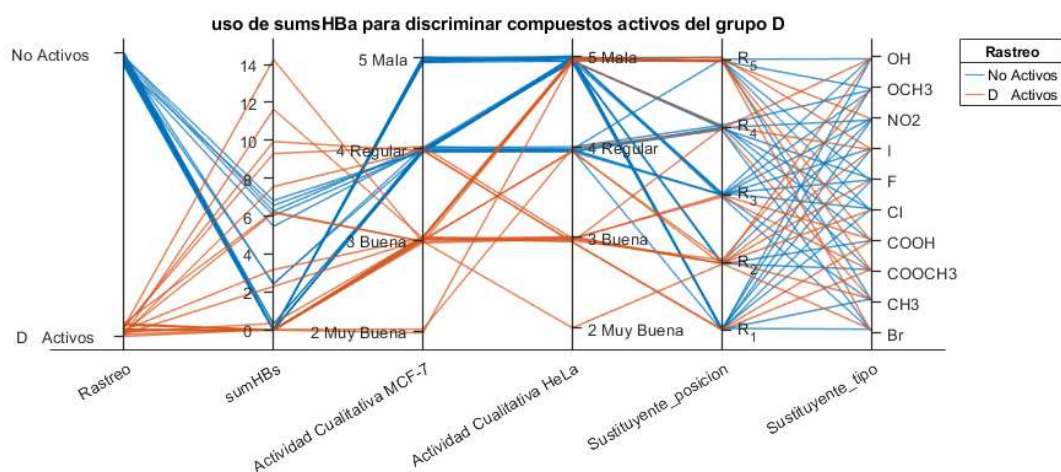


Figura 17. Relación entre compuestos activos del grupo D y el descriptor *sumHBs* propuesto.

En segunda instancia, se analizó el comportamiento de los descriptores del grupo D y se identificó que tanto el grupo A como el grupo D presentan diferencias particulares en descriptores específicos únicas respecto al resto de compuestos. En el Material Suplementario 14 se muestra una representación visual de los descriptores identificados bajo este análisis y la Tabla 9 muestra las conclusiones obtenidas. A partir de los conjuntos de descriptores encontrados, se pudo relacionar el cambio entre una buena o una mala actividad biológica general entre el grupo A y D, y explicar a su vez diferencias entre el grupo B y C ligadas a una buena actividad biológica.

Tabla 9. Descriptores cuya presencia o aumento de valor se relaciona con un cambio positivo o negativo en la actividad biológica

Mejor Representante	Descriptores	Observación	Relacionado con
Grupo A	<i>minssNH</i> , <i>minHssNH</i> , <i>maxHssNH</i> , <i>maxssNH</i> , <i>SssNH</i> , <i>nHssNH</i> , <i>nssNH</i> , <i>SHssNH</i>	Valores $\neq 0$	Buena actividad antiproliferativa general
	<i>nAtomP</i> , <i>MDEN-22</i> , <i>nHBint2</i>	Valores altos	
Grupo D	<i>nHaaNH</i> , <i>naaNH</i> , <i>SHaaNH</i> , <i>SaaNH</i> , <i>minHaaNH</i> , <i>minaaNH</i> , <i>maxHaaNH</i> , <i>maxaaNH</i> , <i>nFRing</i> , <i>nF12Ring</i> , <i>nFG12Ring</i>	Valores $\neq 0$	

<i>nT12Ring, nTG12Ring, nF12HeteroRing, nFG12HeteroRing, nT12HeteroRing, nTG12HeteroRing,</i>	
<i>naaaC, SaaaC, nF9Ring, R_TpiPCTPC, nT9Ring, nF9HeteroRing, nT9HeteroRing</i>	Valores altos
	Mala actividad antiproliferativa general

Los descriptores indicados en la tabla también podrían contribuir a la explicación de por qué el grupo B es menos activo teóricamente que el grupo C para las mismas posiciones (R₁-R₆). El farmacóforo del grupo C es derivado del farmacóforo del grupo B, presentando un anillo aromático adicional. De este modo, el grupo B (color rojo) presenta valores más altos que el grupo C (color amarillo) en los descriptores *nF9Ring, nT9Ring, nF9HeteroRing, nT9HeteroRing, SaaaC, naaaC* y *R_TpiPCTPC*. Descriptores que hemos asociado a una mala actividad biológica general. De ellos, los primeros 4 descriptores solo describen si existen anillos fusionados de 9 miembros o no. Mientras que los descriptores *SaaaC* y *R_TpiPCTPC* se relacionan con información eletrotológica similar a la medida por *minHBa* y el número de bifurcaciones disponibles, respectivamente, por lo que tendría más sentido que sean estos dos descriptores los que definen una buena o mala actividad biológica (ver Material Suplementario 14).

Por otro lado, el grupo C presenta valores más altos que el grupo B para el descriptor MDEN-22, cuya magnitud $\neq 0$ se asocia a una buena actividad biológica general. Es así como, en conjunto, las características medidas principalmente por los descriptores *R_TpiPCTPC* y *MDEN-22* permitirían dar una explicación general del comportamiento biológico para todos los grupos de compuestos estudiados a la vez que permiten una descripción química general de las características deseables al identificar una relación entre una actividad antiproliferativa con la presencia de interacciones entre N secundarios (*MDEN-22*) y el mantener un orden de enlace convencional bajo (*R_TpiPCTPC*).

Por último, se estudió si existía una característica general que volviera a todos los compuestos más activos frente a una línea tumoral. No se encontró algún descriptor 1D o 2D calculado que pudiera ser ligado a un cambio directo de la actividad biológica frente a alguna línea tumoral útil para los 4 grupos de compuestos. La razón de esto podría radicar en la naturaleza del conjunto de compuestos empleados, ya que el estudio partió de compuestos que mostraron una actividad similar frente a ambas líneas tumorales. De este modo, de existir una característica única global que dicte si un compuesto será más selectivo hacia alguna

línea tumoral respecto a otra, estaría ligada a descriptores complejos excluidos en el análisis mostrado en esta sección, a información no calculada por los descriptores simples aquí estudiados o, a información tridimensional ajena al planteamiento de este estudio.

MODELO ABCD. Modelo robusto.

Tanto el **modelo QSAR ABCD** para MCF-7 como para HeLa utilizan variantes de descriptores las familias *ATS* y *MATS*. Este tipo de descriptores miden cómo una propiedad se encuentra distribuida a lo largo de una estructura topológica y que tan similar es una zona respecto a otra vecina. Los descriptores AATSC y ATSC usados en los **modelos QSAR ABCD** excluyen de su cálculo cualquier información que dependa directamente del tamaño molecular, pero depende en gran medida de que tan entrelazada se encuentre su estructura. Cabe señalar que 12 de las 50 mejores combinaciones de modelos QSAR válidos que modelen los 4 grupos en ambas líneas tumorales están conformados por una variante de descriptores de la familia *SpM**_Bh** usados en los **modelos QSAR AB**. Esto indica que, bajo una perspectiva matemática, es posible describir el comportamiento de los grupos C y D de forma similar que con los grupos A y B, ya que son características ligadas al mismo tipo de información (ponderaciones e , p , v , c). Lo que podría indicar que, a pesar de sus diferencias estructurales, químicamente los compuestos presentan mecanismos de acción similares.

Predicción de actividad biológica mediante los modelos QSAR ABCD

El **modelo QSAR ABCD MCF-7** presenta buenas aproximaciones tanto para los compuestos del **conjunto de entrenamiento** como para los del **conjunto de validación**, independientemente de las diferencias estructurales entre todos los compuestos. Obteniendo predicciones con una diferencia máxima de $\pm 21\%$. Sin embargo, el modelo parece presentar problemas al calcular la IC_{50} de los farmacóforos de los grupos A, B y D (compuestos **1**, **13** y **33**), ya que la actividad predicha para estos compuestos mostró una desviación del ajuste con 37.2%, 68.91% y 67.86% de variación respecto a las IC_{50} reales. Esto contrasta con la diferencia promedio respecto al resto de compuestos de sus respectivos grupos del **conjunto de entrenamiento** (17.87%, 19.60% y 20.89%). En el caso del grupo C, esta diferencia

predictiva no resultan tan marcada, ya que la variación máxima para este grupo (19.86%) y su farmacóforo (21.99%) es menor.

El *modelo QSAR ABCD HeLa* presenta menos problemas al predecir los compuestos base de los grupos A, C y D respecto al *modelo QSAR ABCD MCF-7*, ya que sus diferencias absolutas (29.18%, 11.62% y 35.95%) se encuentra más cercana a la diferencia promedio de sus respectivos grupos (20.09%, 23.55% y 25.95%). Sin embargo, el error para el compuesto base del grupo B aumenta significativamente (100.38%) a pesar de ser parte del **conjunto de entrenamiento**. A pesar de ello, el modelo mantiene una buena predicción para el resto del grupo B (diferencia absoluta promedio de 22.06%).

Se estudió si se podía asociar una falla predictiva general a un átomo o sustituyente en particular. El compuesto **32**, miembro del grupo C, presenta en su estructura un cambio significativo respecto a todos los compuestos puesto que cuenta con un grupo tiofeno. Ya que es el único compuesto estudiado que incluye un azufre en su estructura, la falla predictiva se asocia con la presencia del átomo de azufre en este grupo tionilo. Para ambos conjuntos, entrenamiento y validación, varios compuestos con sustituyentes -CH₃ y -OCH₃ mostraron una mayor discrepancia entre el valor predicho vs real de la actividad biológica (ver Mat. Supl. 8 y 9). Siendo esta diferencia generalmente entre el 20%-30% y de 30%-50% para el set de entrenamiento y validación, no obstante, las discrepancias más altas tienden a subestimar la actividad biológica real, por lo que el valor predicho podría considerarse como un referente de la actividad mínima que podría presentar experimentalmente dicho compuesto. De este modo, se identifica como limitante a la presencia de átomos de S en la estructura principal no puede ser modelada con fidelidad por estos modelos QSAR, al menos no hasta incrementar el número de moléculas con azufre en el conjunto de entrenamiento.

Significado de los *modelos QSAR ABCD* y su implicación química

Parte del significado químico de los *modelos QSAR ABCD* se abordó previamente. Contemplando la ecuación matemática para estos modelos, características relacionadas con el tamaño estructural de la molécula, la cantidad de ramificaciones en ella y el posicionamiento de sustituyentes en zonas que favorezcan interacciones entre halógenos con N, O y H, o con C a 3, 6 o 7 enlaces de distancia, contribuyen a una buena actividad biológica.

Si bien, interacciones químicas entre átomos separados por 6 o 7 enlaces debe interpretarse como una comunicación electrónica eficiente en estos compuestos. De ahí que los descriptores moleculares empleados para construir los modelos QSAR midieran los cambios de características químicas relacionadas con la electronegatividad, la carga, o el volumen de van Der Waals, a lo largo de la estructura molecular. De esta forma, sustituyentes que contribuyan con las características, principalmente electrónicas, presentes en la zona donde son adicionados, terminaran generando una contribución positiva global a la actividad biológica del compuesto en cuestión. En esta sección se abordará la relación entre el significado matemático de los descriptores que conforman los modelos QSAR ABCD y su posible interpretación química.

De acuerdo con lo mencionado en la introducción, el cálculo de los descriptores de la *familia ATS* : ATSC y AATSC, depende de los siguientes factores: 1) el número de enlaces de distancia que separan a dos o más átomos, 2) cuantas posibles bifurcaciones existan en la estructura molecular, 3) el número total de contribuciones a considerar en el cálculo de los descriptores, 4) el valor correspondiente de su contribución a la ponderación específica que se esté usando, y 5) la distribución de la contribución de los átomos usadas en el “centrado” de los datos [14].

Para los descriptores *ATSC6e* usado por el **modelo QSAR ABCD MCF-7** la ponderación es en función de la electronegatividad (e), mientras el descriptor *AATSC3v* usado por el **modelo QSAR ABCD HeLa**, la ponderación es el volumen de Van Der Waals (v) de los átomos. Lo que implica que las características químicas de interés para los átomos usados como sustituyentes son aquellos cuya electronegatividad o volumen de Van Der Waals sea más grande o pequeña en función de si los átomos con lo que interactúe a 3, 6 o 7 enlaces de distancia sean N, C y H, respectivamente.

Cabe mantener en mente que el modelo está generado bajo una representación logarítmica de la IC_{50} . Razón por la cual la magnitud de las contribuciones de los descriptores al modelo oscila entre 0.02 y 2.02 en magnitud. Descriptores de la *familia ATSC* muestran una relación con la cantidad de átomos disponibles para cada grupo de moléculas y su actividad biológica general, ya que sus valores siguen el orden $A > C > B > D$. Se identifica que las contribuciones de electronegatividad medidas por el descriptor *ATSC6e* resultar cada vez

menos drástica en función de la cantidad de átomos presentes en la molécula, siendo esto concordante con lo planteado en secciones anteriores. Este comportamiento es más marcado en los descriptores de la familia AATSC que en los de la familia MATS.

El hecho de que los valores de las moléculas base de los 4 grupos resulten muy distintas respecto a las predicciones para moléculas de sus mismos grupos se debe asociar a la variabilidad estructural presente en los conjuntos de moléculas y la limitada capacidad de los descriptores usados para construir los modelos QSAR de describirlas en su totalidad. Una generalización de este comportamiento estaría asociada con que ningún farmacóforo presente en su estructura algún anillo sustituido y/o no unido a un anillo aromático. Lo que implicaría que se estaría identificando una nueva característica limitante de los **modelos QSAR ABCD** para fines predictivos. Y a su vez, podría explicar el por qué la actividad biológica del farmacóforo de los grupos C y D pudieron ser predicho con mayor fiabilidad que la de los grupos A y B, incluso a pesar de formar parte del conjunto de validación, que cuenta con una mayor discrepancia entre IC₅₀ predicha vs real.

Considerando que para el **modelo QSAR ABCD HeLa** la contribución del descriptor AATSC3v es positiva y para el **modelo QSAR ABCD MCF-7** la contribución del descriptor ATSC6e es negativa. Un comportamiento deseable sería obtener moléculas cuyas estructuras generen valores para los descriptores AATSC3v negativos y valores de ATSC6e positivos más grandes posibles. Esto se podría conseguir con moléculas que presenten configuraciones benéficas entre diferencias de volumen de Van Der Waals entre átomos cercanos, electronegatividad entre átomos relativamente alejados, y/o que sean más ramificadas, pero electrónicamente comunicadas, con la intención de presentar más posibilidades de generar interacciones benéficas. De esta forma, se podría dar una explicación de por qué los compuestos del grupo A resultan los más efectivos, relacionar por qué los compuestos del grupo D presentan una actividad biológica baja a pesar de mostrar un comportamiento similar al grupo A en varios descriptores.

Para representar la cantidad de caminos posibles en una molécula nos apoyaremos en la familia de descriptores *MWC*. Esta familia de descriptores mide el número de caminos moleculares posibles en los compuestos a 2, 3 y hasta 10 átomos de distancia, por lo que resultarán útiles como análogos de la cantidad de contribuciones posibles consideradas

durante el cálculo de los descriptores *ATSC6e*, *AATSC3v*, *MATS6c* y *MATS7e*. La Figura 18 OIO muestra visualmente el cambio en los respectivos valores de descriptores *MWC* de los grupos A y D. Sus valores aumentan a mayor número de átomos de distancia por existir mayor cantidad de posibles combinaciones. Cabe señalar que, en cada grupo, los sustituyentes parecen agruparse en 3 o 4 principales comportamientos (3 o 4 líneas de un mismo color) debido a que estos descriptores no contemplan en su cálculo las posibles contribuciones por átomos H y el conteo no depende de características atómicas, sino únicamente de distancia atómica.

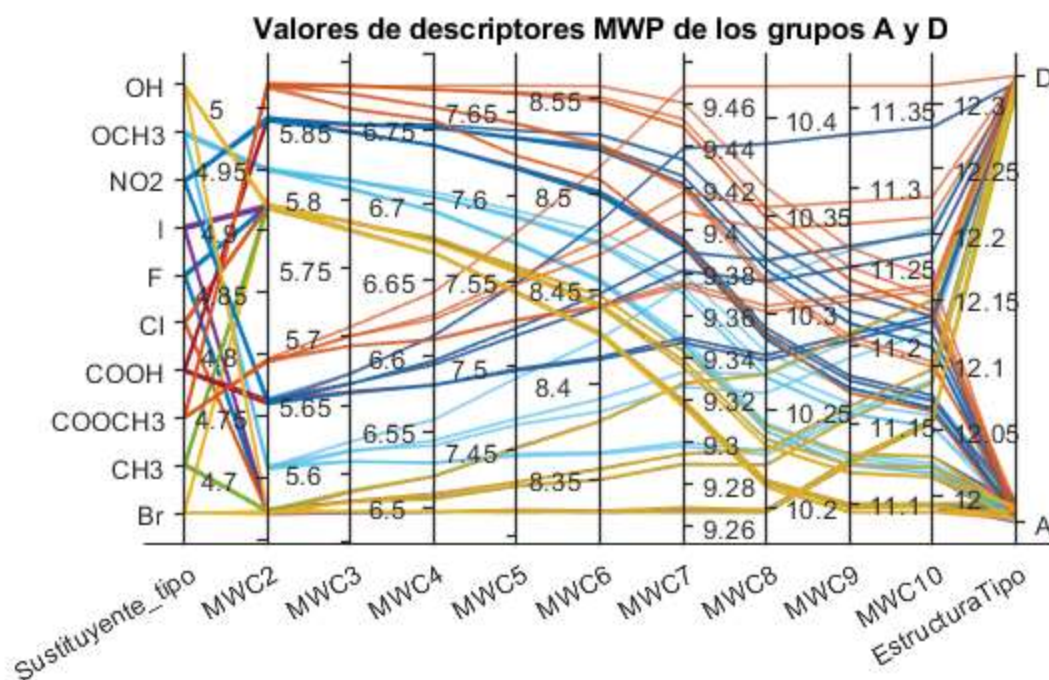


Figura 18. Cambio en los valores de descriptores relacionados con el número de caminos moleculares posibles desde 2 hasta 10 átomos de distancia en función del sustituyente empleado, cada color representa un sustituyente diferente. Se muestra información de

A mayor número de caminos posibles, más elementos son considerados para el cálculo de cada descriptor y, por tanto, mayor será valor del descriptor. Obsérvese que ambos grupos presentan valores similares para *MWC6* y *MWC7*, pero el grupo D presenta valores más bajos en el descriptor *MWC3*. Con esto en mente, el número de contribuciones en el descriptor *AATSC3v* sería menor para compuestos del grupo D que para el grupo A, siendo más difícil generar contribuciones benéficas aprovechables por los sustituyentes de interés. Razón por la cual, la mayoría de los compuestos del grupo D presentan mala actividad biológica. Este mismo criterio podría usarse para explicar la mejora global de

comportamiento del grupo C respecto al del grupo B para las posiciones R₁-R₆ estudiadas y por qué las sustituciones en R₇-R₉ son más benéficas en la actividad antiproliferativa (ver Material Suplementario 15).

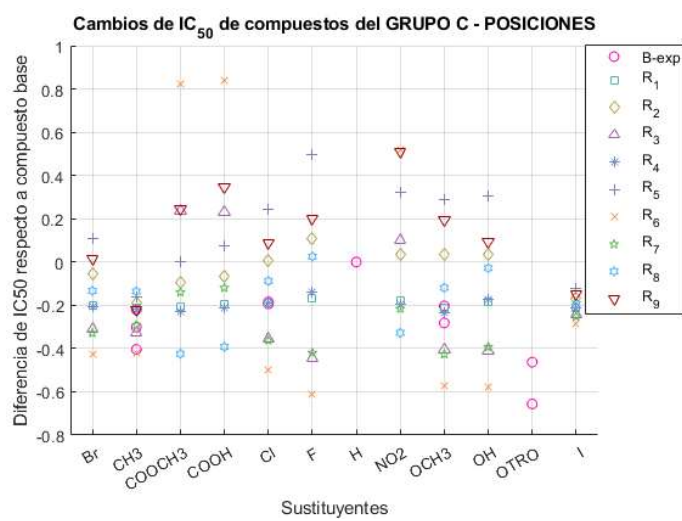
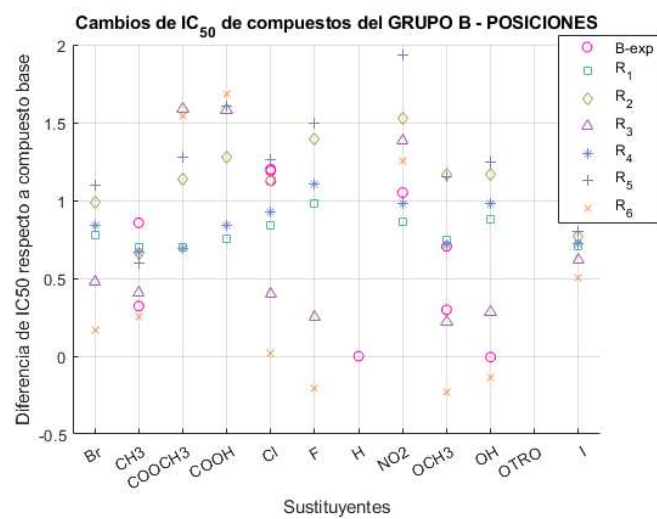
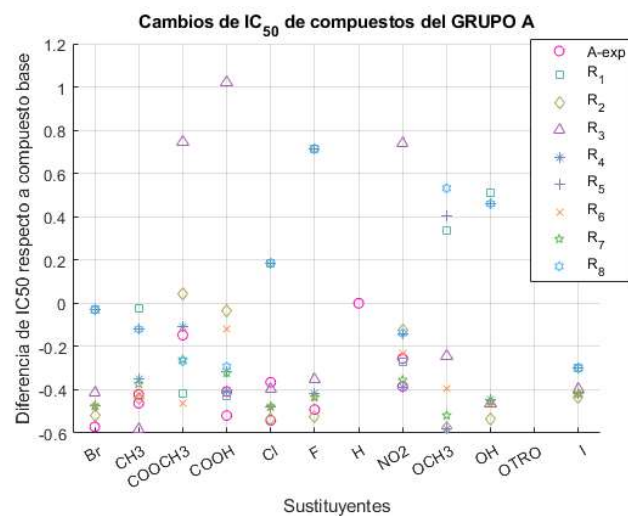
El grupo C presenta, para las posiciones R₁-R₆ valores más altos en estos descriptores que los mostrados por el grupo B. Este aumento de posibles caminos está asociado a la presencia de los anillos aromáticos extra. Si bien, se podrían estar presentando un mayor número de interacciones no benéficas, podemos hipotetizar que parte de esos *caminos* adicionales a 3 6 o 7 enlaces de distancia, serán interacciones benéficas con el N y O de los isoxazoles presentes en su estructura. Del mismo modo, sustituciones en R₇-R₉ son más activas debido a la cercanía con dicha estructura. Específicamente para los compuestos **31** y **32**, la contribución generada por el O y el S a la molécula resulta más benéfica que cualquier contribución generada por anillos aromáticos sustituidos, ya que en el cálculo de los descriptores AATSC el número de caminos posibles para aprovechar el efecto de los sustituyentes se ve afectado por el número de átomos del anillo al que se encuentra unido. Esto implicaría que anillos de furano o tiofeno con alguna sustitución podrían resultar incluso aún más activos que las mejores modificaciones teóricas planteadas basadas en el farmacóforo del grupo C.

DISCUSION BASADA EN ESTRUCTURA

Al momento, se han identificado cualidades generales que vuelven a un compuesto activo, se justificó el uso de los descriptores implementados por los *modelos QSAR ABCD*, se identificaron algunos limitantes predictivos de estos modelos y se relacionaron los descriptores de los *modelos QSAR AB* y *ABCD* con descriptores de mayor simplicidad y significado químico. En esta sección se analizará el efecto específico que sustituciones teóricas en los farmacóforos de los grupos A, B, C y D presentan para la mejora de la actividad biológica frente a cada línea celular.

Efecto de sustituyentes en posiciones específicas frente a MCF-7

La figura 19 muestra los cambios de IC₅₀ frente a MCF-7 generados por un sustituyente en una posición específica. En las siguientes subsecciones se analiza cada grupo independiente basado en la información mostrada por esta figura.



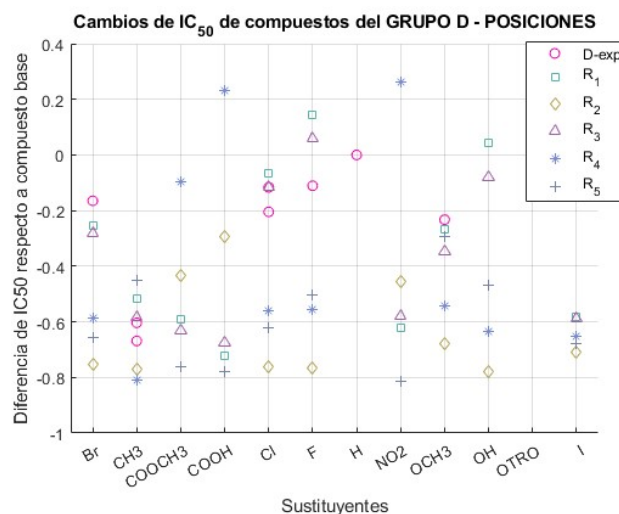


Figura 19 Cambios relativos de IC₅₀ frente a MCF-7 de compuestos de cada grupo con respecto a su molécula base correspondiente (mostrado como el único compuesto con sustituyente H), generados por cada sustituyente evaluado. Valores negativos indican una mejor

Grupo A

Salvo casos particulares para algunos sustituyentes a discutir del grupo A, se nota una clara diferencia de IC₅₀ entre sustituciones en el anillo unido al oxazol (anillo B) y el anillo cercano al nitrógeno secundario (anillo A). Siendo las sustituciones en este último anillo mucho más benéficas que sus contrapartes del anillo unido al oxazol. Así mismo, particularmente para este grupo de compuestos, se identificó que los **modelos QSAR ABCD** tienden a subestimar la IC₅₀ de compuestos que experimentalmente presentaron una IC₅₀ < 20 μM, alrededor de 2 μM (ver Material Suplementario 8 y 9), por lo que en el caso de los compuestos sustituidos con -OCH₃ en R₃, y -CH₃ en R₄, que son los compuestos con mejor IC₅₀ teórica frente a HeLa y MCF-7 respectivamente, se debe considerar la probabilidad de que la actividad biológica predicha sea una subestimación, lo que vuelve a estas sustituciones más atractivas para su evaluación experimental. Abriendo la posibilidad que los mejores representantes de estos compuestos presenten una actividad biológica experimental aún mejor.

Las mejores IC₅₀ teóricas, independientemente de las características del sustituyente, son obtenidas en las posiciones R₂ y R₄ con IC₅₀ buenas para las sustituciones con -Br, -Cl, -F, -NO₂ y -OCH₃ y de igual forma, R₃ para -CH₃ y -OCH₃, obteniendo valores teóricos muy cercanos a los valores experimentales (con una diferencia de ± 2 μM). De tal forma que se pueden fundamentar algunas de las suposiciones de los autores de [34], contemplando la

información teórica de los compuestos evaluados experimentalmente se puede concluir que el efecto benéfico de la actividad biológica depende de sustituyentes halógenos como -F o -Br, lo cual concuerda con literatura ya establecida [43]. En general, las posiciones R₁, R₅ y R₈ son las peores posiciones para poner un sustituyente, independientemente de sus características. Con la excepción de los sustituyentes -COOCH₃, -NO₂ y -COOH, cuyos peores desempeños se generan en la posición R₃. Es en estas posiciones (R₁, R₅, y R₈) donde realizar alguna modificación implica una pérdida de efectividad de entre 40 a más del 100% de actividad biológica respecto al compuesto base correspondiente. De este modo, las IC₅₀ más altas de las moléculas teóricas del grupo A probadas se generan por los sustituyentes -COOCH₃, -COOH, -NO₂ en R₃ y Cl, Br, F, I, OCH₃, CH₃ y OH en R₁, R₅ y R₈. Los sustituyentes -OH, -CH₃ y -OCH₃, presentan un comportamiento similar al de los halógenos atribuible al efecto inductivo que presentan en función de la posición que ocupan en el anillo aromático, aunque se pueden subestimar las IC₅₀.

Particularmente para las sustituciones de un solo átomo de halógeno, las posiciones R₁, R₅ y R₈ generan moléculas teóricas con IC₅₀ intermedias para este grupo ($20 < IC_{50} < 25 \mu M$), mientras que las sustituciones en R₂, R₃, R₆ y R₇ es donde se presentan los mejores valores de IC₅₀ posible para las moléculas sustituidas por algún halógeno, siendo similares entre R₂ y R₃ y entre R₆ y R₇. De manera similar a lo identificado en el caso de sustituciones con halógenos, se obtienen los mismos valores de IC₅₀ calculados por el **modelo QSAR ABCD HeLa**, y similares por el **modelo QSAR ABCD MCF-7** para moléculas sustituidas con un sustituyente -CH₃, -NO₂ o -OH, en posiciones R₅ y R₈ (ver Material Suplementario 10 y 11).

El sustituyente -COOCH₃ obtuvo valores de IC₅₀ relativamente malos ($IC_{50} > 25 \mu M$) para el grupo A. Sin embargo, resulta interesante que la posición R₂, fue la posición probada por el trabajo original ($IC_{50}^{HeLa} = 29.07 \mu M$, $IC_{50}^{MCF-7} = 36.88 \mu M$) para este sustituyente, resulte ser de los peores lugares para posicionar un -COOCH₃ de acuerdo con nuestros **modelos QSAR ABCD** frente a MCF-7 pero de los mejores frente a HeLa. En este sentido, las mejores posiciones teóricas para este sustituyente serían R₄ frente a HeLa ($IC_{50} = 24.7 \mu M$) y R₆ frente a ambas ($IC_{50}^{HeLa} = 27.23 \mu M$, $IC_{50}^{MCF-7} = 23.15 \mu M$), sustituciones no probadas experimentalmente para este sustituyente. R₆ ocupa una posición lo más alejada

posible del N más cercano, lo que haría que la contribución aportada por ese sustituyente sea principalmente interacciones C-C o C-O. Por otro lado, la posición R₄ se encuentra cercana al N del anillo A, pero al presentar un mayor número de átomos O, C y H y encontrarse en una posición donde se pueden generar fácilmente interacciones entre C-C y C-O, de acuerdo con lo planteado en las secciones anteriores, ambas posiciones, R₆ y R₄ resultan provechosas para este tipo de sustituyente. Por lo que valdría la pena realizar su evaluación experimental para corroborar su comportamiento.

Respecto al sustituyente -CH₃, su mejor actividad se observa en R₃, posición no probada experimentalmente, mientras que presenta una actividad decente en R₂, R₄ y R₆. El sustituyente -CH₃ en la posición R₃ presenta una IC₅₀ = 18,11 μM, que resulta ligeramente mejor que cualquier evaluación experimental del grupo A, y es la segunda mejor IC₅₀ teórica solo después de la obtenida por -OCH₃ (IC₅₀ = 18.09 μM) en R₄. Si bien la posición fue estudiada en el trabajo original, este sustituyente no fue evaluado por el trabajo de Gudipati [34], pero los resultados teóricos muestran una IC₅₀ ligeramente mejor que la mejor IC₅₀ experimental probada (IC₅₀ = 18.55 μM) aún incluso contemplando la subestimación para estos sustituyentes. Así mismo, el sustituyente -OH en R₂ presenta una mejora de IC₅₀ comparable con la obtenida por posiciones sustituidas con halógenos. Por lo que el estudio del sustituyente -OCH₃ en R₄, de -CH₃ en R₃ y -OH en R₂ son dos moléculas teóricas más que valdría la pena sintetizar y evaluar experimentalmente.

Grupo B

Los sustituyentes -F, -OCH₃ y -OH son aquellos con mejor valor teórico de mejora relativa de IC₅₀ respecto al compuesto base experimental (**13**) en la línea MCF-7 (IC₅₀^{-F} = 75.309 μM, IC₅₀^{-OCH₃} = 72.074 μM, IC₅₀^{-OH} = 66.997 μM vs IC_{50base_exp}^H = 62.30 μM). De estos descriptores, -F no fue probado experimentalmente en ninguna de las posiciones; -OCH₃ y -OH fueron probados en 2 posiciones, pero no en las identificadas aquí como las más benéficas (ver Figura 4), por lo que valdría la pena sintetizarlos.

Las sustituciones en R₆ y R₃ (posiciones meta- en los anillos laterales y central, respectivamente) son donde se presentan las mejores y segundas mejores IC₅₀ teóricas para casi todos los sustituyentes evaluados, con las excepciones de -NO₂ y -COOH. Para estos sustituyentes el mejor resultado se obtiene para la posición R₁ (sus segundos mejores valores

se presentan en R₄). Para -COOCH₃ la posición R₄ es la mejor y su segundo mejor valor se encuentra en R₁. Las mejores IC₅₀ se obtienen por los sustituyentes -OCH₃, -F y -OH, que llegan a mejorar la IC₅₀ hasta un 20% respecto a la IC₅₀ experimental del compuesto principal (**13**), aunque solo para la posición R₆. Si bien, los mejores valores teóricos de las sustituciones en R₃ no superan el valor de la IC₅₀ experimental del compuesto base (**13**), si superan al valor teórico del compuesto base de su grupo. Por ello, corroborar experimentalmente la efectividad de los sustituyentes -OCH₃, -F y -OH en las posiciones R₃ y R₆ se vuelve de interés. La tendencia general de sustitución para generar los mejores resultados es -orto, -para, y -meta para la mayoría de los sustituyentes (-Br, -Cl, -F, -OCH₃, -OH, -I), incluidos los sustituyentes de interés para este grupo (-OCH₃, -F y -OH), tanto para el anillo central como los laterales. De acuerdo con los resultados de [44], sustituciones con halógeno -Br en la posición -meta de un anillo aromático de sus compuestos estudiados favorece la selectividad y fuerza de enlace hacia canales de calcio dependientes de voltaje (VGCC por sus siglas en ingles), cuyo mal funcionamiento está asociado a múltiples tipos de cáncer, incluido el de mama y de ovario [45]

Grupo C.

Existen varios sustituyentes (-F, -OH, -OCH₃, -Cl, -COOH, -COOCH₃) que llegan a mejorar entre 40-60% la IC₅₀ teórica respecto al compuesto base (compuesto **23**, IC₅₀^{HeLa}= 100.42 μM, IC₅₀^{MCF-7}= 105.28 μM). Dichos sustituyentes resultan muy efectivos en la posición R₆, en segundo lugar, R₇ y en tercer lugar R₃.

El comportamiento observado en el anillo central y laterales observado para el grupo B, se repite en el grupo C. Siendo las sustituciones en R₆ y R₃ de las mejores posiciones para colocar un sustituyente. El efecto de un nuevo par de anillos a los extremos de la molécula presenta una contribución intermedia entre la observada por los anillos central y laterales. Para los anillos extremos, al contrario que para los anillos central y laterales, el efecto observado sobre la posición -para resulta ser la más activa. Presenta las mismas excepciones para los sustituyentes -NO₂, -COOH y -COOCH₃ observadas en el grupo B, sin embargo, sus sustituciones en la posición R₈ (-meta) resultan ser mucho mejores que sus contrapartes en los otros anillos.

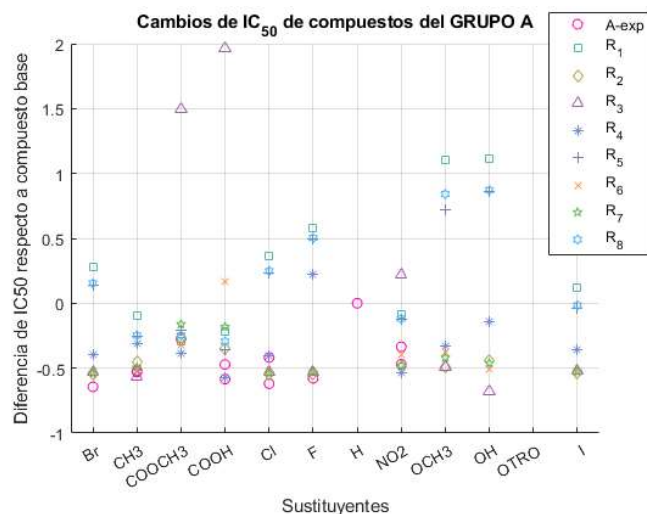
No obstante, ninguna sustitución supera la IC_{50} que se obtiene al modificar los anillos laterales de la molécula por -furilo (furano) o -tienilo (Tiofeno). Por lo que la combinación de una de estas sustituciones con un compuesto cuyos anillos laterales fueran sustituidos parece una opción viable para un estudio experimental.

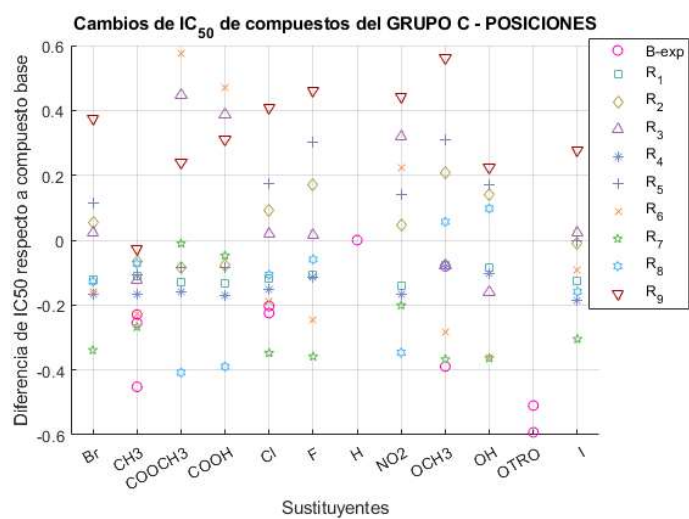
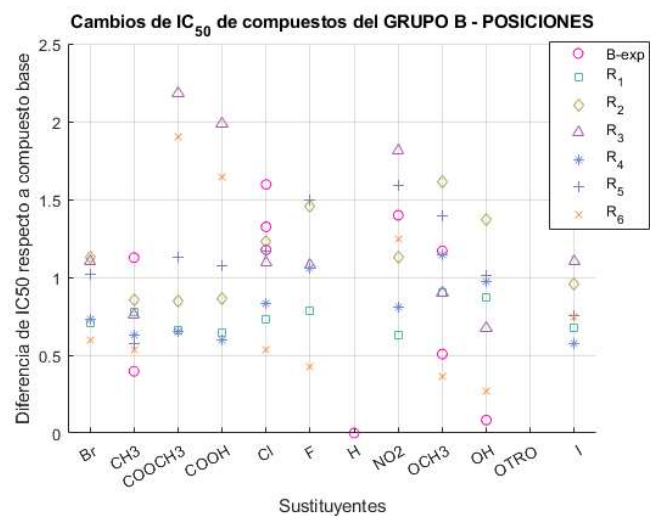
Grupo D

De acuerdo con el **modelo ABCD MCF-7**, el grupo D muestra cambios de IC_{50} más dependientes de la posición sustituida que de las características del sustituyente empleado. La mejor actividad antiproliferativa se obtuvo en R5 ($IC_{50}^{-NO_2} = 38.76 \mu M$), no obstante, la mayoría se obtuvieron en las posiciones R₂ (-Br, -Cl, -F, -I, -OCH₃, -OH) (ver Tabla 7). En general, fueron las sustituciones en R₂ y R₅ las que mejores moléculas generaron en términos de su IC_{50} respecto al valor experimental del compuesto base del grupo D (compuesto **33**, $IC_{50}^{HeLa} = 273.46 \mu M$, $IC_{50}^{MCF-7} = 209.88 \mu M$). Una clara diferencia entre la contribución de halógenos y no halógenos en la actividad antiproliferativa de estos compuestos se observa en este grupo.

Efecto de sustituyentes en posiciones específicas frente a HeLa

De manera similar que con el modelo para MCF-7, se analizó el comportamiento predicho para los compuestos teóricos de acuerdo con lo predicho por el **modelo QSAR ABCD HeLa**. Con la intención de identificar sustituyentes y posiciones específicas que contribuyan a la propuesta de nuevas moléculas teóricas. La Figura 20 muestra el cambio relativo de IC_{50} frente a HeLa de cada una de las moléculas probadas respecto al valor experimental de sus respectivos compuestos base de cada grupo de moléculas.





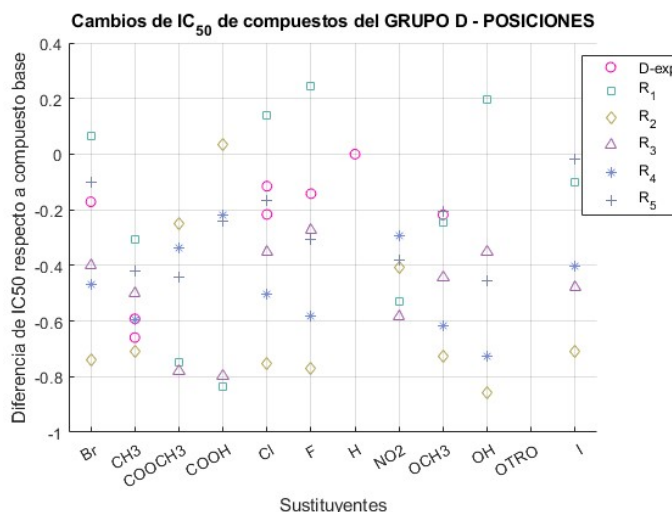


Figura 20 Cambios relativos de IC₅₀ frente a HeLa de compuestos de cada grupo con respecto a su molécula base correspondiente (mostrado como el único compuesto con sustituyente H), generados por cada sustituyente evaluado. Valores negativos indican una mejor

Este modelo muestra comportamientos más similares entre diferentes sustituyentes halógenos, efecto principalmente observable en los grupos A, C y D, donde todos los halógenos presentan casi el mismo nivel de mejora máxima de IC₅₀.

Considerando este modelo, los mejores cambios de IC₅₀ en el grupo A con halógenos se presentan en las posiciones R₂, R₃, R₆ y R₇ mientras la posición R₇ presenta una actividad intermedia. Si bien, las posiciones propuestas por este modelo respecto a su variante para MCF-7 presentan algunas diferencias en la tendencia de actividad antiproliferativa es muy similar para los diferentes halógenos. En este sentido, las dos posiciones probadas experimentalmente, R₂ y R₄, resultaron ser posiciones efectivas para probar sustituciones con halógenos, debido a la contribución generada por la interacción halógeno-N principalmente, interacción descrita también por el **modelo QSAR ABCD MCF-7**. Así mismo, ambos modelos predicen que los sustituyentes -OH, -CH₃ y -OCH₃ generan mejoras en la IC₅₀ similares a las obtenidas por los halógenos teórica y experimentalmente probados, pero siendo sus mejores representantes los sustituidos en la posición R₃ (en lugar de las posiciones R₂, R₃ y R₄, respectivamente). Considerando que se ha visto que la halogenación de anillos aromáticos favorece la inhibición de topoisomerasa I y II, estudios que evalúen la capacidad inhibitoria de topoisomerasas podría ser de interés para los compuestos activos [46].

Respecto al grupo B, los valores predichos se asemejan más al comportamiento experimental observado, en el sentido que ninguna sustitución mejoró sustancialmente la IC_{50} de su compuesto base. No obstante, sigue identificando a las posiciones R_3 y R_6 (posiciones -orto del anillo central o de los laterales) como las mejores posiciones para sustituir y a la posición R_6 como la mejor de ambas. Una de las razones por las que, en ambos casos, la sustitución en -orto de los anillos laterales parece ser más efectiva podría residir en el hecho que esta posición favorece la interacciones con el Na 6 o 7 átomos de distancia. Por otro lado, la sustitución en R_6 representa la adición de dos sustituyentes a la molécula base del grupo B, mientras que la sustitución en R_3 solo añade un único sustituyente. Así mismo, se identifica que -OH, -OCH₃, -F y -Cl, son los que presentan la mejor actividad biológica posible en las posiciones R_6 , nuevamente. Debido a que esta posición ni estos sustituyentes fueron estudiados, siguen siendo sustituciones candidatas a una evaluación experimental, ya que resultarían efectivas para ambas líneas tumorales, MCF-7 y HeLa.

Respecto al grupo C, las sustituciones en posiciones R_8 y R_7 resultan ser de mayor contribución que las sustituciones dobles que involucran sustituir en R_6 . Lo que destaca la importancia de los anillos extremos de la molécula en la contribución a una buena actividad biológica. En este caso, al igual que para la actividad frente a MCF-7, ninguna sustitución simple (ver Figura 20-C) genera mejores resultados teóricos que los experimentalmente obtenidos por el compuesto **32** cuyos anillos extremos fueron sustituidos por un un tiofeno ($IC_{50}^{32-HeLa} = 40.95 \mu M$, $IC_{50}^{32-MCF-7} = 36.19 \mu M$), siendo los más cercanos la sustitución en R_8 con -COOCH₃ ($IC_{50} = 59.425 \mu M$) y la sustitución en R_6 con F ($IC_{50} = 40.6$) frente a HeLa y MCF-7, respectivamente. Por lo que prevalece la idea que una combinación entre una buena sustitución en R_6 junto con un cambio en los anillos extremos por furil o tienil, generaría incluso mejores resultados que únicamente por cambiar los anillos extremos.

Finalmente, respecto al grupo D, se identifica el mismo comportamiento respecto a la dependencia de la mejora de la IC_{50} principalmente debido a la posición sustituida más que a las características del sustituyente. Al igual que para el modelo en MCF-7, se identifica que en general las posiciones R_2 y R_4 son las mejores posiciones para colocar sustituyentes -OH, -OCH₃ o cualquier halógeno. Mientras que las posiciones R_1 y R_3 resultarían mejor para posicionar sustituyentes -COOCH₃, -COOH y -NO₂.

Por último, un comportamiento constante fue el mostrado por el sustituyente -I bajo el **modelo QSAR ABCD MCF-7**. Este sustituyente presentó la variación más pequeña en los cambios de IC_{50} independiente de la posición sustituida en alguno de los 4 grupos de estudio. Su inclusión dentro de cualquiera de las sustituciones mejora la IC_{50} teórica calculada respecto al compuesto base, independientemente de la estructura donde se use. Mientras que en ambos **modelos QSAR ABCD**, presentó un intervalo de modificación más reducido que los de otros sustituyentes. Esto sería una ventaja al enfocar el estudio predictivo de nuevos farmacóforos. Por lo que, la sustitución con -I podría ser una opción para considerar si se desea una leve mejora con seguridad en la actividad biológica o si se planea estudiar el comportamiento de nuevos farmacóforos.

Contraste con autores originales

Contemplando únicamente información de los sustituyentes -Cl, -Br, -F, -CH₃, -COOH, -NO₂ y -COOCH₃ y las sustituciones (R₂ y R₄) evaluados experimentalmente se encuentra una concordancia con la propuesta de Gudapati para las moléculas del grupo A considerando el análisis basado en el **modelo ABCD** [34]. Este trabajo muestra que las posiciones R₂ y R₄ con diferentes sustituyentes pueden generar una mejor actividad antiproliferativa.

Los compuestos del grupo A son oxazoles, mientras que los compuestos del grupo B son isoxazoles. Se ha identificado que los compuestos basados en isoxazoles, aplicados como anticancerígenos, pueden actuar por diversos mecanismos, incluyendo inhibidores de aromatasas, de quinasas, de timidilato, de histona desacetilasa (HDAC), de fosfolipasa a₂, receptor 2 del factor de crecimiento epidérmico (HER2), receptores de estrógeno alfa (ER- α), así como inductores de apoptosis o disruptores de tubulina [47]. El Grupo A presenta en su estructura a la isatina, que presenta diversas aplicaciones, entre ellas como agente anticancerígeno, [34] actividad asociada a su capacidad de formar grupos ionizables positivos, su estructura plana y el carácter hidrofóbico de sus derivados [48].

En los trabajos de [35] y [37] publicados en 2011 y 2012 [grupo C y B], solo se estudió al -Cl como único sustituyente halógeno, el cual mostró una baja actividad biológica. De acuerdo con los valores de los descriptores respecto a otros halógenos (-F, -Br, -I) probados

en este trabajo, el -Cl muestra una actividad muy diferente para las contribuciones electrotopológicas utilizadas en los modelos QSAR aquí propuestos (Ver Material Suplementario 17). Además, las únicas posiciones evaluadas resultar ser posiciones que no favorecen a los halógenos de acuerdo con nuestros *modelos QSAR ABCD*. Por estas razones, los resultados obtenidos por el uso de -Cl como sustituyente no pueden ser utilizados para generalizar el comportamiento del resto de halógenos. Y no podría asegurarse que el uso de halógenos como sustituyentes sea perjudicial para todos los casos donde se desee usar en los farmacóforos de los grupos B y C, como lo plantea el autor en su trabajo.

Las posiciones estudiadas experimentalmente en los compuestos del grupo C favorecen interacciones de tipo C-C a 3, 6 o 7 enlaces de distancia de acuerdo con nuestros modelos. Razón que se puede asociar con el hecho que la sustitución con -CH₃ en R₁ del anillo intermedio es la sustitución que mejor IC₅₀ experimental dio en este grupo (IC₅₀^{HeLa} = 55.01 μM y IC₅₀^{MCF-7} = 62.75 μM), excluyendo a los compuestos **31** y **32**). Es probable que, de haber sido estudiada experimentalmente, esta misma posición también hubiera generado una buena actividad antiproliferativa para compuestos del grupo B. Por último, la diferencia principal entre las estructuras del grupo B y el C, que es el tipo de anillo unido a los isoxazoles, se identifica como la razón de que el grupo C sea más propenso a mejorar la IC₅₀ de sustituyentes en R₁-R₆ respecto a la obtenible por el grupo B.

Los compuestos del grupo D [36] muestran un comportamiento distinto al presentado por los compuestos de los grupos A, B o C, cuya naturaleza se ha discutido en secciones anteriores. Las predicciones generadas por los *modelos QSAR ABCD* para este grupo se logran predecir con baja variación la actividad biológica experimental (ver Material Suplementario 8 y 9), salvo casos donde la predicción resulta ser subestimada. En este sentido, este compuesto presenta más potenciales caminos que, de haber sido considerados durante la selección de las sustituciones experimentales a evaluar y de acuerdo con nuestro modelo, se hubiesen obtenido valores de IC₅₀ experimentales mucho más activas que las obtenidas en su respectivo trabajo original. En este sentido, no se puede usar el mismo criterio de una posición -para, -meta u -orto respecto al N más para identificar la mejor actividad antiproliferativa, criterio que parecen haber usados en los trabajos originales del grupo A y D.

Otro de los trabajos de [37] con compuestos en cuya estructura presentan al menos un isoxazol muestra una tendencia similar respecto a la presencia de sustituyentes propuesta por nuestro modelo. Tal es el caso del trabajo [49], en donde grupos electroattractores generan una actividad anticancerígena mejor en contraste con aquellos sustituidos con grupos electrodonadores. En este trabajo, no se estudió su actividad en la línea tumoral HeLa ni fue realizado algún ensayo para identificar el mecanismo de acción, por lo que no cumplió los criterios de inclusión del presente estudio ni se puede sugerir que el mecanismo de acción sea similar al que rige los compuestos estudiados. Si bien el mecanismo de acción podría ser distinto, la tendencia de los halógenos para mejorar la actividad antitumoral se mantiene en posiciones donde haya más interacciones entre C-N a 3, 6 o 7 enlaces de distancia. Curiosamente, para los compuestos en este ejemplo, el -Cl es el que tiene el mejor desempeño. Al igual que con [35], la actividad mejora significativamente si se sustituye un benceno con un -furano o tiofeno.

Otros estudios teóricos de los compuestos reportados

En el conocimiento actual, este es el primer estudio que aborda el análisis descriptivo de los compuestos aquí analizados, para dar así una explicación de la actividad antiproliferativa en dos líneas tumorales de forma individual y plante la posibilidad de basarse en el mismo mecanismo de acción frente a células tumorales HeLa y MCF-7, basada en un análisis cuantitativo estructura-actividad (QSAR).

Acercamientos previos de la explicación del comportamiento de moléculas evaluadas por los trabajos originales solo es abordado en el análisis SAR realizado en la revisión de [50] para moléculas de [37]. En su análisis de relación estructura-actividad (SAR) recupera la propuesta abordada por el autor original que da importancia a sustituciones de grupos electrodonadores como metil y metoxi aumentan la actividad biológica, mientras que grupos electroattractores como cloro y nitro disminuyen la actividad anticancerígena. Sin embargo, al no haber sido probados otros tipos de grupos halógenos sustituyentes por parte del trabajo original tales como -Br o -F, generó que la propuesta planteada en la revisión parezca sesgada.

ALCANCE DE MODELOS QSAR y PRPUESTA DE NUEVAS MOLÉCULAS

En la literatura existen casos donde se reporta que los modelos matemáticos no concuerdan con los datos experimentalmente observados. Por lo general, en estos casos se atribuye a que el modelo no resultó experimentalmente utilizable por no representar con fiabilidad el comportamiento químico-biológico de los compuestos, y generalmente se termina desechando los modelos propuestos. Sin embargo, antes de desechar por completo un modelo QSAR se debería considerar, bajo criterio del autor, al menos lo siguiente:

- 1) Que el modelo cuente con validez matemática, es decir, que la información con la que se haya realizado el modelo QSAR cumpla los criterios matemáticos supuestos por la herramienta matemática usada (por ejemplo, normalidad en la distribución de los datos),
- 2) Que las diferencias observadas experimentalmente sean mayores a las equivalentes previstas por la correlación del ajuste y los datos del modelo (R^2) tanto por el conjunto de entrenamiento, como por el conjunto de validación,
- 3) Identificar la particularidad y alcance del comportamiento químico-biológico del modelo matemático aplicado y,
- 4) Considerar la solidez de los criterios tanto químicos como matemáticos y la interpretación química dado al modelo QSAR durante la proposición de las nuevas moléculas.

A continuación, se discuten algunos ejemplos.

PREDICCIÓN DE NUEVOS COMPUESTOS. ALCANCE DE LOS MODELOS PROPUESTOS

La Figura 21 muestra una representación visual de todos los cambios generados en cada estructura en función del sustituyente y su posición ocupada.

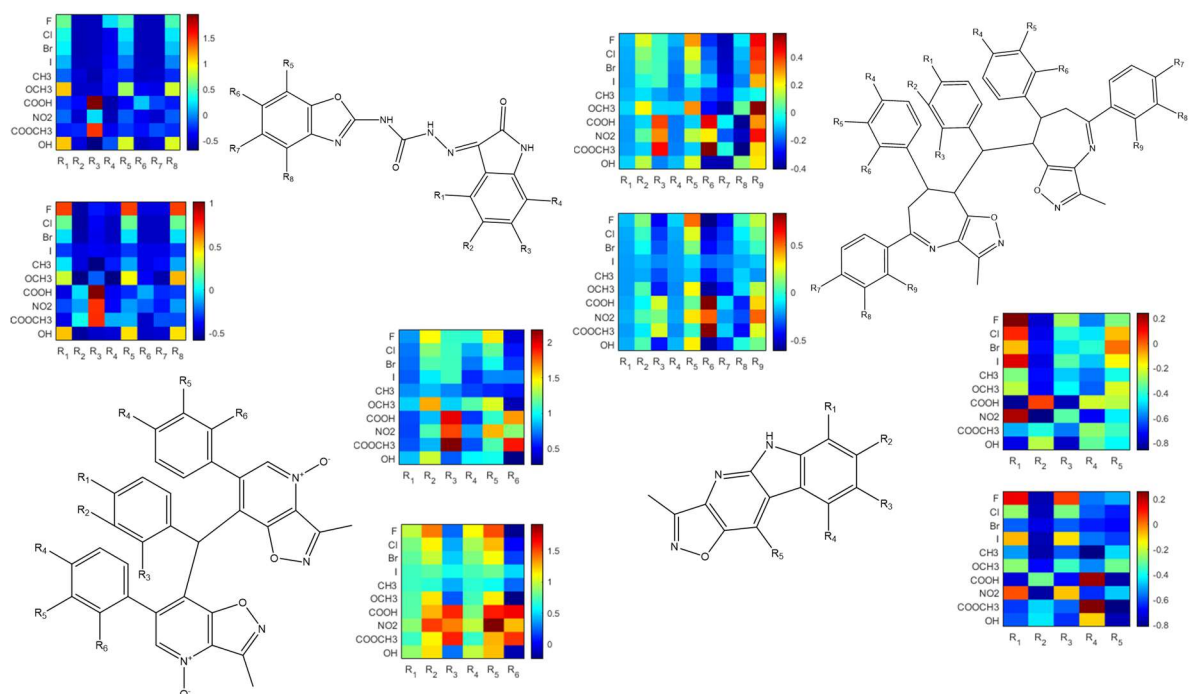


Figura 21 Representación visual de modificaciones de IC50 de los farmacóforos de los grupos A, B, C y D con cada uno de los sustituyentes en posiciones específicas. Valores más bajos representan una mejora en la potencia de los compuestos y valores negati

Si lo que se desea es seguir estudiando el comportamiento de los farmacóforos en los grupos A, B, C y D, se recomienda priorizar sustituyentes y posiciones mostradas en color azul en la Figura 21. Con base en el análisis realizado, la Tabla 10 muestra las moléculas a priorizar para una evaluación experimental frente a las líneas celulares de interés. Así mismo, se recomendaría realizar dobles sustituciones en anillos alejados a átomos N u O ya que se ha visto que zonas con sustituyentes ricos en electrones favorecen la potencia de compuestos basados en azoles [31].

Tabla 10 Propuesta de evaluación de moléculas teóricas basadas en los grupos A, B, C y D.

MCF-7			HeLa	
Grupo	Posición	Sustituyente	Posición	Sustituyente
A	R ₁ , R ₆	-COOCH ₃	R ₃	-OH, -OCH ₃ , C-H ₃
	R ₃	-CH ₃		
	R ₄	-OCH ₃		
B	R ₃ , R ₆	-OCH ₃ , -F, -OH	R ₃ , R ₆	Mismas que para MCF-7, -Cl
C	R ₈	-COOH, -COOCH ₃	R ₈	Mismas que para MCF-7, -NO ₂
	R ₆ , R ₇	-F, -Cl, -OH, -OCH ₃	R ₇	Mismas que para MCF-7, -Br, -I
D	R ₂	Halógenos	R ₂	Halógenos, -OH, -OCH ₃
	R ₅	-COOH, -COOCH ₃ , -NO ₂	R ₁ , R ₃	-COOH, -COOCH ₃ , -NO ₂

RECOMENDACIONES PARA LA BUSQUEDA DE NUEVOS FARMACÓFOROS.

Una potencial ventaja de los *modelos QSAR AB complejos* radica en que utiliza información estructural para el descriptor de mayor contribución, lo cual podría propiciar un análisis con mayor libertad de modificación de estructuras teóricas más allá de sustituyentes únicos. Por otro lado, permite dar una descripción que potencialmente puede aplicarse a la descripción del comportamiento y predicción de compuestos con estructuras diferentes que actúen de forma similar a los compuestos del grupo A y B frente a células anticancerígenas de HeLa y MCF-7. Sin embargo, la aplicación este modelo más allá de cambio de sustituyentes simples debe ser tomada con cautela, debido a que, en primer lugar, no se tienen datos experimentales de posibles mecanismos de acción obtenidos por los mismos ensayos para los 4 grupos de moléculas originales.

En segundo lugar, todos los modelos QSAR fueron realizados con moléculas con presencia de uno o más átomos de C, N, O, o halógenos. Por lo que tratar de predecir el comportamiento de moléculas con grupos funcionales distintos a estos, podría generar resultados no efectivos. Como ejemplo de esta posibilidad se cuenta con los compuestos **31** y **32**, pertenecientes al grupo C. Estos compuestos presentaron una modificación en el farmacóforo intercambiando un anillo aromático por un -furilo o un -tienilo, respectivamente. Los *modelos QSAR ABCD* pudieron predecir el comportamiento del compuesto **31** con una buena precisión respecto a su valor experimental para ambas líneas celulares. Por otro lado, estos modelos no mostraron buena precisión en la actividad biológica del compuesto **32**. Lo cual se asocia a la presencia de los átomos de S que conforman el -tienilo, siendo este compuesto el único con -S en su estructura. De este modo, se recomienda limitar el uso de los *modelos QSAR ABCD* para fines predictivos de nuevos farmacóforos únicamente a estudiar aquellos cuya estructura principal no presente átomos distintos a C, N u O.

Respecto a los modelos QSAR planteados en este trabajo. Dado que la actividad biológica experimentalmente cuantificada es similar entre las líneas tumorales HeLa y MCF-7, se podría usar como criterio empírico que las IC₅₀ predichas de moléculas teóricas propuestas sean similares en ambas líneas tumorales. Así mismo, debido a que las ramificaciones presentes en las moléculas contribuyen con el cálculo de ambos descriptores ATSC6e y AATSC3v al mismo tiempo, contemplar directamente valores benéficos de ambos

descriptores durante la propuesta de nuevas moléculas podría ser un segundo criterio. Utilizar estos criterios podría mejorar la confiabilidad del modelo teórico aplicado a la predicción de IC_{50} de compuestos con diferentes estructuras base. Reduciendo así el riesgo de una mala predicción derivada de una errónea aplicación de estos modelos en compuestos cuyo comportamiento antiproliferativo sea diferente al de los grupos A, B, C o D (no selectividad pronunciada entre líneas tumorales HeLa y MCF-7).

Uno de los temas discutidos en la sección **DISCUSION > MODELOS ABCD** fue que los *modelos QSAR ABCD* presentan particular problema al modelar el comportamiento de los farmacóforos de los 4 grupos. Las principales causas a las que se puede atribuir son debido a la naturaleza de los descriptores usados. A lo largo de las secciones **DISCUSION > MODELO QSAR. Relación con descriptores simples** y **DISCUSIÓN > MODELOS ABCD** se presentaron evidencias que permiten concluir que el principal problema para modelar el comportamiento de los farmacóforos sin ninguna sustitución se debe tanto a la naturaleza de los descriptores usados, como a la variabilidad estructural presente en cada uno de los grupos.

Al encontrarse una explicación matemática de la discrepancia con de la actividad biológica predicha para estos compuestos, se descarta la posibilidad de encontrarnos con compuestos que resulten ser *acantilados de actividad*. Con base en este análisis, se puede sugerir que, si se desea evaluar nuevas potenciales estructuras bajo los modelos QSAR ABCD tanto para HeLa como para MCF-7, la mejor opción es realizar la evaluación conjunta de una familia entera de compuestos sustituidos en zonas de interés. De tal forma que pueda ser considerada al promedio de la IC_{50} de dicho conjunto como un estimado de la actividad biológica de dicho farmacóforo teórico. Siendo esto particularmente válido para moléculas similares a las del grupo A y D.

Es así como queda fundamentada el alcance y limitaciones de los modelos QSAR presentados en este manuscrito. Por último, la Tabla 11 muestra los modelos QSAR con las mejores estadísticas posibles para cada enfoque de análisis QSAR realizado: *modelos QSAR AB* y *modelos QSAR ABCD*, que han de ser usados preferentemente bajo las recomendaciones aquí descritas.

Tabla 11 Modelos QSAR de actividad biológica frente a líneas tumoral HeLa y MCF-7, obtenidos por diferentes aproximaciones.

ID (Grupos Usados)	Modelo QSAR generado
AB 1.1 (A y B)	QSAR HeLa
	$-Log\left(\frac{1}{IC_{50}}\right) = -0.404 + 1.388 SpMin5_{Bhv} - 0.047 minHBa$ $R_1 = 0.427, \quad R^2 = 0.952, \quad R^2_{aj} = 0.945, \quad Fstat = 128.812,$ $p - valor = 2.69309 E^{-9}, RMSE = 0.080$ $R^2_{test} = 0.788 R^2_{aj_{test}} = 0.735$
	QSAR MCF-7
	$-Log\left(\frac{1}{IC_{50}}\right) = -0.285 + 1.328 SpMin5_{Bhv} - 0.047 minHBa$ $R_1 = 0.427, \quad R^2 = 0.962, R^2_{aj} = 0.956, \quad Fstat = 162.973,$ $p - valor = 6.234 E^{-10}, RMSE = 0.068$ $R^2_{test} = 0.817 R^2_{aj_{test}} = 0.771$
ABCD (A, B, C y D)	QSAR HeLa
	$-Log\left(\frac{1}{IC_{50}}\right) = 2.0151 + 0.028377 AATSC3v - 1.7778 MATS6e$ $R_1 = 0.221, \quad R^2 = 0.907, R^2_{aj} = 0.900, \quad Fstat = 117.296,$ $p - valor = 4.085 E^{-13}, RMSE = 0.1211$ $R^2_{test} = 0.748 R^2_{aj_{test}} = 0.725$
	QSAR MCF-7
	$-Log\left(\frac{1}{IC_{50}}\right) = 1.9306 - 0.10924 ATSC6e + 1.0401 MATS7e$ $R_1 = 0.465, \quad R^2 = 0.906, R^2_{aj} = 0.898, \quad Fstat = 115.44,$ $p - valor = 4.858 E^{-13}, RMSE = 0.1015$ $R^2_{test} = 0.852 R^2_{aj_{test}} = 0.838$

Capítulo 5.

Conclusiones

Se propone un total de 30 modelos QSAR independientes de 2 descriptores que permiten evaluar el comportamiento específico de 2 grupos de compuestos con mejor actividad antiproliferativa dando prioridad a la capacidad matemática del modelo (**modelos QSAR AB complejos**), a un análisis químicamente simple (**modelos QSAR AB simples**) y; el comportamiento de los 4 grupos de compuestos de estudio (**modelos QSAR ABCD**). Estos modelos describen la actividad antiproliferativa (medida como IC₅₀) evaluada experimentalmente frente a las líneas tumorales MCF-7 o HeLa, de **40** compuestos sintetizados y evaluados experimentalmente y **+200** compuestos teóricos basados en los farmacóforos de estos grupos. Estos modelos QSAR tienen las mejores características posibles considerando: validez de conjuntos usados para la generación de modelos QSAR, la no colinealidad de descriptores usados en un mismo análisis QSAR, mínimo uso de descriptores, un excelente ajuste lineal entre el modelo matemático y los datos experimentales, una excelente capacidad predictiva ($R^2_{\text{test}} > 0.85$ en los modelos principales) y, descripción química.

Se generó una descripción matemática y química basada en información de los descriptores usados por los **modelos QSAR ABCD**, *AATSC3v*, *ATSC6e*, *MATS6c* y *MATS7e*. que describe el comportamiento general y específico de los compuestos estudiados. La descripción puede ser aplicada en el estudio de nuevos fármacos con excelente potencial

antitumoral frente a HeLa y MCF-7, y se definió un intervalo de seguridad donde sea válido aplicar estos modelos QSAR. Así mismo, las conclusiones generadas por cada modelo QSAR son independientes una de otra, sin embargo, pueden ser complementarias entre ellas. Y, dado que los compuestos aquí estudiados presentan una actividad biológica similar en ambas líneas tumorales, se podría contemplar al mismo tiempo los modelos QSAR de ambas líneas tumorales durante el planteamiento de nuevos compuestos.

Se identificaron regiones benéficas y perjudiciales para la IC_{50} de cada grupo de moléculas, identificando las mejores posiciones específicas para cada sustituyente. De igual forma, se identificaron posiciones y sustituyentes específicos a priorizar para una evaluación experimental. Con base en el análisis químico y matemático de los descriptores discutido para los **modelos QSAR ABCD** de HeLa y MCF-7, se propusieron nuevas modificaciones a las estructuras de los farmacóforos de cada grupo (ver Tabla 10). Finalmente, se estudió la capacidad de la aproximación proporcionada por dichos modelos QSAR para estudiar, y proponer, nuevos farmacóforos totalmente distintos a los usados por los grupos de compuestos estudiados.

Capítulo 6.

Referencias

- [1] N. Y. Guerrero-Pepinosa, M. C. Cardona-Trujillo, S. C. Garzón-Castaño, L. A. Veloza, and J. C. Sepúlveda-Arias, “Antiproliferative activity of thiazole and oxazole derivatives: A systematic review of in vitro and in vivo studies,” *Biomed. Pharmacother.*, vol. 138, no. March, 2021, doi: 10.1016/j.biopha.2021.111495.
- [2] WHO, “Cancer,” *WHO Portal*, 2022. <https://www.who.int/news-room/fact-sheets/detail/cancer#:~:text=Key facts,and rectum and prostate cancers> (accessed Aug. 10, 2022).
- [3] R. E. Marklew, A. A. Jackson, M. J. Wiseman, and S. A. Wootton, “ICONIC: An international task force supporting collaboration in nutrition and cancer globally,” *Trends Food Sci. Technol.*, vol. 130, no. August, pp. 3–10, 2022, doi: 10.1016/j.tifs.2022.08.019.
- [4] J. G. Hernández-Pérez *et al.*, “Metabolic Syndrome and Prostate Cancer Risk in Mexican Men: A Population Case-control Study,” *Arch. Med. Res.*, vol. 53, pp. 594–602, 2022, doi: 10.1016/j.arcmed.2022.07.003.
- [5] C. Loveday *et al.*, “Analysis of rare disruptive germline mutations in 2,135 enriched BRCA-negative breast cancers excludes additional high-impact susceptibility genes,” *Ann. Oncol.*, vol. xxx, no. xxx, 2022, doi: 10.1016/j.annonc.2022.09.152.
- [6] S. Winters, A. Alomari, G. Shokar, C. Martin, A. Dwivedi, and N. K. Shokar, “Breast cancer screening outcomes among Mexican-origin Hispanic women participating in a breast cancer screening

- program,” *Prev. Med. Reports*, vol. 24, p. 101561, 2021, doi: 10.1016/j.pmedr.2021.101561.
- [7] B. J. Ager *et al.*, “Advancing clinical research globally: Cervical cancer research network from Mexico,” *Gynecol. Oncol. Reports*, vol. 25, no. June, pp. 90–93, 2018, doi: 10.1016/j.gore.2018.06.007.
 - [8] E. Lazcano-Ponce and B. Allen-Leigh, “Innovation in Cervical Cancer Prevention and Control in Mexico,” *Arch. Med. Res.*, vol. 40, no. 6, pp. 486–492, 2009, doi: 10.1016/j.arcmed.2009.07.007.
 - [9] J. H. Heimovaara *et al.*, “Ten-year experience of a national multidisciplinary tumour board for cancer and pregnancy in the Netherlands,” *Eur. J. Cancer*, vol. 171, pp. 13–21, Aug. 2022, doi: 10.1016/j.ejca.2022.04.040.
 - [10] U. T. Kadam *et al.*, “Conceptualizing multiple drug use in patients with comorbidity and multimorbidity: proposal for standard definitions beyond the term polypharmacy,” *J. Clin. Epidemiol.*, vol. 106, pp. 98–107, 2019, doi: 10.1016/j.jclinepi.2018.10.014.
 - [11] J. E. Welsh, *Animal Models for Studying Prevention and Treatment of Breast Cancer*. Elsevier, 2013.
 - [12] E. Connect, “Fases de desarrollo de un nuevo fármaco,” *Elsevier*, 2020. <https://www.elsevier.com/es-es/connect/medicina/edu-fases-de-desarrollo-de-un-nuevo-farmaco>.
 - [13] I. Union, O. F. Pure, and A. Chemistry, “Glossary of Terms Used in Medicinal Chemistry,” vol. 70, no. 5, pp. 1129–1143, 1998, doi: 10.1351/pac199870051129.
 - [14] C. T. Korch and A. Capes-Davis, “The Extensive and Expensive Impacts of HEp-2 [HeLa], Intestine 407 [HeLa], and Other False Cell Lines in Journal Publications,” *SLAS Discov.*, vol. 26, no. 10, pp. 1268–1279, 2021, doi: 10.1177/24725552211051963.
 - [15] P. Nambisan, “Relevance of Intellectual Property Rights in Biotechnology,” *An Introd. to Ethical, Saf. Intellect. Prop. Rights Issues Biotechnol.*, pp. 291–309, 2017, doi: 10.1016/b978-0-12-809231-6.00013-2.
 - [16] C. Korch and M. Varella-Garcia, “Tackling the Human Cell Line and Tissue Misidentification Problem Is Needed for Reproducible Biomedical Research,” *Adv. Mol. Pathol.*, vol. 1, no. 1, pp. 209–228.e36, 2018, doi: 10.1016/j.yamp.2018.07.003.
 - [17] M. C. Alley *et al.*, “Feasibility of Drug Screening with Panels of Human Tumor Cell Lines Using a Microculture Tetrazolium Assay,” *Cancer Res.*, vol. 48, no. 3, pp. 584–588, 1988.
 - [18] F. Ghasemi, A. Mehridehnavi, A. Pérez-Garrido, and H. Pérez-Sánchez, “Neural network and deep-

- learning algorithms used in QSAR studies: merits and drawbacks,” *Drug Discovery Today*, vol. 23, no. 10. Elsevier Ltd, pp. 1784–1790, Oct. 01, 2018, doi: 10.1016/j.drudis.2018.06.016.
- [19] S. R. Khan, D. Al Rijjal, A. Piro, and M. B. Wheeler, “Integration of AI and traditional medicine in drug discovery,” *Drug Discov. Today*, vol. 26, no. 4, pp. 982–992, 2021, doi: 10.1016/j.drudis.2021.01.008.
- [20] S. Y. Liao, J. C. Chen, L. Qian, Y. Shen, and K. C. Zheng, “QSAR, action mechanism and molecular design of flavone and isoflavone derivatives with cytotoxicity against HeLa,” *Eur. J. Med. Chem.*, vol. 43, no. 10, pp. 2159–2170, 2008, doi: 10.1016/j.ejmech.2007.10.033.
- [21] R. Todeschini and V. Consonni, “Handbook of Molecular Descriptors (Google eBook),” p. 688, 2008, [Online]. Available: <http://books.google.com/books?id=TCuHqbgvMbEC&pgis=1>.
- [22] L. Y. Xia, Y. W. Wang, D. Y. Meng, X. J. Yao, H. Chai, and Y. Liang, “Descriptor selection via log-sum regularization for the biological activities of chemical structure,” *Int. J. Mol. Sci.*, vol. 19, no. 1, pp. 1–15, 2018, doi: 10.3390/ijms19010030.
- [23] G. Sliwoski, J. Mendenhall, and J. Meiler, “Autocorrelation descriptor improvements for QSAR: 2DA_Sign and 3DA_Sign,” *J. Comput. Aided. Mol. Des.*, vol. 30, no. 3, pp. 209–217, Mar. 2016, doi: 10.1007/s10822-015-9893-9.
- [24] Yapcwsoft, “PaDEL Descriptor.” 2022, [Online]. Available: <http://www.yapcwsoft.com/dd/padeldescriptor/>.
- [25] B. Hollas, “An analysis of the autocorrelation descriptor for molecules,” 2003. doi: 10.1023/A:1023247831238.
- [26] V. Facchinetti, C. R. B. Gomes, and M. V. N. de Souza, “Application of nitriles on the synthesis of 1,3-oxazoles, 2-oxazolines, and oxadiazoles: An update from 2014 to 2021,” *Tetrahedron*, vol. 102, p. 132544, 2021, doi: 10.1016/j.tet.2021.132544.
- [27] K. D. Katariya, D. R. Vennapu, and S. R. Shah, “Synthesis and molecular docking study of new 1,3-oxazole clubbed pyridyl-pyrazolines as anticancer and antimicrobial agents,” *J. Mol. Struct.*, vol. 1232, May 2021, doi: 10.1016/j.molstruc.2021.130036.
- [28] F. E. Bennani *et al.*, “Overview of recent developments of pyrazole derivatives as an anticancer agent in different cell line,” *Bioorg. Chem.*, vol. 97, p. 103470, 2020, doi: 10.1016/j.bioorg.2019.103470.
- [29] V. Balaraju, S. Kalyani, G. Sridhar, and E. Laxminarayana, “Design, synthesis and biological assessment of 1,3,4-oxadiazole incorporated oxazole-triazole derivatives as anticancer agents,” *Chem.*

Data Collect., vol. 33, pp. 1–6, 2021, doi: 10.1016/j.cdc.2021.100695.

- [30] P. Edukondalu, R. Sireesha, C. M. Bandaru, M. V. B. Rao, P. Kala, and R. R. Raju, “Design, Synthesis and anticancer evaluation of 2-(5-(Benzo[d]thiazol-2-yl)-1H-imidazol-1-yl)-5-aryl-1H-benzo[d]imidazole derivatives,” *Chem. Data Collect.*, vol. 35, no. July, 2021, doi: 10.1016/j.cdc.2021.100753.
- [31] K. Donthiboina *et al.*, “Synthesis and biological evaluation of substituted N-(2-(1H-benzo[d]imidazol-2-yl)phenyl)cinnamides as tubulin polymerization inhibitors,” *Bioorg. Chem.*, vol. 103, no. January, p. 104191, 2020, doi: 10.1016/j.bioorg.2020.104191.
- [32] R. R. Bhandare *et al.*, “Multistep synthesis and screening of heterocyclic tetrads containing furan, pyrazoline, thiazole and triazole (or oxadiazole) as antimicrobial and anticancer agents,” *J. Saudi Chem. Soc.*, vol. 26, no. 3, p. 101447, 2022, doi: 10.1016/j.jscs.2022.101447.
- [33] N. D. Tangellamudi, S. B. Shinde, V. Pooladanda, C. Godugu, and S. Balasubramanian, “Facile synthesis of 2-aryl 5-hydroxy benzo[d]oxazoles and their in vitro anti-proliferative effects on various cancer cell lines,” *Bioorganic Med. Chem. Lett.*, vol. 28, no. 23–24, pp. 3639–3647, 2018, doi: 10.1016/j.bmcl.2018.10.038.
- [34] R. Gudipati, R. N. Reddy Anreddy, and S. Manda, “Synthesis, anticancer and antioxidant activities of some novel N-(benzo[d]oxazol-2-yl)-2-(7- or 5-substituted-2-oxoindolin-3-ylidene)hydrazinecarboxamide derivatives,” *J. Enzyme Inhib. Med. Chem.*, vol. 26, no. 6, pp. 813–818, Dec. 2011, doi: 10.3109/14756366.2011.556630.
- [35] E. Rajanarendar, M. Nagi Reddy, K. Rama Murthy, P. Surendar, R. N. Reddy, and Y. N. Reddy, “Synthesis and in vitro and in vivo anticancer activity of novel phenylmethylene bis-isoxazolo[4,5-b]azepines,” *Bioorganic Med. Chem. Lett.*, vol. 22, no. 1, pp. 149–153, 2012, doi: 10.1016/j.bmcl.2011.11.044.
- [36] E. Rajanarendar *et al.*, “Synthesis and in vitro and in vivo anticancer activity of novel 3-methyl-5H-isoxazolo[5',4':5,6]pyrido[2,3-b]indoles,” *Bioorganic Med. Chem. Lett.*, vol. 22, no. 21, pp. 6677–6680, 2012, doi: 10.1016/j.bmcl.2012.08.098.
- [37] E. Rajanarendar *et al.*, “Multi-component synthesis and in vitro and in vivo anticancer activity of novel arylmethylene bis-isoxazolo[4,5-b]pyridine-N-oxides,” *Eur. J. Med. Chem.*, vol. 50, pp. 274–279, 2012, doi: 10.1016/j.ejmech.2012.02.004.
- [38] N. Adhikari, S. K. Baidya, and T. Jha, “Effective anti-aromatase therapy to battle against estrogen-mediated breast cancer: Comparative SAR/QSAR assessment on steroidal aromatase inhibitors,”

- European Journal of Medicinal Chemistry*, vol. 208. Elsevier Masson s.r.l., Dec. 15, 2020, doi: 10.1016/j.ejmech.2020.112845.
- [39] BIOVIA, "BioViaDraw." Dassault Systèmes, San Diego, [Online]. Available: <https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/>.
- [40] J. Li, D. Luo, T. Wen, Q. Liu, and Z. Mo, "Representative feature selection of molecular descriptors in QSAR modeling," *J. Mol. Struct.*, vol. 1244, p. 131249, 2021, doi: 10.1016/j.molstruc.2021.131249.
- [41] W. M. Eldehna, A. Altoukhy, H. Mahrous, and H. A. Abdel-Aziz, "Design, synthesis and QSAR study of certain isatin-pyridine hybrids as potential anti-proliferative agents," *Eur. J. Med. Chem.*, vol. 90, pp. 684–694, 2015, doi: 10.1016/j.ejmech.2014.12.010.
- [42] I. Saidi, M. Manachou, M. Znati, J. Bouajila, and H. Ben Jannet, "Synthesis of new halogenated flavonoid-based isoxazoles: in vitro and in silico evaluation of α -amylase inhibitory potential, a SAR analysis and DFT studies," *J. Mol. Struct.*, vol. 1247, p. 131379, 2022, doi: 10.1016/j.molstruc.2021.131379.
- [43] M. F. Chan *et al.*, "Halogen Substitution at the isoxazole ring Enhances the Activity of N-(ISOXAZOLYL)Sulfonamide Endothelin Antagonists," vol. 6, no. 20, pp. 2393–2398, 1996, doi: 10.1016/0960-894X(96)00441-6.
- [44] S. A. Steiger, C. Li, D. S. Backos, P. Reigan, and N. R. Natale, "Dimeric isoxazolyl-1,4-dihydropyridines have enhanced binding at the multi-drug resistance transporter," *Bioorganic Med. Chem.*, vol. 25, no. 12, pp. 3223–3234, 2017, doi: 10.1016/j.bmc.2017.04.008.
- [45] F. A. E. Ragab Ibrahim *et al.*, "Insights on possible interplay between epithelial-mesenchymal transition and T-type voltage gated calcium channels genes in metastatic breast carcinoma," *Heliyon*, vol. 8, no. 8, p. e10160, 2022, doi: 10.1016/j.heliyon.2022.e10160.
- [46] S. Y. Hwang *et al.*, "Identification of new halogen-containing 2,4-diphenyl indenopyridin-5-one derivative as a boosting agent for the anticancer responses of clinically available topoisomerase inhibitors," *Eur. J. Med. Chem.*, vol. 227, p. 113916, 2022, doi: 10.1016/j.ejmech.2021.113916.
- [47] G. C. Arya, K. Kaur, and V. Jaitak, "Isoxazole derivatives as anticancer agent: A review on synthetic strategies, mechanism of action and SAR studies," *Eur. J. Med. Chem.*, vol. 221, p. 113511, 2021, doi: 10.1016/j.ejmech.2021.113511.
- [48] M. Maldonado-Santiago, Á. Santiago, N. Pastor, L. Alvarez, and R. S. Razo-Hernández, "Isatin derivatives as DNA minor groove-binding agents: a structural and theoretical study," *Struct. Chem.*,

vol. 31, no. 4, pp. 1289–1307, 2020, doi: 10.1007/s11224-020-01497-w.

- [49] E. Rajanarendar, M. Nagi Reddy, S. Rama Krishna, K. Govardhan Reddy, Y. N. Reddy, and M. V. Rajam, “Design, synthesis, in vitro antimicrobial and anticancer activity of novel methylenebis-isoxazolo[4,5-b]azepines derivatives,” *Eur. J. Med. Chem.*, vol. 50, pp. 344–349, 2012, doi: 10.1016/j.ejmech.2012.02.013.
- [50] J. Akhtar, A. A. Khan, Z. Ali, R. Haider, and M. Shahar Yar, “Structure-activity relationship (SAR) study and design strategies of nitrogen-containing heterocyclic moieties for their anticancer activities,” *Eur. J. Med. Chem.*, vol. 125, pp. 143–189, 2017, doi: 10.1016/j.ejmech.2016.09.023.

Material Suplementario

Material Suplementario 1. IC₅₀ experimental extraída directamente de los trabajos originales de los compuestos evaluados. En el último renglón se muestra la IC₅₀ de su respectivo fármaco de referencia Cisplatino. MSO

Molécula, grupo A	IC ₅₀ HeLa (μ M)	IC ₅₀ MCF-7 (μ M)	Molécula, grupo B	IC ₅₀ HeLa (μ g/mL)	IC ₅₀ MCF-7 (μ g/mL)	Molécula, grupo C	IC ₅₀ HeLa (μ g/mL)	IC ₅₀ MCF-7 (μ g/mL)	Molécula, grupo D	IC ₅₀ HeLa (μ g/mL)	IC ₅₀ MCF-7 (μ g/mL)
1	44.61	41.71	13	28.61	31.87	23	68.7	68.9	33	54.48	53.44
2	18.82	21.21	14	62.40	60.70	24	60.2	61.1	34	49.26	49.06
3	16.97	19.12	15	65.52	57.35	25	55.1	55.9	35	61.11	60.40
4	15.88	17.89	16	79.03	72.06	26	53.3	50.1	36	19.72	18.82
5	21.35	24.15	17	70.74	74.26	27	45.6	53.8	37	48.32	46.51
6	29.67	31.05	18	42.04	44.29	28	38.4	41.9	38	50.51	51.34
7	23.58	24.68	19	47.91	45.96	29	65.9	57.3	39	23.62	22.52
8	32.39	35.57	20	32.83	33.55	30	55.9	59.0	40	55.64	54.50
9	25.95	26.48	21	70.22	79.05	31	32.7	35.9			
10	23.65	25.67	22	80.05	76.22	32	28.5	24.1			
11	20.97	22.42									
12	18.52	20.06									
Cisplatino	14.18	15.58		3.84	4.19		3.8	4.2		3.42	5.05

Material Suplementario 2. Actividad antiproliferativa calculada a partir de los **modelos QSAR AB** de información compleja (1.1-1.4) para el conjunto de entrenamiento.

Mol.	HeLa [μ M]				MCF-7 [μ M]				HeLa [μ M]	MCF-7 [μ M]
	1.1	1.2	1.3	1.4	1.1	1.2	1.3	1.4		
2	16.361	16.379	17.032	17.128	21.065	21.066	21.829	21.942	18.82	21.21
4	15.686	15.909	16.886	17.026	20.488	20.248	21.631	21.796	15.88	17.89
5	18.223	21.886	23.238	21.556	27.384	23.205	28.921	27.013	21.35	24.15
6	28.218	24.028	21.135	21.038	29.850	34.560	26.566	26.456	29.67	31.05
8	31.708	32.693	28.520	30.044	39.469	38.397	34.864	36.553	32.39	35.57
10	21.897	18.723	17.491	17.459	23.792	27.452	22.368	22.332	23.65	25.67
11	15.938	17.233	18.306	17.368	22.036	20.547	23.282	22.197	20.97	22.42
12	17.112	16.941	18.450	19.974	21.720	21.941	23.475	25.232	18.52	20.06
13	69.290	69.994	70.460	70.158	76.173	75.451	76.630	76.330	52.93	58.96
14	68.987	69.690	70.158	69.859	75.844	75.124	76.303	76.006	112.52	109.45
16	125.609	126.441	126.404	125.439	137.142	136.113	137.075	136.113	137.44	125.32
17	123.126	124.118	124.108	123.173	134.636	133.459	134.598	133.666	123.02	129.15
18	74.149	75.094	74.529	74.345	81.173	80.211	80.613	80.431	73.93	77.89
19	77.668	75.900	75.360	75.941	82.409	84.110	81.871	82.443	79.77	76.52
20	71.663	70.626	70.958	71.173	76.846	77.842	77.171	77.382	57.34	58.6
21	120.315	121.581	124.727	125.274	132.111	130.669	135.186	135.715	115.22	129.7

Material Suplementario 3. Actividad antiproliferativa calculada a partir de los **modelos QSAR AB** de información compleja (1.1-1.4) para el conjunto de validación.

Molécula	HeLa [μ M]				MCF-7 [μ M]				HeLa [μ M]	MCF-7 [μ M]
	1.1	1.2	1.3	1.4	1.1	1.2	1.3	1.4		
1	15.971	16.223	16.879	16.917	20.862	20.590	21.631	21.676	44.61	41.71
3	18.873	19.082	20.098	20.242	24.542	24.315	25.728	25.898	16.97	19.12
7	22.382	21.393	23.396	25.297	26.853	27.998	29.132	31.276	23.58	24.68
9	19.170	19.442	20.194	20.228	24.976	24.675	25.855	25.896	25.95	26.48
15	69.543	70.248	70.712	70.408	76.448	75.724	76.903	76.601	114.83	100.51
22	80.367	78.572	74.545	74.542	84.678	86.390	80.718	80.712	126.95	120.88

Material Suplementario 4. IC₅₀ teórica frente a HeLa del conjunto de entrenamiento usado en los *modelos QSAR simples*

Molécula	Modelo									HeLa [μM]
	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	
2	19.68	19.67	19.26	20.35	20.07	19.43	19.12	22.19	22.19	18.82
4	18.80	21.44	21.88	20.03	19.76	19.13	18.82	21.86	21.86	15.88
5	22.56	22.13	21.26	20.08	19.80	19.17	18.86	21.90	21.90	21.35
6	22.00	21.54	22.24	23.88	23.46	25.07	25.24	22.23	22.23	29.67
8	27.08	25.65	26.05	25.25	25.14	28.26	30.79	22.07	22.07	32.39
10	22.00	21.54	22.24	23.88	24.38	25.07	25.24	22.23	22.23	23.65
11	22.56	22.13	21.26	20.08	20.58	19.17	18.86	21.90	21.90	20.97
12	22.96	22.69	23.03	23.82	24.32	25.01	25.18	22.18	22.18	18.52
13	57.50	61.52	60.57	61.40	61.18	57.29	59.66	71.64	71.64	52.93
14	67.89	69.58	68.56	67.22	68.35	64.65	64.07	71.34	71.34	112.52
16	126.28	124.13	121.48	122.07	125.92	118.15	118.64	126.86	126.86	137.44
17	123.20	121.66	119.27	119.82	118.89	115.96	116.39	124.61	124.61	123.02
18	80.61	79.07	77.93	73.91	71.03	73.30	69.13	71.34	71.34	73.93
19	99.10	90.49	92.32	89.39	88.42	100.55	109.30	75.86	75.86	79.77
20	65.58	66.26	67.89	74.78	77.63	74.17	69.97	72.15	72.15	57.34
21	122.55	129.25	133.06	131.50	128.15	131.22	125.34	124.39	124.39	115.22

Material Suplementario 5. IC₅₀ teórica frente a HeLa del conjunto de validación usado en los *modelos QSAR simples*

Molécula	Modelo									HeLa [μM]
	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	
1	19.06	19.52	18.74	18.30	18.37	16.94	17.57	21.95	21.95	44.61
3	23.99	25.10	24.89	23.96	23.73	22.92	22.64	25.98	25.98	16.97
7	22.96	22.69	23.03	23.82	23.40	25.01	25.18	22.18	22.18	23.58
9	24.21	25.28	25.05	24.12	24.83	23.08	22.79	26.15	26.15	25.95
15	72.91	72.54	72.89	72.18	74.42	73.89	78.56	71.89	71.89	114.83
22	74.81	73.48	83.85	102.72	97.90	123.00	120.43	72.24	72.24	126.95

Material Suplementario 6. IC₅₀ teórica frente a MCF-7 del conjunto de entrenamiento usado en los *modelos QSAR simples*

Molécula	Modelo									MCF-7 [μM]
	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	
2	21.73	21.68	21.25	22.40	22.14	21.46	21.15	24.31	24.31	21.21
4	20.79	23.53	23.99	22.06	21.79	21.13	20.82	23.95	23.95	17.89
5	24.74	24.26	23.34	22.11	21.84	21.18	20.86	24.00	24.00	24.15
6	24.16	23.65	24.38	26.10	25.68	27.37	27.55	24.35	24.35	31.05
8	29.45	27.93	28.35	27.53	27.43	30.67	33.29	24.18	24.18	35.57
10	24.16	23.65	24.38	26.10	26.64	27.37	27.55	24.35	24.35	25.67
11	24.74	24.26	23.34	22.11	22.65	21.18	20.86	24.00	24.00	22.42
12	25.16	24.86	25.21	26.04	26.57	27.30	27.49	24.30	24.30	20.06
13	59.51	63.48	62.55	63.36	63.14	59.31	61.63	73.45	73.45	58.96
14	69.70	71.39	70.39	69.07	70.16	66.53	65.95	73.14	73.14	109.45
16	128.45	126.67	124.09	124.58	128.16	120.57	120.92	129.54	129.54	125.32
17	125.38	124.19	121.86	122.31	121.25	118.36	118.66	127.26	127.26	129.15
18	82.09	80.66	79.55	75.62	72.78	74.99	70.90	73.14	73.14	77.89

19	100.15	91.95	93.73	90.85	89.86	101.56	109.89	77.74	77.74	76.52
20	67.47	68.16	69.76	76.49	79.24	75.87	71.75	73.97	73.97	58.6
21	124.74	131.57	135.27	133.66	130.22	133.15	127.33	127.03	127.03	129.7

Material Suplementario 7. IC₅₀ teórica frente a MCF-7 del conjunto de validación usado en los *modelos QSAR simples*

Molécula	Modelo									MCF-7 [μM]
	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	
1	21.06	21.52	20.69	20.24	20.34	18.82	19.51	24.05	24.05	41.71
3	26.38	27.52	27.30	26.34	26.11	25.26	24.97	28.43	28.43	19.12
7	25.16	24.86	25.21	26.04	25.62	27.30	27.49	24.30	24.30	24.68
9	26.62	27.72	27.48	26.51	27.27	25.43	25.15	28.61	28.61	26.48
15	74.63	74.31	74.65	73.95	76.11	75.59	80.10	73.70	73.70	100.51
22	76.49	75.24	85.36	103.56	98.85	122.85	120.32	74.06	74.06	120.88

Material Suplementario 8. Modelos QSAR ABCD. IC₅₀ set de entrenamiento. Valores representados como IC₅₀ (μM)

Molécula	HeLa			MCF-7			Actividad biológica Real	
	Predicción	Diferencia	Fracción	Predicción	Diferencia	Fracción	HeLa	MCF-7
1	28.354	11.685	0.292	27.16	16.09	0.37	40.04	43.25
2	18.239	-1.347	-0.080	20.57	1.42	0.06	16.89	21.99
4	18.090	-3.837	-0.269	20.84	-2.29	-0.12	14.25	18.55
5	21.793	-2.630	-0.137	24.23	0.81	0.03	19.16	25.04
6	20.614	6.016	0.226	37.78	-5.58	-0.17	26.63	32.20
8	29.392	-0.321	-0.011	45.07	-8.18	-0.22	29.07	36.89
10	18.517	2.710	0.128	26.47	0.15	0.01	21.23	26.62
11	27.574	-8.753	-0.465	28.16	-4.91	-0.21	18.82	23.25
12	16.999	-0.376	-0.023	25.77	-4.97	-0.24	16.62	20.80
13	105.871	-53.036	-1.004	105.23	-42.93	-0.69	52.84	62.30
14	92.803	19.515	0.174	87.77	27.88	0.24	112.32	115.65
16	110.706	26.488	0.193	87.19	45.23	0.34	137.19	132.42
17	91.302	31.498	0.256	114.42	22.05	0.16	122.80	136.47
18	85.990	-12.193	-0.165	103.83	-21.53	-0.26	73.80	82.30
19	113.187	-33.560	-0.421	106.96	-26.1	-0.32	79.63	80.86
20	66.997	-9.760	-0.171	53.93	7.99	0.13	57.24	61.92
21	96.811	18.203	0.158	119.94	17.11	0.12	115.01	137.05
24	59.656	20.354	0.254	66.17	18.71	0.22	80.01	84.88
25	66.911	10.486	0.135	73.38	8.7	0.11	77.40	82.08
26	78.213	-3.345	-0.045	81.81	-8.25	-0.11	74.87	73.56
29	86.956	5.358	0.058	83.40	0.5	0.01	92.31	83.90
30	83.485	-5.657	-0.073	84.43	1.44	0.02	77.83	85.87
34	176.684	37.521	0.175	186.06	-19.14	-0.11	214.21	166.92
35	311.217	-69.268	-0.286	150.51	24.76	0.14	226.64	175.27
36	164.619	62.021	0.274	87.08	-17.54	-0.25	93.13	69.54
37	136.364	-43.231	-0.464	136.73	24.27	0.15	213.79	161.00
40	152.169	61.616	0.288	195.67	-10.24	-0.06	241.95	185.43

Material Suplementario 9. Modelos QSAR ABCD. Set de validación.

Molécula	HeLa			MCF-7			Actividad biológica Real	
	Predicción	Diferencia	Fracción	Predicción	Diferencia	Fracción	HeLa	MCF-7
3	18.065	-2.834	-0.186	20.13	-0.30	-0.02	15.23	19.83
7	26.241	-5.077	-0.24	41.80	-16.21	-0.63	21.16	18.55
9	23.822	-0.531	-0.023	22.57	4.89	0.18	23.29	25.04
15	100.407	14.218	0.124	76.02	30.19	0.28	114.62	32.20
22	95.467	31.256	0.247	123.40	4.33	0.03	126.72	25.59
23	88.751	11.672	0.116	82.13	23.15	0.22	100.42	36.88
27	87.185	-25.866	-0.422	80.80	-5.18	-0.07	61.32	27.46
28	83.007	-27.993	-0.509	82.35	-19.61	-0.31	55.01	62.75
31	62.219	-12.986	-0.264	52.74	3.76	0.07	49.23	56.50
32	64.355	-23.410	-0.572	71.39	-35.20	-0.97	40.95	36.19
33	175.153	98.312	0.36	67.47	142.42	0.68	273.46	209.88
38	198.292	36.337	0.155	222.82	-36.22	-0.19	234.63	186.59
39	189.657	-78.106	-0.7	101.07	-17.86	-0.21	111.55	83.22

Material Suplementario 10. Identificador de compuestos teóricos y experimentales.

BASE	Posición	Sustituyente									
		F	Cl	Br	I	CH3	OCH3	COOH	COOCH3	OH	NO2
A	R1	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
	R2	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
	R3	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30
	R4	T31	T32	T33	T34	T35	T36	T37	T38	T39	T40
	R5	T41	T42	T43	T44	T45	T46	T47	T48	T49	T50
	R6	T51	T52	T53	T54	T55	T56	T57	T58	T59	T60
	R7	T61	T62	T63	T64	T65	T66	T67	T68	T69	T70
	R8	T71	T72	T73	T74	T75	T76	T77	T78	T79	T80
B	R1	T81	T82	T83	T84	T85	T86	T87	T88	T89	T90
	R2	T91	T92	T93	T94	T95	T96	T97	T98	T99	T100
	R3	T101	T102	T103	T104	T105	T106	T107	T108	T109	T110
	R4	T111	T112	T113	T114	T115	T116	T117	T118	T119	T120
	R5	T121	T122	T123	T124	T125	T126	T127	T128	T129	T130
	R6	T131	T132	T133	T134	T135	T136	T137	T138	T139	T140
C	R1	T141	T142	T143	T144	T145	T146	T147	T148	T149	T150
	R2	T151	T152	T153	T154	T155	T156	T157	T158	T159	T160
	R3	T161	T162	T163	T164	T165	T166	T167	T168	T169	T170
	R4	T171	T172	T173	T174	T175	T176	T177	T178	T179	T180
	R5	T181	T182	T183	T184	T185	T186	T187	T188	T189	T190
	R6	T191	T192	T193	T194	T195	T196	T197	T198	T199	T200
	R7	T201	T202	T203	T204	T205	T206	T207	T208	T209	T210
	R8	T211	T212	T213	T214	T215	T216	T217	T218	T219	T220
	R9	T221	T222	T223	T224	T225	T226	T227	T228	T229	T230
D	R1	T231	T232	T233	T234	T235	T236	T237	T238	T239	T240
	R2	T241	T242	T243	T244	T245	T246	T247	T248	T249	T250
	R3	T251	T252	T253	T254	T255	T256	T257	T258	T259	T260
	R4	T261	T262	T263	T264	T265	T266	T267	T268	T269	T270
	R5	T271	T272	T273	T274	T275	T276	T277	T278	T279	T280

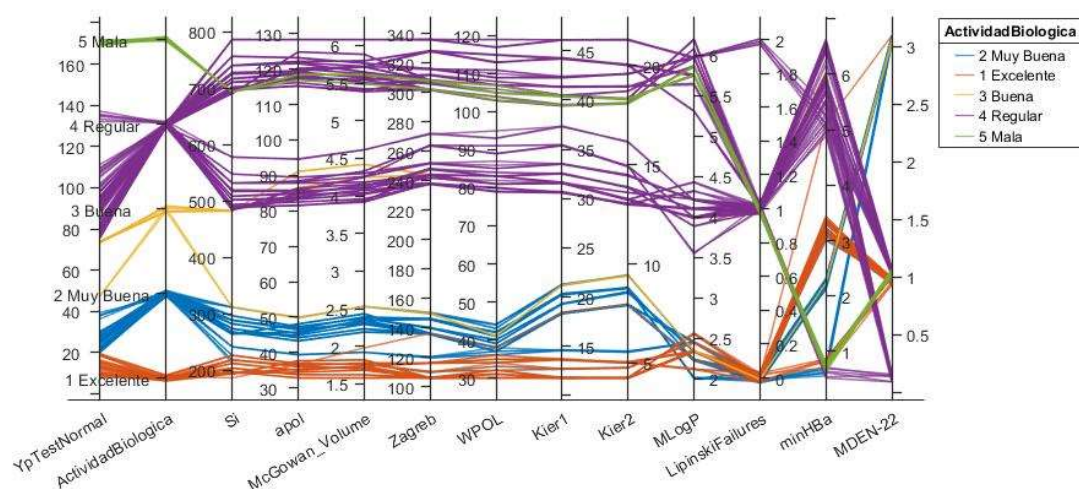
Material Suplementario 11. Actividad antiproliferativa (IC_{50} [μM]) predicha por los mejores modelos QSAR AB y ABCD

ID	HeLa AB	MCF-7 AB	HeLa ABCD	MCF-7 ABCD	ID	HeLa AB	MCF-7 AB	HeLa ABCD	MCF-7 ABCD
T1	16.33	21.16	63.43	74.16	T141	83.82	92.15	89.81	87.87
T11	16.36	21.06	18.24	20.57	T151	83.92	92.28	117.38	116.66
T21	16.21	21.03	18.82	28.10	T161	84.04	92.47	101.91	58.28
T31	5.71	7.74	49.05	25.28	T171	84.59	92.54	88.89	90.83
T41	16.30	21.01	59.91	74.16	T181	84.76	92.74	130.63	157.67
T51	16.09	21.00	18.74	24.44	T191	84.96	93.02	75.86	40.60
T61	16.16	21.00	18.71	24.44	T201	82.61	92.16	64.43	61.07
T71	16.36	21.01	60.19	74.16	T211	82.73	92.30	94.47	107.92
T2	19.17	24.97	54.68	51.35	T221	82.88	92.48	146.53	126.51
T12	18.87	24.54	18.07	20.13	T142	151.83	169.53	88.70	85.10
T22	18.89	24.61	18.67	25.98	T152	152.76	170.56	109.49	105.79
T32	19.17	24.98	23.82	22.57	T162	153.96	171.91	102.53	67.96
T42	18.98	24.66	49.40	51.35	T172	151.03	168.71	85.19	85.23
T52	18.93	24.64	18.47	22.57	T182	151.81	169.58	117.98	131.02
T62	18.87	24.61	18.45	22.57	T192	152.79	170.69	81.42	53.00
T72	18.90	24.62	49.99	51.35	T202	152.24	169.55	65.49	67.13
T3	15.75	20.60	51.39	41.95	T212	153.01	170.40	89.68	96.02
T13	15.69	20.49	18.09	20.84	T222	154.04	171.54	141.40	114.70
T23	15.82	20.68	18.66	25.35	T143	79.34	89.90	88.26	84.14
T33	15.90	20.77	24.33	22.66	T153	79.25	89.80	106.01	99.39
T43	15.71	20.50	45.51	41.95	T163	79.08	89.66	102.74	72.45
T53	15.92	20.77	18.45	22.66	T173	78.80	89.43	83.67	83.53
T63	15.85	20.75	18.43	22.66	T183	78.66	89.29	111.87	116.99
T73	15.64	20.48	46.20	41.95	T193	78.43	89.08	84.41	60.35
T4	15.68	20.57	44.86	30.31	T203	80.38	90.46	66.38	70.72
T14	15.58	20.42	18.70	24.47	T213	80.29	90.36	87.59	91.26
T24	15.77	20.65	19.15	25.99	T223	80.14	90.22	137.90	106.62
T34	15.90	20.77	25.62	25.12	T144	78.04	89.45	87.98	83.10
T44	15.60	20.42	38.62	30.31	T154	77.91	89.35	99.58	87.58
T54	15.91	20.75	18.96	25.12	T164	77.64	89.18	102.69	79.70
T64	15.81	20.73	18.94	25.12	T174	77.30	88.90	81.83	82.71
T74	15.49	20.40	39.41	30.31	T184	77.13	88.75	100.28	92.20
T5	17.15	25.02	36.08	42.24	T194	76.80	88.52	91.02	75.34
T15	18.22	27.38	21.79	24.23	T204	79.76	90.22	69.81	78.78
T25	17.17	25.58	17.40	18.11	T214	79.64	90.12	84.49	84.96
T35	15.94	22.04	27.57	28.16	T224	79.41	89.96	128.21	89.78
T45	17.69	26.26	29.79	38.04	T145	84.71	94.07	89.45	82.85
T55	16.28	25.86	19.48	25.01	T155	84.60	93.95	93.72	85.02
T65	17.22	26.02	20.28	27.11	T165	84.34	93.69	87.96	71.06
T75	18.13	26.35	30.11	38.04	T175	85.49	94.84	83.58	82.32
T6	19.35	29.28	84.38	57.82	T185	85.33	94.67	89.61	88.13
T16	20.63	30.48	20.45	18.50	T195	85.01	94.33	77.70	60.97
T26	19.58	29.58	20.26	32.74	T205	82.14	92.06	73.57	74.59
T36	17.34	28.10	26.87	18.09	T215	82.06	91.98	93.11	90.92

T46	19.79	29.71	68.81	60.65	T225	81.93	91.83	97.67	82.15
T56	19.58	30.03	24.99	26.10	T146	91.81	100.73	92.59	82.92
T66	20.05	30.06	23.30	20.84	T156	89.59	98.35	121.37	108.80
T76	20.20	29.74	73.71	66.22	T166	86.24	94.80	92.75	62.73
T7	20.42	24.71	31.16	24.65	T176	93.02	101.52	92.41	80.51
T17	22.38	26.85	26.24	41.80	T186	90.96	99.33	131.64	135.79
T27	20.75	25.04	118.66	87.37	T196	87.84	96.05	72.15	45.32
T37	17.11	21.72	17.00	25.77	T206	91.64	102.13	63.52	60.36
T47	20.91	25.80	25.80	29.57	T216	90.05	100.39	106.09	92.73
T57	21.08	25.06	46.58	38.22	T226	87.66	97.78	157.00	125.51
T67	21.67	25.24	32.78	29.33	T147	86.61	94.40	86.93	84.44
T77	21.47	25.90	28.26	30.57	T157	86.56	94.42	92.63	98.23
T10	25.46	27.15	36.63	31.44	T167	86.33	94.37	139.21	129.43
T20	28.22	29.85	20.61	37.78	T177	88.28	95.47	83.19	82.94
T30	26.04	27.67	48.86	75.23	T187	88.25	95.52	92.12	113.01
T40	21.90	23.79	18.52	26.47	T197	88.01	95.49	147.74	193.46
T50	25.79	28.20	35.17	37.24	T207	83.12	92.75	95.66	92.73
T60	28.08	28.38	24.13	33.24	T217	83.15	92.81	61.27	64.02
T70	28.21	28.49	20.88	27.91	T227	83.15	92.85	131.59	141.94
T80	26.17	28.30	35.29	37.24	T150	88.79	96.15	86.49	86.42
T8	30.05	37.94	28.46	25.23	T160	88.89	96.14	105.05	109.24
T18	31.71	39.47	29.39	45.07	T170	89.02	95.97	132.31	115.91
T28	30.37	38.23	99.91	75.48	T180	94.02	98.17	83.69	84.77
T38	28.06	36.44	24.70	38.56	T190	93.78	98.19	114.48	139.14
T48	30.21	38.08	31.55	38.69	T200	92.87	97.99	122.72	159.89
T58	41.43	47.59	27.23	23.15	T210	83.94	93.34	80.22	82.52
T68	40.70	47.48	33.62	32.04	T220	83.97	93.42	65.61	70.80
T78	30.19	38.05	30.13	31.68	T230	83.92	93.47	144.63	158.98
T9	16.31	21.14	84.52	65.44	T148	101.11	109.67	87.60	83.59
T19	16.42	21.20	22.39	20.08	T158	99.27	107.80	91.95	95.43
T29	16.21	21.05	12.75	23.15	T168	96.50	105.09	145.25	129.92
T39	16.08	21.01	34.41	23.93	T178	103.55	111.32	84.46	80.92
T49	16.35	21.16	74.43	63.16	T188	101.80	109.58	91.79	105.52
T59	16.06	20.98	19.54	22.91	T198	99.10	107.02	158.27	192.02
T69	16.17	20.99	21.55	23.64	T208	97.92	108.98	99.40	90.57
T79	16.45	21.17	74.91	63.16	T218	96.68	107.65	59.42	60.60
T81	69.67	76.59	94.29	123.57	T228	94.90	105.76	124.64	131.15
T91	69.77	76.70	130.00	149.14	T149	83.10	91.73	91.94	85.88
T101	69.92	76.86	109.95	78.03	T159	83.15	91.80	114.46	108.84
T111	72.77	77.14	108.93	131.22	T169	83.21	91.89	84.25	62.05
T121	72.78	77.29	131.93	155.34	T179	83.59	91.92	90.01	86.90
T131	72.59	77.47	75.31	49.45	T189	83.67	92.02	117.40	137.57
T82	123.13	134.64	91.30	114.42	T199	83.77	92.16	64.22	44.52
T92	124.15	135.70	117.91	132.57	T209	82.20	91.78	63.89	63.93
T102	125.61	137.14	110.71	87.19	T219	82.25	91.85	110.21	102.30
T112	120.31	132.11	96.81	119.94	T229	82.33	91.94	122.90	115.10
T122	121.39	133.29	114.60	140.69	T231	5.64	10.06	340.50	239.73
T132	123.11	135.16	81.19	63.63	T241	6.11	9.79	62.28	48.77

T83	68.47	75.88	90.04	110.90	T251	5.79	9.86	198.29	222.82
T93	68.44	75.80	112.59	124.11	T261	5.48	9.97	114.37	93.03
T103	68.46	75.69	110.96	91.85	T271	5.54	9.98	189.05	104.33
T113	65.96	73.88	91.47	114.67	T233	5.18	9.45	291.42	156.23
T123	66.01	73.91	106.56	130.64	T243	4.57	8.65	70.74	51.98
T133	66.28	74.11	84.47	72.61	T253	4.88	9.36	164.62	150.51
T84	67.70	75.88	88.72	106.50	T263	5.29	9.58	144.83	86.95
T94	67.70	75.80	103.27	110.34	T273	5.21	9.65	245.27	72.34
T104	67.81	75.68	111.07	100.80	T234	5.15	9.41	245.97	87.54
T114	63.97	73.23	83.19	107.42	T244	4.29	8.43	79.72	61.42
T124	64.11	73.30	92.80	112.06	T254	4.70	9.28	142.96	86.51
T134	64.63	73.64	92.35	93.48	T264	5.29	9.56	163.38	72.62
T85	69.11	75.98	93.76	106.09	T274	5.17	9.64	268.93	67.09
T95	69.06	75.92	98.23	103.26	T232	6.93	12.57	311.22	195.67
T105	68.99	75.84	92.80	87.77	T242	6.32	11.61	67.98	49.62
T115	74.15	81.17	85.99	103.83	T252	6.60	12.35	176.68	186.06
T125	73.78	80.62	83.24	99.48	T262	7.02	12.68	135.36	92.51
T135	72.86	79.26	81.25	78.19	T272	7.04	12.92	228.45	79.87
T86	72.96	80.16	100.39	108.87	T235	10.72	20.20	189.66	101.07
T96	71.62	78.70	138.21	134.83	T245	11.46	18.44	79.19	48.31
T106	69.54	76.45	100.41	76.02	T255	9.23	14.91	136.36	87.08
T116	77.67	82.41	113.19	106.96	T265	8.24	15.71	110.81	39.71
T126	76.26	81.03	126.67	133.85	T275	8.05	13.55	158.18	114.93
T136	73.82	78.77	72.07	47.78	T236	15.62	22.81	206.78	153.96
T87	69.53	76.44	86.70	109.41	T246	14.03	20.84	74.32	67.28
T97	69.58	76.49	98.47	141.71	T256	11.99	18.49	152.17	136.73
T107	69.65	76.56	157.59	160.66	T266	12.02	19.57	104.95	96.24
T117	76.67	81.51	84.60	114.38	T276	12.48	19.10	216.67	148.13
T127	76.37	81.18	109.45	162.12	T237	15.88	20.08	45.19	58.07
T137	75.38	80.16	139.86	167.18	T247	13.88	18.57	283.45	148.16
T90	69.64	76.56	86.12	116.27	T257	11.18	14.99	55.46	68.18
T100	69.72	76.64	112.31	157.51	T267	10.93	15.80	213.83	258.62
T110	69.82	76.75	148.68	148.62	T277	11.22	13.65	207.20	46.39
T120	80.37	84.68	95.47	123.40	T239	5.83	10.28	327.13	218.72
T130	79.98	84.33	137.03	182.54	T249	6.92	10.91	39.45	46.50
T140	78.59	82.60	118.61	140.50	T259	6.28	10.26	177.60	193.17
T88	77.35	84.92	87.65	105.71	T269	5.46	10.06	74.24	76.49
T98	76.38	83.86	97.74	133.01	T279	5.63	9.95	148.71	111.66
T108	74.97	82.34	168.16	161.10	T240	16.66	22.28	128.10	79.00
T118	86.40	91.67	87.37	105.58	T250	14.57	19.34	161.69	114.55
T128	85.04	90.19	112.44	141.90	T260	11.79	15.39	113.31	88.42
T138	82.44	87.44	153.28	158.38	T270	11.76	16.46	193.75	265.07
T89	69.48	76.38	98.70	117.28	T280	12.36	14.12	169.17	38.76
T99	69.54	76.44	125.08	135.22	T238	16.48	23.21	69.29	85.42
T109	69.61	76.52	88.37	79.67	T248	15.58	22.29	205.59	118.85
T119	71.93	76.69	104.22	123.15	T258	14.31	21.20	60.14	77.22
T129	71.89	76.76	106.33	140.09	T268	14.61	21.82	180.93	189.53
T139	71.66	76.85	67.00	53.93	T278	15.22	21.82	152.91	49.85

Material Suplementario 12. Descriptores relacionados con los **modelos QSAR AB complejos** y su actividad biológica cualitativa. Se observa que modelos QSAR AB catalogan como activos a los compuestos que presenten bajos valores en la mayoría de los descriptores mostrados.



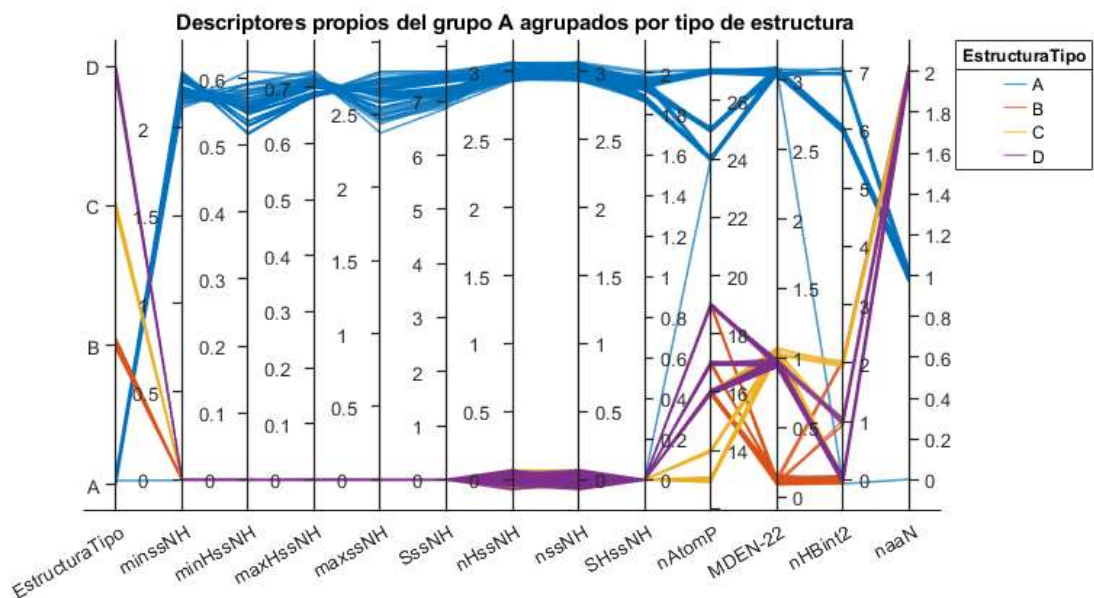
Material Suplementario 13. Valores de los descriptores usados en los modelos QSAR AB 1.1 y ABCD para HeLa y MCF-7. Redondeado a 5 decimales. MS2

Grupo	Mol	MATS6c	MATS7e	ATSC6e	AATSC3v	SpMin5_Bhp	SpMin5_Bhv	SpMin5_Bhe	SpMin5_Bhi	minHBa
A	1	0.28248	-0.42340	0.51549	-2.12528	1.33586	1.27754	1.00812	0.97574	2.26582
	2	0.37788	-0.43145	1.54417	-2.90137	1.33923	1.27706	1.00731	0.97627	2.16283
	3	0.36367	-0.47852	1.18256	-3.93827	1.33356	1.27397	1.00891	0.97886	0.66487
	4	0.35530	-0.48343	0.99676	-4.44160	1.33224	1.27304	1.00987	0.97986	2.30437
	5	0.26954	-0.49561	0.28252	-6.96372	1.37104	1.36616	1.13182	1.07104	2.28528
	6	0.34569	-0.17950	1.52563	-3.04415	1.48174	1.38910	1.08953	1.05568	2.14466
	7	0.27135	-0.14365	1.46500	-4.00756	1.42158	1.35569	1.12953	1.12849	2.16709
	8	0.23798	-0.04307	2.12309	-4.36289	1.51429	1.48177	1.20754	1.19742	2.21232
	9	0.33358	-0.47852	0.72632	-1.58920	1.33586	1.27748	1.00818	0.97600	0.60290
	10	0.39300	-0.30787	1.71843	-1.72224	1.41519	1.31578	1.01679	0.98298	2.14466
	11	0.24284	-0.46486	-0.02252	-5.03562	1.33588	1.29592	1.04011	0.98682	2.28528
	12	0.40820	-0.29097	1.98527	-2.07946	1.35112	1.28711	1.03824	1.03638	2.16709
B	13	-0.00105	-0.08219	-1.62044	0.27469	1.82710	1.82522	1.70995	1.68411	5.94935
	14	0.02050	-0.10711	-1.13634	-0.39127	1.82710	1.82522	1.70995	1.68411	5.98922
	15	0.01975	-0.01536	0.30880	0.76700	1.82710	1.82522	1.70995	1.68411	5.91611
	16	-0.00593	-0.03952	-0.46655	0.65223	1.82692	1.82522	1.70995	1.68411	0.52345
	17	0.01438	-0.09011	-2.02871	-1.02448	1.82658	1.82522	1.70995	1.68411	0.69362
	18	-0.01324	-0.06673	-1.42001	-3.67191	1.84603	1.84715	1.73316	1.70836	5.98880
	19	-0.02782	-0.05794	-1.45409	-0.37965	1.84141	1.83162	1.71330	1.69231	5.40597
	20	0.10614	-0.08173	1.04178	-0.01278	1.83400	1.82570	1.70987	1.68690	5.88193
	21	-0.02222	-0.07356	-2.05828	-2.42094	1.82101	1.81969	1.71256	1.69141	0.71064
	22	0.00050	-0.07034	-2.14065	-1.21096	1.86374	1.85666	1.72834	1.70443	5.87019
C	23	0.00552	-0.03867	-0.22034	-2.01302	1.88637	1.90463	1.82240	1.79849	6.58891
	24	0.08297	-0.05342	0.49827	-3.24082	1.88525	1.90338	1.82303	1.80001	0.81450
	25	0.04039	-0.06458	-0.01947	-4.15203	1.88942	1.90794	1.82599	1.80219	6.61921
	26	0.00226	-0.03706	-0.18983	-4.15203	1.90005	1.91779	1.83732	1.81417	6.62836
	27	0.00747	-0.05918	-0.35117	-2.16352	1.89591	1.90784	1.82416	1.80360	5.68157
	28	0.00410	-0.03599	-0.20586	-3.12619	1.89738	1.91479	1.83369	1.81033	6.61263
	29	0.00904	-0.04093	-0.30307	-2.10549	1.89416	1.90723	1.82381	1.80251	5.74706
	30	0.01091	-0.04410	-0.38227	-2.61147	1.88320	1.90170	1.82402	1.80266	0.78607
	31	0.09322	-0.09822	0.97304	-1.95430	1.77395	1.76105	1.64041	1.61465	6.02788
	32	0.06675	-0.05975	0.13586	-3.09638	1.74636	1.75031	1.64613	1.62370	-1.28366
D	33	-0.18040	-0.16054	-0.59896	-3.25636	1.07873	1.06905	0.91112	0.88457	3.30657
	34	-0.23709	0.25438	-0.68155	-6.67489	1.06366	1.05895	0.91253	0.89891	0.86432
	35	-0.23478	0.17008	-0.64122	-7.61238	1.05616	1.05664	0.91461	0.90268	3.35977
	36	-0.20434	-0.07314	-0.78144	-8.58778	1.22284	1.20629	0.99530	0.98150	3.33380

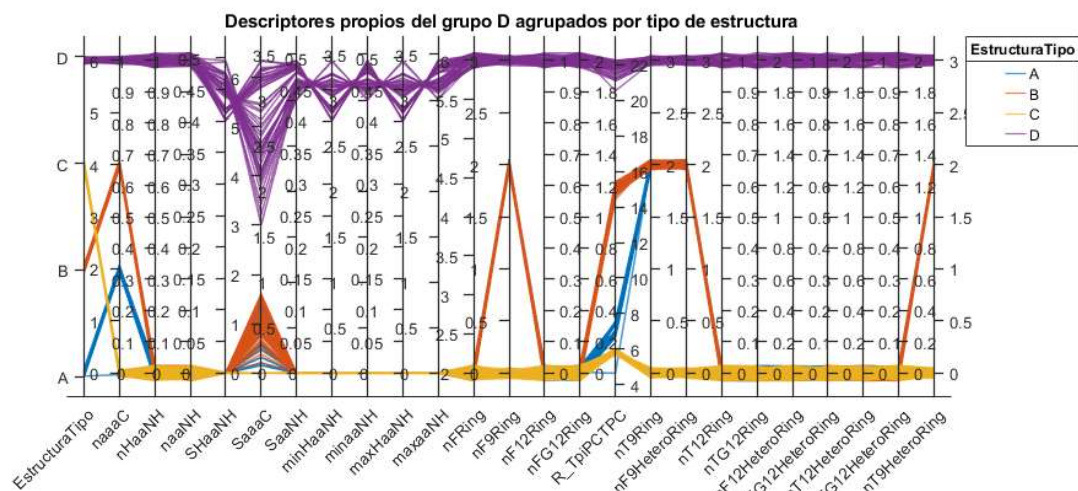
37	-0.16174	0.14638	-0.48529	-4.24000	1.29050	1.27491	1.01166	0.98490	3.30233
38	-0.23422	0.31658	-0.80610	-4.72934	1.09566	1.06670	0.90877	0.88702	3.16713
39	-0.24558	-0.00696	-0.74372	-6.12234	1.26316	1.30599	1.04127	0.96592	3.37486
40	-0.32631	0.25438	-0.88175	-3.59976	1.07537	1.06349	0.91173	0.88751	0.82770

Material Suplementario 14. Descriptores de características presentes principalmente en el a) grupo A y b) grupo D. Los descriptores son asociados a características propias de compuestos con buena o mala actividad biológica.

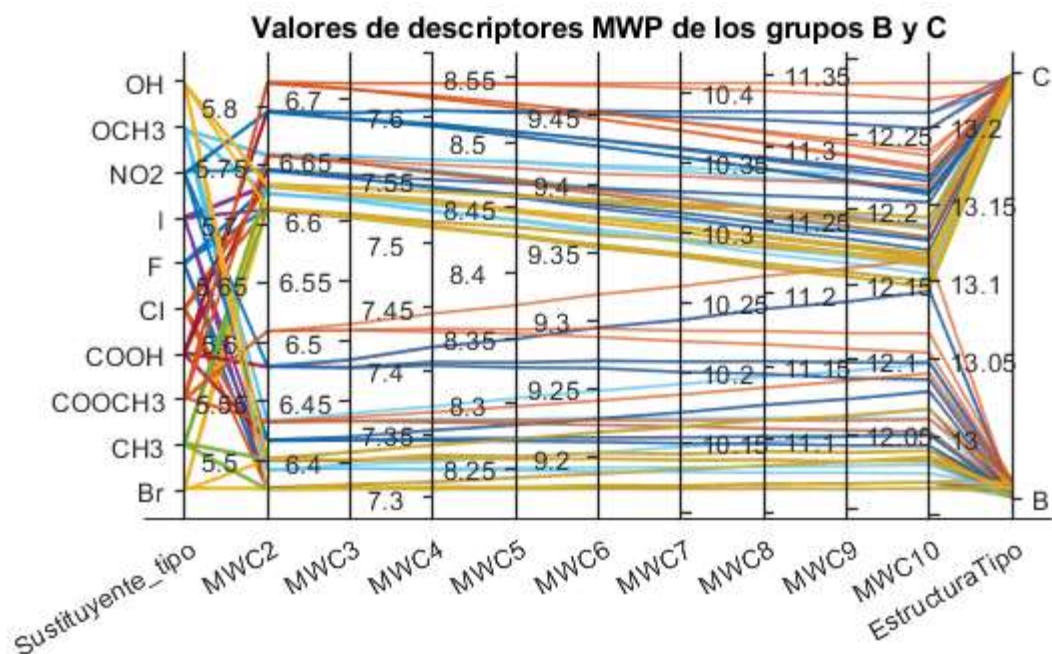
a)



b)

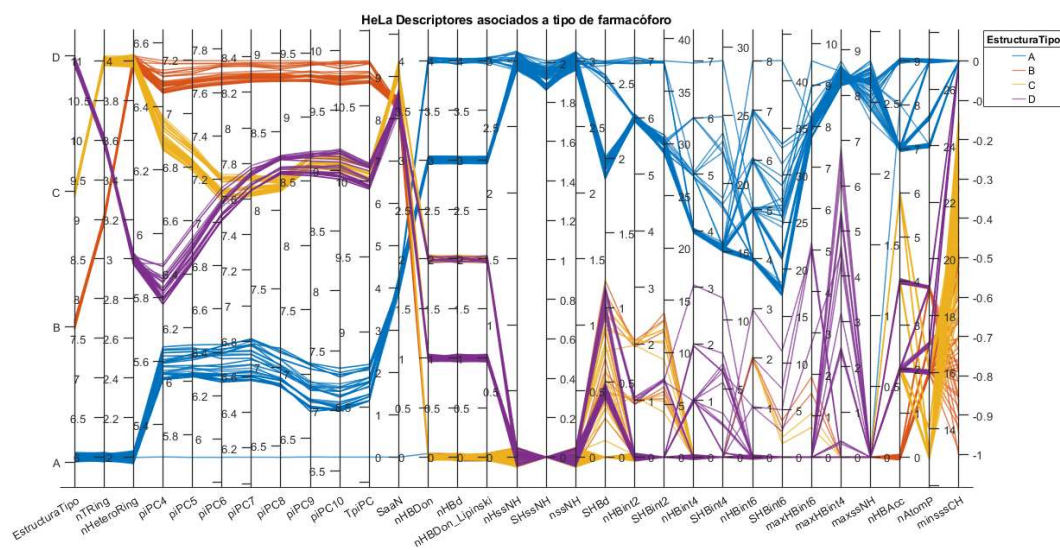


Material Suplementario 15. Cambio en los valores de descriptores relacionados con el número de caminos moleculares posibles desde 2 hasta 10 átomos de distancia en función del sustituyente (izq.) y la posición (der.) empleada. A mayor número de caminos posibles, más elementos son considerados para el cálculo de cada descriptor y, por tanto, mayor será valor del descriptor. Los valores de los descriptores aquí mostrados corresponden a los compuestos del grupo B y C.

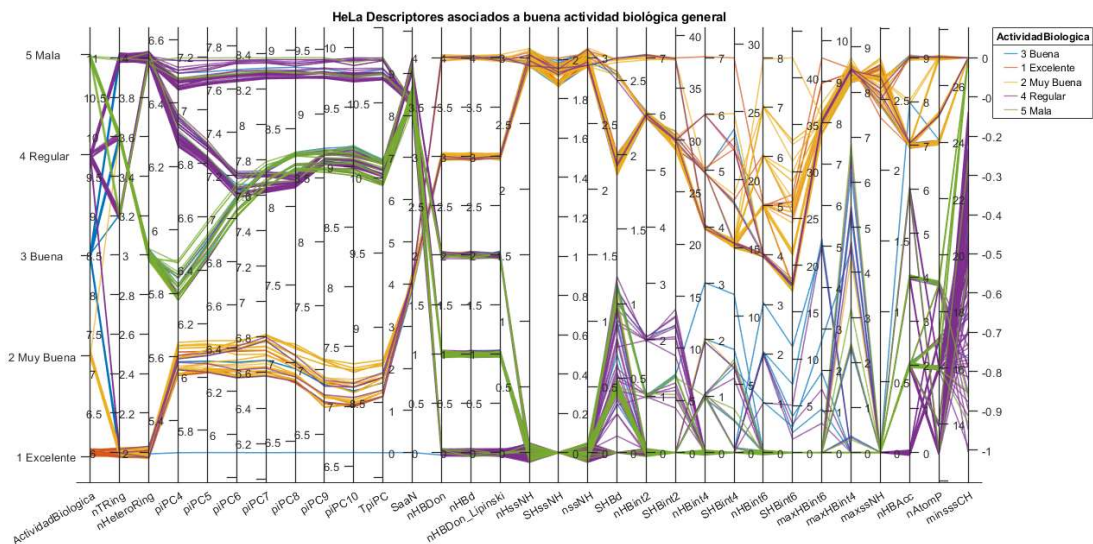


Material Suplementario 16. Descriptores característicos de cada grupo usado en la generación de los **modelos QSAR ABCD complejos**. a) Cada color representa un grupo de compuestos; b) agrupados por tipo de actividad biológica cualitativa para HeLa y c) MCF-7

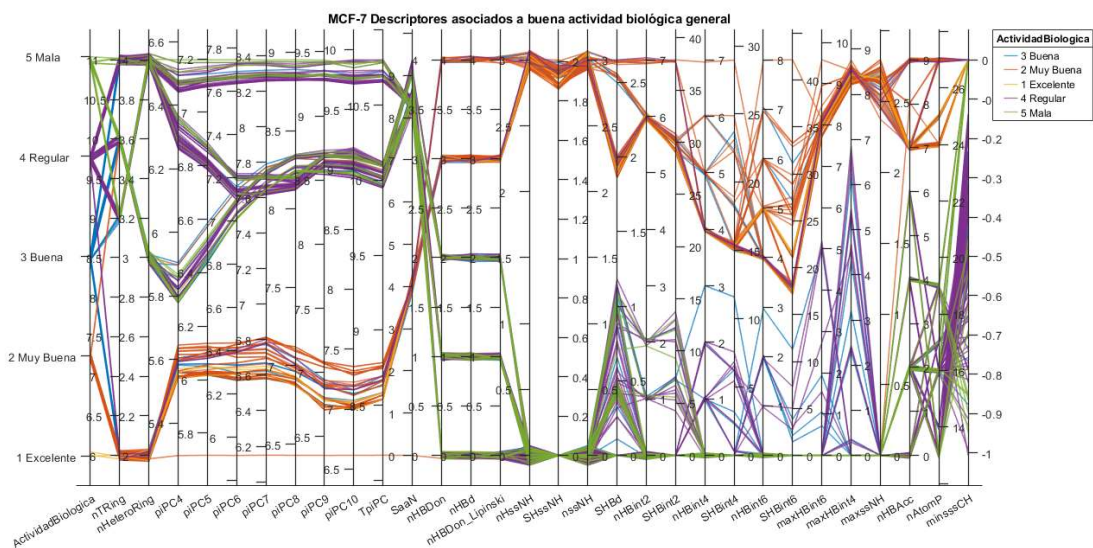
a)



b)



c)



Material Suplementario 17. Valor del descriptor minHBa redondeado a 2 dígitos, para todas las moléculas basadas en los farmacóforos de los grupos A-D. Los valores correspondientes a moléculas sustituidas con Cl son las únicas con valores debajo de 1.

Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor	Mol.	Valor
T1	2.13	T33	2.29	T65	2.20	T97	6.02	T129	4.95	T161	6.59	T193	5.70	T225	6.49	T257	3.30
T2	2.16	T34	2.29	T66	2.21	T98	6.04	T130	5.06	T162	6.59	T194	5.94	T226	6.48	T258	3.30
T3	2.19	T35	2.28	T67	2.22	T99	5.99	T131	5.23	T163	6.61	T195	6.50	T227	6.47	T259	3.30
T4	5.11	T36	2.29	T68	2.21	T100	5.99	T132	4.89	T164	6.63	T196	6.49	T228	6.50	T260	3.30
T5	2.19	T37	2.28	T69	2.22	T101	6.01	T133	4.98	T165	6.58	T197	6.49	T229	6.49	T261	3.08
T6	2.20	T38	2.28	T70	2.23	T102	6.01	T134	5.12	T166	6.59	T198	6.48	T230	6.48	T262	3.14
T7	2.20	T39	2.28	T71	2.23	T103	6.02	T135	5.92	T167	6.60	T199	6.47	T231	3.00	T263	3.18
T8	2.19	T40	2.28	T72	2.22	T104	6.04	T136	5.92	T168	6.57	T200	6.45	T232	3.11	T264	3.14
T9	0.56	T41	2.26	T73	2.18	T105	5.97	T137	5.91	T169	6.58	T201	6.49	T233	3.17	T265	3.14
T10	0.66	T42	2.26	T74	2.20	T106	5.98	T138	5.91	T170	6.59	T202	6.48	T234	3.11	T266	3.12
T11	0.70	T43	2.26	T75	2.22	T107	5.99	T139	5.90	T171	6.59	T203	6.47	T235	3.11	T267	3.19
T12	0.60	T44	2.26	T76	2.20	T108	5.99	T140	5.88	T172	6.61	T204	6.49	T236	3.43	T268	3.22
T13	0.62	T45	2.26	T77	2.22	T109	6.00	T141	6.48	T173	6.63	T205	6.48	T237	3.38	T269	3.19
T14	0.71	T46	2.26	T78	2.23	T110	6.01	T142	6.47	T174	6.58	T206	6.46	T238	3.36	T270	3.19
T15	0.72	T47	2.26	T79	2.23	T111	5.48	T143	6.45	T175	6.59	T207	6.45	T239	3.38	T271	3.01
T16	0.63	T48	2.26	T80	2.22	T112	5.65	T144	6.45	T176	6.60	T208	6.44	T240	3.38	T272	3.10
T17	2.32	T49	2.14	T81	5.90	T113	5.92	T145	6.43	T177	6.56	T209	6.42	T241	3.44	T273	3.15
T18	2.30	T50	2.17	T82	5.89	T114	5.41	T146	6.40	T178	6.56	T210	6.47	T242	3.39	T274	3.10
T19	2.30	T51	2.19	T83	5.87	T115	5.55	T147	6.47	T179	6.57	T211	6.46	T243	3.36	T275	3.10
T20	2.30	T52	2.17	T84	5.87	T116	5.77	T148	6.45	T180	6.57	T212	6.44	T244	3.39	T276	3.19
T21	2.30	T53	2.19	T85	5.84	T117	5.92	T149	6.44	T181	6.58	T213	5.14	T245	3.39	T277	3.22
T22	2.29	T54	2.20	T86	5.82	T118	5.91	T150	0.75	T182	6.60	T214	5.30	T246	0.83	T278	3.24
T23	2.29	T55	2.20	T87	0.69	T119	5.90	T151	0.70	T183	6.56	T215	5.52	T247	0.87	T279	3.22
T24	2.30	T56	2.19	T88	0.62	T120	5.89	T152	0.62	T184	6.57	T216	5.10	T248	0.86	T280	3.22
T25	2.32	T57	2.11	T89	0.52	T121	5.88	T153	0.78	T185	6.58	T217	5.24	T249	0.82		
T26	2.31	T58	2.14	T90	0.71	T122	5.87	T154	0.73	T186	5.71	T218	5.43	T250	0.70		
T27	2.30	T59	2.17	T91	0.64	T123	5.90	T155	0.67	T187	5.93	T219	5.01	T251	3.37		
T28	2.31	T60	2.14	T92	0.55	T124	5.89	T156	0.80	T188	6.27	T220	5.12	T252	3.35		
T29	2.30	T61	2.17	T93	5.98	T125	5.88	T157	0.75	T189	5.66	T221	5.28	T253	3.33		
T30	2.29	T62	2.19	T94	5.99	T126	5.87	T158	0.69	T190	5.86	T222	6.51	T254	3.35		
T31	2.29	T63	2.19	T95	6.01	T127	5.85	T159	6.57	T191	6.17	T223	6.50	T255	3.35		
T32	2.30	T64	2.17	T96	6.01	T128	5.83	T160	6.58	T192	5.54	T224	6.49	T256	3.30		

Mol = molécula.