

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Orgánica

Desarrollo y aplicación de metodologías *in silico* para la predicción, diseño y síntesis de nuevos compuestos bioactivos inspirados en productos naturales

Tesis presentada para optar por el título de doctor de la Universidad de Buenos Aires en el área de Química Orgánica

Lic. Julián Francisco Fernández

Director de Tesis: Dr. Jorge A. Palermo
Codirector: Dr. Martin Lavecchia
Consejera de estudios: Dra. Gabriela M. Cabrera

Lugar de Trabajo: UMYMFOR - Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Ciudad Autónoma de Buenos Aires, 21 de Julio de 2025

Resumen

La búsqueda de nuevos fármacos a partir de productos naturales ha dado como resultado numerosos medicamentos a lo largo de la historia, como la penicilina y otros antibióticos, los agentes anticancerígenos de origen vegetal, y varios analgésicos. Actualmente, la naturaleza sigue siendo una fuente esencial de compuestos con potencial terapéutico, y la exploración de nuevas bioactividades es un área de gran relevancia. Sin embargo, uno de los principales desafíos en este campo es la escasa cantidad de compuesto puro que suele obtenerse de los productos naturales, lo que limita las opciones de ensayos biológicos que pueden realizarse. En este trabajo se presenta el desarrollo de Target Fisher, una innovadora metodología *in silico* que permite predecir blancos moleculares para cualquier tipo de molécula pequeña, incluidos los compuestos naturales, mediante la combinación de docking molecular e inteligencia artificial. Por último, se presentan casos de aplicación experimental de esta herramienta que involucraron tanto el aislamiento de productos naturales para confirmar sus predicciones, llevando al descubrimiento de nuevas bioactividades, como también su uso para el diseño y posterior síntesis de nuevos inhibidores que puedan actuar sobre blancos terapéuticos relevantes en enfermedades neurodegenerativas.

Development and application of *in silico* methodologies for the prediction, design, and synthesis of new bioactive compounds inspired by natural products.

The search for new drugs from natural products has historically led to the discovery of numerous medications, such as penicillin and other antibiotics, plant-derived anticancer agents, and various analgesics. Today, nature remains a vital source of compounds with therapeutic potential, and the exploration of new bioactivities is an area of great relevance. However, one of the main challenges in this field is the limited quantity of pure compound typically obtained from natural products, which restricts the range of biological assays that can be performed. This work presents the development of Target Fisher, an innovative in silico methodology that enables the prediction of molecular targets for any type of small molecule, including natural products, by combining molecular docking and artificial intelligence. Finally, experimental applications of this tool are presented, involving both the isolation of natural products to validate its predictions—leading to the discovery of new bioactivities—and its use in the design and subsequent synthesis of new inhibitors that can act on therapeutically relevant targets in neurodegenerative diseases.

Agradecimientos

En primer lugar me gustaría agradecer a mis dos directores, el Dr. Jorge Palermo y el Dr. Martin Lavecchia, quienes me dieron la oportunidad de llevar adelante mi tesis de doctorado en sus laboratorios. A ambos les agradezco por el esfuerzo y la paciencia a lo largo de estos años. Gracias por la comunicación transparente y por generar un excelente clima de trabajo, fomentando la creatividad y el trabajo en equipo. Gracias, además, por todas las charlas y la excelente relación por fuera del ámbito laboral.

A mis compañeros del L12 a lo largo de todos estos años, la Dra. Marianela Sanchez, la Dra. Lucía Fernández, al Dr. Cristian Pis Diez, a quien además le agradezco por ser mi primer tutor, al Dr. Ivan Prieto, la Dra. Romina Ronchi, la Dra. María Belén Valdez y el Lic. Juan Robirosa y a la nueva camada, cuya incorporación no hizo más que agregarle risas y energía al laboratorio: Lic. Rodrigo Suárez, Nicolás Ramírez, Iara Laborde, a quién además le agradezco por su activa participación en el último capitulo de esta tesis, a Daniel Esses, a Patricia y a Manuel. Todos ustedes, de una forma directa o indirecta, han contribuido a mi formación. Con muchos de ustedes me llevo además una invaluable amistad.

A mis compañeros de La Plata, en especial al Lic. Leandro Martínez-Heredia y a la Lic. Patricia Quispe, por armar un hermoso grupo de trabajo a pesar de la distancia, siempre apoyandonos y ayudandonos a llevar a cabo muchas de las locuras que se nos fueron ocurriendo.

Además de Jorge, en el Departamento de Química Orgánica encontré otras personas que fueron pilares fundamentales a lo largo de este camino. En primer lugar quiero volver a agradecer a la Dra. Lucía Fernández por toda su ayuda, por las charlas, por los mates, por los consejos, por todo. Por otro lado, a la Dra. Gabriela Cabrera y al Dr. Fernando Durán, por siempre abrirme las puertas a sus laboratorios y estar disponibles para lo que necesitase desde el primer día que pise el laboratorio de Jorge. Muchas gracias por toda la confianza en mí, los consejos y la ayuda en los momentos que la necesité.

No quiero dejar de agradecer a los integrantes de ambos laboratorios, la Dra. Adriana Cirigliano, el Dr. Gastón Siless, la Dra. Fernanda D'Jonsiles y los Lic. Fernando Caracciolo y Agustina Giudice, la Dra. Mayra Rengifo Carrillo, al Dr. Alejandro Dana, a los Lic. Lisandro Ronconi y Luis Rivas. Gracias por los trabajos en equipo, las salidas fuera del laboratorio, las charlas y los cafés compartidos.

A la Dra. María Isabel Rodríguez-Franco, por abrirme las puertas de su laboratorio y depositar su absoluta confianza en mí. También a su grupo de trabajo, en especial a la Dra. Concepción Perez, por los ensayos biológicos realizados.

A la Dra. Alicia Couto y su grupo, la Dra. Malena Landoni, y Lic. Maximiliano Rodriguez, con quienes compartí laboratorio durante cinco años. A la Dra. Rosa Erra-Balsells. Por el café y las charlas compartidas.

A UMYMFOR (CONICET-FCEN) y a sus autoridades por facilitar las tareas de mi tesis al poner a disposición todos sus equipos y personal.

Al plantel docente del Departamento de Química Orgánica y a sus autoridades, con quienes tuve el placer de llevar adelante mis tareas de docencia, cursar materias y com-

partir muchos gratos momentos.

Al personal no docente del Departamento de Química Orgánica, por haber estado siempre disponible para cualquier pedido, trámite o eventualidad.

A los Dres. Cristian Rodriguez, Gernot Eskuche y el Ing. José Gallardo por los espectros de RMN realizados a lo largo de estos cinco años.

Al Dr. Reinaldo Pis Diez (CEQUINOR-UNLP) por abrirme las puertas de su grupo de investigación y proveer las computadoras utilizadas para los cálculos llevados a cabo en esta Tesis

Al Dr. Maximilian Helf, por la oportunidad brindada, la enorme cantidad de consejos y oportunidades y por abrir las puertas a una amistad. A la compañia Novartis por brindarme la oportunidad de participar del programa Next Generation Scientist, que contribuyó significativamente a mi formación, a Akiko Keller por llevarlo adelante y a todos mis compañeros del programa.

A los Dres. Nahuel Schenone y Esteban Avigliano (Fundación Bosques Nativos Argentinos para la Biodiversidad) y a todo el personal del CIAR, especialmente a Javier Unizony, por permitirnos aprovechar las instalaciones del CIAR, y por toda la colaboración y el asesoramiento recibidos en los viajes de campaña a la Selva Misionera.

A la Universidad de Buenos Aires, y a la Facultad de Ciencias Exactas y Naturales, por permitirme acceder a formación de primer nivel de forma gratuita tanto de grado como de doctorado, y otorgarme tanto mis dos primeras becas de investigación como el financimiento para realizar una estadía en el exterior.

Al CONICET por haber financiado la beca de finalización de doctorado con la cual

pude llevar adelante las tareas correspondientes a la última etapa de este trabajo. A la Universidad de Buenos Aires, CONICET, MINCyT y a la Agencia Nacional de Promoción Científica y Tecnológica por el financiamiento de este trabajo.

A mis amigos de toda la vida, Bada, Eze, Baby, Richard, Ariel, Nico, gracias por estar siempre.

A los amigos que hice durante todos estos años, Paula, Valen, Sheila, Nacho, Magui, Juan, Agus, Buho, Pablo, Nati, Lauti, el grupo de básquet, gracias por todas las charlas, las idas a la cancha, los cafés, los turnos de docencia, las materias compartidas.

Al Sr. Marcelo Gallardo y al Club Atlético River Plate por todas las alegrías a lo largo de todos estos años.

A Sebastián Kiss, que se encargó de recuperar unos archivos indispensables para la finalización de esta Tesis luego de haberlos perdido de manera accidental.

Quisiera agradecer a mis padres, Daniel y Graciela, porque sin ellos jamás hubiese llegado hasta este punto. Durante el doctorado logré combinar la proactividad y creatividad heredada de mi padre y la capacidad analítica propia de mi madre, características que creo están impresas en todo el trabajo. Gracias por darme la mejor educación del mundo, gracias por los consejos, la paciencia, el interés. Gracias.

A mis hermanos, Florencia y Santiago, por todo su apoyo y su genuino interés durante todos estos años. Gracias por la compañía y por estar siempre a lo largo de todo el proceso. El mismo agradecimiento va para toda la familia, que por suerte es mucha: abuelos, tíos, primos, futuros suegros y cuñada, gracias por el cariño y el apoyo durante todos estos años.

Al Pity, por siempre recibirme con alegría y por su fiel compañía.

A Euge, el amor de mi vida, gracias por ser mi cable a tierra durante todos estos años, gracias por el constante apoyo en mis decisiones. Jamás hubiese logrado esto sin vos.

Dedicado a Euge,

por caminar a mi lado en este viaje,

y por ser parte de todos los que vendrán.

Prólogo

El camino que llevó a la realización de esta Tesis comenzó con una mezcla de intuición y azar. Aunque siempre supe que quería dedicarme a la investigación, el rumbo concreto se fue delineando a través de encuentros fortuitos y decisiones guiadas más por la intuición que por la certeza. Al año de haber ingresado a la carrera de Química —con más dudas que definiciones sobre qué área me interesaba realmente, si síntesis, productos naturales, nanoparticulas u otra cosa—, decidí acercarme al laboratorio de Jorge, quien más tarde sería mi director de tesis. Me atrajeron tanto los temas de investigación como su forma de ser: accesible, humilde, abierto. Características que, aunque parezcan simples, no eran fáciles de encontrar. Sin pensarlo demasiado, le envié un mail y a la semana estaba trabajando en su laboratorio. Así comenzó una aventura que me llevó a recorrer distintos proyectos, desde la síntesis orgánica hasta los productos naturales, aprendiendo a trabajar en la mesada y preguntando (y cuestionando) absolutamente todo lo que se me cruzaba por la cabeza.

Con Jorge hubo una conexión inmediata. Desde el primer día me sentí en confianza, libre para proponer, para equivocarme, para aprender. Creo que él también vio algo en mí, porque desde el inicio depositó una confianza absoluta en mis capacidades. Esa libertad fue, quizás, lo más importante que recibí en esos primeros años: la posibilidad de pensar por cuenta propia y de imaginar sin límites.

Un momento clave fue cuando Jorge comenzó a dictar la materia "Seminarios de

Química Orgánica". Los miércoles, justo los días que iba al laboratorio, se convertían en jornadas completas entre la práctica experimental y la escucha atenta de charlas de investigadores de distintas ramas de la química. En esas clases descubrí por primera vez el mundo de la química computacional aplicada al descubrimiento de fármacos. Recuerdo con nitidez las sensaciones al salir de esas charlas: una mezcla de fascinación y perplejidad. ¿Cómo podía ser que herramientas tan poderosas, actuales y accesibles no fueran de uso cotidiano en la química orgánica? Veía un potencial inmenso, una posibilidad de achicar distancias con los grandes laboratorios del mundo sin necesidad de infraestructura costosa. Desde entonces supe que quería explorar ese cruce.

Mi relación con la computación, sin embargo, venía de mucho antes. Entre los tres y cuatro años, mucho antes de aprender a leer o escribir, ya jugaba en la computadora de mi abuelo, intentando escribir palabras y explorando por intuición. El *juego* consistía en adivinar la combinación de teclas que hiciesen que el buscador de archivos de la PC arrojara algún resultado. Desde entonces siempre mostré cierta facilidad e interés por la computación, lo que generó cierta sorpresa en mi entorno cuando decidí estudiar Química. Incluso durante los últimos años de la carrera me pregunté si no me habría equivocado de camino. Por eso, el doctorado me dio algo que considero valiosísimo: la oportunidad de reconciliar y combinar mis dos grandes pasiones.

Nunca dudé en hacer un doctorado, pero sí dudé dónde hacerlo. En el último año de la carrera, me dediqué a buscar laboratorios de química medicinal en el exterior. En esa búsqueda encontré al grupo de neurofármacos del Instituto de Química Médica en Madrid, dirigido por la Dra. María Isabel Rodríguez Franco —una figura que, sin saberlo

entonces, sería muy importante en esta tesis. Sin embargo, decidí quedarme en Argentina por motivos profundamente personales: sentía que dos de mis abuelos estaban atravesando su última etapa, y no quería perdérmela. Lamentablemente, tenía razón.

Ya con la decisión tomada, supe que quería seguir trabajando con Jorge. Tengo la filosofía —y la convicción— de que uno siempre tiene que esforzarse por ser la mejor versión posible de uno mismo, y en su laboratorio sentí que ese camino era posible. Pero le pedí una condición: el proyecto debía tener una rama computacional fuerte. Sabía que era algo fuera de su zona de confort, pero no tuvo reparos. Le propuse hacer una herramienta de predicción de actividad biológica, pero luego de pensarlo un poco, nos pareció una idea demasiado "volada" como para pedir una beca. A los pocos días Jorge apareció con un plan excelente, que involucraba el diseño racional de compuestos, inspirados en productos naturales, para inhibir un blanco molecular.

Faltaba un codirector que pudiera guiarme en el aspecto computacional. No teníamos claro a quién acudir, hasta que, en un pequeño congreso al que asistimos casi por azar, Belén, en ese entonces becaria del laboratorio, se topó con un póster de una base de datos de productos naturales, llamada *NaturAr*. Así conocimos a Martín, quien resultó ser el complemento ideal: paciente, claro para explicar, abierto a nuevas ideas, y tan generoso con la libertad como Jorge. Con él aprendí lo esencial de la química computacional y me animé a explorar y construir cosas nuevas. Además, compartimos cierta ingenuidad —ese impulso casi infantil de meternos en terrenos desconocidos con la convicción de que podemos aprender o hacer cualquier cosa. Esa ingenuidad, a veces desconcertante y riesgosa, fue también un combustible que nos permitió ir mucho más allá de los objetivos

inicialmente planteados en los proyectos. Nos permitió enfrentarnos a problemas complejos que, de haberlos pensado con más detenimiento y racionalidad, probablemente nos habrían hecho retroceder antes de empezar.

Mirando hacia atrás, me doy cuenta de que este camino está lleno de conexiones que, aunque pequeñas en apariencia, fueron determinantes. La materia Seminarios, el pequeño congreso, guardar el sitio web del grupo de Maribel mientras buscaba grupos de investigación en el exterior, incluso un *workshop* internacional sobre química medicinal dictado durante un congreso SINAQO en Mendoza al que asistí, apenas comenzado el doctorado, inscribiéndome de casualidad. Allí pude ver de cerca cómo trabajan los grupos líderes en *drug discovery*, tanto en la academia como en la industria y me permití soñar con algún día ser parte de ese mundo, extremadamente lejano en aquellos primeros días.

El inicio del doctorado estuvo marcado por la pandemia. Me agarró muy temprano, con apenas algunas reacciones hechas, y me mantuvo alejado de la mesada por más de un año y medio. Lo que en principio parecía ser una etapa angustiante y poco productiva, terminó siendo todo lo contrario: pude aprovechar el tiempo para formarme a fondo en los aspectos de química computacional que iba a necesitar a lo largo de la tesis. Hasta nos dimos el lujo de publicar un trabajo en ese contexto.

Tuve la suerte de realizar dos viajes de investigación que marcaron profundamente esta Tesis. Uno al laboratorio de Maribel, donde pudimos validar experimentalmente algunos de los hallazgos computacionales y cuyos resultados están directamente reflejados en el trabajo. El otro, a Suiza, para trabajar en la farmacéutica Novartis durante los últimos meses del doctorado. Allí, en uno de esos ambientes que alguna vez había admirado desde

lejos en aquella SINAQO en Mendoza, tuve el privilegio de cumplir un sueño: trabajar en temas de frontera y discutir ciencia con profesionales formados al más alto nivel.

En ambos casos, pude comprobar que la formación que recibimos en el país es de excelencia. Las bases adquiridas nos permiten insertarnos rápidamente en entornos de investigación exigentes y abordar nuevos temas con muy poca necesidad de adaptación. A pesar de la abismal diferencia económica, entendí que esa brecha no debería traducirse en una limitación a nuestra ambición científica, especialmente en el área computacional. Por el contrario, es necesario redoblar esfuerzos para incorporar temáticas novedosas a nuestras líneas de investigación de manera inteligente: a veces podemos anticiparnos a los temas trascendentes del futuro, otras veces podemos sumarnos a ellos en etapas tempranas, y también podemos encontrar pequeños huecos en el conocimiento donde avanzar a un ritmo saludable y sostenible en el contexto local.

Además, estos viajes reafirmaron la importancia de establecer colaboraciones internacionales fructíferas, no solo como vía de acceso a recursos o tecnología, sino como espacios de intercambio que enriquecen nuestra ciencia, amplían nuestra mirada y nos permiten aportar al conocimiento global desde una perspectiva propia.

Gran parte del trabajo presentado en esta Tesis nace de una idea que surgió un fin de semana, de esas que llegan como un relámpago, cargadas de una adrenalina difícil de describir, la sensación de que algo distinto, innovador y factible está tomando forma dentro tuyo. Esa idea podría haber quedado simplemente en eso, un impulso pasajero, si no hubiera sido por la libertad que me dieron ambos directores para explorar, para probar, para confiar en lo nuevo. Esa confianza es algo que valoro profundamente y por lo

que estaré siempre agradecido. Quedará en el lector juzgar la originalidad y la calidad de lo presentado, pero el recorrido que permitió que esa idea se transformara en un cuerpo de trabajo concreto ya representa, para mí, una recompensa invaluable.

El doctorado fue para mí mucho más que una etapa académica. Fue el espacio donde convergieron la pasión por la química, la fascinación por la computación y el deseo profundo de crear algo que tuviera sentido. Donde aprendí a investigar, pero también a confiar en mi intuición, a aceptar la incertidumbre como parte del camino, y a valorar profundamente la libertad de pensamiento. Y todo eso fue posible gracias a las personas que me acompañaron, me enseñaron y me impulsaron a ir un poco más allá.

Esta Tesis concluye una etapa muy importante en mi vida, la más importante a nivel formativo. Un recorrido en el que encontré mi lugar en la ciencia, pero también una forma de habitarla: abierta, honesta, interdisciplinaria, guiada por la curiosidad y la libertad de pensamiento. Una ciencia en la que la creatividad no es un lujo, sino una necesidad. Y es precisamente esta creatividad la que debemos proteger y cultivar en nuestros espacios académicos: necesitamos entornos donde las ideas puedan respirar, donde el pensamiento lateral sea valorado, donde la curiosidad genuina encuentre tierra fértil. Porque las grandes transformaciones científicas no ocurren solo por acumulación de datos, sino por saltos intuitivos que requieren contextos que los propicien y los celebren.

Creo profundamente en la investigación como un camino para expandir lo posible, y en la búsqueda de ser la mejor versión de uno mismo como una forma de honrar ese camino. Si algo espero haber logrado en estas páginas, es transmitir esa convicción.

Índice general

Glosario XXII

1 Introducción **1**
1.1. La naturaleza como fuente de medicamentos 2
1.2. La riqueza química de los productos naturales 4
1.3. Naturalmente bioactivos 7
1.4. Desarrollo de nuevas moléculas bioactivas basadas en productos naturales 10
1.4.1. Síntesis total de productos naturales 10
1.4.2. Diversificación estructural de productos naturales 11
1.5. De lo natural a lo digital: herramientas computacionales para descubrir nuevas bioactividades 15
1.5.1. El uso de docking en la predicción de blancos moleculares basada en el blanco molecular 19
1.5.2. La descripción de la interacción entre el compuesto y la proteína . 25
1.5.3. La incorporación de la inteligencia artificial 27
1.6. Desarrollo de una metodología de predicción de blancos moleculares para productos naturales 29

2 Desarrollo de una nueva metodología de predicción de blancos moleculares **41**
2.1. Integrando docking inverso y clasificación por IA: Diseño de la metodología Target Fisher 42
2.1.1. La Caseína Quinasa II como prueba piloto 42
2.1.2. Independencia respecto del programa de docking 57
2.1.3. Comparación con otros algoritmos de Inteligencia Artificial 58
2.2. Extensión de la metodología a un mayor número de blancos moleculares . 59
2.3. Cambio hacia Autodock Vina 62
2.3.1. Creación de una herramienta independiente para obtener huellas de interacción basadas en mecánica molecular 63
2.3.2. Comparación de la metodología con la de referencia 64
2.3.3. Desafíos a la hora de definir los conjuntos de activos y de inactivos 66
2.3.4. Selección del mejor modelo para cada blanco molecular 77
2.4. Target Fisher 81
2.4.1. Funcionamiento de la herramienta 82
2.4.2. Código Fuente 84
2.4.3. Servidor Web 84

3 Cribado virtual de una quimioteca *in house* de productos naturales utilizando Target Fisher, con foco en enzimas relevantes para enfermedades neurodegenerativas **91**
3.1. Blancos moleculares relevantes en enfermedades neurodegenerativas 92
3.1.1. Acetilcolinesterasa (AChE) y Butirilcolinesterasa (BuChE) 94
3.1.2. Monoamino Oxidasas A y B (MAO-A y MAO-B) 96
3.1.3. Glucógeno sintasa quinasa - 3β y la hipótesis de τ 97

3.1.4. β-Secretasa 1 (BACE-1) y la hipótesis amiloide 99
3.2. Armado de una quimioteca *in-house* de productos naturales 100
3.3. Predicciones de actividad usando Target Fisher 103
3.4. Ensayos de actividad biológica 107
3.5. Estudio del modo de unión de los compuestos activos 108
3.5.1. Simulaciones de dinámica molecular 112
3.5.2. Cálculos de energía de interacción 114
3.6. Comparación con otras metodologías 117

4 Cribado virtual multiblanco en una quimioteca sintética: identificación de inhibidores de GSK-3β 127
4.1. Análisis fisicoquímico de la quimioteca del grupo de neurofármacos . . . 129
4.2. Predicción de actividad biológica y selección de compuestos para evaluación experimental 131
4.3. Ensayos *in vitro* de una familia de benzotiazepinas presentes en la quimioteca 138
4.4. Relaciones estructura-actividad y análisis del modo de unión 141
4.4.1. Comparación entre 23 y 10 141
4.4.2. Modificación de 23 a 28 142
4.4.3. Otras relaciones estructura-actividad (SAR) identificadas 143
4.5. Puesta a punto de la síntesis de los compuestos de la serie benzotiazepínica 144
4.6. Estudio del modo de unión 147
4.6.1. Análisis de las energías de interacción ΔH entre los compuestos y la proteína 149

5 Síntesis de derivados de triptamina inspirados en alcaloides naturales 155
5.1. Síntesis de análogos de Aspidostomida G 160
5.1.1. Formación de amidas utilizando triptaminas comerciales 160
5.1.2. Diversificación estructural del anillo indólico mediante la síntesis de triptaminas 161
5.1.3. Síntesis *one pot* de aspidostomidas a partir de la síntesis de indol de Fisher 162
5.1.4. Reemplazo isostérico del grupo amida 164
5.2. Diseño racional de análogos con potencial actividad sobre CK2 167
5.2.1. Estudio de la posible interacción entre la aspidostomida G y CK2 167
5.2.2. Estudios por dinámica molecular de la interacción entre la aspidostomida G y CK2 169
5.2.3. Armado de una base de datos de todos los análogos sintetizables de aspidostomida G 170
5.2.4. Estudios de docking molecular 171
5.3. Exploración de la actividad biológica 173
5.4. Diseño de nuevos inhibidores de MAO-B 178

6 Materiales y métodos 183
6.1. Métodos computacionales - Generalidades 184
6.1.1. Docking molecular 184
6.1.2. Construcción de modelos de aprendizaje automático 187
6.1.3. Simulaciones de dinámica molecular 188
6.1.4. Simulaciones de dinámica molecular de los compuestos **10**, **23**, **32**, **32**, **34** y **33**. 190

6.2. Métodos experimentales - Generalidades 192
6.2.1. Equipos y condiciones generales 192
6.2.2. Aislamiento, purificación e identificación de cumarinas 193
6.2.3. Medición y simulación de Dicrosismo Circular Electrónico (ECD) 194
6.2.4. Protocolos de síntesis 205
6.2.5. Síntesis clásica de triptaminas 209
6.2.6. Información estructural de los compuestos sintetizados 215
6.3. Ensayos de actividad biológica 259
6.3.1. Ensayos de actividad enzimática 259
6.3.2. Ensayos de actividad antifúngica 262
6.3.3. Evaluación de la actividad tripanocida 263

7 Conclusiones **271**
7.1. Perspectivas futuras 276
7.2. Producción científica 277
7.3. Conclusión final 278

Glosario

Acrónimo	Significado
12-MMV	12-metoxi-N-metil-voachalotina
ACID	Herramienta de docking inverso
ACE	Enzima convertidora de angiotensina 1
ACHE	Acetilcolinesterasa
AD	Enfermedad de Alzheimer (Alzheimer's Disease)
AfroDB	Base de datos de productos naturales de África
AUC-ROC	Área bajo la curva ROC (Receiver Operating Characteristic)
Aβ	β-amiloide
BACE1	β-secretasa 1
BCHE	Butirilcolinesterasa
BHE	Barrera hematoencefálica
AutoDock4	Programa de docking molecular
Bz	Benznidazol
BBB^+	Permeable a la barrera hematoencefálica
BBB^-	No permeable a la barrera hematoencefálica
BRAF	Proteína quinasa Serina/Treonina B-Raf
CDK1	Quinasa dependiente de ciclina 1
CDK2	Quinasa dependiente de ciclina 2
CDK5	Quinasa dependiente de ciclina 5
ChEI	Inhibidor de colinesterasa (Cholinesterase Inhibitor)
CK2	Caseína Quinasa II
COCONUT	Base de datos de productos naturales
COX-2	Ciclooxigenasa-2
CSIC	Consejo Superior de Investigaciones Científicas
CSNK2A1	Subunidad alfa de la quinasa de caseína II
DHFR	Dihidrofolato reductasa
DHODH	Dihidroorotato deshidrogenasa
DMSO	Dimetilsulfóxido
DPDR-CPI	Herramienta de docking inverso con aprendizaje automático
DrugBank	Base de datos de moléculas aprobadas para uso médico
DUDE-Z	Servidor para generación de decoys
DYRK1A	Quinasa de tirosina regulada por fosforilación de doble especificidad 1A
ECFP	Extended Connectivity Fingerprints
FAD	Flavín adenín dinucleótido
F10	Factor X (Protrombinasa)
FLT1	Receptor 1 del factor de crecimiento endotelial vascular
FCEN.UBA	Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
FRED	Programa de docking molecular de OpenEye
FXa	Factor X (Protrombinasa)
GridSearchCV	Búsqueda en grilla con validación cruzada

Continúa en la página siguiente...

Tabla 1 – Continuación

Acrónimo	**Significado**
GSK3B	Glicógeno sintasa quinasa 3 beta
HDAC2	Histona deacetilasa 2
HDAC6	Histona deacetilasa 6
HDAC8	Histona deacetilasa 8
HBA	Aceptores de enlace de hidrógeno (Hydrogen Bond Acceptors)
HBD	Donores de enlace de hidrógeno (Hydrogen Bond Donors)
IA	Inteligencia Artificial
IFP	Huella de Interacción Proteína-Ligando (Interaction Fingerprint)
Hybrid	Programa de docking molecular de OpenEye
IGF1R	Receptor del factor de crecimiento similar a la insulina 1
IQUIBICEN	Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales
IQM	Instituto de Química Médica
INVDOCK	Herramienta de docking inverso
JAK3	Janus quinasa 3
LBTP	Predicción de Blancos basada en el Ligando (Ligand-Based Target Prediction)
MAOA	Monoaminooxidasa A
MAOB	Monoaminooxidasa B
MAOs	Monoamino oxidasas
MIC	Concentración inhibitoria mínima (Minimum Inhibitory Concentration)
MIEF	Huellas de Energía de Interacción Molecular (Molecular Interaction Energy Fingerprints)
ML	Aprendizaje Automático (Machine Learning)
MLPC	Clasificador de Perceptrón Multicapa (Multi-Layer Perceptron Classifier)
Morgan Fingerprints	Descriptores moleculares utilizados para análisis de similitud
MTNR1A	Receptor de melatonina tipo 1A
MTNR1B	Receptor de melatonina tipo 1B
MW	Peso molecular (Molecular Weight)
ND	Enfermedad neurodegenerativa (Neurodegenerative Disease)
NaturAr	Base de datos de productos naturales de Argentina
NFT	Ovillos neurofibrilares (Neurofibrillary Tangles)
NPASS	Base de datos especializada en registro de actividades biológicas de productos naturales
NPAtlas	Base de datos de productos naturales de microorganismos
NuBBE	Base de datos de productos naturales de Brasil
NVL	Número de Violaciones de las Reglas de Lipinski
NER	Número de Enlaces Rotables
Nx	Nifurtimox
OpenEye	Suite comercial de programas de modelado molecular

Continúa en la página siguiente...

Tabla 1 – Continuación

Acrónimo	**Significado**
PADIF	Descriptores de interacción proteína-ligando
PCA	Análisis de Componentes Principales
PD	Enfermedad de Parkinson (Parkinson's Disease)
PDB	Protein Data Bank
PDE5	Fosfodiesterasa específica de cGMP tipo 5
PDE5A	Fosfodiesterasa 5A
PDE7A	Fosfodiesterasa 7A
PDE10A	Fosfodiesterasa 10A
PDPK1	Piruvato deshidrogenasa quinasa-1
PI3Kg	Fosfoinositida-3-quinasa gamma
PIK3CG	Fosfoinositida-3-quinasa gamma
PIM-2	(Protooncogén) Proteína quinasa Serina/Treonina 2
PLA2B	Fosfolipasa A2, asociada a membranas
PPB2	Herramienta de predicción de blancos
PRSS1	Tripsina
Ptgs2	Ciclooxigenasa-2
PLAT	Activador tisular del plasminógeno
PyBOP	Benzotriazol-1-il-oxitripirrolidinfosfonio hexafluorofosfato
Pyraclostrobin	Fungicida comercial utilizado como control positivo
RDKit	Librería de quimioinformática de código abierto
RF	Random Forest (Bosques Aleatorios)
RMSD	Desviación cuadrática media (Root Mean Square Deviation)
ROC	Característica Operativa del Receptor (Receiver Operating Characteristic)
ROS	Especies reactivas de oxígeno (Reactive Oxygen Species)
SAR	Relaciones estructura-actividad (Structure-Activity Relationship)
SBTP	Predicción de Blancos basada en la Estructura (Structure-Based Target Prediction)
SEA	Herramienta de predicción de blancos basada en similitud
SA score	Puntuación de accesibilidad sintética (Synthetic Accessibility score)
Smina	Programa de docking molecular
SmiLib	Software de química combinatoria
SNC	Sistema Nervioso Central
SuperPred	Herramienta de predicción de blancos
SVM	Máquinas de Vectores de Soporte (Support Vector Machines)
SwissBioisostere	Herramienta para análisis de bioisosterismo
TargetHunter	Herramienta de predicción de blancos
TargetNet	Herramienta de predicción de blancos
TarFisDock	Herramienta de docking inverso
THF	Tetrahidrofurano
TIE2	Receptor de angiopoyetina-1
TLC	Cromatografía en capa fina (Thin Layer Chromatography)

Continúa en la página siguiente...

Tabla 1 – Continuación

Acrónimo	**Significado**
TPSA	Área de Superficie Polar Topológica (Topological Polar Surface Area)
TPA	Activador tisular del plasminógeno
Thrombin	Trombina
Trypsin	Tripsina
UMYMFOR	Unidad Microbiología y Micología, Facultad de Química, Bioquímica y Farmacia
Vinardo	Programa de docking molecular
VEGFR1	Receptor 1 del factor de crecimiento endotelial vascular
SwissTargetPrediction	Herramienta de predicción de blancos

Capítulo 1

Introducción

1.1 La naturaleza como fuente de medicamentos

El uso de productos naturales en el tratamiento de enfermedades ha sido una constante desde los albores de la humanidad. Los primeros registros médicos, como el papiro Ebers del antiguo Egipto, datan de alrededor del año 2900 a.C. y documentan el uso de 328 ingredientes naturales en 876 prescripciones médicas para tratar diversas enfermedades (1). Otras culturas, como la griega y la china, también poseen registros antiguos del uso de este tipo de sustancias en medicina como la Materia Médica China (1100 a.C., 52 formulaciones), y los registros de Dioscorides (100 a.C.) y Theophrastus (300 a.C.). Estos primeros descubrimientos, a menudo formulados en medicinas tradicionales y pociones, contenían mezclas de compuestos que fueron identificadas muchos años después (2).

El avance de la química y la biología en los últimos siglos ha permitido una comprensión más profunda de estos compuestos naturales. En particular, la identificación de su estructura química y sus rutas biosintéticas ha llevado a la clasificación de estos compuestos en metabolitos primarios y secundarios. Mientras que los metabolitos primarios, como la glucosa y los ácidos nucleicos, son esenciales para la vida de cualquier organismo, los metabolitos secundarios, usualmente referenciados como productos naturales, no son indispensables para la supervivencia inmediata del organismo productor, pero juegan un papel crucial en su adaptación al entorno (3). Las rutas biosintéticas de estos últimos son entonces, por lo general, específicas para la especie, familia o género que las produce. De esta manera, las distintas especies a lo largo de la evolución fueron meticulosamente creando moléculas con estructuras químicas únicas y de muy alta complejidad que

les permiten defenderse de otras especies invasoras de manera específica, permitiendo su supervivencia en el entorno.

A lo largo de la historia, varios productos naturales han demostrado ser de gran importancia en la medicina y la farmacología. Un ejemplo emblemático es el ácido acetilsalicílico (aspirina, (figura 1.1d)), compuesto derivado de la salicina aislada de la corteza del sauce *Salix alba* L., y cuyo aporte a la medicina ha sido fundamental para el tratamiento de la inflamación y el dolor. Otro ejemplo destacado es la morfina (figura 1.1a), un alcaloide aislado de *Papaver somniferum* L., que ha sido esencial en el tratamiento del dolor desde su descubrimiento en 1803. La morfina, que históricamente fue utilizada como extracto crudo, dio lugar en el siglo XIX a la creación de derivados como la heroína (figura 1.1b) y la codeína (figura 1.1c), que tienen propiedades analgésicas. Finalmente, un descubrimiento revolucionario en la historia de la medicina fue la penicilina (figura 1.1e), aislada en 1928 por Alexander Fleming a partir del hongo *Penicillium notatum*. Este compuesto, que marcó el inicio de la era de los antibióticos, ha sido fundamental en el tratamiento de infecciones bacterianas y ha salvado innumerables vidas desde su introducción en la medicina.

Hoy en día, los productos naturales siguen siendo una fuente fundamental de nuevos medicamentos, ya sea a través de su uso directo o mediante su modificación para dar lugar a derivados semisintéticos. Un análisis reciente de Newmann y Cragg revela que un porcentaje significativo de los fármacos no biológicos aprobados en la actualidad tiene algún vínculo con compuestos naturales (4). Esta tendencia demuestra la relevancia continua de este tipo de compuestos en farmacología y refuerza la importancia de enfoques

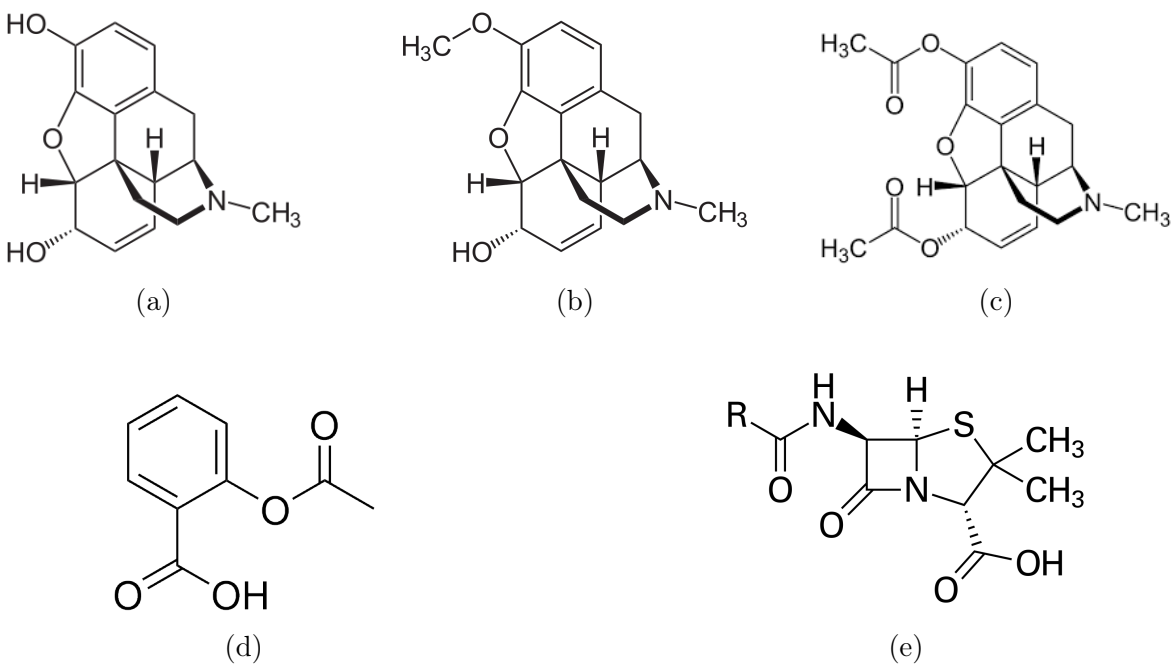


Figura 1.1: Estructuras químicas de morfina (a), heroína (b) y codeína (c), ácido acetilsalicílico (d) y penicilina (e).

como el que se propone en esta tesis, que integra la identificación de blancos moleculares con la optimización de derivados a partir de estos compuestos.

1.2 La riqueza química de los productos naturales

Los avances tecnológicos han facilitado la recopilación y digitalización de vastos conjuntos de datos sobre productos naturales. La base de datos más completa hasta la fecha, COCONUT (5), incluye más de 400.000 productos naturales únicos. Existen también bases de datos especializadas que se centran en productos de una fuente específica, como NPAtlas (6), que contiene productos derivados de microorganismos, o en regiones geográficas, como AfroDB (África) (7), NuBBE (Brasil) (8), y NaturAr (Argentina) (9). Estas bases de datos permiten una exploración exhaustiva del espacio químico ocupado por los productos naturales.

La síntesis orgánica de nuevos candidatos a fármacos ha sido profundamente influenciada por las reglas de Lipinski (10), también conocidas como la "Regla de 5". Estas reglas establecen criterios fisicoquímicos clave que las moléculas deben cumplir para ser consideradas como posibles candidatos a fármacos orales, incluyendo un peso molecular inferior a 500, no más de 5 donores de enlaces de hidrógeno, no más de 10 aceptores de enlaces de hidrógeno y un coeficiente de partición octanol-agua (log P) menor a 5. Su implementación ha orientado el diseño de nuevas moléculas hacia el cumplimiento de estos parámetros, ya que hacerlo reduce potencialmente los riesgos, tiempos y costos asociados al desarrollo de moléculas candidatas que no los satisfagan. Sin embargo, la mayoría de los productos naturales no cumplen estas reglas y, de hecho, fueron considerados por el propio Lipinski como excepciones a las mismas. Esto se debe a que los productos naturales han evolucionado en la naturaleza con características específicas, y su ingreso a las células generalmente depende de transportadores específicos en lugar de la difusión pasiva (10).

El espacio químico teórico ocupado por sólo aquellas moléculas que cumplan las reglas de Lipinski se estima en un número de moléculas del orden de 10^{60}. Esto permite concluir que el espacio químico total, definido como el número total de moléculas teóricas que podrían existir, es aún mayor. Dentro de este vasto espacio, el subespacio correspondiente a los productos naturales representa sólo una pequeña fracción, la cual se superpone parcialmente con los compuestos que cumplen la "Regla de 5" de Lipinski. No obstante, esta fracción es excepcionalmente diversa y, en comparación con los compuestos sintéticos, exhibe propiedades fisicoquímicas únicas (11).

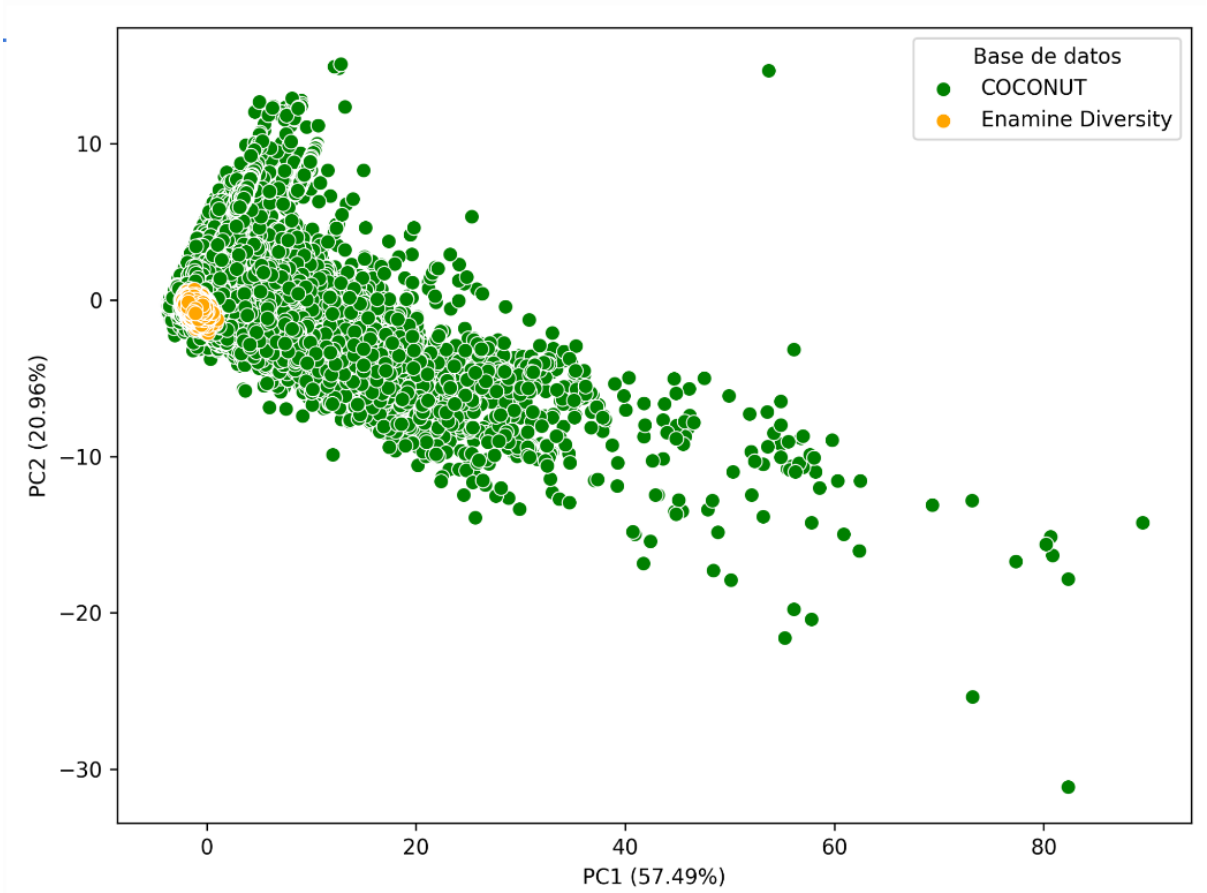


Figura 1.2: Representación en 2D de los componentes principales (PC1 y PC2) obtenidos mediante PCA de las bases de datos "Enamine Diversity" (naranja) y "COCONUT" (verde). Los puntos se distribuyen en el espacio según las características fisicoquímicas de las moléculas, mostrando la variabilidad entre ambas bases de datos.

Una comparación entre el espacio químico ocupado por los productos naturales y compuestos sintéticos se puede observar en la figura 1.2. En ella, se muestra un análisis de componentes principales, técnica que permite reducir la dimensionalidad de los datos y visualizar similitudes y diferencias en la distribución estructural de los compuestos, utilizando COCONUT, la base de datos que contiene la mayor cantidad de productos naturales y el conjunto de compuestos sintéticos de alta diversidad vendido por Enamine (12), empresa dedicada a la generación de quimiotecas listas para su uso en proyectos de descubrimiento de fármacos. La diferencia en el espacio químico ocupado por ambos conjuntos es notoria, y es similar a lo observado cuando se hace la misma comparación

pero considerando sólo drogas aprobadas (13).

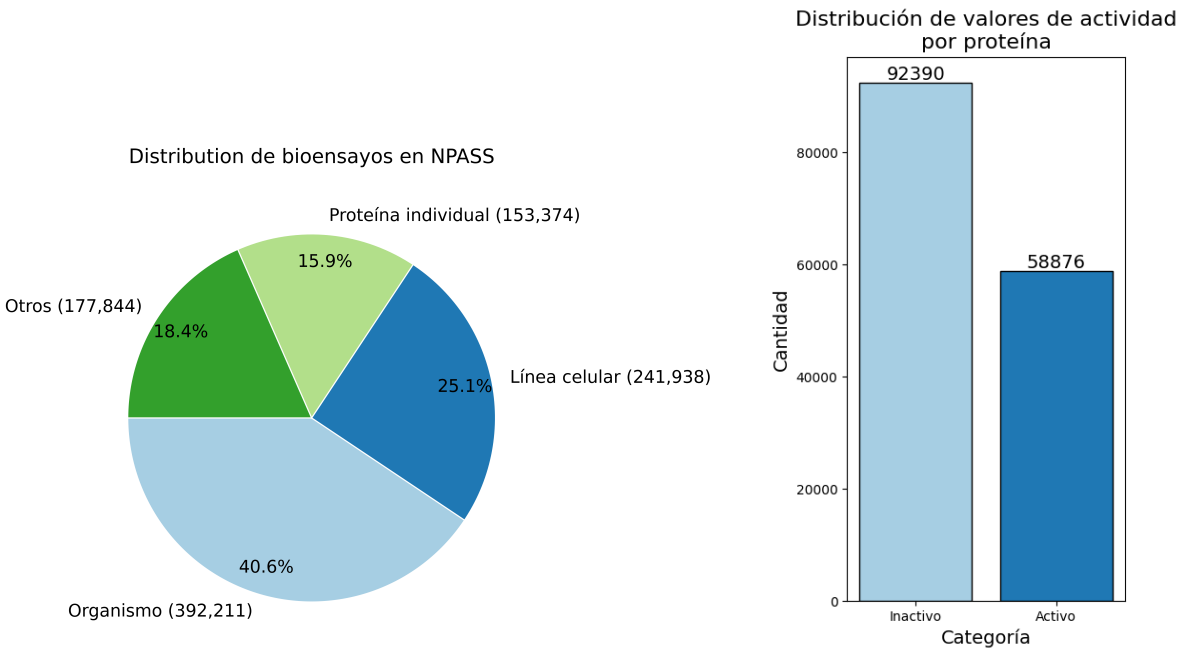
Esta complejidad estructural, producto de millones de años de evolución, ha transformado a los productos naturales en no sólo valiosos candidatos a fármacos, sino también en una fuente inagotable de inspiración para quienes se dedican a la química medicinal. Además, un análisis reciente de la base de datos ChEMBL destaca que los motivos estructurales derivados de productos naturales continúan desempeñando un papel fundamental en la innovación farmacéutica (14).

1.3 Naturalmente bioactivos

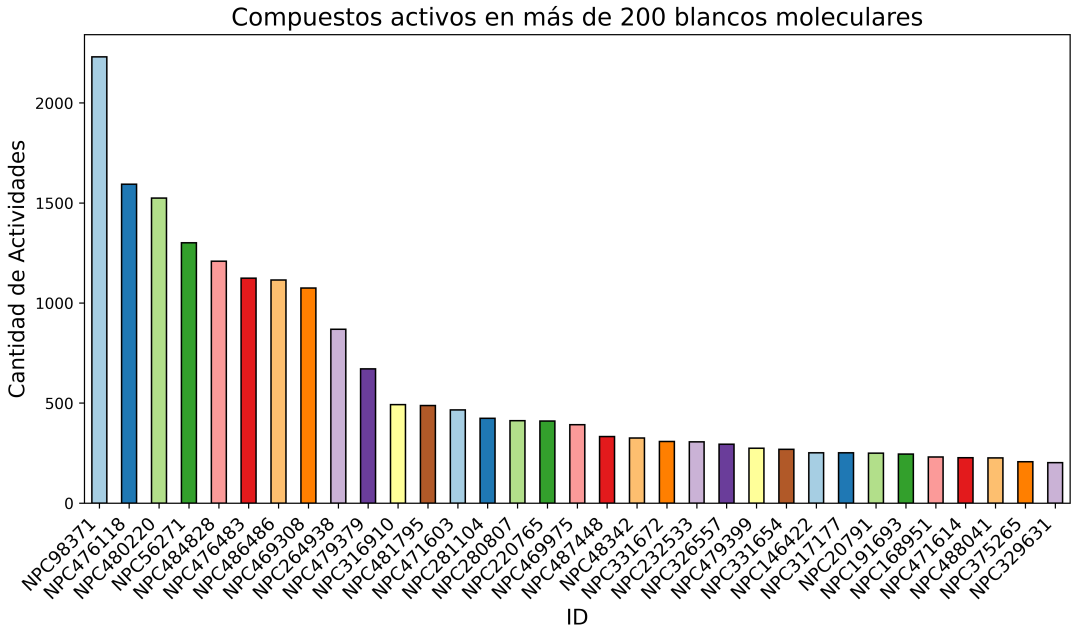
Los productos naturales son el resultado de un proceso evolutivo extenso, que ha seleccionado moléculas biológicamente relevantes (15). Este hecho sugiere que muchos de estos compuestos están diseñados para interactuar con blancos biológicos específicos, lo que los convierte en excelentes candidatos para el desarrollo de nuevas terapias. Sin embargo, la identificación de su bioactividad plantea desafíos significativos.

En primer lugar, la cantidad de compuesto que se obtiene de la fuente natural suele ser reducida, permitiendo solamente un número limitado de ensayos biológicos a realizar. Por otro lado, el número de ensayos que puede realizar el grupo que aisló el compuesto habitualmente es limitado y muy poco variado. Como consecuencia, la exploración de la bioactividad de un producto natural suele ser sumamente incompleta, siendo la elección de experimentos a realizar un proceso crítico.

Esto puede verse reflejado cuantitativamente mediante el análisis de la base de datos NPASS (16), especializada en el registro de actividades biológicas de productos naturales.



(a) (b)



(c)

Figura 1.3: Análisis de la base de datos NPASS. (a) Distribución general de bioensayos registrados. (b) Distribución específica de bioensayos contra proteínas individuales. (c) Compuestos asociados a múltiples blancos moleculares.

La misma, en su versión más reciente, contiene 96481 compuestos únicos y 958866 ensayos de actividad recopilados. Analizando la distribución de estos últimos, que se muestra en la figura 1.3, se observa que la gran mayoría (65,7 %) es del tipo fenotípico, buscando principalmente actividades frente a organismos enteros y líneas celulares. Esto pone de manifiesto que, por lo general, la búsqueda de la actividad biológica ha estado restringida principalmente a este tipo de ensayos, que suelen estar más al alcance de los grupos de investigación.

La otra clase de ensayos importante es aquella en donde se estudia la potencial inhibición de una proteína individual o blanco molecular por parte del compuesto en estudio. Este tipo de experimentos tienen la potencialidad de brindar una hipótesis del posible mecanismo de acción del compuesto, permitiendo el estudio de su actividad biológica en mayor profundidad. Además, el poder identificar el blanco molecular de un producto natural permite el uso de estrategias de química computacional para el modelado del complejo formado con la proteína, abriendo las puertas al diseño racional de derivados que pueden ser sintetizados químicamente para maximizar esa interacción. De acuerdo a lo recopilado por NPASS, sólo el 15,9 % de los ensayos son de este tipo, siendo 2581 el número de proteínas distintas utilizadas. De estos 153374 ensayos, el 38,4 % (58876) dio como resultado la inhibición de la proteína en estudio a una concentración menor a 10 µM. Finalmente, si se tiene en cuenta que un compuesto puede tener más de un blanco molecular, del análisis de NPASS se llega a que solo el 5,7 % (5509) de los productos naturales recopilados tienen identificado al menos un blanco molecular. En conclusión, a pesar de que el número de productos naturales descubiertos es extenso, y de que se

presume que la gran mayoría de ellos presentan alguna bioactividad, de sólo una pequeña parte se han comprobado experimentalmente las proteínas con las que interactúan.

1.4 Desarrollo de nuevas moléculas bioactivas basadas en productos naturales

El aislamiento y uso directo de productos naturales no es la única manera en que la naturaleza puede aprovecharse para el desarrollo de nuevas terapias. Las estructuras únicas de estos compuestos han sido una fuente inagotable de inspiración para la síntesis orgánica, dando lugar a dos estrategias ampliamente utilizadas en el diseño de nuevos fármacos inspirados en la naturaleza: la **síntesis total** y la **diversificación estructural de productos naturales**.

1.4.1. Síntesis total de productos naturales

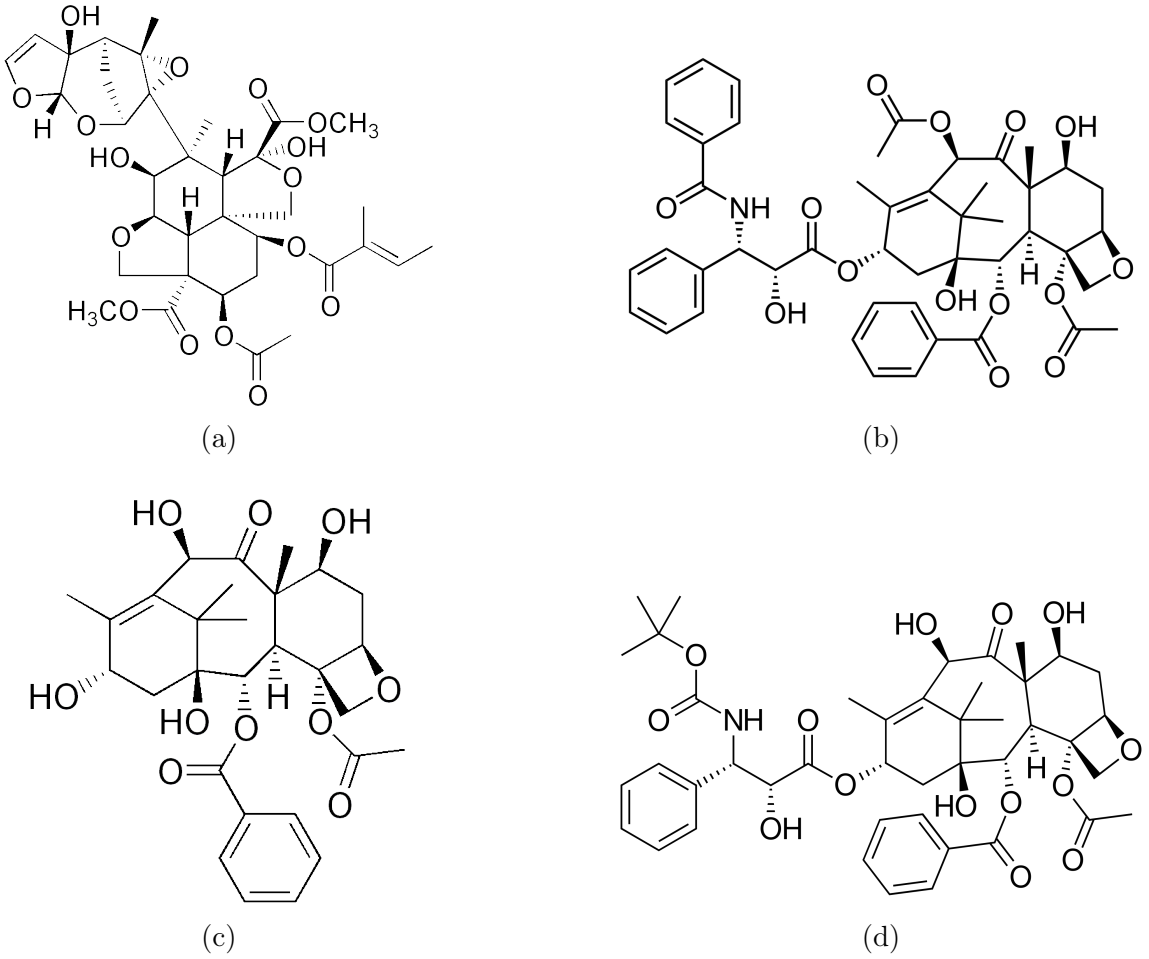
La síntesis total consiste en la obtención de un compuesto objetivo partiendo de sustancias más simples y comerciales mediante rutas sintéticas diseñadas cuidadosamente. La riqueza química de los productos naturales, caracterizada por la diversidad de grupos funcionales y la abundancia de estereocentros definidos, ha desafiado e inspirado a químicos sintéticos. En el proceso de buscar la síntesis total de estos compuestos, se ha impulsado el desarrollo y descubrimiento de nuevas técnicas y reacciones en el campo de la química orgánica.

Una de las principales ventajas de esta estrategia es la posibilidad de diseñar y sintetizar análogos del compuesto objetivo. Estas variaciones estructurales permiten explorar

propiedades fisicoquímicas distintas o incluso descubrir nuevas bioactividades. Sin embargo, un desafío significativo radica en los bajos rendimientos finales que suelen obtenerse en la síntesis total y el impacto ecológico asociado al escalado industrial de múltiples reacciones. El balance entre las ventajas y desventajas de este enfoque dependerá en gran medida de la complejidad del compuesto en cuestión. Por ejemplo, el caso de la azadiractina, un limonoide insecticida aislado de *Azadirachta indica* (árbol del paraíso), ilustra una situación donde la síntesis total, aunque técnicamente lograda tras décadas de esfuerzo y con rendimientos extremadamente bajos, resulta poco práctica debido a la posibilidad de obtener el compuesto en grandes cantidades directamente de su fuente natural. Lo mismo ocurre con numerosos productos naturales que pueden ser obtenidos mediante fermentación. En contraste, existen casos como el del paclitaxel (comercializado como Taxol®), un antitumoral inicialmente aislado en cantidades mínimas de *Taxus brevifolia*, cuya baja disponibilidad natural impulsó esfuerzos en síntesis total y, más exitosamente, en semisíntesis a partir de precursores abundantes como la 10-deacetilbaccatina III. Este enfoque no solo permitió la producción a escala comercial del fármaco, sino que también condujo al desarrollo de derivados clínicamente relevantes como el docetaxel (Taxotere®).

1.4.2. Diversificación estructural de productos naturales

Otra estrategia destacada es la generación de derivados sintéticos a partir de un producto natural original, también conocida como diversificación estructural. En este enfoque, se aísla el compuesto de su fuente natural y, a partir de este, se realizan modificaciones químicas específicas. Este método requiere que el compuesto en cuestión pueda ser



(a) (b)

(c) (d)

Figura 1.4: Estructuras químicas de la azadiractina (a), paclitaxel (Taxol, b), 10-deacetilbaccatina III (c) y docetaxel (d).

obtenido en cantidades significativas a partir de su organismo productor.

Generalmente, en este tipo de estrategias, se planifican rutas sintéticas de pocos pasos de reacción que introduzcan cambios relevantes en la estructura molecular, lo que permite generar quimiotecas con alta diversidad y complejidad. Un ejemplo reciente desarrollado en el país es el caso de la 12-metoxi-N-metil-voachalotina (12-MMV), aislada de la corteza de *Tabernaemontana catharinensis* (horquetero), recolectada en la provincia de Misiones. A partir de este compuesto, se sintetizaron los 12 derivados que se muestran en la figura 1.5, cuya actividad antifúngica fue evaluada frente a distintas cepas de hongos fitopatógenos. La variación en la capacidad de estos compuestos en inhibir el crecimiento fúngico evidencia que es posible inducir una actividad biológica inexistente en la molécula original mediante modificaciones químicas específicas.

Al igual que en el caso más tradicional del aislamiento y purificación de nuevos productos naturales, estos dos enfoques también comparten beneficios similares respecto a la búsqueda de nuevas bioactividades. En particular, independientemente de la estrategia utilizada, los proyectos se verían beneficiados de poder conocer actividades biológicas alternativas para el producto natural estudiado, que por lo visto en la sección anterior, mayormente se desconocen.

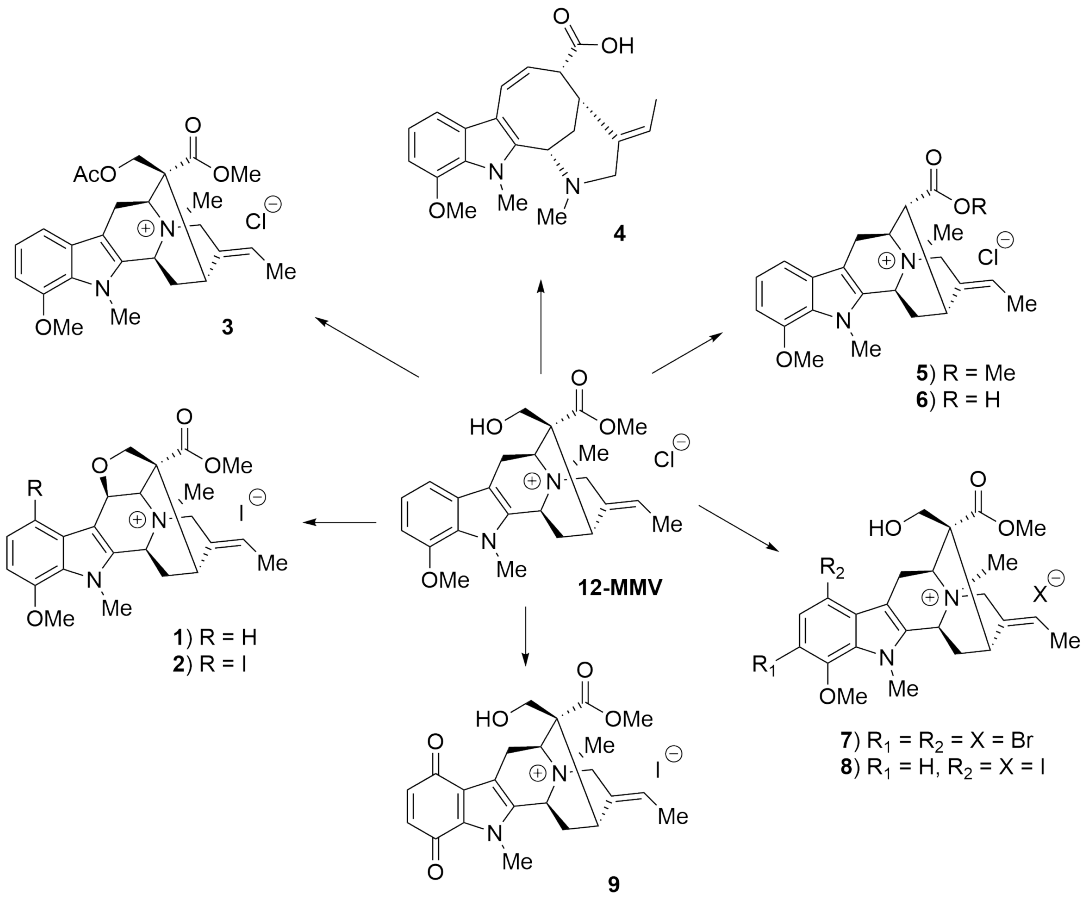


Figura 1.5: Derivados del 12-metoxi-N-metil-voachalotina (12-MMV), aislada de la corteza de *Tabernaemontana catharinensis* (horquetero).

1.5 De lo natural a lo digital: herramientas computacionales para descubrir nuevas bioactividades

Como se desprende del análisis anterior, el descubrimiento de nuevas actividades biológicas de los productos naturales y sus derivados es un área rica en estudio, y de allí surge la necesidad de buscar nuevas estrategias para su descubrimiento. En los últimos años, el crecimiento exponencial de la capacidad de cálculo de las computadoras, en conjunto con su costo decreciente, ha provocado que los métodos computacionales emerjan como un enfoque inicial para la predicción de bioactividades. Estos métodos tienen la ventaja de ser mucho más rápidos y menos costosos que sus contrapartes experimentales. Además, permiten la exploración de un gran número de bioactividades para una cantidad potencialmente ilimitada de compuestos que no necesariamente tienen que encontrarse físicamente disponibles para su ensayo, permitiendo guiar su producción y optimizar la selección de ensayos biológicos.

Las metodologías computacionales para la predicción de bioactividades generalmente buscan predecir la actividad potencial de un compuesto frente a un blanco molecular específico. Dependiendo del tipo de información utilizada, se pueden clasificar en tres enfoques: aquellos basados en la estructura del ligando (LBTP, *ligand-based target prediction*), los que se basan en la estructura del blanco molecular (SBTP, *structure-based target prediction*) y los enfoques híbridos, que integran información de ambos tipos.

Las metodologías LBTP se basan en la premisa de que compuestos similares tendrán bioactividades similares. Por este motivo, este tipo de estrategias se basan en la compa-

ración de la estructura química del compuesto en estudio con las de aquellos compuestos que han sido estudiados para un determinado blanco molecular. Si la estructura es similar, entonces el compuesto será potencialmente activo frente a dicho blanco, y en caso contrario, el compuesto será predicho como inactivo. Este tipo de estrategias puede ser útil para una rápida evaluación frente a un gran número de blancos, debido a su bajo costo computacional. Sin embargo, presenta la desventaja del sesgo introducido por las estructuras de los compuestos que ya han sido ensayados frente a un determinado blanco. Como la gran mayoría de los compuestos ensayados son de origen sintético, y el espacio químico ocupado por los mismos es distinto al de los productos naturales, estas herramientas no son las más adecuadas para este tipo de compuestos. Algunos ejemplos de este tipo de enfoques son SEA (17), SwissTargetPrediction (18), PPB2 (19), TargetNet (20), SuperPred (21) y TargetHunter (22).

Por otro lado, las metodologías SBTP utilizan información proveniente de la interacción entre el compuesto y el blanco molecular para elaborar predicciones. Esto se realiza mediante el uso de alguna herramienta (por lo general, docking molecular) para ubicar el ligando en el sitio activo del blanco molecular. De allí se extrae información que luego puede ser comparada con la obtenida para compuestos con actividad conocida frente a ese blanco molecular. De esta manera, al no utilizar sólo la estructura química del ligando en estudio para elaborar las predicciones, el sesgo introducido por los compuestos sintéticos es potencialmente menor.

En la tabla 1.1 se muestra un resumen de aquellas herramientas publicadas en la actualidad, junto con su disponibilidad al momento de escritura de esta tesis. Del análisis

Tabla 1.1: Herramientas disponibles actualmente para la predicción de blancos moleculares.

Nombre	**Principio**	**URL**	**Disp.**	**Código**
INVDOCK	Docking Inverso	http://bidd.group/group/softwares/invdock.htm	No	Cerrado
idTarget	Docking Inverso	http://idtarget.rcas.sinica.edu.tw/	No	Cerrado
TarFisDock	Docking Inverso	http://www.dddc.ac.cn/tarfisdock	No	Cerrado
DPDR-CPI	Docking Inverso + ML	http://cpi.bio-x.cn/dpdr	No	Cerrado
ACID	Docking Inverso	http://chemyang.ccnu.edu.cn/ccb/server/ACID	No	Cerrado
DIA DB	Docking Inverso + Búsqueda por Similitud	http://bio-hpc.eu/software/dia-db	Si	Cerrado
GUT-DOCK	Docking Inverso	https://gut-dock.miningmembrane.com/	No	Cerrado
Drug ReposER	Farmacóforos	http://mfrlab.org/drugreposer/	Si	Cerrado
LigAdvisor	Interacciones ligando receptor 2D	https://ligadvisor.unimore.it/	Si	Cerrado
PharmMapper	Farmacóforos	http://59.78.96.61/pharmmapper	Si	Cerrado

de las mismas se desprende que su incorporación a proyectos de descubrimiento de fármacos presenta ciertas dificultades. En primer lugar, la gran mayoría se ofrecen únicamente desde un servidor web, que, si bien facilita su uso para usuarios de distintos campos, trae ciertas consecuencias negativas. Por un lado, la disponibilidad de la herramienta queda supeditada a que el servidor web perdure en el tiempo, hecho que depende de la estabilidad del grupo de investigación que lo mantiene y se vuelve especialmente difícil cuando la metodología implica algún tipo de cálculo relativamente costoso y que requiera algún tipo de infraestructura de cálculo compleja. Por otro lado, el servidor web no permite su uso para aquellos usuarios que requieran privacidad en cuanto a la información de las moléculas que son evaluadas. Asimismo, dificulta la posibilidad de comparar dicha metodología con otra diferente, puesto que hay una dependencia con el funcionamiento y las posibilidades que permite el servidor. Como se puede observar, la gran mayoría de estas herramientas no se encuentran disponibles para su uso desde su servidor web original, no son de código abierto, ni proveen alguna forma de instalación local. Esto marca una gran dificultad en el campo, puesto que, si bien los detalles del desarrollo de las mismas se encuentran descritos en las publicaciones respectivas, la imposibilidad de utilizarlas provoca que el área se encuentre en una etapa temprana de desarrollo, muy lejos de su incorporación en proyectos de química medicinal no computacionales.

La otra problemática que se desprende de estas metodologías, y que es independiente del tipo de enfoque que utilizan, es la falta de métricas de rendimiento de los modelos utilizados. La predicción de bioactividad requiere de la evaluación de la molécula en estudio en cada uno de los blancos moleculares presentes en la herramienta, lo que implica la

existencia de un gran número de modelos, cada uno con su respectivo poder clasificatorio. Lamentablemente, solo algunas de estas herramientas (PPB2 y TargetNet en particular) reportan este tipo de métricas para cada uno de los modelos individuales. En ese sentido, otra falla que se desprende de este tema es que tampoco es posible encontrar con qué compuestos fueron entrenados dichos modelos.

1.5.1. El uso de docking en la predicción de blancos moleculares basada en el blanco molecular

La tabla 1.1 muestra además que por lo general, como se mencionó con anterioridad, la técnica más utilizada para predecir el modo de unión del compuesto hacia su potencial blanco molecular es el *docking molecular*, también conocido como acoplamiento molecular. De forma sencilla y conceptual se puede explicar esta técnica a partir del modelo de llave cerradura introducido por Fischer en 1894 (23). En este modelo, la proteína o blanco molecular es considerada la cerradura y la pequeña molécula la llave, teniendo que ambas interactuar correctamente para producir el efecto biológico. El docking molecular realiza cálculos computacionales para buscar la conformación u orientación relativa de la llave que mejor se adapte a la cerradura, prediciendo de esta forma el posible modo de unión de la molécula pequeña al blanco molecular. Vale aclarar que en la práctica esta teoría es demasiado simplista, y el problema se asemeja más al modelo de "mano en guante" introducido por Koshland Jr. en 1958 (24) y que considera la flexibilidad tanto del ligando como del receptor.

La búsqueda del modo de unión más probable de la molécula al blanco molecular

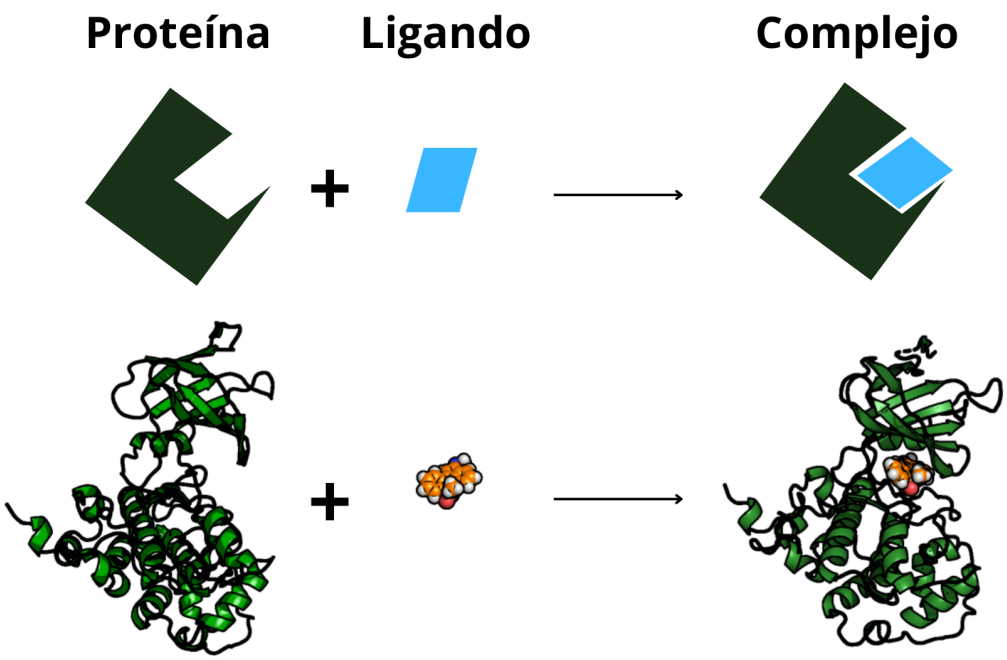


Figura 1.6: Representación esquemática del proceso de docking molecular entre un ligando (molécula pequeña) y un blanco molecular (proteína).

se realiza mediante el uso de un algoritmo de búsqueda conformacional. Por lo general, la molécula es ubicada en el sitio de unión y, mediante la rotación de sus enlaces, se evalúan distintos confórmeros y orientaciones, denominadas poses, mediante una función de puntuación. Aquella pose que tenga la mejor puntuación se plantea como el modo de unión más probable de la molécula en estudio al blanco molecular. El vasto desarrollo de funciones de puntuación y algoritmos de búsqueda, sumado a la posibilidad de abordar problemas de mayor complejidad, como considerar la flexibilidad del blanco, la falta de información sobre sitios de unión, y el uso de moléculas con características especiales como macrociclos y péptidos, hacen del docking molecular un campo de gran desarrollo en la actualidad y con un amplio número de herramientas disponibles. Algunos ejemplos a destacar son Autodock Vina (25), Autodock4 (26), Glide (27), DOCK6.9 (28), Gnina (29), Vinardo (30), GOLD (31), entre otros.

En principio, cualquiera de estos programas puede ser utilizado para el desarrollo de metodologías del tipo SBTP. La elección del software a utilizar dependerá entonces de sus prestaciones y rendimiento, además de la licencia provista por el desarrollador, en caso de que se busque crear una metodología de uso libre.

Estos programas permiten explorar de forma rápida y efectiva la interacción entre la molécula y el sitio activo del blanco molecular, lo que posibilita su uso en campañas de *screening* virtual. En ellas, grandes bibliotecas virtuales de compuestos son evaluadas frente a un único blanco molecular, seleccionando aquellos que presentan mayor afinidad para su evaluación en etapas experimentales posteriores. El proceso contrario, es decir el estudio de un único compuesto frente a más de un blanco molecular en simultaneo, es conocido como *screening* virtual inverso, es el utilizado preferentemente en las herramientas del tipo SBTP.

Uno de los principales desafíos en la incorporación de programas de *docking* para el desarrollo de este tipo de metodologías radica en la complejidad de modelar sistemas que pueden presentar una alta variabilidad entre sí. En particular, el cribado virtual de una quimioteca de compuestos sobre un blanco molecular requiere una serie de pasos previos al uso del programa de *docking*, los cuales son altamente dependientes de la información disponible sobre el sistema en estudio. La calidad del modelo final está directamente relacionada con el nivel de atención y el detalle aplicado en cada una de estas etapas, las cuales se describen a continuación.

1. **Obtención de la estructura tridimensional de la proteína.** Este primer paso tiene una alta dependencia con la información conocida de la misma, que en caso

de estar cristalizada se debe seleccionar la estructura adecuada (de existir varias) del Protein Data Bank, y en caso de no estarlo se debe recurrir a un modelado por homología. Es interesante remarcar que en los últimos años la predicción de estructuras tridimensionales de proteínas ha mejorado, pudiendo ahora obtener un modelo altamente preciso de un amplio número de proteínas, debido a los últimos avances en inteligencia artificial y el consecuente desarrollo de AlphaFold (32), lo que ha dado lugar al Premio Nobel de química en el año 2024 a los científicos David Baker, Demis Hassabis y John Jumper por la predicción de estructuras proteicas (33)

2. **Preparación de la estructura tridimensional de la proteína para su uso en docking molecular.** Este paso implica el agregado de hidrógenos y de carga a la proteína, proceso no trivial debido a variaciones en el microentorno proteico que pueden generar cambios en la protonación de ciertos residuos.

3. **Identificación del sitio de unión.** Por lo general, los algoritmos de docking evalúan a cada compuesto en un sector definido de la proteína, un sitio de unión a ligando. La identificación de dicho sitio de unión es altamente dependiente de la información que se tenga de la proteína, dado que pueden ser reconocidos por la ubicación de algún ligando en alguna estructura cristalina, por homología con proteínas similares o mediante programas de detección de sitios de unión como FPocket (34).

4. **Obtención de compuestos evaluados experimentalmente en la proteína.** De forma tal de evaluar el poder clasificatorio del algoritmo de docking, en los casos

donde sea posible, un paso común es recurrir a una base de datos como ChEMBL, PubChem y BindingDB (35; 36; 37) y descargar de allí una selección de compuestos que cuenten con información experimental de la actividad frente al blanco molecular en estudio.

5. **Preparación de los compuestos para su uso en docking molecular.** Los ligandos descargados en el paso anterior, y aquellos nuevos a evaluar, deben ser protonados acordemente, se debe decidir qué hacer con aquellos que presentan incertidumbre estructural y/o tautómeros y luego deben ser convertidos a su forma tridimensional y optimizar su estructura. Algunos programas de docking, como AutoDock Vina, requieren además que los compuestos a evaluar estén en un formato en particular.

6. **Docking molecular.** Una vez cumplidos todos los pasos anteriores, el usuario está listo para ejecutar el programa de docking y armar un modelo inicial, que luego podrá ir siendo refinado mediante la iteración con los pasos anteriores y comparado entre distintos programas y funciones de puntuación.

La creación exitosa de un modelo requiere tanto el conocimiento y la experiencia del usuario como la disponibilidad de programas confiables para cada una de las tareas en el proceso: la predicción del sitio de unión, la protonación de proteínas y ligandos, la obtención de estructuras tridimensionales y su minimización, entre otros. Aquí es donde se hace evidente la diferencia entre los programas de uso gratuito y aquellos de pago, ya que estos últimos suelen ofrecer soluciones integradas y metodologías validadas para cada

paso. Toda esta problemática es aún más relevante en el caso de usar docking inverso, puesto que involucra el modelado de un conjunto amplio de proteínas, escalando todas estas dificultades.

Por otro lado, si bien los avances en los algoritmos de docking han generado que los actuales programas se desempeñen razonablemente bien en la búsqueda de modo de unión, la incorporación de sus funciones de *score* trae consigo ciertas dificultades (38). En primer lugar, existe una variación inevitable entre los sitios de unión de los distintos blancos moleculares a modelar, provocando que la distribución de puntajes obtenidas por la función de puntuación del programa de docking sean muy distintas entre blancos moleculares. En consecuencia, se debe tener especial consideración al comparar las puntuaciones entre los distintos blancos moleculares, dado que un muy buen valor para uno puede ser muy malo para otro blanco. A modo de ejemplo: supongamos que la molécula **1** al evaluarse en el blanco molecular **A** tiene como resultado una puntuación de interacción de -9.0 kcal/mol. Al ser evaluada sobre el blanco molecular **B** se obtiene una puntuación de -11.0 kcal/mol. En principio, uno podría pensar que **1** es más probable que interactúe sobre **B** que sobre **A**, pero si uno mira el valor respecto de la distribución de compuestos ya evaluados previamente sobre ambos blancos, podría encontrar que dentro del rango de valores manejados para **A**, **1** se encuentra en el 1 % mejor y, al mismo tiempo estaría en la mitad en el rango de **B**. Esta situación es de vital importancia al evaluar las distintas plataformas para su uso, e inclusive es independiente del uso de docking molecular: cualquiera sea la función de puntuación utilizada para clasificar, se debe tener en cuenta el comportamiento de esa función dentro de cada sistema.

1.5.2. La descripción de la interacción entre el compuesto y la proteína

Otra dificultad que trae consigo el uso de docking molecular es que las funciones de puntuación provistas son de índole general. Para cada sistema propuesto, el sitio activo de la proteína será distinto y también las propiedades fisicoquímicas de los ligandos evaluados experimentalmente. Por lo tanto, una misma función de puntuación puede funcionar mejor o peor dependiendo del caso e inevitablemente tendrá un desempeño distinto dependiendo del sistema en estudio. Además, otra consideración a tener en cuenta y que se desprende del punto anterior, es que por lo general las puntuaciones calculadas, y que en general se asocian a energías de unión, frecuentemente no tienen correlación con la actividad del compuesto, por lo que tampoco son una buena métrica para saber si un compuesto puede ser activo en un determinado blanco molecular o no.

Las limitaciones en las funciones de puntuación de los programas de docking hacen que, si bien el algoritmo llega a un modo de unión preferentemente confiable, el programa por sí solo no sea suficiente para clasificar entre compuestos activos e inactivos. Por este motivo, la búsqueda de funciones de filtrado posteriores al docking que permitan mejorar la clasificación entre compuestos activos e inactivos es un campo de mucho interés en la actualidad.

Una de las estrategias para este fin es la incorporación de huellas de interacción proteína-ligando (IFPs por sus siglas en inglés). Partiendo de la base de que un compuesto muy activo interactúa de manera distinta al de un compuesto menos activo, se puede

presumir que, si se consideran ambos conjuntos completos, pueden existir patrones de interacción entre los compuestos más activos que no estén presentes en los demás y sean lo que marcan la diferencia de actividad entre los compuestos. Los IFPs son definidos en base a los átomos interactuantes entre la proteína y el ligando en el complejo que ambos forman. Los mismos son almacenados en vectores unidimensionales, que pueden ser del tipo booleano (1 o 0), entero o decimal, proveyendo una representación concisa e informativa de los patrones de interacción presentes en el complejo en estudio.

En la actualidad, existen diversas formas de obtener este tipo de representaciones, descritas en la tabla 1.2. Las mismas varían en su manera de describir interacciones y en la longitud del vector resultante, que puede contener más o menos información.

El uso de IFPs para *virtual screening* en la actualidad tiene como una de sus problemáticas la insuficiente incorporación de términos energéticos a la descripción de las interacciones en un complejo proteína-ligando. Como se observa en la tabla anterior, por lo general los distintos enfoques se centran en identificar y describir las interacciones presentes, pero sin realizar ningún cálculo de energías, que potencialmente podría proveer más información. Otra problemática presente es que existen pocas herramienta de uso libre que permitan llevar estos enfoques a la práctica de forma escalable (sin necesidad de depender de un servidor web). Algunas de ellas son LUNA (47), ProLIF (48), IChem (49) y ODDT (50).

Tabla 1.2: Diferentes enfoques de huellas de interacción proteína-ligando (IFPs).

Nombre	**Característica**	**Longitud**
Structural IFP (39; 40)	Utiliza tipos de interacción bien definidos, como enlaces de hidrógeno, enlaces de halógeno y acoplamiento π-π	Cada residuo se representa con una cadena de bits de siete posiciones
PyPLIF-HIPPOS (41)	Utiliza tipos de interacción bien definidos, como enlaces de hidrógeno, enlaces de halógeno y acoplamiento π-π	Siete bits representan siete interacciones diferentes por cada residuo
Triplet IFP (42)	Utiliza dos átomos interactuantes y un pseudoátomo de interacción posicionado en tres ubicaciones potenciales	210 enteros
APIF (43)	Considera las posiciones relativas de los pares de átomos en lugar de las ubicaciones absolutas de las interacciones	294 bits
SILIRID (44)	Agrupa las interacciones por tipo de residuo	168 enteros
SPLIF (45)	Codifica fragmentos interactuantes de proteínas y ligandos representándolos como huellas digitales circulares	La longitud depende del número de fragmentos interactuantes identificados
PLECFP (46)	Empareja y transforma los entornos ECFP del ligando y los átomos de la proteína interactuantes	La huella digital plegada en bruto consiste en enteros entre 0 y 2^{32}

1.5.3. La incorporación de la inteligencia artificial

En los últimos años, la inteligencia artificial ha experimentado un crecimiento explosivo. De manera sencilla, la inteligencia artificial (IA) puede entenderse como la capacidad de una máquina para realizar tareas que emulan la inteligencia humana. Esto incluye desde reconocer patrones y tomar decisiones hasta aprender de la experiencia y resolver problemas complejos. Su avance ha sido impulsado por la enorme disponibilidad de datos, el desarrollo de algoritmos más sofisticados y la creación de software que permite aplicar estos modelos de manera masiva.

Existen distintas formas de implementar la inteligencia artificial. Un enfoque tradicional consiste en programar un conjunto de reglas fijas que definen cómo debe comportarse un sistema ante determinadas situaciones. Siguiendo esta lógica, un programa de ajedrez clásico podría estar diseñado con estrategias predefinidas y una base de datos de movimientos calculados manualmente por expertos. Aunque este sistema juega bien, su conocimiento está limitado a lo que sus creadores le han indicado.

Por otro lado, el aprendizaje automático (*machine learning*) permite que la máquina aprenda por sí sola a partir de datos, sin necesidad de que se le programen reglas explícitas. En este caso, un sistema de ajedrez basado en aprendizaje automático analizaría millones de partidas, identificaría patrones en los movimientos y optimizaría su estrategia con el tiempo, sin intervención directa de un programador que le indique qué hacer en cada situación.

Esta capacidad de encontrar y diferenciar patrones complejos, muchas veces imperceptibles para el ojo humano, hace que los algoritmos de inteligencia artificial sean una herramienta interesante para explorar los perfiles de IFPs. En el contexto de la predicción de actividad biológica, se espera que los compuestos activos frente a una proteína presentan un patrón de interacción característico que los distinga de aquellos que no lo son. Un algoritmo de IA puede aprovechar estas diferencias para clasificar nuevos compuestos de actividad desconocida, aprendiendo de datos previos y generalizando a nuevas moléculas. Este tipo de enfoques ha sido estudiado en el pasado (51; 52) y discutida su aplicación en metodologías de predicción de blancos moleculares (53), concluyendo que, si bien es una estrategia viable, su implementación requiere precaución en el manejo de

decoys (moléculas que potencialmente no interactúan con el blanco molecular) y en la escalabilidad del proceso a múltiples blancos, ya que cada modelo demanda un tiempo considerable de desarrollo.

Por otro lado, una de las características de los enfoques basados en IA es que los modelos obtenidos pueden ser específicos para cada blanco molecular, a diferencia de las clásicas funciones de puntuación de los programas de *docking*, que suelen ser más generales y no están tan afinadas a las interacciones particulares entre ligandos y una proteína en particular.

1.6 Desarrollo de una metodología de predicción de blancos moleculares para productos naturales

Del análisis anterior se desprende que las limitaciones de las herramientas actuales, junto con la complejidad estructural de los productos naturales, dificultan la incorporación de metodologías de predicción de blancos moleculares en proyectos de investigación en este campo. En respuesta a este desafío, la presente Tesis se enfoca en desarrollar una metodología de predicción de blancos moleculares, denominada *Target Fisher*, diseñada específicamente para ser aplicable a este tipo de compuestos. Este enfoque, basado en un consenso entre diversas técnicas, busca aprovechar la precisión del *docking* en la determinación de modos de unión y la capacidad de la inteligencia artificial para identificar patrones en las interacciones ligando-blanco molecular. El objetivo es optimizar la predicción de actividad, evitando depender exclusivamente de la estructura química de los

compuestos previamente evaluados y, al mismo tiempo, permitir que los resultados guíen la síntesis y validación experimental de nuevos compuestos.

A lo largo de esta Tesis, en el capítulo 2 se profundizará en el desarrollo de la metodología propuesta y en los desafíos inherentes a la creación de herramientas de predicción de blancos moleculares, haciendo especial énfasis en la importancia de desarrollar un software íntegro, de código abierto, y en las dificultades que implica entrenar modelos de inteligencia artificial en entornos con muy poca cantidad de datos. Además, en los capítulos subsiguientes se presentará el uso de esta metodología a través de diversos casos de estudio que no sólo validaron su aplicabilidad, sino que también permitieron modelar detalladamente las interacciones entre compuestos y blancos moleculares mediante simulaciones de dinámica molecular. Estos estudios facilitaron el diseño y la síntesis de nuevos compuestos inspirados en productos naturales, demostrando el potencial de esta herramienta para identificar nuevas bioactividades y su aplicación en la química de diversificación estructural de productos naturales.

Bibliografía

[1] Metwaly, A. M.; Ghoneim, M. M.; Eissa, I. H.; Elsehemy, I. A.; Mostafa, A. E.; Hegazy, M.-E. F.; Afifi, W. M.; Dou, D. Traditional Ancient Egyptian Medicine: A Review. *Saudi Journal of Biological Sciences* **2021**, *28*, 5823–5832.

[2] Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

[3] Dewick, P. M. *Medicinal Natural Products*; Wiley, 2009.

[4] Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products* **2020**, *83*, 770–803.

[5] Sorokina, M.; Merget, B.; Méndez-Lucio, O.; Sturm, N.; Moritz, F.; Valleriani, A.; Steinbeck, C. COCONUT 2.0: A Comprehensive Overhaul and Curation of the Collection of Open Natural Products Database. *Nucleic Acids Research* **2024**, *53*, D634–D642.

[6] Poynton, E. F.; van Santen, J. A.; Pin, M.; Contreras, M. M.; McMann, E.; Parra, J.; Showalter, B.; Zaroubi, L.; Duncan, K. R.; Linington, R. G. The Natural Products Atlas 3.0: Extending the Database of Microbially Derived Natural Products. *Nucleic Acids Research* **2025**, *53*, D691–D699.

[7] Ntie-Kang, F.; Zofou, D.; Babiaka, S. B.; Meudom, R.; Scharfe, M.; Lifongo, L. L.; Mbah, J. A.; Mbaze, L. M.; Sippl, W.; Efange, S. M. N. AfroDb: A Select Highly Potent and Diverse Natural Product Library from African Medicinal Plants. *PLOS ONE* **2013**, *8*, e78085.

[8] Pilon, A. C.; Valli, M.; Dametto, A. C.; Pinto, M. E. F.; Freire, R. T.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S. NuBBEDB: An Updated Database to Uncover Chemical and Biological Information from Brazilian Biodiversity. *Scientific Reports* **2017**, *7*, 7215.

[9] Fernández, J. A.; Mori, M.; Goossens, H.; Nicolás, G.; Carbognani, F.; Lamberti, M. J.; Guzmán, P.; Cravero, F.; Pagani, M. R.; Prieto, M.; others NaturAr: Chemistry, Bioactivity, and Toxicity of Natural Products from Argentina. *Journal of Chemical Information and Modeling* **2025**, *64*, 2575–2581.

[10] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced Drug Delivery Reviews* **1997**, *23*, 3–25.

[11] Wetzel, S.; Schuffenhauer, A.; Roggo, S.; Ertl, P.; Waldmann, H. Cheminformatic Analysis of Natural Products and Their Chemical Space. *CHIMIA International Journal for Chemistry* **2007**, *61*, 355–355.

[12] Grygorenko, O. O.; Volochnyuk, D. M.; Ryabukhin, S. V.; Judd, D. B. Enamine Ltd: The Science and Business of Organic Chemistry and Beyond. *European Journal of Organic Chemistry* **2021**, *2020*, 6401–6422.

[13] Chen, Y.; Garcia de Lomana, M.; Friedrich, N.-O.; Kirchmair, J. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products. *Journal of Chemical Information and Modeling* **2018**, *58*, 1518–1532.

[14] Over, B. *et al.* Occurrence of "Natural Selectionïn Successful Small Molecule Drug Discovery. *Journal of Medicinal Chemistry* **2024**, *67*, 6926–6948.

[15] Grigalunas, M.; Brakmann, S.; Waldmann, H. Chemical Evolution of Natural Product Structure. *Journal of the American Chemical Society* **2022**, *144*, 3314–3329.

[16] Zeng, X. *et al.* NPASS Database Update 2023: Quantitative Natural Product Activity and Species Source Database for Biomedical Research. *Nucleic Acids Research* **2023**, *51*, D621–D629.

[17] Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nature Biotechnology* **2007**, *25*, 197–206.

[18] Daina, A.; Michielin, O.; Zoete, V. SwissTargetPrediction: Updated Data and New Features for Efficient Prediction of Protein Targets of Small Molecules. *Nucleic Acids Research* **2019**, *47*, W357–W364.

[19] Awale, M.; Reymond, J.-L. Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning. *Journal of Chemical Information and Modeling* **2019**, *59*, 10–17.

[20] Yao, Z.-J.; Dong, J.; Che, Y.-J.; Zhu, M.-F.; Wen, M.; Wang, N.-N.; Wang, S.;

Lu, A.-P.; Cao, D.-S. TargetNet: A Web Service for Predicting Potential Drug–Target Interaction Profiling via Multi-Target SAR Models. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 413–424.

[21] Nickel, J.; Gohlke, B.-O.; Erehman, J.; Banerjee, P.; Rong, W. W.; Goede, A.; Dunkel, M.; Preissner, R. SuperPred: Update on Drug Classification and Target Prediction. *Nucleic Acids Research* **2014**, *42*, W26–W31.

[22] Wang, L.; Ma, C.; Wipf, P.; Liu, H.; Su, W.; Xie, X.-Q. TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database. *The AAPS Journal* **2013**, *15*, 395–406.

[23] Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **1894**, *27*, 2985–2993.

[24] Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences* **1958**, *44*, 98–104.

[25] Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling* **2021**, *61*, 3891–3898.

[26] Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **2009**, *30*, 2785–2791.

[27] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47*, 1739–1749.

[28] Balius, T. E. *et al.* DOCK 6: Incorporating Hierarchical Traversal through Precomputed Ligand Conformations to Enable Large-Scale Docking. *Journal of Computational Chemistry* **2024**, *45*, 288–299.

[29] McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D. R. GNINA 1.0: Molecular Docking with Deep Learning. *Journal of Cheminformatics* **2021**, *13*, 43.

[30] Quiroga, R.; Villarreal, M. A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLOS ONE* **2016**, *11*, e0155183.

[31] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein–Ligand Docking Using GOLD. *Proteins: Structure, Function, and Bioinformatics* **2003**, *52*, 609–623.

[32] Ahdritz, G. *et al.* OpenFold: Retraining AlphaFold2 Yields New Insights into Its Learning Mechanisms and Capacity for Generalization. *Nature Methods* **2024**, *21*, 1514–1524.

[33] The Nobel Prize in Chemistry 2024. 2024; https://www.nobelprize.org/prizes/chemistry/2024/press-release/.

[34] Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, *10*, 168.

[35] Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Research* **2012**, *40*, D1100–D1107.

[36] Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 Update. *Nucleic Acids Research* **2023**, *51*, D1373–D1380.

[37] Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Research* **2016**, *44*, D1045–D1053.

[38] Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry* **2006**, *49*, 5912–5931.

[39] Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional ProteinLigand Binding Interactions. *Journal of Medicinal Chemistry* **2004**, *47*, 337–344.

[40] Vass, M.; Kooistra, A. J.; Ritschel, T.; Leurs, R.; de Esch, I. J. P.; de Graaf, C. Molecular Interaction Fingerprint Approaches for GPCR Drug Discovery. *Current Opinion in Pharmacology* **2016**, *30*, 59–68.

[41] Istyastono, E. P.; Radifar, M.; Yuniarti, N.; Prasasty, V. D.; Mungkasi, S. PyPLIF HIPPOS: A Molecular Interaction Fingerprinting Tool for Docking Results of AutoDock Vina and PLANTS. *Journal of Chemical Information and Modeling* **2020**, *60*, 3697–3702.

[42] Da, C.; Kireev, D. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling* **2014**, *54*, 2552–2561.

[43] Perez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening. *Journal of Chemical Information and Modeling* **2009**, *49*, 1245–1260.

[44] Chupakhin, V.; Marcou, G.; Baskin, I.; Varnek, A. Simple Ligand–Receptor Interaction Descriptor (SILIRID) for Alignment-Free Binding Site Comparison. *Computational and Structural Biotechnology Journal* **2014**, *10*, 33–37.

[45] Da, C.; Kireev, D. Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study. *Journal of Chemical Information and Modeling* **2014**, *54*, 2555–2561.

[46] Jafari, S. M.; Kadivar, M.; Heydari, M.; Delavari, S.; Emerson, I. A. Development of a Protein–Ligand Extended Connectivity (PLEC) Fingerprint and Its Application for Binding Affinity Predictions. *Bioinformatics* **2019**, *35*, 1334–1342.

[47] Fassio, A. V.; Shub, L.; Ponzoni, L.; McKinley, J.; O'Meara, M. J.; Ferreira, R. S.; Keiser, M. J.; de Melo Minardi, R. C. Prioritizing Virtual Screening with Interpretable Interaction Fingerprints. *Journal of Chemical Information and Modeling* **2022**, *62*, 4300–4318.

[48] Bouysset, C.; Fiorucci, S. ProLIF: A Library to Encode Molecular Interactions as Fingerprints. *Journal of Cheminformatics* **2021**, *13*, 72.

[49] Da Silva, F.; Desaphy, J.; Rognan, D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem* **2018**, *13*, 507–510.

[50] Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): A New Open-Source Player in the Drug Discovery Field. *Journal of Cheminformatics* **2015**, *7*, 26.

[51] Sato, T.; Honma, T.; Yokoyama, S. Combining Machine Learning and Pharmacophore-Based Interaction Fingerprint for in Silico Screening. *Journal of Chemical Information and Modeling* **2010**, *50*, 170–185.

[52] Yan, Y.; Wang, W.; Sun, Z.; Zhang, J. Z. H.; Ji, C. Protein–Ligand Empirical Interaction Components for Virtual Screening. *Journal of Chemical Information and Modeling* **2017**, *57*, 1793–1806.

[53] Réau, M.; Langenfeld, F.; Zagury, J.-F.; Lagarde, N.; Montes, M. The Development of Target-Specific Machine Learning Models as Scoring Functions for Docking-Based

Target Prediction. *Journal of Chemical Information and Modeling* **2018**, *58*, 2469–2477.

Capítulo 2

Desarrollo de una nueva metodología de predicción de blancos moleculares

2.1 Integrando docking inverso y clasificación por IA: Diseño de la metodología Target Fisher

El funcionamiento inicial propuesto para la metodología Target Fisher se muestra en el esquema de la figura 2.1. Al ser una metodología de predicción de blancos moleculares bajo un enfoque SBTP que utiliza docking inverso, el primer paso consiste en evaluar la molécula en cada sitio de unión utilizando un software de docking molecular. Una vez obtenido cada complejo proteína-ligando, la interacción entre ambos será descrita mediante una **huella de interacción** (IFP) que será luego evaluada en un modelo de clasificación generado con inteligencia artificial específicamente para cada blanco molecular, obteniendo para cada par molécula/blanco una **predicción** de actividad.

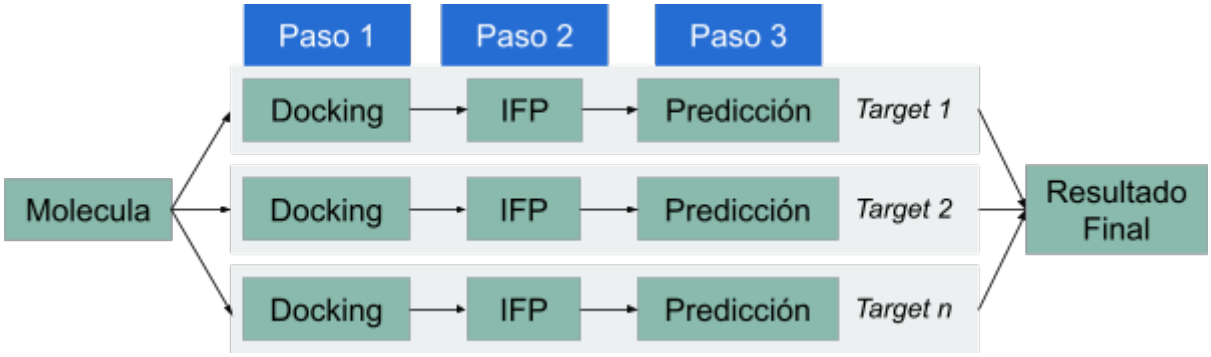


Figura 2.1: Representación esquemática del procedimiento empleado por Target Fisher para la identificación de blancos moleculares potenciales a partir de una molécula de interés.

2.1.1. La Caseína Quinasa II como prueba piloto

A modo de prueba inicial, se buscó una plataforma adecuada de docking molecular que permita obtener una huella de interacción de forma sencilla, sin necesidad de tener que integrar distintas herramientas. Dentro de aquellas herramientas a las que se puede

acceder de forma gratuita, se destaca DOCK6 (1). Este software, pionero en el desarrollo de algoritmos de docking, tiene la ventaja de brindar al usuario la posibilidad de, una vez *dockeado* el compuesto, describir su interacción mediante distintas funciones de puntuación, entre las que se incluye el "Footprint score"(2; 3). Esta opción permite calcular puentes de hidrógeno intermoleculares, en conjunto con energías de interacción básicas (van der Waals y electrostáticas) entre el ligando en estudio y cada residuo de la proteína, permitiendo su comparación posterior con un compuesto de referencia.

El blanco molecular seleccionado para la prueba inicial fue la Caseína Quinasa II (CK2), una quinasa del tipo serina/treonina altamente conservada. Esta enzima posee más de 300 sustratos conocidos (4). La misma presenta una estructura de heterotetrámero con dos subunidades catalíticas (α y α') y dos subunidades regulatorias β idénticas. Esta proteína tiene roles muy importantes en el crecimiento, supervivencia y proliferación celular, y se ha comprobado que su expresión anormal está asociada a diversas patologías de distinto tipo, como infecciosas, inflamatorias y carcinogénicas (5). Como ocurre en general con las quinasas, la mayoría de los inhibidores interactúan con el sitio de unión a ATP. Esto puede verse reflejado en varias de las estructuras cristalinas reportadas en el *Protein Data Bank* (PDB) para esta proteína, tales como 3BQC, 2RKP, 1PJK, 5B0X y 3H30, entre otras, que muestran distintos tipo de inhibidores de este sitio activo.

Para el armado del sistema de docking, como punto de partida se seleccionó la estructura 3H30, puesto que consta de una buena resolución (1.56 Å) y no presenta mutaciones. Los compuestos experimentalmente ensayados fueron obtenidos de ChEMBL,(6) reuniéndose así un total de 1167 moléculas. Los detalles tanto de la preparación del blanco

molecular y de los ligandos, como también de los parámetros utilizados para el docking, se encuentran descritos en el capítulo 6.

Una vez obtenidos todos los complejos proteína/ligando, los mismos fueron procesados a modo tal de obtener una huella de interacción para cada complejo. Esta huella consiste en un vector donde cada componente se corresponde a la suma de las energías de interacción de van de Waals y electrostática entre la pose de mejor puntuación del ligando y un residuo en particular. El conjunto de todas las huellas de interacción se convirtió así en el conjunto de datos utilizado para la creación de los modelos de aprendizaje automático. Este conjunto se encuentra ejemplificado en la Tabla 2.1, la cual contiene como columnas a cada uno de los N residuos de la proteína y como filas los valores de interacción de cada uno de los M ligandos.

Ligando	**Residuo 1**	**Residuo 2**	**...**	**Residuo N**	**Actividad**
Compuesto 1	$Energía_{L1R1}$	$Energía_{L1R2}$	...	$Energía_{L1RN}$	1
Compuesto 2	$Energía_{L2R1}$	$Energía_{L2R2}$	...	$Energía_{L2RN}$	1
Compuesto 3	$Energía_{L3R1}$	$Energía_{L2R2}$	...	$Energía_{L3RN}$	0
...	...	...	...	...	...
Compuesto M	$Energía_{LMR1}$	$Energía_{LMR2}$	...	$Energía_{LMRN}$	0

Tabla 2.1: Conjunto de datos con las huellas de interacción de los M ligandos y sus energías por residuo. Cada fila representa los valores de interacción de un ligando con cada uno de los N residuos de la proteína. La columna Actividad presenta el valor 1 si es activo, 0 si es inactivo.

Como se mencionó en la introducción, el aprendizaje automático se basa en la construcción de modelos, que son representaciones matemáticas entrenadas con grandes volúmenes de datos para identificar relaciones y hacer predicciones o tomar decisiones. Una vez recolectados los datos que van a utilizarse para su entrenamiento, el armado del modelo consta de las siguientes etapas:

1. **Preprocesamiento y limpieza de los datos**: Una vez que los datos han sido recolectados, es crucial prepararlos adecuadamente. Esto implica la eliminación de valores atípicos, el manejo de datos faltantes, la normalización o estandarización de las variables y, en algunos casos, la transformación de los datos en un formato adecuado para el modelo. Este paso es fundamental porque la calidad de los datos influye directamente en la capacidad del modelo para aprender patrones útiles.

2. **Selección del modelo**: Dependiendo del tipo de problema que se quiera resolver (clasificación, regresión, clustering, etc.), se elige una arquitectura de modelo adecuada. Por ejemplo, para problemas de clasificación se pueden utilizar algoritmos como máquinas de vectores de soporte (SVM) o redes neuronales, mientras que para problemas de regresión se suelen emplear modelos como regresión lineal o árboles de decisión.

3. **División del conjunto de datos**: Para evaluar correctamente el desempeño del modelo, los datos se dividen en tres conjuntos principales:

    a) **Conjunto de entrenamiento**: Se utiliza para ajustar los parámetros internos del modelo y enseñarle los patrones de los datos.

    b) **Conjunto de validación**: Se emplea para evaluar el modelo durante el entrenamiento y ajustar hiperparámetros sin que el modelo vea estos datos en el entrenamiento directo.

    c) **Conjunto externo de prueba**: Se usa para medir el rendimiento final del modelo en datos completamente nuevos y no influye en el entrenamiento.

4. **Entrenamiento del modelo**: En esta etapa, el modelo ajusta sus parámetros internos iterativamente mediante un proceso de optimización. Los algoritmos de aprendizaje supervisado, por ejemplo, buscan minimizar una función de error que mide la diferencia entre las predicciones del modelo y los valores reales. Durante este proceso, el modelo aprende a reconocer patrones en los datos de entrenamiento.

5. **Ajuste de hiperparámetros**: Los hiperparámetros son configuraciones externas al modelo que afectan su aprendizaje, como la tasa de aprendizaje en redes neuronales o los valores de regularización en modelos de regresión. Para optimizarlos, se emplean técnicas como **GridSearchCV**, que permiten explorar distintas combinaciones de hiperparámetros mediante validación cruzada (concepto que se explicará más adelante), buscando el mejor equilibrio entre precisión y generalización.

6. **Evaluación del modelo**: Una vez entrenado el modelo con los mejores hiperparámetros, se evalúa su rendimiento en el conjunto de prueba mediante métricas como la precisión, la exactitud, el área bajo la curva ROC y la puntuación F1. Estas métricas permiten verificar si el modelo es capaz de generalizar correctamente y no se ha sobreajustado a los datos de entrenamiento.

El entrenamiento de este primer modelo, así como todos los subsiguientes, fue realizado mediante la utilización de la librería scikit-learn (7), de amplio uso en el campo y que permite el acceso a una gran variedad de algoritmos para distintas tareas. En este caso, el primer paso consistió en dividir al conjunto de interacciones proteína-ligando obtenido en un conjunto de entrenamiento (70 %) y un conjunto de prueba o testeo externo(30 %).

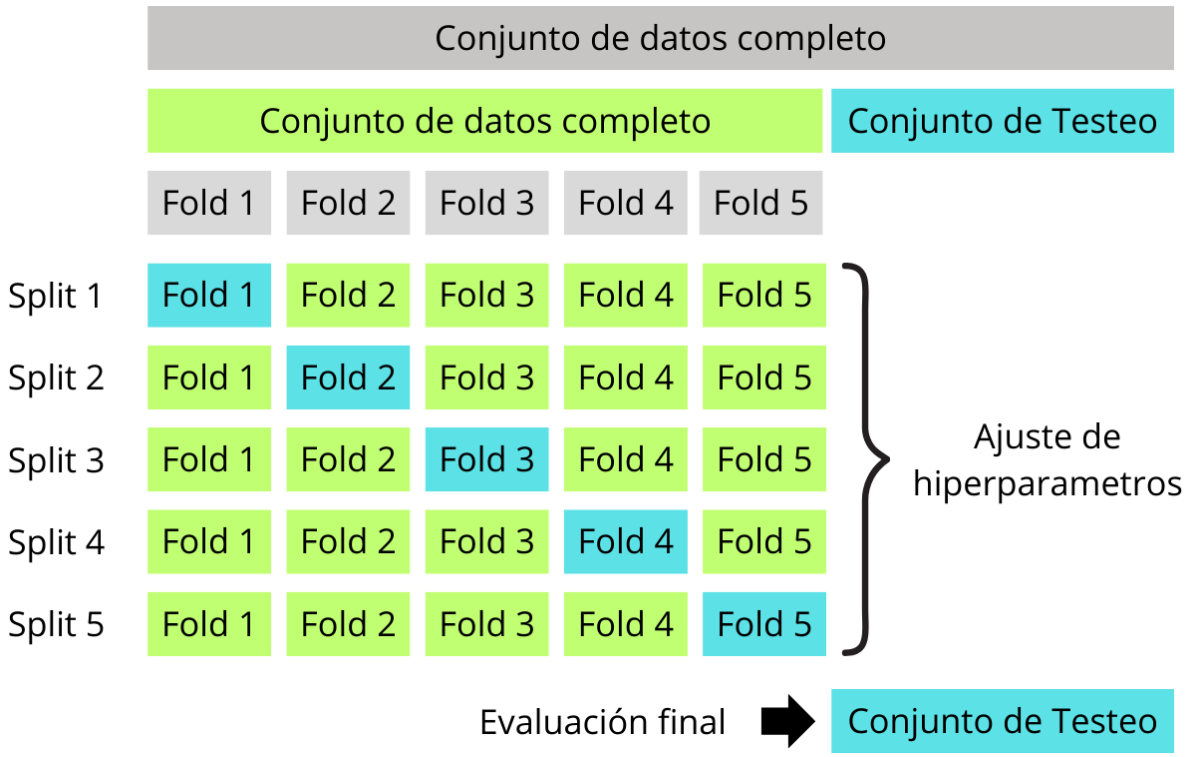


Figura 2.2: Ilustración del concepto de validación cruzada. El conjunto de datos se divide en K subconjuntos (*folds*), y en cada iteración, $K-1$ subconjuntos se utilizan para entrenar el modelo y uno para validarlo. Adaptado de https://scikit-learn.org/stable/modules/cross_validation.html.

No se generó un conjunto de validación separado, ya que se implementó una validación cruzada para la búsqueda de hiperparámetros.

La validación cruzada es una técnica que mejora la estimación del desempeño del modelo al dividir el conjunto de entrenamiento en múltiples subconjuntos o **folds** (en este caso, 5). En cada iteración del proceso de entrenamiento, uno de estos subconjuntos se reserva temporalmente como conjunto de validación, mientras que los restantes se usan para entrenar el modelo. Luego, el proceso se repite hasta que cada subconjunto haya sido utilizado una vez como conjunto de validación. Una descripción gráfica de cómo funciona se puede observar en la figura 2.2.

De esta manera, esta estrategia permite evaluar el rendimiento del modelo en distintos subconjuntos de los datos, reduciendo la dependencia de una única partición de entre-

namiento y ayudando a optimizar los hiperparámetros de manera más robusta. Además, minimiza el riesgo de sobreajuste y proporciona una estimación más precisa del desempeño del modelo en datos no vistos.

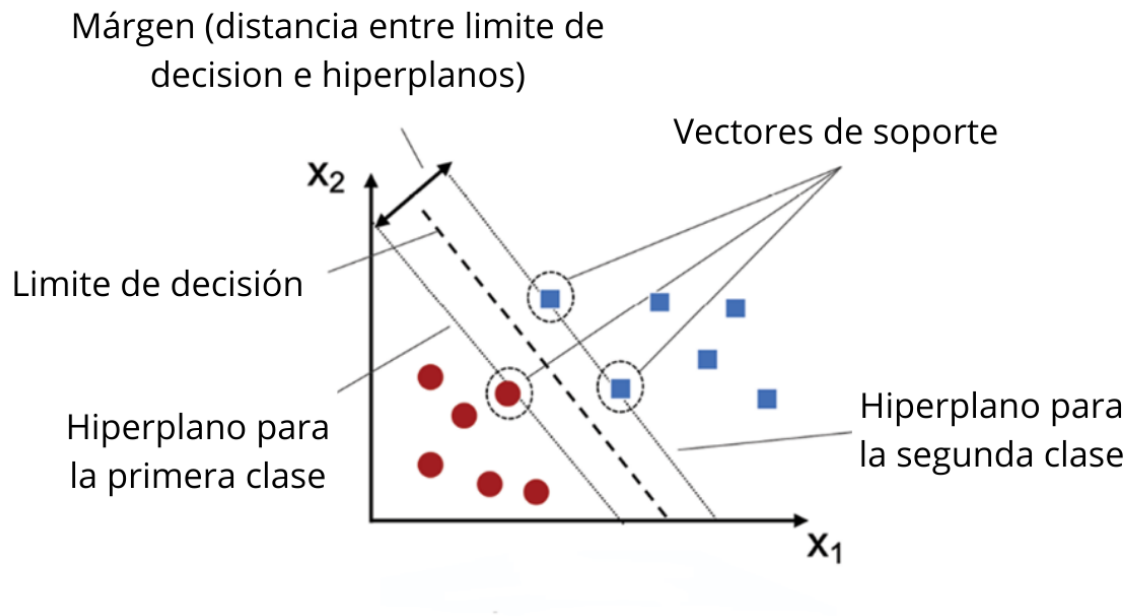


Figura 2.3: Representación gráfica del funcionamiento de las máquinas de vectores de soporte (SVM). El algoritmo busca el hiperplano que maximiza el margen entre las dos clases de datos.

Una vez procesado y dividido el conjunto de datos inicial, el siguiente paso consiste en definir la arquitectura del modelo. En este caso, como algoritmo de clasificación inicialmente se utilizaron las máquinas de vectores de soporte (**SVMs**, por sus siglas en inglés). Las SVM buscan encontrar el **hiperplano** que mejor separa las diferentes clases de datos en un espacio de características, maximizando el margen entre las clases, como se observa en la figura 2.3. Las mismas tienen como característica el ser adecuadas para conjuntos con un número de datos pequeño como el propuesto aquí, y además, su uso para este tipo de estrategias ha sido reportado previamente (8; 9; 10). Una dificultad que trae este algoritmo es que en aquellos conjuntos de datos que presentan desbalances entre clases,

el algoritmo tiende a sesgar hacia la clase más abundante. Por este motivo, se utilizó una línea de corte de IC_{50} de 100 nM para diferenciar 501 compuestos “activos"de otros 666 con una actividad menor, con IC50 por encima de dicho valor, quedando de esta forma los datos balanceados.

Uno de los cuidados más importantes que se debe tener al entrenar un modelo de inteligencia artificial es prestar especial cuidado al sobreentrenamiento. Este proceso ocurre cuando los hiperparámetros del modelo son ajustados de forma tal que el modelo pierde generalización y solo puede predecir correctamente datos muy similares a aquellos con los que fue entrenado, como se ilustra en la figura 2.4.

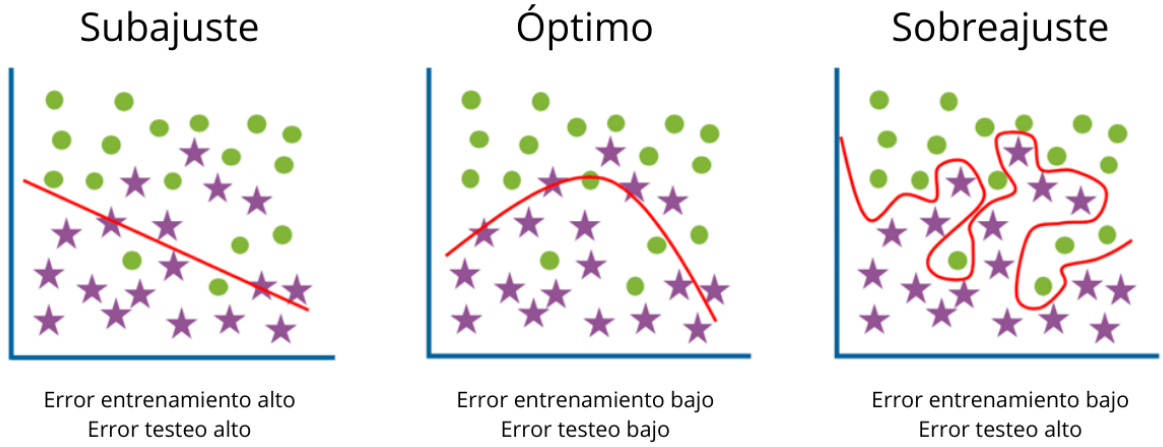


Figura 2.4: Ilustración del sobreajuste (overfitting) en modelos de aprendizaje automático. A la izquierda, un modelo bien balanceado; a la derecha, un modelo sobreajustado que se adapta excesivamente a los datos de entrenamiento.

Para evitar esto, se utilizan diversas técnicas que promueven la generalización y mejoran el rendimiento del modelo en datos no evaluados previamente por el mismo. Una estrategia, por ejemplo, puede ser modificar los hiperparámetros de forma aleatoria y ver qué ocurre con el rendimiento del modelo. En este caso, se empleó una búsqueda de hiperparámetros en forma de grilla, utilizando la herramienta *GridSearchCV*, disponible

en scikit-learn. La misma realiza una búsqueda exhaustiva probando todas las combinaciones posibles de un rango predefinido de valores para los hiperparámetros del modelo. Para evaluar cada combinación, utiliza validación cruzada, es decir, divide los datos de entrenamiento en varios subconjuntos y entrena el modelo repetidamente en diferentes particiones. Luego, selecciona la combinación de hiperparámetros que logra el mejor rendimiento promedio en estos subconjuntos. De esta manera, se evalúa todo el conjunto de hiperparametros obtenidos en el rango indicado, y se pueden analizar los resultados para evaluar si se está sobreentrando el modelo o no.

En el caso del modelo en estudio, se evaluaron los hiperparámetros **C** y **gamma**, propios de las **máquinas de vectores de soporte (SVM)**. Una representación esquemática del efecto de estos hiperparámetros semuestra en la figura 2.5. El hiperparámetro **C** controla la penalización por errores en la clasificación: valores altos hacen que el modelo busque clasificar correctamente cada punto de entrenamiento, lo que puede llevar al sobreajuste, mientras que valores bajos permiten mayor flexibilidad pero pueden generar subajuste. Por otro lado, el hiperparámetro **gamma** influye en la forma en que los puntos de datos contribuyen a la clasificación: valores altos hacen que cada punto tenga un radio de influencia más pequeño (modelo más complejo y propenso al sobreajuste), mientras que valores bajos generan una separación más suave entre clases.

Para evaluar el desempeño del modelo con distintos valores de estos hiperparámetros, se utilizó **GridSearchCV**, generando diferentes conjuntos de entrenamiento y testeo. Se midió el rendimiento mediante varias métricas, entre ellas:

- **Área bajo la curva ROC (AUC-ROC):** mide la capacidad del modelo para dis-

tinguir entre clases en distintos umbrales de clasificación. Una curva ROC (Receiver Operating Characteristic) grafica la tasa de verdaderos positivos frente a la tasa de falsos positivos, y un AUC cercano a 1 indica una excelente separación entre clases.

- **Exactitud:** proporción de muestras clasificadas correctamente sobre el total.
- **Puntuación F1:** métrica que equilibra precisión y exhaustividad (recall), calculada como la media armónica entre ambas. Es útil cuando hay clases desbalanceadas.

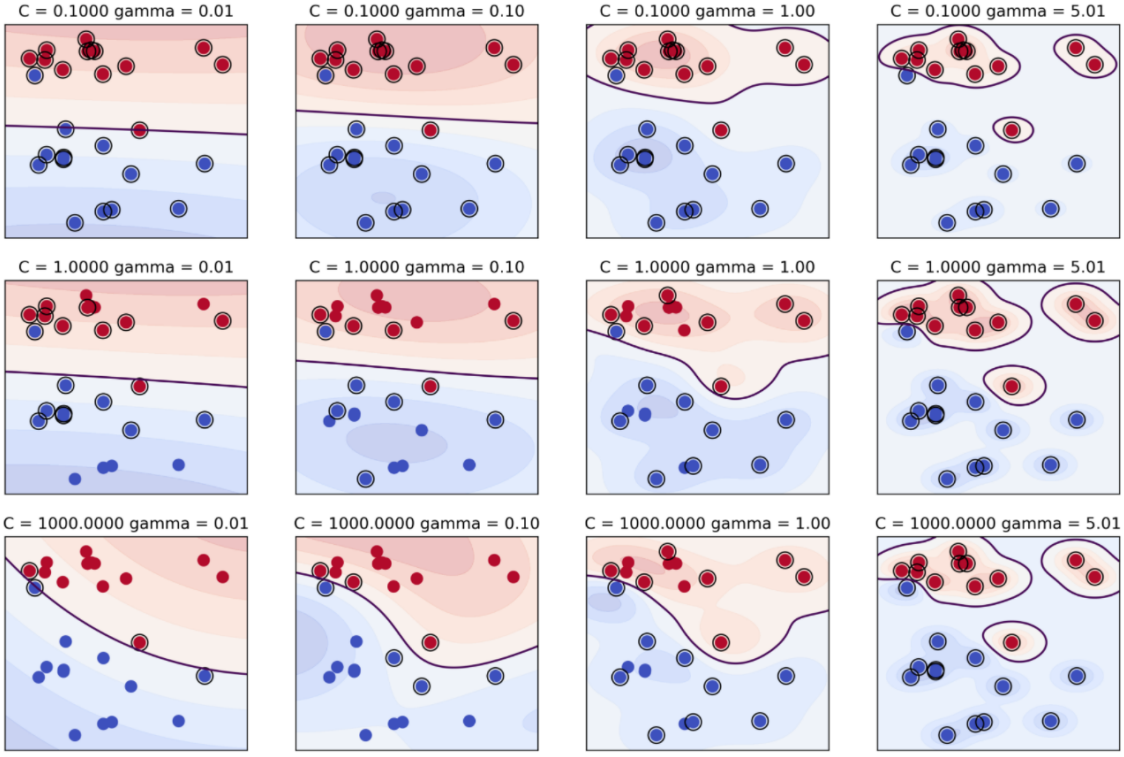


Figura 2.5: Efecto de los hiperparámetros `C` y `gamma` en el comportamiento de las SVM. A la izquierda, diferentes valores de `C`; a la derecha, diferentes valores de `gamma`. Imagen adaptada de https://amueller.github.io/aml/02-supervised-learning/07-support-vector-machines.html.

Un ejemplo de este análisis se muestra en la figura 2.6, donde para $\mathbf{C = 10}$, se exploraron valores de **gamma** entre **0 y 0.05**. La selección del modelo final debe realizarse en el conjunto de hiperparámetros donde se logre el mejor poder clasificatorio con el

menor sobreentrenamiento posible, punto marcado con una cruz verde. Un ejemplo de este análisis se ve en la figura 2.6, en donde para el valor C=10, se barrió el hiperparámetro gamma entre 0 y 0.05. De esta manera, la figura muestra el desempeño de distintos modelos entrenados con el mismo conjunto de datos pero con distintos hiperparámetros, realizando validación cruzada. Se puede observar que la exactitud y la métrica F1 alcanzan su valor máximo en simultáneo y luego decrecen al aumentar gamma. En el caso del área bajo la curva ROC, la misma tiene un máximo en la misma zona. De esta manera, bajo este valor de 0, el valor de gamma óptimo se encuentra cercano a 0.005. Valores superiores llevan al sobreentrenamiento y a la consecuente pérdida de rendimiento.

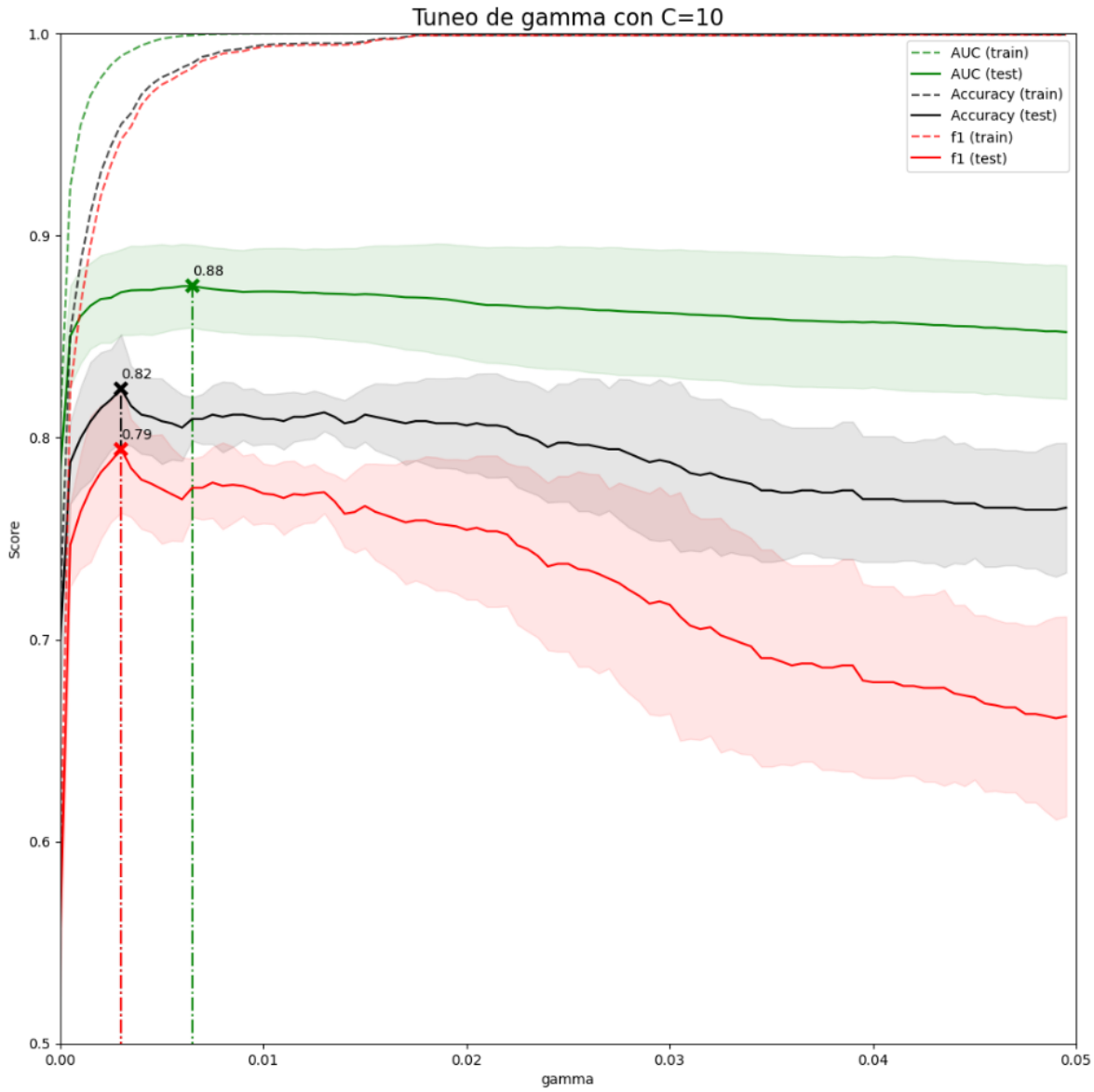


Figura 2.6: Análisis del rendimiento del modelo SVM con C=10 y diferentes valores de gamma. La cruz verde indica el punto de equilibrio óptimo entre rendimiento y generalización.

El primer objetivo planteado para esta prueba fue ver si el modelo generado tenía un poder clasificatorio significativo y cómo era su desempeño respecto a la función de puntuación del docking molecular. La métrica utilizada para esta comparación es el área bajo la curva ROC, muy utilizada en las campañas de **screening virtual**, especialmente en aquellas que buscan nuevos ligandos utilizando docking, ya que permite evaluar la

capacidad del modelo para distinguir correctamente entre compuestos activos e inactivos en un rango amplio de clasificaciones. Los resultados mostraron que el modelo generado mediante aprendizaje automático supera significativamente al docking en capacidad de clasificación, incrementando el área bajo la curva ROC de **0.696** a **0.926**, lo que evidencia una mejora sustancial en la discriminación entre compuestos activos e inactivos.

Otro de los grandes desafíos en el armado de modelos de inteligencia artificial se encuentra en la interpretabilidad de los mismos. La complejidad de los patrones encontrados por estos algoritmos suele ser tan alta que su funcionamiento se asemeja al de una "caja negra", impidiendo al usuario entender en qué se está basando el modelo para elaborar una predicción.

Una estrategia para abordar esta problemática es el cálculo de las importancias de los **features** (características o atributos), que son las variables de entrada que el modelo utiliza para tomar decisiones, en este caso el valor energético de la interacción del ligando con cada residuo. Una técnica utilizada para evaluar la relevancia de estos **features** es la **permutación de features**. En esta técnica, una métrica (en este caso **ROC AUC**) se reevalúa en el conjunto de testeo externo luego de que los valores de una columna específica de **features** son aleatoriamente alterados. Si el impacto de esta alteración es significativo en el rendimiento del modelo, la métrica medida disminuirá, lo que indicaría que ese **feature** tiene una alta importancia en el poder clasificatorio del modelo.

El resultado de esta prueba se muestra en la figura 2.7. Como se observa, casi ninguno de los residuos más relevantes se encuentra cerca del sitio activo de CK2, lo que sugiere que el modelo basa sus predicciones en interacciones biológicamente irrelevantes. Una

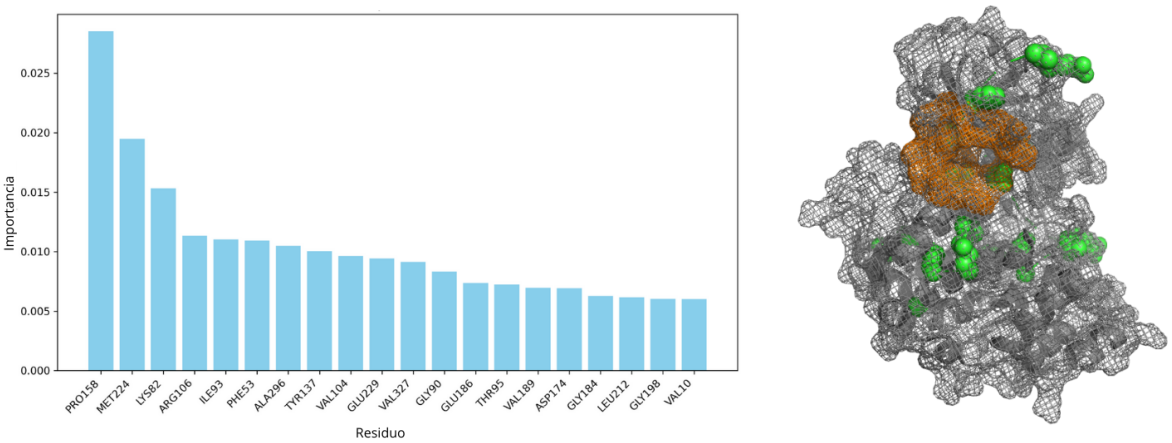


Figura 2.7: (izq.) Análisis de la importancia de los residuos en el modelo de CK2. (der.) Imagen de la proteína con los residuos considerados por el modelo como los más importantes destacados en verde, en marrón se destaca el sitio activo.

posible causa de este hecho es que el modelo fue entrenado con huellas de interacción que incluyen la energía de interacción del ligando con todos los residuos de la proteína, no solo con los del sitio activo. De esta manera, aunque el poder clasificatorio del modelo es elevado, parece estar basado en gran medida en ruido, lo que reduce su valor interpretativo desde una perspectiva biológica.

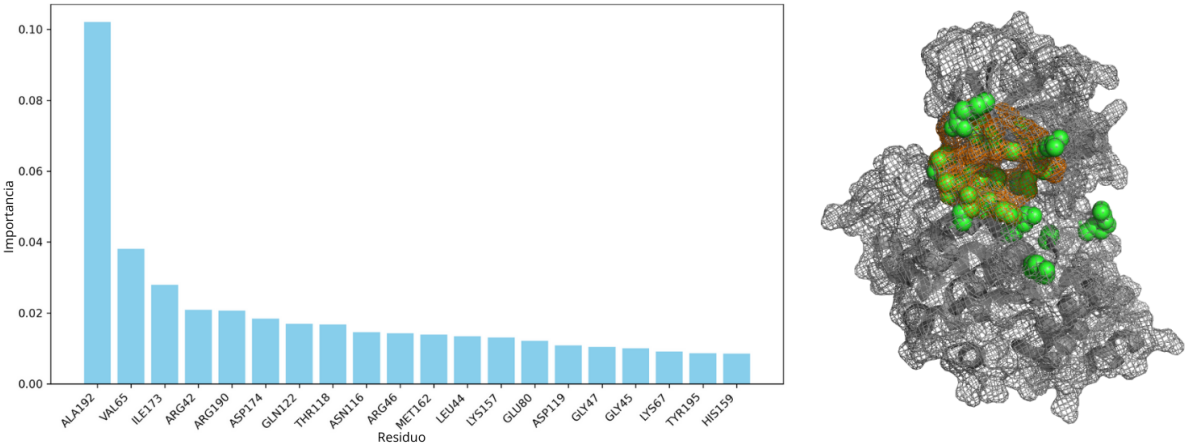


Figura 2.8: (izq.) Análisis de la importancia de los residuos en el modelo de CK2. (der.) Imagen de la proteína con los residuos considerados por el modelo como los más importantes marcados.

Una estrategia para forzar al algoritmo a centrarse en los residuos biológicamente más relevantes consistió en restringir el entrenamiento únicamente a aquellos residuos

que presentaran una interacción significativa con al menos uno de los ligandos activos. Para ello, se reentrenó el modelo aplicando un doble filtro: por un lado, se seleccionaron únicamente los residuos cuya interacción superara un umbral mínimo ajustado en ±0.5 kcal/mol, y por otro, se incluyeron únicamente aquellos que se encontraran a una distancia menor a 5 Å de cualquiera de los ligandos activos. Como resultado, el número de residuos considerados se redujo de 334 a 47, sin cambios significativos en el rendimiento del modelo, que mantuvo un valor de AUC ROC de 0.92. Esto demuestra que el algoritmo conserva su poder clasificatorio, incluso al enfocarse únicamente en los residuos que presentan interacciones significativas y proximidad espacial al sitio activo. Tal como se muestra en la figura 2.8, estos residuos se localizan, como era de esperarse, en la región del sitio de unión.

CK2	**SVM**	**SVM + FI**
Residuos	47	23
ROC AUC	0.92	0.90

Tabla 2.2: Comparación del rendimiento del modelo SVM con todos los residuos filtrados (SVM) y con solo los residuos más importantes (SVM + FI).

Por último, el cálculo de importancias permite eliminar los residuos irrelevantes y reentrenar el modelo, enfocándose exclusivamente en aquellos que tienen mayor impacto en la predicción. Este enfoque resultó en la conservación de aproximadamente la mitad de los residuos, sin una pérdida significativa en el rendimiento clasificatorio, como se muestra en la tabla 2.2. Además, en la figura 2.9, se observan las importancias relativas de los residuos más significativos, confirmando que es posible obtener un modelo con un excelente poder clasificatorio utilizando solo los residuos más importantes alrededor del

sitio activo.

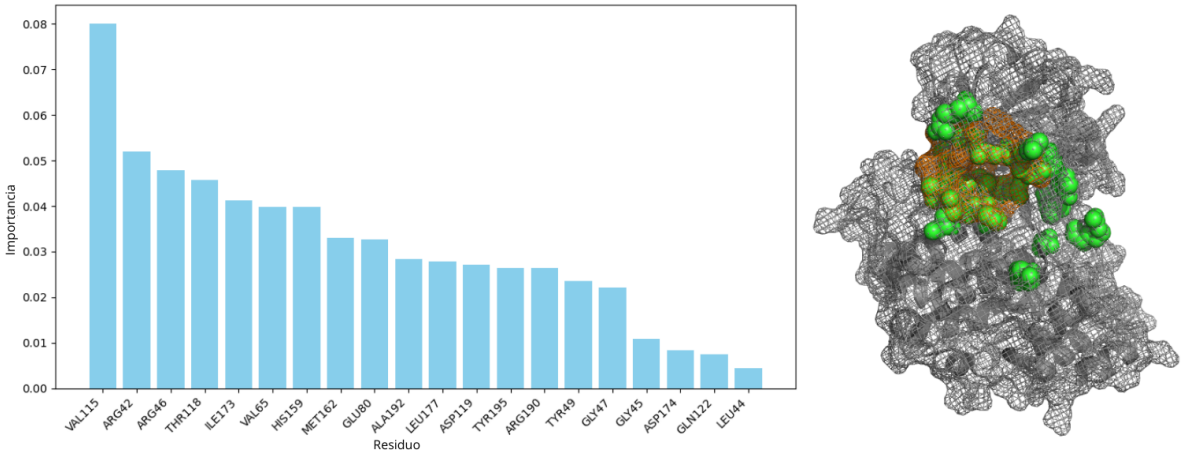


Figura 2.9: (izq.) Análisis de la importancia de los residuos en el modelo de CK2. (der.) Imagen de la proteína con los residuos considerados por el modelo como los más importantes marcados.

2.1.2. Independencia respecto del programa de docking

Con el fin de evaluar la dependencia del modelo obtenido respecto al programa de docking utilizado, se entrenaron nuevos modelos utilizando huellas de interacción obtenidas a partir de las mejores poses de docking tanto con FRED (11) como con Autodock Vina (12). Dichas huellas de interacción fueron obtenidas mediante el reprocesado de las poses con el módulo de descomposición de interacción de DOCK6.9. Los resultados se muestran en la tabla 2.3 y evidencian que la mejora en el poder clasificatorio es independiente del programa de docking utilizado.

	Autodock Vina	**Fred**	**DOCK 6.9**
Docking	0.78	0.79	0.70
Docking + SVM	**0.88**	**0.91**	**0.93**

Tabla 2.3: Comparación del rendimiento (ROC AUC) entre diferentes programas de docking y la combinación con modelos SVM.

2.1.3. Comparación con otros algoritmos de Inteligencia Artificial

Una vez establecida la metodología para la creación de modelos, se procedió a estudiar el efecto de cambiar el algoritmo de entrenamiento. Para ello se utilizó **Random Forest** y **MLPC**, también disponibles en scikit learn para entrenar nuevos modelos utilizando CK2.

El método **Random Forest** funciona combinando muchos pequeños árboles de decisión, como si fueran varios caminos posibles para tomar una decisión final. En lugar de depender de una sola respuesta, este método se basa en la opinión conjunta de muchos modelos más pequeños, lo que hace que sus predicciones sean más estables y confiables.

Por otro lado, **MLPC** es un modelo de redes neuronales, cuyo funcionamiento se inspira en el funcionamiento del cerebro. Aprende a reconocer patrones en los datos mediante varias capas de procesamiento, donde cada una transforma la información de manera progresiva hasta encontrar la mejor forma de clasificarla. Este tipo de modelo mejora con la práctica, ajustando sus cálculos en cada intento para reducir los errores y hacer mejores predicciones.

En la Tabla 2.4, se pueden ver los resultados de esta comparación. No se encontraron diferencias significativas en el desempeño de estos modelos en comparación con el método SVM utilizado previamente.

	SVM	**RF**	**MLPC**
ROC AUC	0.90	0.91	0.90
F1	0.82	0.82	0.80
Precisión	0.89	0.90	0.80
Recall	0.82	0.78	0.80

Tabla 2.4: Comparación del rendimiento entre diferentes algoritmos de aprendizaje automático para CK2.

2.2 Extensión de la metodología a un mayor número de blancos moleculares

Una vez armado el primer modelo satisfactoriamente, el siguiente paso para la confección de una metodología de predicción de blancos moleculares es expandir este mismo enfoque hacia un número mayor de proteínas. Con este fin, se tomó como referencia la publicación de Nogueira et al. mencionada previamente (8). En este trabajo, los autores evalúan la factibilidad de utilizar la información obtenida por los vectores de interacción desarrollados por el mismo grupo, llamados PADIFs, para crear modelos específicos para cada blanco molecular utilizando SVMs y redes neuronales. Para la evaluación de la metodología utilizan 20 blancos moleculares obtenidos a partir de los trabajos de Schomburg et al. y Bauer et al. (13; 14), en donde proponen distintos conjuntos de blancos para hacer una evaluación comparativa entre métodos de predicción de blancos moleculares y de docking. La lista de blancos se describe en la tabla 2.5.

A modo de realizar una comparación inicial entre la metodología propuesta y la utilizada en el mencionado trabajo, se seleccionó un subconjunto de 10 blancos moleculares de la lista anterior. Utilizando los conjuntos de activos e inactivos provistos en dicho trabajo,

Tabla 2.5: Lista de blancos moleculares utilizados como referencia para la metodología.

Target	**Uniprot ID**	**Descripción**	**PDB**
ACE	P12821	Enzima convertidora de angiotensina 1	1uze
BRAF	P15056	Proteína quinasa Serina/Treonina B-Raf	3skc
COX-2	Q05769	Ciclooxigenasa-2	1cx2
DHFR	P00374	Dihidrofolato reductasa	1s3v
FXa	P00742	Factor X (Protrombinasa)	1for
GSK3B	P49841	Glicógeno sintasa quinasa 3 beta	3i4b
HDAC2	Q92769	Histona deacetilasa 2	3max
HDAC6	Q9UBN7	Histona deacetilasa 6	5edu
HDAC8	Q9BY41	Histona deacetilasa 8	3sff
IGF1R	P08069	Receptor del factor de crecimiento similar a la insulina 1	3nw7
JAK3	P52333	Janus quinasa 3	3lxl
PDE5	O76074	Fosfodiesterasa específica de cGMP tipo 5	1xp0
PDK1	O15530	Piruvato deshidrogenasa quinasa-1	2xch
PI3Kg	P48736	Fosfoinositida-3-quinasa gamma	3dbs
PIM-2	Q9P1W9	(Protooncogén) Proteína quinasa Serina/Treonina 2	2iwi
Thrombin	P00734	Trombina	3rm2
TIE2	Q02763	Receptor de angiopoyetina-1	2oo8
TPA	P00750	Activador tisular del plasminógeno	1a5h
Trypsin	P00760	Tripsina	2g5n
VEGFR1	P17948	Receptor 1 del factor de crecimiento endotelial vascular	3hng

se armaron los modelos de docking y aprendizaje automático para cada blanco utilizando DOCK6.9. A modo de sumar simplicidad al presente estudio, y además permitir la comparación directa con el citado trabajo, se decidió entrenarlos utilizando únicamente SVMs.

Los resultados se muestran en la figura 2.10, y se observa que los modelos de aprendizaje automático generados mostraron, al igual que con CK2, un poder clasificatorio mayor al de los modelos que solo usaban docking molecular. Sin embargo, en casi todos los casos, los modelos de aprendizaje automático tuvieron un rendimiento menor a los

obtenidos en el trabajo de referencia, en el que utilizan GOLD y PADIF para la obtención de los vectores de interacción.

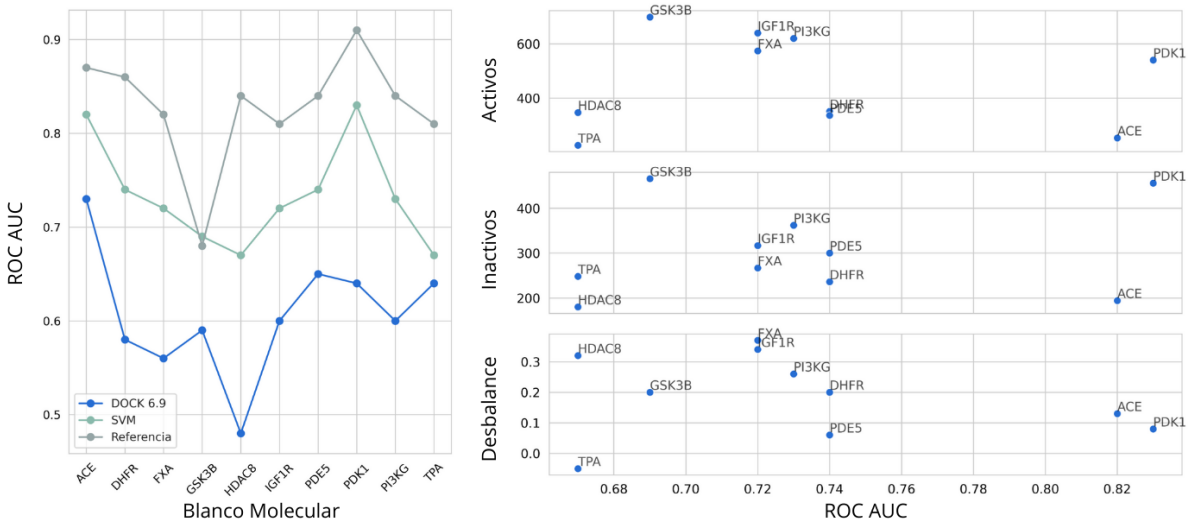


Figura 2.10: Comparación del rendimiento (ROC AUC) entre distintas metodologías de predicción de blancos moleculares para diversos blancos: (izq.) Comparación del desempeño de DOCK6.9 + SVM frente a GOLD + PADIF + SVM. (der.) Cantidad de activos e inactivos, junto con el desbalance entre clases, calculado como (inactivos - activos) / total, para los casos estudiados.

El análisis de los conjuntos de activos e inactivos utilizados para el entrenamiento muestra que no hay una clara tendencia entre el efecto de la cantidad de activos, inactivos, ni el balance entre ellos y el rendimiento final del modelo. En este punto, cabe aclarar que este análisis es parcial, producto de que la cantidad de modelos entrenados es baja y la cantidad de datos entre ellos es similar.

2.3 Cambio hacia Autodock Vina

Si bien, como se vió en la sección 2.1.2, la mejora en el poder clasificatorio parece ser independiente del programa de docking utilizado, sigue existiendo una dependencia en la calidad de la pose de mejor puntuación obtenida, puesto que es la utilizada para entrenar el modelo. Si bien el impacto en las métricas de rendimiento puede no verse reflejado, el hecho de que esta primera pose sea o no un buen reflejo de lo que ocurre biológicamente afecta la calidad de los patrones de interacción que encontrará el modelo. En el trabajo de Wang et al. (15) se realiza una comparación exhaustiva de distintos programas de docking de uso gratuito y de pago. En dicho análisis concluyen que GOLD es el programa en el cual la mejor pose obtenida es la que con mayor frecuencia es la correcta, mientras que DOCK es uno de los de peor rendimiento, hecho que podría explicar el menor rendimiento obtenido en el poder clasificatorio de la metodología actual y la de referencia.

Entre los programas gratuitos evaluados en dicho trabajo, se destaca Autodock Vina, el cual, si bien presenta un desempeño promedio en la predicción correcta de la pose mejor puntuada, sobresale en la correlación entre su valor de puntuación y el valor de afinidad obtenido experimentalmente. Autodock Vina es uno de los programas de código abierto más rápidos y utilizados. Además de su alta velocidad, ofrece ventajas adicionales, como la facilidad para preparar los blancos moleculares utilizando las mismas herramientas que Autodock. Su más reciente actualización permite su integración con programas basados en Python y el uso de otras funciones de puntuación de forma más sencilla (12). Además, al ser un software de código abierto ampliamente distribuido, continuamente se publican

nuevas herramientas para mejorar su funcionalidad y facilidad de uso. Por estos motivos, en este punto del desarrollo se decidió cambiar de motor de docking por Autodock Vina.

2.3.1. Creación de una herramienta independiente para obtener huellas de interacción basadas en mecánica molecular

El cambio de motor de docking trae como dificultad principal la inexistencia de programas o librerías que permitan el cálculo de huellas de interacción basadas en un campo de fuerzas de mecánica molecular como se realizaba en DOCK6.9. Este tipo de interacción trae como ventaja la posibilidad de evaluar energéticamente el ligando con cada residuo de la proteína, obteniéndose un valor decimal como dato. El uso de valores continuos para el entrenamiento del modelo permite luego estudiar el impacto del cambio de dicho valor en la predicción, con técnicas como los gráficos de dependencia parcial y expectativa condicional individual (16; 17), como veremos más adelante.

La carencia de esta herramienta motivó a la creación de una propia, o mejor dicho, la adaptación de una existente para su uso en esta Tesis. Este trabajo, realizado en colaboración con el Lic. Leandro Martinez Heredia del Grupo de Química Computacional del CEQUINOR, consistió en una modificación de la herramienta OBEnergy, desarrollada por Tim Vandermeersch e incluida en la plataforma OpenBabel (18). La misma, escrita en el lenguaje de programación C++, fue creada con el fin de medir la energía de un sistema (por ejemplo, una molécula en el vacío) de forma interatómica, permitiendo el uso de distintos campos de fuerzas como MMFF94, UFF, GAFF o Ghemical.

La adaptación de dicha herramienta involucró la identificación del ligando dentro de la

proteína, el cálculo de la interacción de cada átomo del sistema con todos los demás, y la posterior recomposición de la energía de interacción tomando solo la interacción entre los átomos del ligando y los de cada uno de los residuos, eliminando las interacciones intra-ligando e intra-proteína. La nueva herramienta, llamada MIEF (Molecular Interaction Energy Fingerprints) (19) permite obtener el vector de interacción de un ligando con su proteína desde cualquier complejo en formato ".mol2" utilizando distintos campos de fuerza, que fueron modificados para calcular las contribuciones de las energías de van der Waals y electrostática. La misma es de código abierto y licencia *"GNU General Public License v3.0"*.

2.3.2. Comparación de la metodología con la de referencia

Una vez disponible la herramienta de descomposición de interacciones, fue posible armar nuevos modelos para cada uno de los blancos utilizados. Definidas las dos clases a diferenciar, se procedió al armado de los modelos utilizando Autodock Vina como motor de docking. Al igual que con DOCK6.9, la protonación del blanco molecular fue realizada utilizando Chimera (20) y su conversión al formato PDBQT, requerido por Autodock, se realizó con la ayuda del script *prepare_receptor4.py* disponible en la herramienta MGLTools. La caja fue definida con la ayuda de la interfaz gráfica del software AutodockTools (21), siendo sus coordenadas las obtenidas utilizando como referencia al ligando de la estructura cristalina en cada caso.

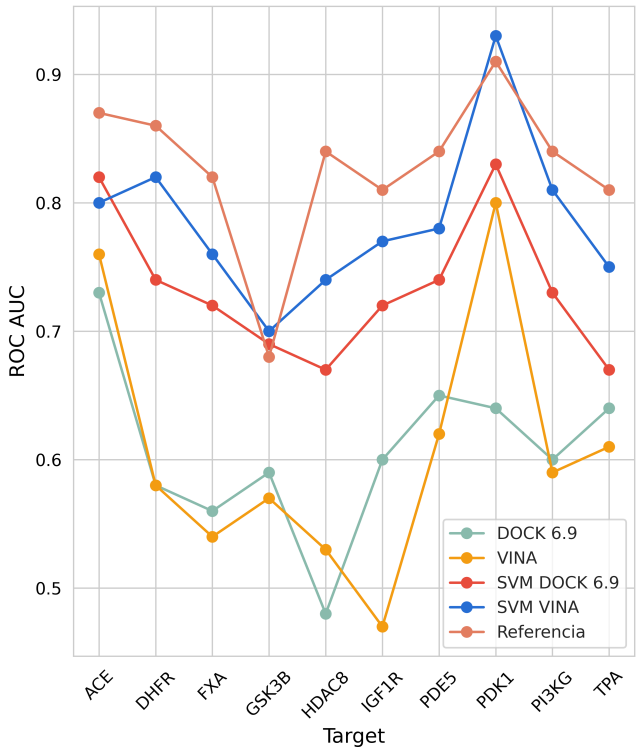


Figura 2.11: Comparación del rendimiento (ROC AUC) entre diferentes metodologías: DOCK6.9, Autodock VINA, DOCK6.9 + SVM, Autodock Vina + SVM y GOLD + PADIF + SVM (Referencia).

Los ligandos predefinidos en el trabajo de referencia fueron minimizados utilizando Balloon (22; 23). Finalmente, la conversión al formato PDBQT fue realizada con la librería meeko (24). Una vez obtenidos los complejos proteína-ligando, se obtuvieron los MIEFs para cada caso. De ellos, se tuvieron en cuenta solo los residuos en un entorno al sitio de unión, ubicados en un radio de 8 Å de distancia respecto a la mejor pose obtenida de cada uno de los ligandos. Los modelos fueron entrenados de la forma anteriormente utilizada, obteniendo un modelo final luego de seleccionar aquellas interacciones de mayor relevancia luego del cálculo de importancias.

La comparación con el subconjunto de modelos utilizados anteriormente, que se observa en la figura 2.11, muestra que, con excepción del target ACE, el uso de Autodock

Vina lleva a una mejora en el poder clasificatorio de los modelos en comparación con lo obtenido con DOCK6.9. Esta mejora es tal, que incluso llega a superar ligeramente los resultados obtenidos para GSK3B y PDK1 en el modelo de referencia.

Este resultado promisorio motivó a la expansión de la metodología al resto de los blancos moleculares listados en la tabla 2.5. El resultado de dicha comparación se muestra en la figura 2.12. Allí se pueden observar dos fenómenos: por un lado, se reafirma la tendencia de que el uso de modelos de aprendizaje automático específicos para cada blanco molecular lleva a una mejora respecto al solo uso de la función de puntuación docking, en este caso Vina. Por otro lado, la comparación con los resultados publicados por *Nogueira et. al.* muestra que el rendimiento de los nuevos modelos es a nivel global ligeramente inferior, con varios modelos alcanzando rendimientos similares e incluso superiores.

2.3.3. Desafíos a la hora de definir los conjuntos de activos y de inactivos

Si bien el trabajo de *Nogueira et al.* provee los compuestos activos e inactivos para un conjunto de 20 blancos moleculares, la creación de una nueva metodología implica el agregado de un mayor número de proteínas de interés. Para eso, es necesario establecer criterios comunes para la conformación de los conjuntos de activos e inactivos. En dicho trabajo, a lo largo de la creación de los modelos, los autores se han debido enfrentar con dos grandes dificultades en este punto:

- La identificación de compuestos activos e inactivos: El rango de concentraciones de inhibición informado en literatura es considerablemente distinto para cada blanco.

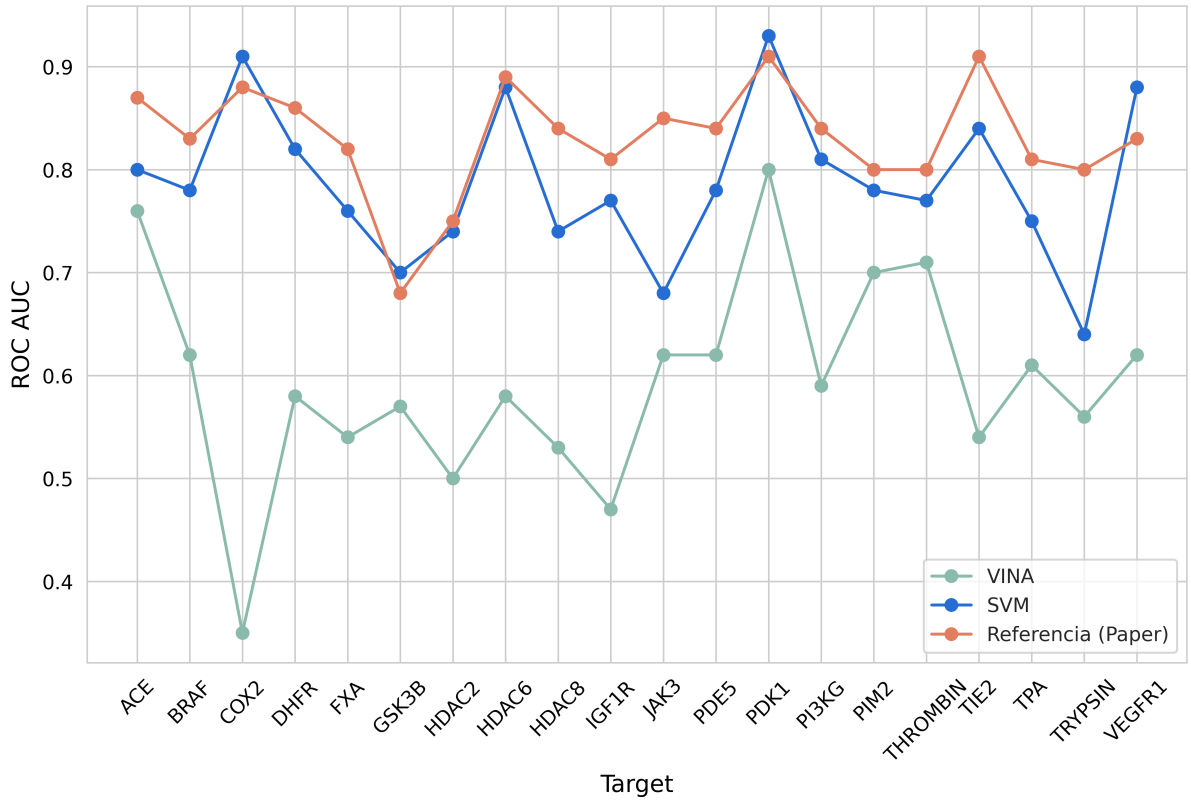


Figura 2.12: Comparación del rendimiento (ROC AUC) entre diferentes metodologías para todos los blancos moleculares estudiados: Autodock Vina, Autodock Vina + SVM y GOLD + PADIF + SVM.

Esto provoca que en la elaboración de conjuntos de activos e inactivos para cada blanco no pueda utilizarse una misma línea de corte, puesto que esto generaría conjuntos extremadamente desbalanceados. Este hecho se discutirá más ampliamente más adelante en la sección 2.3.3.

- El uso de decoys en un enfoque basado en docking: Se entiende por *decoy* a una molécula con características fisicoquímicas similares a las del conjunto activo, pero al no estar *diseñadas* para interactuar con el blanco se presumen inactivas. Esto tiene como problema el hecho de que se genera una pose artificial entre un compuesto y una proteína que en la realidad simplemente no ocurre. Dado que en líneas generales la función de puntuación del programa de docking no puede distinguir

entre la pose real y la artificial, se le termina pidiendo al modelo que cumpla con esta tarea, asumiendo que las poses son significativamente distintas.

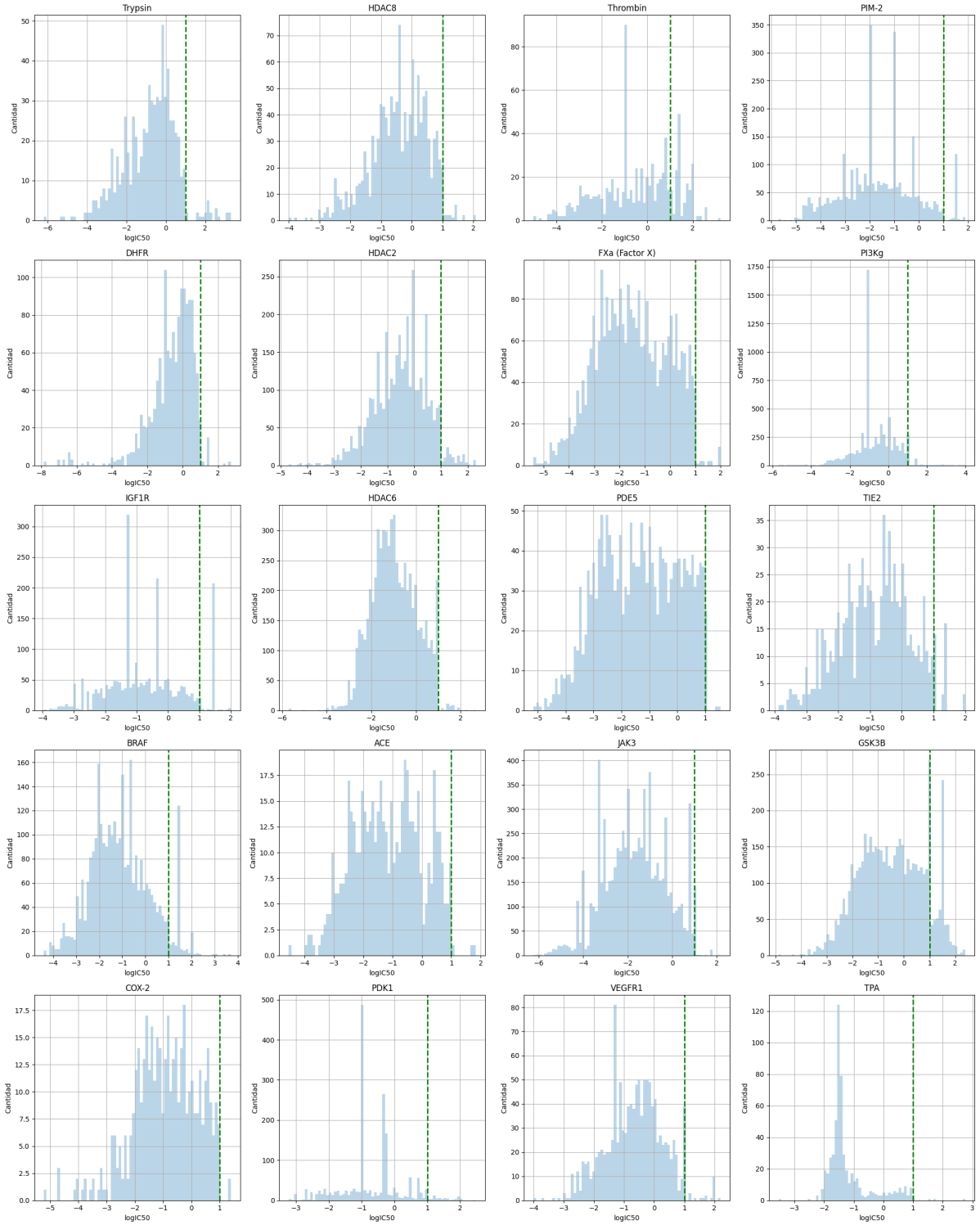


Figura 2.13: Distribución de los valores de actividad (IC_{50}) para los 20 blancos moleculares estudiados, según datos de PubChem. La línea punteada vertical indica una actividad de 10 μM para cada distribución.

Previo al armado de los nuevos modelos con Vina, se procedió a estudiar estos dos problemas en detalle.

Distribución de actividades para blancos moleculares seleccionados

En primer lugar, se estudió la distribución de los valores de actividad (IC_{50}) en la base de datos PubChem para los 20 modelos propuestos en el trabajo de Nogueira. Los resultados, que se muestran en la figura 2.13, mostraron que las distribuciones no son uniformes, teniendo muchas de ellas un salto abrupto en el valor 10 μM. Este resultado era esperado, puesto que este es un valor frecuentemente utilizado en la literatura como valor de corte para definir si un compuesto es activo o no. Además, el uso de este valor de 10 μM responde a una estrategia experimental orientada a la eficiencia y reducción de costos. Dado que los ensayos de inhibición enzimática pueden ser costosos y laboriosos, generalmente se realiza el estudio del compuesto a esta concentración inicial. Si en esta prueba preliminar la inhibición de la enzima no alcanza un umbral significativo, se considera que el compuesto no posee actividad relevante, y en consecuencia, no se procede a determinar el IC50 de manera más precisa. Este enfoque permite optimizar los recursos y focalizar los esfuerzos en compuestos con una respuesta inicial prometedora, evitando el gasto innecesario en aquellos que no demuestran potencial de actividad.

Evaluación del uso de decoys

En las campañas de *screening* virtual, los decoys son utilizados para aumentar la cantidad de datos del modelo, y analizar si el mismo es capaz de diferenciar entre compuestos activos y *decoys* mediante alguna función de puntuación, como puede ser el valor de la

función puntuación o *score* provista por el programa de docking. Como se mencionó anteriormente, resulta problemática la incorporación de *decoys* si se tiene como objetivo el entrenamiento de un modelo de inteligencia artificial, puesto que las interacciones entre la proteína y los decoys son ficticias, y además los compuestos son presuntamente inactivos, sin tener la certeza de que algunos de ellos puedan efectivamente tener actividad contra la proteína en estudio. Una problemática adicional que trae consigo la incorporación de este tipo de moléculas es la posibilidad de que las mismas tengan un sesgo oculto en su generación, efecto que ya se ha reportado en literatura (25). Como consecuencia de esta combinación de factores, el algoritmo podría aprender a diferenciar entre compuestos activos y decoys y su poder clasificatorio perderse cuando se le pida evaluar un conjunto de prueba conteniendo activos e inactivos reales.

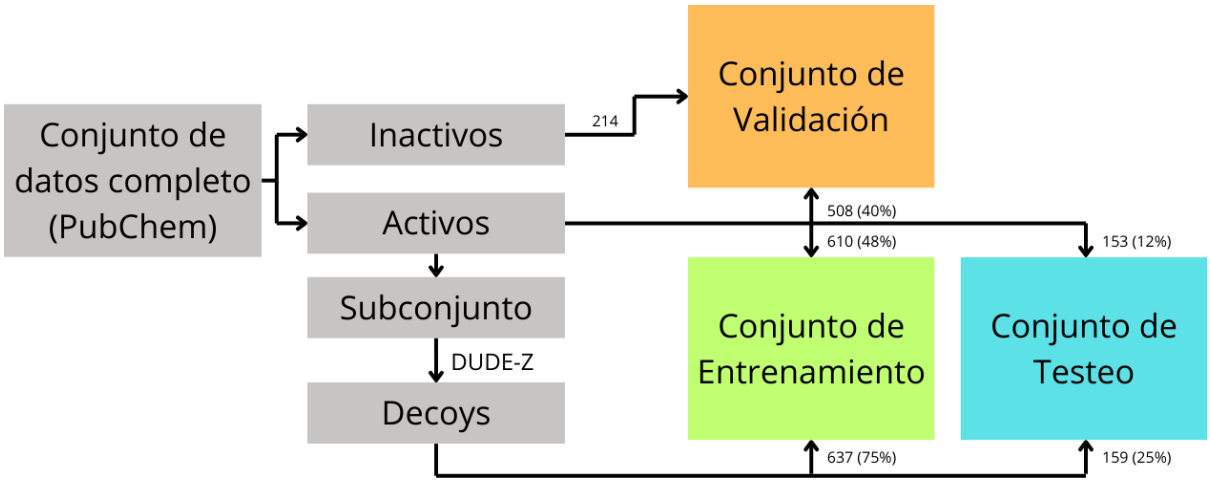


Figura 2.14: Diseño del experimento para evaluar el impacto del uso de decoys en el entrenamiento de modelos de inteligencia artificial.

A modo de evaluar esto último, se seleccionó como blanco de prueba la Acetilcolinesterasa (ACHE) y se generaron *decoys* utilizando el servidor web DUDE-Z (26), a partir de una selección de 50 compuestos activos estructuralmente distintos. El diseño del experimento se muestra en la figura 2.14, donde se esquematiza la comparación entre modelos

entrenados y evaluados con *decoys* frente a aquellos evaluados con inactivos reales. Los resultados, ver figura 2.15, evidenciaron una caída significativa en todas las métricas utilizadas para evaluar el rendimiento del modelo cuando la clasificación se realizó sobre un conjunto "real", es decir, con compuestos verdaderamente inactivos, destacando así el impacto negativo del uso de decoys artificiales en la generalización del modelo.

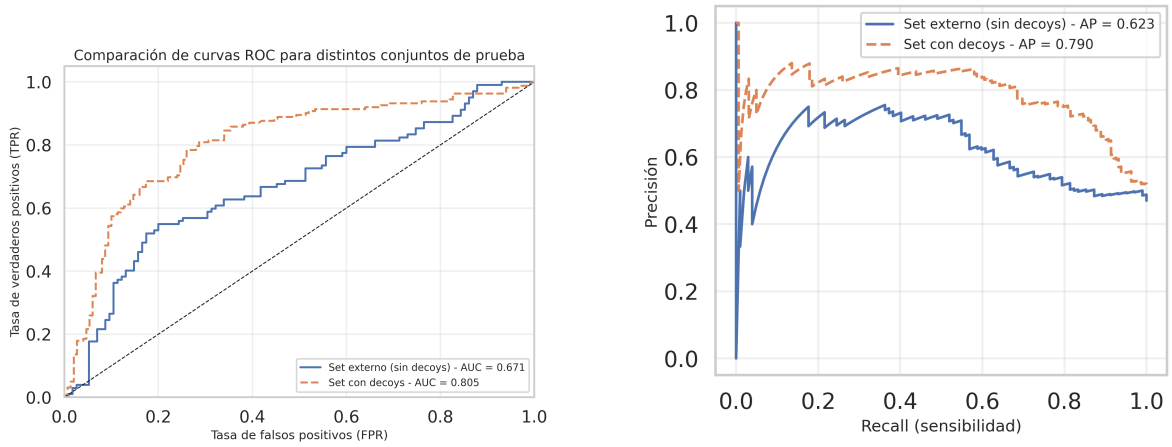


Figura 2.15: Resultados del experimento con decoys. (izq.) Curvas ROC obtenidas al evaluar el modelo con un conjunto de decoys y con inactivos reales. (der.) Curvas Precision-Recall para ambos escenarios.

Lo observado en estos dos análisis pone de manifiesto una de las problemáticas de mayor relevancia en la actualidad y que es quizás una consecuencia de la tendencia en el sistema científico global de no publicar resultados negativos. La calidad de los modelos de inteligencia artificial tiene una muy alta dependencia en la calidad de los datos utilizados para su entrenamiento, y la falta de reporte de datos inactivos reales es actualmente uno de los grandes obstáculos para el desarrollo de modelos predictivos de alta calidad.

Para poder continuar con el desarrollo de la metodología de predicción de blancos moleculares, se decidió entonces tomar la mediana de la distribución observada en la figura 2.13 (indicada con una línea punteada vertical) como valor de corte. De esta manera,

se obtienen dos clases de compuestos balanceadas y que, además, se conoce experimentalmente su interacción con la proteína, evitando entrenar el modelo con patrones de interacción sin ninguna correlación con la realidad experimental.

Selección de nuevos blancos moleculares

Una vez establecidas las estrategias a abordar para las dos problemas anteriores, se procedió a seleccionar nuevos blancos moleculares para su incorporación a la herramienta. Para ello se buscó conformar un conjunto de proteínas de distintas familias, a modo de evaluar la metodología en blancos moleculares de distinta naturaleza, y además se tuvo en cuenta aquellos blancos a los que se tenga acceso y posibilidad de realizar comprobaciones experimentales.

Dicha búsqueda de nuevos *targets* arrojó como resultado la selección de 18 blancos moleculares adicionales, que se listan en la tabla 2.6.

Si bien en este punto ya se cuenta con una metodología confiable que lleva a la creación de modelos con un poder clasificatorio satisfactorio, previo al armado final de los modelos existen dos puntos adicionales a considerar: la similitud estructural entre los compuestos de cada clase y la determinación del punto de corte en el valor de actividad biológica, ya que el uso de un valor puntual puede hacer que los compuestos que se encuentren a un lado o al otro del valor sean parecidos, llevando a huellas de interacción similares que dificulten el rendimiento del modelo. En otras palabras, tener compuestos similares a ambos lados de la línea de corte puede implicar exigirle al modelo que aprenda a distinguir entre huellas de interacción muy similares y cuyas diferencias no implican un impacto brusco

en la actividad. Ambos efectos serán estudiados en las secciones a continuación.

ID	UniProtID	Nombre	N° Comp.	Tipo
ACHE	P22303	Acetilcolinesterasa	3647	Hidrolasa
BACE1	P56817	Beta-secretasa 1	8614	Proteasa
BCHE	P06276	Butirilcolinesterasa	2116	Hidrolasa
CDK1	P06493	Quinasa dependiente de ciclina 1	2726	Quinasa
CDK2	P24941	Quinasa dependiente de ciclina 2	6359	Quinasa
CDK5	Q00535	Quinasa dependiente de ciclina 5	1038	Quinasa
PLA2B	P14555	Fosfolipasa A2, asociada a membranas	251	Fosfolipasa
CSNK2A1	P68400	Subunidad alfa de la quinasa de caseína II	2141	Quinasa
Cruzipain	P25779	Cruzipaina	430	Proteasa
DHODH	Q02127	Dihidroorotato deshidrogenasa	1571	Oxidorreducatasa
DYRK1A	Q13627	Quinasa de tirosina regulada por fosforilación de doble especificidad 1A	3368	Quinasa
MAOA	P21397	Monoaminooxidasa A	2471	Oxidorreductasa
MAOB	P27338	Monoaminooxidasa B	3892	Oxidorreductasa
MTNR1A	P48039	Receptor de melatonina tipo 1A	1053	Receptor de Membrana
MTNR1B	P49286	Receptor de melatonina tipo 1B	995	Receptor de Membrana
PDE5A	O76074	Fosfodiesterasa 5A	1797	Fosfodiesterasa
PDE7A	Q13946	Fosfodiesterasa 7A	635	Fosfodiesterasa
PDE10A	Q9Y233	Fosfodiesterasa 10A	6558	Fosfodiesterasa

Tabla 2.6: Lista de nuevos blancos moleculares incorporados a la metodología.

Estudio del uso de filtros de similitud

Un procedimiento habitual en las campañas de cribado virtual es la reducción del conjunto de compuestos activos mediante filtros de similitud. El objetivo de este proceso es obtener un conjunto de datos heterogéneo, buscando evaluar que la función de puntuación utilizada sea abarcativa y diferencie compuestos activos de distinto origen, y no

se encuentre sesgada únicamente hacia algunos de cierto tipo.

En lo que refiere al armado de modelos de inteligencia artificial, en este punto surge un dilema. Por un lado, el uso de filtros de este tipo lleva a la reducción en el número de datos, aspecto crucial para el correcto entrenamiento del modelo y que además se encuentra ya en valores muy bajos. Otra ventaja que trae consigo realizar un filtrado por similitud estructural es que los patrones de interacción con la proteína resultan más heterogéneos, evitando que el algoritmo de entrenamiento se sesgue hacia ciertos patrones en particular. Por otro lado, es importante tener en cuenta en el análisis que una reducción drástica en el número de compuestos puede ocasionar que el algoritmo directamente no encuentre patrones comunes entre los compuestos activos que los diferencien de los inactivos (y viceversa), provocando una potencial caída en el poder clasificatorio del modelo. El correcto armado del mismo dependerá del correcto balance entre estos efectos.

Estudio del efecto del uso de una ventana para diferenciar las clases

Otro aspecto a considerar en el armado de los modelos es la exigencia que se le da al algoritmo de clasificación. En el caso de los modelos anteriores, se le pide al modelo que sea capaz de clasificar entre compuestos con diferencias mínimas en su valor de actividad, por lo que es de esperar que el poder clasificatorio varíe al modificar esta exigencia.

Para evaluar esta posibilidad, se re-entrenaron los modelos pero en este caso diferenciando ambas clases con una ventana de 5 y 10 % de diferencia a ambos lados del valor de corte original.

El área bajo la curva ROC para estos modelos, que además incluyeron la presencia o

no del filtro por similitud, se muestra en la tabla 2.7. Del análisis de la misma se pueden elaborar conclusiones respecto a ambos efectos. En primer lugar, en líneas generales, la no utilización de filtros por similitud lleva a un poder clasificatorio superior, como se puede ver al comparar los promedios de las últimas 3 columnas respecto a los de las primeras dos. Por otro lado, el uso de una ventana que diferencie activos de inactivos no parece tener un efecto notorio en líneas generales, si bien se observan diferencias significativas para algunos blancos moleculares en particular. Esto pareciera indicar que efectivamente en algunos casos los patrones de interacción de los compuestos pertenecientes a la ventana son similares entre sí y perjudican al desempeño del modelo final, mientras que en otros este efecto es insignificante o incluso opuesto.

Tabla 2.7: ROC AUC (conjunto de prueba) para diversos modelos utilizando filtrado por similitud y ventanas de diferentes tamaños

Target	**Filtro por similitud**		**Sin filtro por similitud**		
	Sin Ventana	**Ventana 5 %**	**Sin Ventana**	**Ventana 5 %**	**Ventana 10 %**
ACE	0.76	0.76	0.71	0.74	0.70
ACHE	0.74	0.74	0.79	0.81	0.81
BACE1	0.67	0.76	0.74	0.73	0.78
BCHE	0.63	0.69	0.74	0.77	0.78
BRAF	0.68	0.73	0.75	0.76	0.77
CDK1	0.66	0.69	0.71	0.72	0.74
CDK2	0.71	0.70	0.78	0.78	0.79
CDK5	0.63	0.63	0.66	0.65	0.66
Cruzipain	0.71	0.89	0.68	0.62	0.56
CSNK2A1	0.68	0.78	0.75	0.74	0.70
DHFR	0.90	0.88	0.92	0.84	0.87
DHODH	0.78	0.80	0.81	0.83	0.83
DYRK1A	0.84	0.89	0.88	0.90	0.87
F10	0.77	0.68	0.78	0.76	0.76
FLT1	0.73	0.78	0.78	0.73	0.78
GSK3B	0.72	0.70	0.79	0.79	0.80
HDAC2	0.64	0.67	0.72	0.69	0.70
HDAC6	0.71	0.71	0.70	0.73	0.75
HDAC8	0.60	0.68	0.69	0.59	0.56
IGF1R	0.67	0.68	0.74	0.75	0.78
JAK3	0.74	0.76	0.80	0.82	0.82
MAOA	0.70	0.71	0.75	0.74	0.74
MAOB	0.79	0.80	0.81	0.81	0.80
MTNR1A	0.69	0.66	0.69	0.77	0.80
MTNR1B	0.60	0.64	0.70	0.78	0.83
PDE5A	0.70	0.69	0.77	0.80	0.81
PDE7A	0.81	0.84	0.82	0.84	0.88
PDE10A	0.77	0.79	0.81	0.81	0.84
PDPK1	0.72	0.74	0.83	0.81	0.81
PIK3CG	0.76	0.71	0.77	0.77	0.76
PIM2	0.78	0.83	0.87	0.88	0.89
PLA2B	0.90	0.91	0.89	0.81	0.81
PLAT	0.77	0.84	0.84	0.85	0.90
Ptgs2	0.76	0.76	0.72	0.82	0.75
PRSS1	0.63	0.68	0.71	0.84	0.80
THROMBIN	0.67	0.75	0.83	0.82	0.77
TIE2	0.75	0.75	0.81	0.82	0.79
Promedio	0.72	0.75	0.77	0.78	0.78

2.3.4. Selección del mejor modelo para cada blanco molecular

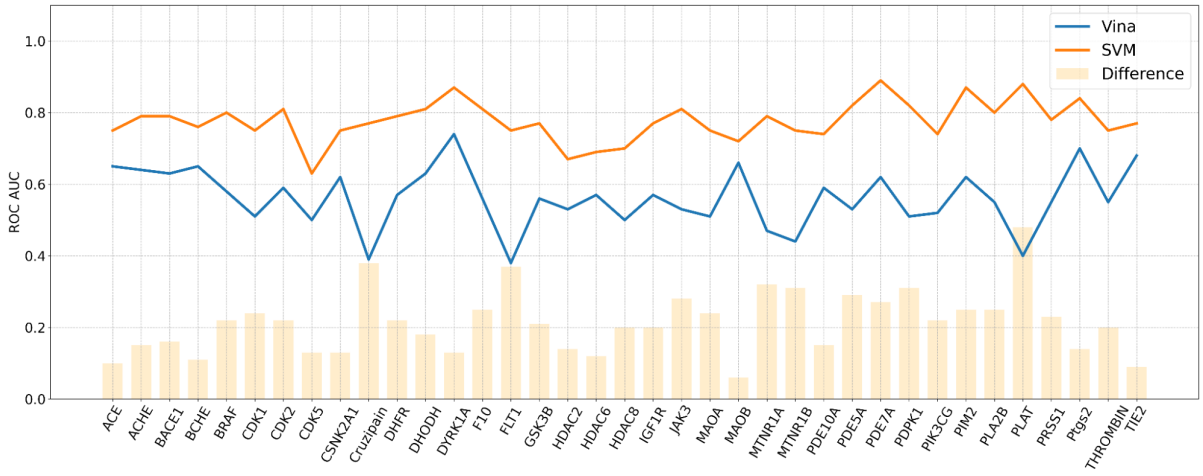


Figura 2.16: Comparación del rendimiento (ROC AUC) entre Autodock Vina y los mejores modelos Autodock Vina + SVM para cada blanco molecular.

A partir del análisis anterior, fue posible seleccionar el mejor modelo para cada blanco molecular, a modo tal de incluirlo en la nueva herramienta de predicción de actividad a desarrollar en la presente tesis. En la figura 2.16 se puede observar el poder clasificatorio de todos estos modelos respecto al de la función de puntuación de Autodock Vina, utilizando nuevamente el área bajo la curva ROC como instrumento de medición.

Naturalmente, la correcta evaluación e interpretación de los modelos por parte del usuario requiere de más de una métrica. En la tabla 2.8 se muestran los valores de precisión, exactitud y F1 para los modelos finales.

La posibilidad de tener un conjunto de modelos seleccionados abre la puerta no sólo a su compilación para el desarrollo de una metodología de predicción, sino también para su caracterización en mayor profundidad, de forma tal de entender cómo realizan los diferentes modelos sus predicciones y si éstas están hechas con fundamento biológico.

Tabla 2.8: Métricas de rendimiento para los mejores modelos de cada blanco molecular

Target	**Precisión**	**Recall**	**F1**	**Exactitud**	**ROC AUC**
ACE	0.71	0.75	0.73	0.69	0.77
ACHE	0.71	0.63	0.67	0.73	0.81
BACE1	0.68	0.66	0.67	0.71	0.75
BCHE	0.71	0.57	0.63	0.71	0.76
BRAF	0.74	0.65	0.69	0.7	0.77
CDK1	0.68	0.68	0.68	0.7	0.74
CDK2	0.78	0.68	0.73	0.72	0.79
CDK5	0.7	0.67	0.69	0.67	0.68
PLA2B	0.85	0.79	0.81	0.85	0.91
Cruzipain	0.63	0.83	0.72	0.76	0.84
CSNK2A1	0.74	0.71	0.72	0.72	0.77
DHFR	0.82	0.78	0.8	0.84	0.91
DHODH	0.76	0.81	0.78	0.77	0.82
DYRK1A	0.86	0.81	0.84	0.81	0.87
F10	0.73	0.69	0.71	0.69	0.76
FLT1	0.79	0.71	0.75	0.75	0.77
GSK3B	0.72	0.61	0.66	0.7	0.78
HDAC2	0.64	0.63	0.63	0.66	0.71
HDAC6	0.71	0.69	0.7	0.7	0.75
HDAC8	0.51	0.83	0.63	0.57	0.65
IGF1R	0.71	0.81	0.76	0.71	0.77
JAK3	0.79	0.66	0.72	0.74	0.82
MAOA	0.67	0.73	0.7	0.71	0.77
MAOB	0.73	0.77	0.75	0.75	0.81
MTNR1A	0.74	0.7	0.72	0.74	0.77
MTNR1B	0.89	0.64	0.75	0.78	0.84
PDE10A	0.82	0.69	0.75	0.76	0.83
PDE5A	0.63	0.87	0.73	0.71	0.82
PDE7A	0.85	0.86	0.86	0.81	0.89
PDPK1	0.78	0.78	0.78	0.77	0.84
PIK3CG	0.76	0.63	0.69	0.68	0.76
PIM2	0.88	0.79	0.83	0.83	0.89
PLAT	0.95	0.87	0.91	0.86	0.88
PRSS1	0.68	0.71	0.7	0.74	0.83
Ptgs2	0.74	0.76	0.75	0.78	0.82
THROMBIN	0.62	0.81	0.7	0.73	0.8
TIE2	0.74	0.66	0.7	0.73	0.8

Significancia de los patrones de interacción

Un aspecto fundamental para la correcta caracterización de los modelos es entender si los mismos capturan efectivamente patrones de interacción significativos en lugar de ruido aleatorio. Para responder a esta pregunta, se puede realizar una prueba de reorganización de etiquetas (*label shuffling test*) para cada modelo. Esta técnica consiste en aleatorizar las etiquetas (1 o 0, dependiendo de si el compuesto es activo o inactivo) del conjunto de datos, lo que permite evaluar si los modelos son capaces de identificar patrones relevantes incluso cuando la información que se les proporciona es aleatoria. En este proceso, los modelos son reentrenados y evaluados con las etiquetas reorganizadas. Como se muestra en la Tabla 2.9, los puntajes de AUC-ROC para todos los modelos cayeron a un promedio de 0.49, lo que confirma que su rendimiento predictivo original se basaba en patrones reales que diferenciaban a los compuestos activos de los inactivos, en lugar de correlaciones fortuitas.

Intercambio de conjuntos de ligando y blancos moleculares

Otro desafío importante en los modelos de aprendizaje automático para el cribado virtual basado en estructuras es el poder evaluar en qué medida la estructura de la proteína está realmente influyendo en las predicciones del modelo, y si el poder clasificatorio no es simplemente consecuencia de que los compuestos activos e inactivos son distintos entre sí.

Este efecto es potencialmente más significativo en conjuntos de datos pequeños, dado que los mismos pueden introducir sesgos. Para evaluarlo, primero se calcularon diversas

Tabla 2.9: Puntajes de AUC-ROC obtenidos al realizar una prueba de reorganización de etiquetas para evaluar la capacidad de los modelos para identificar patrones relevantes.

Target	ROC AUC	Target	ROC AUC	Target	ROC AUC
ACE	0.43	CSNK2A1	0.53	IGF1R	0.46
PIK3CG	0.48	ACHE	0.50	DHFR	0.52
JAK3	0.50	PIM2	0.51	BACE1	0.51
DHODH	0.50	MAOA	0.48	PLAT	0.57
BCHE	0.49	DYRK1A	0.49	MAOB	0.41
PRSS1	0.50	BRAF	0.51	F10	0.47
MTNR1A	0.58	Ptgs2	0.48	CDK1	0.48
FLT1	0.50	MTNR1B	0.44	THROMBIN	0.49
CDK2	0.52	GSK3B	0.48	PDE10A	0.45
TIE2	0.52	CDK5	0.49	HDAC2	0.58
PDE5A	0.47	PLA2B	0.35	HDAC6	0.50
PDE7A	0.50	Cruzipain	0.46	HDAC8	0.48
PDPK1	0.53				
Average	**0.49**				

propiedades fisicoquímicas de los conjuntos de activos e inactivos tales como el peso molecular, la carga, el logP y el número de enlaces rotacionales. El análisis de estas propiedades se encuentra en el capítulo 6. No se observaron diferencias significativas en las propiedades para la mayoría de los objetivos, y en los casos donde existían diferencias, estas no se correlacionaron con el rendimiento del modelo.

Por otro lado, para explorar en mayor profundidad la influencia de la proteína, se realizó una prueba de intercambio de ligandos en tres blancos moleculares no relacionados (ACHE, MAO-B y DYRK1A). Los resultados, ilustrados en la figura 2.17, muestran que aunque los modelos con los ligandos incorrectos aún pueden diferenciar compuestos activos de inactivos, su rendimiento disminuyó al utilizar pares de ligandos y proteínas que no coincidían. Esto sugiere que la correcta interacción proteína-ligando juega un papel crucial en el rendimiento del modelo.

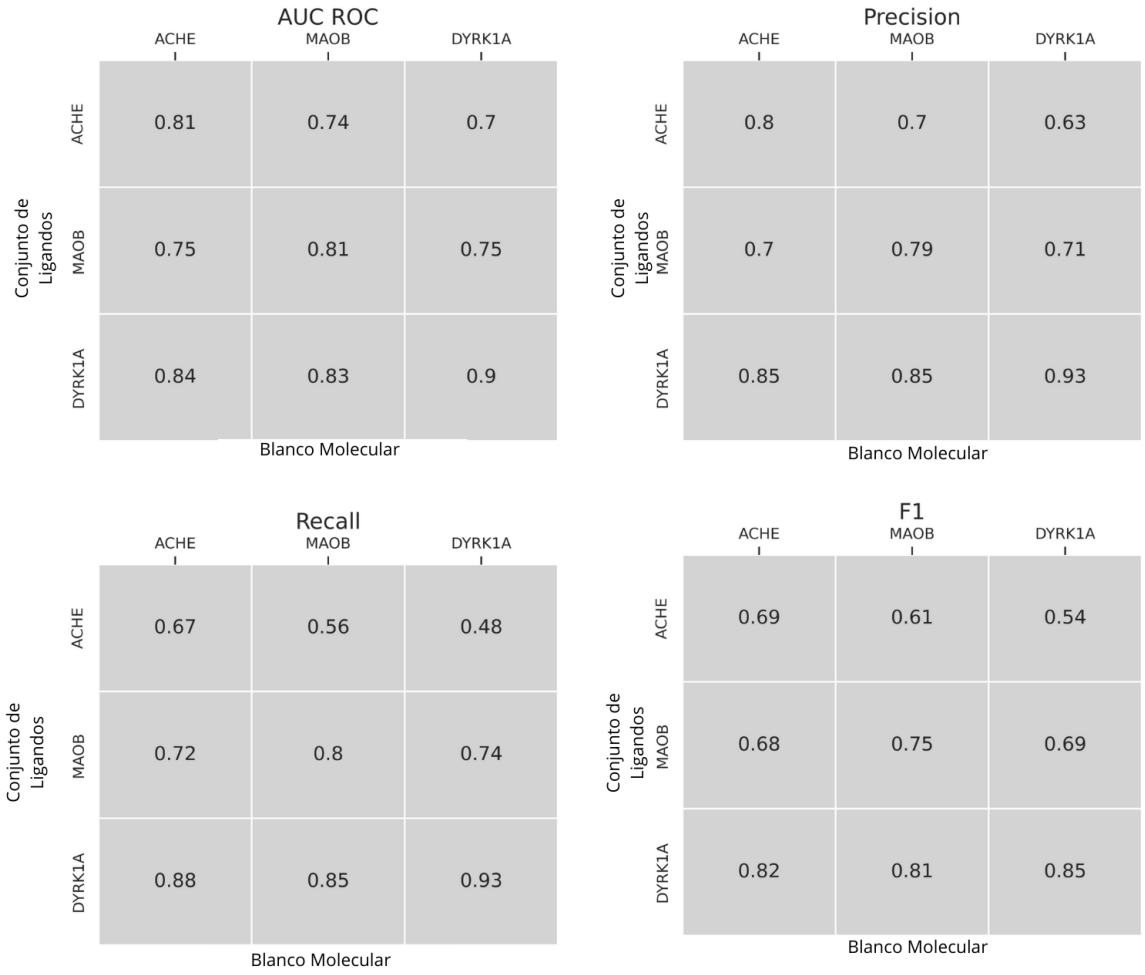


Figura 2.17: Resultados del experimento de intercambio de conjuntos de ligandos y blancos moleculares.

2.4 Target Fisher

El efectivo armado y estudio de los modelos de aprendizaje automático que se detalló en las secciones anteriores motivó a cumplir con el objetivo propuesto en el inicio del presente capítulo que es el armado de una metodología de predicción de blancos moleculares basada en la estructura del blanco molecular. A la misma se la llamó Target Fisher, y busca compilar los modelos anteriores en una plataforma de fácil acceso, con disponibilidad de métricas de rendimiento y resultados simples de interpretar.

2.4.1. Funcionamiento de la herramienta

En la figura 2.18 se presenta un esquema detallado de cómo funciona Target Fisher, un sistema diseñado para predecir la interacción de moléculas con diferentes blancos moleculares. El proceso comienza cuando el usuario suministra una molécula, ya sea a través de su representación en formato SMILES o bien con su estructura tridimensional en formatos estándar como .mol2 o .sdf. Una vez recibida, la molécula es evaluada en los distintos blancos moleculares disponibles, utilizando el software Autodock Vina. Para garantizar la consistencia y fiabilidad de los resultados, se emplean los mismos parámetros que se utilizaron durante la creación y entrenamiento de cada modelo, lo que asegura una evaluación precisa basada en condiciones previas definidas.

Este proceso no solo permite evaluar la actividad de nuevos compuestos, sino que también facilita una comprensión más profunda sobre cómo las interacciones entre las moléculas y sus blancos pueden influir en la predicción de la actividad biológica. Además, la flexibilidad del sistema para adaptarse a diferentes tipos de entrada y su capacidad para procesar múltiples blancos moleculares lo hacen una herramienta poderosa en el ámbito de la predicción de actividad farmacológica.

Finalmente, el programa genera un informe con las predicciones de actividad del compuesto en estudio frente a cada uno de los blancos seleccionados. Este informe incluye además gráficos de cuadrantes, como los ilustrados en la figura 2.18, que permiten una visualización comparativa entre la predicción del modelo y la puntuación de docking, observando además la distribución del conjunto de testeo empleado durante el entrena-

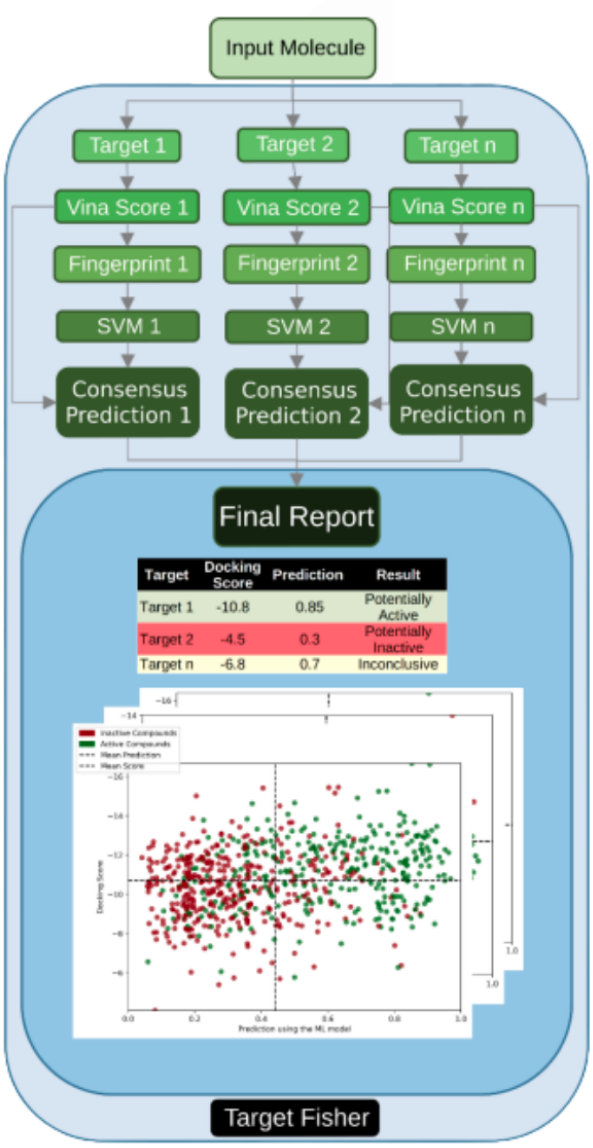


Figura 2.18: Esquema del funcionamiento de Target Fisher. La molécula ingresada por el usuario es evaluada mediante docking en los blancos moleculares disponibles, y las huellas de interacción resultantes son analizadas por los modelos de aprendizaje automático específicos para cada blanco.

Posteriormente, del complejo formado entre el ligando (la molécula en estudio) y la proteína (el blanco molecular) se extrae un patrón de interacción mediante el uso de los Molecular Interaction Energy Fingerprints (MIEFs). Este patrón sirve como entrada para la predicción de actividad, que se evalúa a través de modelos de inteligencia artificial específicamente entrenados para cada blanco molecular. Finalmente, la predicción de la actividad de la molécula puede presentarse de dos formas: gráfica, mediante representaciones visuales que comparan el rendimiento del modelo con un conjunto de testeo, o analítica, donde se proporciona un valor numérico que refleja el desempeño del modelo en función de las moléculas evaluadas.

miento. De este modo, el usuario puede evaluar el desempeño de su molécula en relación con los compuestos del set de entrenamiento, e interpretar rápidamente que aquellos ubicados en el cuadrante superior derecho presentan una mayor probabilidad de ser activos, al combinar altos valores tanto en la predicción del modelo como en la puntuación de docking.

2.4.2. Código Fuente

Target Fisher fue escrito íntegramente en lenguaje de programación Python 3, e integra el uso de OpenBabel, Autodock Vina, MIEFs y scikit-learn para elaborar predicciones de blancos moleculares utilizando los modelos finales para cada caso. Acepta moléculas en distintos formatos como .smi, .sdf o .mol2 y entrega como resultado un informe en formato .pdf. El código fuente puede encontrarse en `https://gitlab.com/gqc/target-fisher` y puede distribuirse respecto a la condiciones de la licencia GNU GPL v3.0.

2.4.3. Servidor Web

A modo de facilitar su uso a investigadores no especializados en técnicas computacionales, se procedió al armado de un servidor web, el cual puede accederse a través de la dirección `https://gqc.quimica.unlp.edu.ar/targetfisher`. Su desarrollo fue realizado también en colaboración con el Lic. Leandro Martinez Heredia, y basado en Django (27), *framework* conocido por su robustez y flexibilidad, permitiendo la integración de Python al desarrollo web.

La interfaz de usuario, que se puede observar en la figura 2.19, fue diseñada para ser

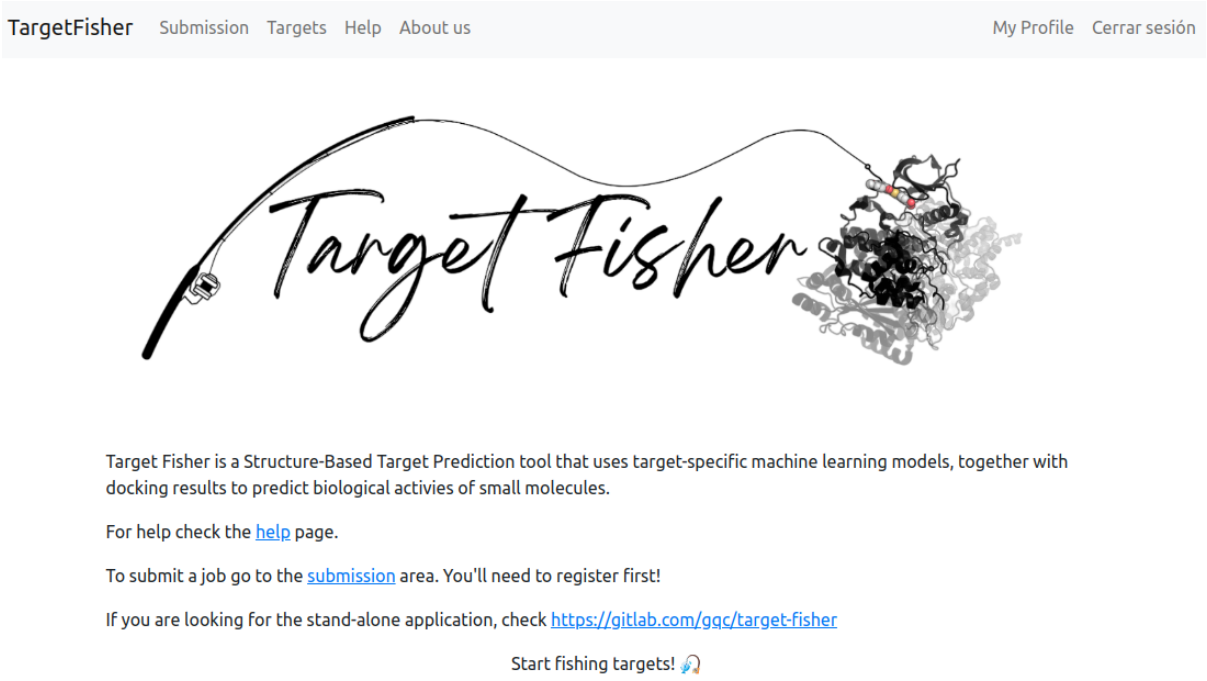


Figura 2.19: Interfaz del servidor web de Target Fisher. El usuario puede dibujar o subir la molécula a evaluar y seleccionar los blancos moleculares de interés.

amigable y simple. El usuario, una vez registrado, puede dibujar fácilmente la estructura química de la molécula que desee y seleccionar los blancos moleculares en los que la quiera evaluar. Una vez enviado el trabajo al servidor, el mismo lo gestiona con el sistema de colas Celery (28), permitiendo el manejo de las tareas requeridas por distintos usuarios de forma ordenada y secuencial. Al finalizar, el usuario puede acceder desde la plataforma web tanto a un resumen visual de los datos obtenidos, como así también descargar el informe y los complejos proteína-ligando obtenidos para cada caso.

De forma tal de tener la mayor transparencia posible, dentro del servidor web pueden encontrarse las distintas métricas de rendimiento de los modelos, siendo esta información complementaria a la comparación con los conjuntos de testeo que pueden hacerse de forma gráfica al mirar el informe.

Bibliografía

[1] Balius, T. E.; Chakrabarti, M. DOCK 6: Incorporating Hierarchical Traversal through Precomputed Ligand Conformations to Enable Large-Scale Docking. *Journal of Computational Chemistry* **2024**, *45*, 47–63.

[2] Balius, T. E.; Rizzo, R. C. Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons. *Journal of Computational Chemistry* **2011**, *32*, 2273–2289.

[3] Mukherjee, S.; Rizzo, R. C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *Journal of Chemical Information and Modeling* **2010**, *50*, 1986–2000.

[4] Meggio, F.; Pinna, L. A. One-Thousand-and-One Substrates of Protein Kinase CK2? *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* **2003**, *17*, 349–368.

[5] Chua, M. M. J.; Dominguez, I. CK2 in Cancer: Cellular and Biochemical Mechanisms and Potential Therapeutic Target. *Pharmaceuticals* **2017**, *10*, 18.

[6] Davies, M.; Overington, J. P. ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* **2023**, *52*, D1180–D1190.

[7] Pedregosa, F.; Duchesnay, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

[8] Nogueira, J. J.; Jorgensen, W. L. The Development of Target-Specific Machine Learning Models as Scoring Functions for Docking-Based Target Prediction. *Journal of Chemical Information and Modeling* **2018**, *58*, 1688–1700.

[9] Jorissen, R. N.; Gilson, M. K. Virtual Screening of Molecular Databases Using a Support Vector Machine. *Journal of Chemical Information and Modeling* **2005**, *45*, 549–561.

[10] Lavecchia, A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today* **2015**, *20*, 318–331.

[11] McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling* **2011**, *51*, 578–596.

[12] Eberhardt, J.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling* **2021**, *61*, 3891–3898.

[13] Schomburg, K. T.; Rarey, M. Benchmark Data Sets for Structure-Based Computational Target Prediction. *Journal of Chemical Information and Modeling* **2014**, *54*, 2261–2274.

[14] Bauer, M. R.; Boeckler, F. M. Evaluation and Optimization of Virtual Screening

Workflows with DEKOIS 2.0 – A Public Library of Challenging Docking Benchmark Sets. *Journal of Chemical Information and Modeling* **2013**, *53*, 1447–1462.

[15] Wang, Z.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics* **2016**, *18*, 12964–12975.

[16] Goldstein, A.; Pitkin, E. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. *arXiv preprint arXiv:1309.6392* **2014**,

[17] Molnar, C. *Interpretable Machine Learning*; Lulu.com, 2020.

[18] O'Boyle, N. M.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 33.

[19] Martínez Heredia, L.; Fernández, J. MIEF: Molecular Interaction Energy Fingerprints. https://gitlab.com/gqc/mief, 2023.

[20] Pettersen, E. F.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.

[21] Forli, S.; Olson, A. J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* **2016**, *11*, 905–919.

[22] Puranen, J. S.; Johnson, M. S. Accurate conformation-dependent molecular electros-

tatic potentials for high-throughput in silico drug discovery. *Journal of Computational Chemistry* **2010**, *31*, 1444–1454.

[23] Vainio, M. J.; Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *Journal of chemical information and modeling* **2007**, *47*, 2462–2474.

[24] ForliLab Meeko: A Molecular Docking and Scoring Package. 2025; https://github.com/forlilab/Meeko, Accessed: 2025-05-01.

[25] Chen, L.; Cruz, A.; Ramsey, S.; Dickson, C. J.; Duca, J. S.; Hornak, V.; Koes, D. R.; Kurtzman, T. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PloS one* **2019**, *14*, e0220113.

[26] Stein, R. M.; Yang, Y.; Balius, T. E.; O'Meara, M. J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B. K.; Irwin, J. J. Property-unmatched decoys in docking benchmarks. *Journal of chemical information and modeling* **2021**, *61*, 699–714.

[27] Foundation, D. S. Django: The Web framework for perfectionists with deadlines. 2025; https://www.djangoproject.com/, Accessed: 2025-05-01.

[28] Project, C. Celery: Distributed Task Queue. 2025; https://github.com/celery/celery, Accessed: 2025-05-01.

Capítulo 3

Cribado virtual de una quimioteca *in house* de productos naturales utilizando Target Fisher, con foco en enzimas relevantes para enfermedades neurodegenerativas

El paso siguiente al desarrollo de la metodología de predicción de blancos moleculares explicada en el capítulo anterior es su validación experimental. Este proceso permite evaluar su aplicabilidad en proyectos de descubrimiento de fármacos y reutilizar los resultados obtenidos para refinar los modelos. Para ello, es fundamental contar con ensayos biológicos que permitan evaluar la inhibición de alguna de las enzimas presentes en Target Fisher.

En el marco de nuestro proyecto de investigación, y gracias al financiamiento otorgado por la Universidad de Buenos Aires, se estableció una colaboración con el grupo de la Dra. María Isabel Rodríguez-Franco del Instituto de Química Médica en Madrid, España. Este grupo se especializa en la búsqueda de inhibidores de enzimas clave en trastornos neurodegenerativos y dispone de ensayos frente a varios blancos moleculares disponibles en Target Fisher. De esta manera, la colaboración establecida abrió las puertas a la validación experimental del programa.

3.1 Blancos moleculares relevantes en enfermedades neurodegenerativas

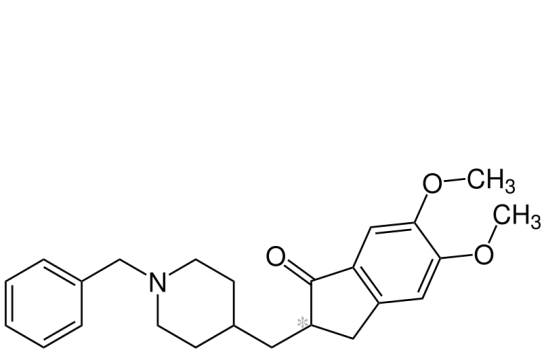
Las enfermedades neurodegenerativas (NDs), causadas por la pérdida progresiva de la estructura o función de las neuronas, representan una de las principales causas de discapacidad en personas mayores. Entre ellas, la enfermedad de Alzheimer (AD) y el mal de Parkinson (PD) son las más prevalentes, caracterizadas por la creciente muerte celular de neuronas específicas y la acumulación de agregados proteicos anómalos. Estas enfer-

medades generan discapacidades intelectuales y/o motoras que deterioran drásticamente la calidad de vida de los pacientes y acortan su esperanza de vida.

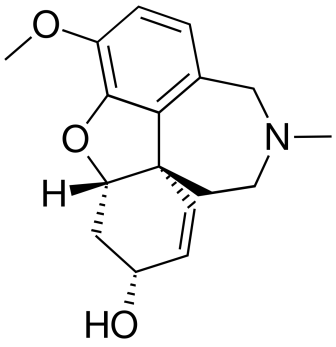
El tratamiento actual para estas patologías es mayormente sintomático, ya que carecen de terapias curativas. En este contexto, el descubrimiento de fármacos dirigidos a blancos moleculares clave es una estrategia esencial para el desarrollo de nuevas terapias neuroprotectoras. En el marco de esta tesis, se seleccionaron seis enzimas relacionadas con AD y PD, cuya inhibición o modulación ha demostrado potencial terapéutico:

- Acetilcolinesterasa (AChE)
- Butirilcolinesterasa (BuChE)
- Monoamino Oxidasa A (MAO-A)
- Monoamino Oxidasa B (MAO-B)
- Glucógeno Sintasa Quinasa 3 Beta (GSK3-β)
- β-Secretasa 1 (BACE-1)

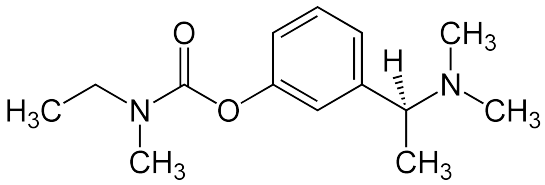
3.1.1. Acetilcolinesterasa (AChE) y Butirilcolinesterasa (BuChE)



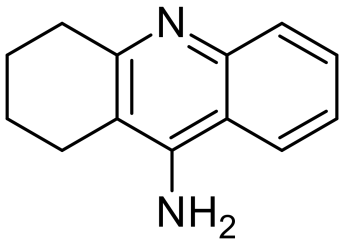
(a) Donepezilo



(b) Galantamina



(c) Rivastigmina



(d) Tacrina

Figura 3.1: Inhibidores de colinesterasas.

La **hipótesis colinérgica**, propuesta en 1976 por Peter Davies y A. J. F. Maloney,(1) fue la primera en intentar explicar la fisiopatología de la enfermedad de Alzheimer (AD). Según esta teoría, la reducción de los niveles de acetilcolina (ACh) en las terminales presinápticas del hipocampo y la neocorteza, causada por la degeneración de neuronas colinérgicas en los núcleos basales del cerebro anterior, desempeña un papel clave en la patogénesis de la AD.(2)

En sujetos sanos, la acetilcolinesterasa (AChE) es la principal enzima responsable de la hidrólisis de ACh en el sistema nervioso central. Sin embargo, en estadios avanzados de la enfermedad se observa un aumento en la actividad de la butirilcolinesterasa (BuChE).(3) Este fenómeno sugiere que BuChE podría contribuir a la progresión del deterioro cogniti-

vo, ya que su mayor actividad acentuaría la deficiencia de ACh, exacerbando la disfunción sináptica y la pérdida neuronal.

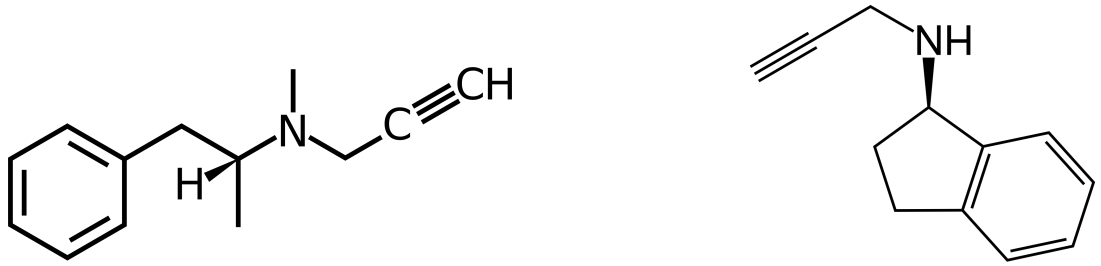
Dado que la reducción de ACh se ha correlacionado con el deterioro cognitivo en AD, la inhibición de las colinesterasas se convirtió en una estrategia terapéutica clave.(4) La tacrina, el primer inhibidor de la colinesterasa (ChEI) aprobado para AD, fue retirado del mercado debido a su hepatotoxicidad.(5) Actualmente, los fármacos **donepezilo, galantamina** y la **rivastigmina**, figura 3.1, son los tres ChEIs aprobados para el tratamiento de AD.(6) Estos medicamentos prolongan la acción de ACh en la sinapsis y pueden mejorar temporalmente la función cognitiva. Sin embargo, su eficacia disminuye con el tiempo y presentan efectos adversos como náuseas, vómitos y bradicardia. En particular, la **galantamina** es un ejemplo de producto natural utilizado en clínica, ya que es un alcaloide que se aísla de los bulbos y flores de varias especies de la familia *Amaryllidaceae*, como *Galanthus nivalis* (campanilla común), *Galanthus caucasicus*, *Galanthus woronowii*, así como también de especies de *Narcissus* (narciso), *Leucojum aestivum* (campanilla de verano) y *Lycoris*, incluyendo *Lycoris radiata* (lirio araña rojo).

Debido a que la actividad de BuChE aumenta en etapas avanzadas del AD, la inhibición dual de AChE y BuChE ha sido propuesta como una estrategia más efectiva para contrarrestar la progresión de la enfermedad. Rivastigmina es el único ChEI aprobado que inhibe ambas enzimas, mostrando un efecto más sostenido en la preservación de la neurotransmisión colinérgica.(7)

Aunque tradicionalmente se ha considerado que los ChEIs solo alivian los síntomas de AD de manera modesta,(8) estudios recientes sugieren que el sistema colinérgico está

interconectado con otros mecanismos patológicos, como la acumulación de β-amiloide (Aβ), la hiperfosforilación de τ, la inflamación y la apoptosis.(9) Estas interacciones podrían ofrecer nuevas oportunidades terapéuticas más allá de la simple restauración de los niveles de ACh.

3.1.2. Monoamino Oxidasas A y B (MAO-A y MAO-B)



(a) Selegilina (b) Rasagilina

Figura 3.2: Ejemplos de Inhibidores de monoamino oxidasas.

Las monoamino oxidasas (**MAOs**) desempeñan un papel fundamental en la regulación de neurotransmisores al catalizar la degradación de monoaminas como la dopamina, serotonina y noradrenalina.(10) Existen dos isoformas: **MAO-A**, que degrada preferentemente serotonina y noradrenalina, y **MAO-B**, que metaboliza dopamina y otras aminas biógenas.(11)

En el contexto de enfermedades neurodegenerativas, especialmente en la enfermedad de Alzheimer (AD), se ha observado que el estrés oxidativo es un factor relevante; por ello, se exploran compuestos con propiedades antioxidantes y de captura de radicales para contrarrestar dicho estrés.(12; 13) Además, se han estudiado inhibidores de MAO, ya que estas enzimas, en particular la isoforma MAO-B que se encuentra sobreexpresada

en pacientes con AD, localizadas en la membrana externa de las mitocondrias, no solo degradan neurotransmisores, sino que también generan peróxido de hidrógeno y otras especies reactivas de oxígeno (ROS) durante la oxidación de las monoaminas.(14) Recientemente, se ha evidenciado que la actividad de MAO-B puede regular la producción de Aβ en las neuronas a través de la acción de la γ-secretasa, abriendo la posibilidad de utilizar inhibidores de MAO-B tanto para mitigar el daño oxidativo como para reducir la generación del péptido Aβ42.(15)

En la enfermedad de Parkinson (PD), la degeneración de las neuronas dopaminérgicas conduce a una marcada disminución de la dopamina. La inhibición de MAO-B prolonga la vida media de la dopamina remanente, lo que se traduce en mejoras en la sintomatología motora.(10) De hecho, fármacos como selegilina y rasagilina, figura 3.2, que actúan como inhibidores irreversibles de MAO-B, se utilizan en terapia coadyuvante junto con la levodopa, aunque sus efectos adversos y contraindicaciones han impulsado la búsqueda de nuevos inhibidores selectivos y reversibles con mejores perfiles de seguridad.

3.1.3. Glucógeno sintasa quinasa - 3β y la hipótesis de τ

La **glucógeno sintasa quinasa 3 beta (GSK3β)** es una enzima clave en múltiples procesos celulares, incluido el control de la fosforilación de la proteína τ, un componente esencial del citoesqueleto neuronal. En condiciones fisiológicas, τ estabiliza los microtúbulos dentro de los axones neuronales, garantizando la correcta función del transporte intracelular. Sin embargo, en la enfermedad de Alzheimer (AD), la hiperactivación de GSK3β induce una fosforilación excesiva de τ, lo que disminuye su afinidad por los mi-

crotúbulos y favorece su agregación en ovillos neurofibrilares (NFTs, por sus siglas en inglés), estructuras insolubles que contribuyen a la degeneración neuronal y la progresión de la enfermedad.

La hipótesis de τ, propuesta en 2009,(6) está estrechamente relacionada con la hipótesis amiloide, que postula que la secuencia patológica en AD comienza con la acumulación de Aβ, seguida por la hiperfosforilación de τ, la formación de NFTs y la muerte neuronal. No obstante, el fracaso de múltiples terapias antiamiloides ha llevado a algunos investigadores a cuestionar si Aβ es realmente el factor determinante en la neurotoxicidad mediada por τ.(16) En este contexto, se ha propuesto que la patogénesis de AD podría estar impulsada por τ de manera independiente de Aβ.

Dado su papel central en la fosforilación de τ, GSK3β se ha convertido en un blanco terapéutico atractivo.(2; 17) Diversos estudios han demostrado que su inhibición puede reducir la agregación de τ y mejorar la función cognitiva en modelos preclínicos de AD. Además, se ha sugerido que el aumento en los niveles de GSK3β podría regular la γ-secretasa, promoviendo la formación de Aβ y exacerbando la toxicidad neuronal. Por estas razones, el desarrollo de inhibidores de GSK3β ha sido una estrategia ampliamente explorada. Sin embargo, debido a que esta quinasa también participa en otros procesos celulares esenciales, su inhibición plantea desafíos relacionados con la selectividad y la seguridad, lo que dificulta el desarrollo de fármacos eficaces y bien tolerados.

3.1.4. β-Secretasa 1 (BACE-1) y la hipótesis amiloide

La hipótesis amiloide, propuesta en 1991 por John Hardy y David Allsop,(18; 19) sostiene que los oligómeros solubles de Aβ son los principales agentes patogénicos en AD. Estos péptidos, compuestos por 40 a 42 aminoácidos, se generan a partir de la escisión de la proteína precursora de amiloide (APP) mediante la acción secuencial de dos enzimas: β-secretasa 1 (BACE-1) y γ-secretasa.(20) BACE-1 es la principal enzima que determina la tasa de producción de Aβ y ha demostrado actividad tanto *in vitro* como *in vivo*, consolidándose como un blanco terapéutico prometedor en AD.(21; 22)

Dado el papel clave de BACE-1 en la producción de Aβ, su inhibición ha sido una estrategia ampliamente explorada para el desarrollo de terapias modificadoras de la enfermedad. Sin embargo, los ensayos clínicos con inhibidores de BACE-1 han revelado efectos adversos significativos, lo que ha llevado a la suspensión de múltiples estudios y a un replanteamiento de su viabilidad terapéutica. A pesar de estos desafíos, la búsqueda de inhibidores más seguros sigue siendo un área activa de investigación, así como otras estrategias complementarias, como el uso de agentes antiagregantes, inmunoterapia dirigida a los oligómeros de Aβ y anticuerpos catalíticos contra Aβ.

En el presente capítulo se utilizará el Target Fisher para la búsqueda de inhibidores de origen natural de estas enzimas, haciendo especial foco en MAO-B, producto de los resultados obtenidos.

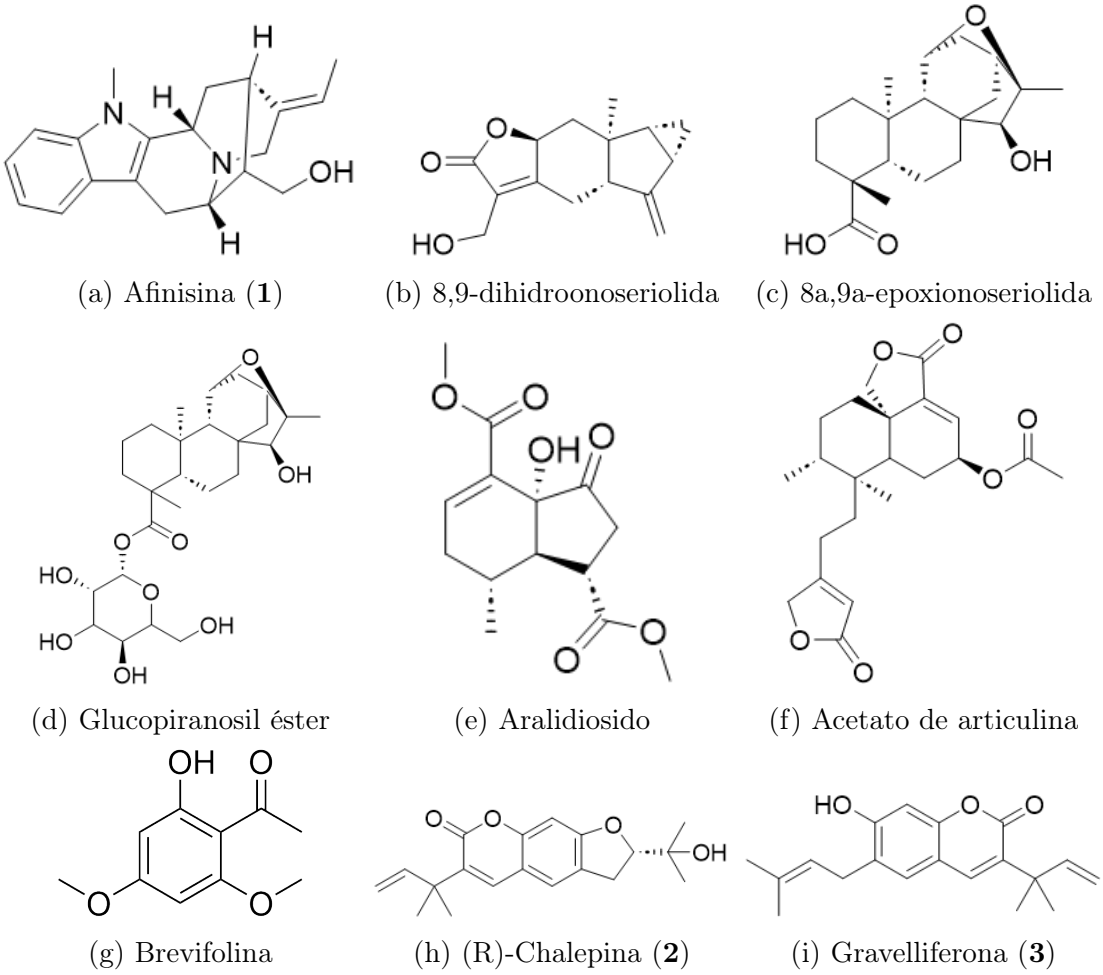
3.2 Armado de una quimioteca *in-house* de productos naturales

A lo largo de su historia, nuestro grupo de investigación se ha abocado al descubrimiento y diversificación estructural de productos naturales de orígenes tanto marinos como terrestres. Esto ha generado un extenso conocimiento sobre los metabolitos secundarios disponibles en el país, y también de las fuentes naturales a las que se tiene acceso, ya sea por estar disponibles en el laboratorio o mediante colaboraciones.

Con el objetivo de buscar productos naturales que puedan inhibir alguna de estas enzimas, se aprovechó este conocimiento para recopilar tanto los metabolitos secundarios disponibles en el laboratorio, como también aquellos que puedan obtenerse de las distintas fuentes naturales a las que se tiene rápido acceso. De esta manera se compiló una quimioteca de 23 compuestos, cuyas estructuras se pueden observar en las figuras 3.3 y 3.4, entre los que se incluyen terpenos de *Hyalis argentea* (compuestos b y c de la figura 3.3 y b y c de la figura 3.4), *Baccharis trimera* (compuesto f de la figura 3.3) y *Verbena bonariensis* (compuestos e de la figura 3.3 y a de la figura 3.4), alcaloides de *Tabernaemontana catharinensis* (compuestos *1* y n de la figura 3.4) y *Balfourodendron riedelianum* (compuestos g, **6**, i, j y k de la figura 3.4) y cumarinas de *Helietta apiculata* (compuestos **2**, **3**, **4** y **5**).

Como se puede apreciar, la mayor parte de los compuestos provienen de plantas colectadas en la provincia de Misiones. Esto se debe a que esta provincia tiene uno de los reservorios mejor conservados de la selva tropical atlántica. Este valioso y a la vez ame-

nazado ecosistema, es el de mayor biodiversidad del país, y es compartido con Brasil y Paraguay.



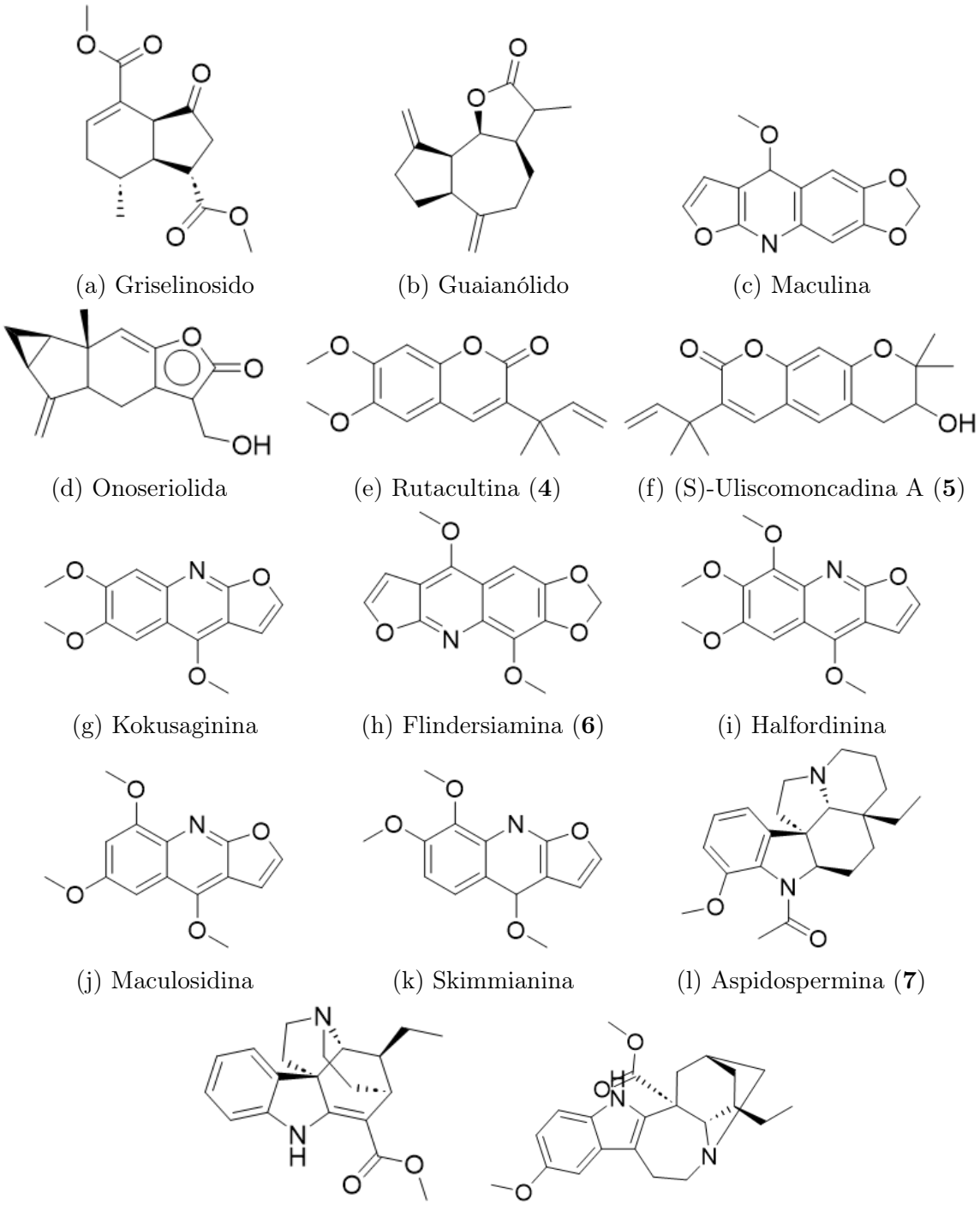
(a) Afinisina (**1**) (b) 8,9-dihidroonoseriolida (c) 8a,9a-epoxionoseriolida

(d) Glucopiranosil éster (e) Aralidiosido (f) Acetato de articulina

(g) Brevifolina (h) (R)-Chalepina (**2**) (i) Gravelliferona (**3**)

Figura 3.3: Compuestos incluidos en la primera parte de la quimioteca de productos naturales.

La fundación Bosques Nativos Argentinos para la Biodiversidad, como vía para ganar apoyo tanto económico como político para sus esfuerzos de preservación y reforestación, tiene un proyecto enfocado al descubrimiento y el uso sustentable de recursos naturales de la selva atlántica. Dicha fundación posee una reserva de más de 1000 Ha en plena Selva Misionera, a 35 km de Oberá, donde se desarrollan proyectos de reforestación con especies autóctonas, de zonas parcialmente degradadas de este ecosistema. En dicha reserva, la Fundación posee un vivero de especies autóctonas, y allí también se encuentra el CIAR,



(a) Griselinosido (b) Guaianólido (c) Maculina

(d) Onoseriolida (e) Rutacultina (**4**) (f) (S)-Uliscomoncadina A (**5**)

(g) Kokusaginina (h) Flindersiamina (**6**) (i) Halfordinina

(j) Maculosidina (k) Skimmianina (l) Aspidospermina (**7**)

(m) 11-metoxitubotaiwina (n) Voacangina

Figura 3.4: Compuestos incluidos en la segunda parte de la quimioteca de productos naturales.

Centro de Investigaciones Antonia Ramos. Este centro dispone de un laboratorio con equipamiento para la realización de investigaciones biológicas *in situ* y para la realización de colectas, y también cuenta con instalaciones y alojamiento con capacidad para ocho investigadores.(23)

Dicho centro es ofrecido para su uso a diferentes grupos de investigación que desarrollan estudios en áreas tan variadas como botánica, genética, calidad del agua, zoología y uso sustentable de la biodiversidad, entre otras. En este contexto, el descubrimiento de nuevos metabolitos bioactivos y derivados es uno de los objetivos para generar valor agregado a los recursos naturales. Por este motivo, desde hace más de 10 años nuestro grupo trabaja en colaboración con el CIAR, hecho que ha resultado en múltiples publicaciones.(24; 25; 26) En el desarrollo del presente trabajo de Tesis, se tuvo la oportunidad de participar en una de las campañas de recolección, llevada a cabo en el mes de febrero de 2022. En ella se recolectaron varias de las especies mencionadas.

3.3 Predicciones de actividad usando Target Fisher

Los compuestos seleccionados fueron convertidos a su forma tridimensional a partir de sus estructuras en formato SMILES y minimizados utilizando el software Balloon,(27) y luego fueron evaluados por Target Fisher. Las predicciones obtenidas se observan en las figuras 3.5 a 3.10.

Como se puede observar en los resultados de Target Fisher presentados en las figuras 3.9 y 3.10, los únicos compuestos que se ubicaron en el cuadrante superior derecho fueron las cumarinas **2**, **3**, **4** y **5**. En particular, se identificó a la gravelliferona como un potencial

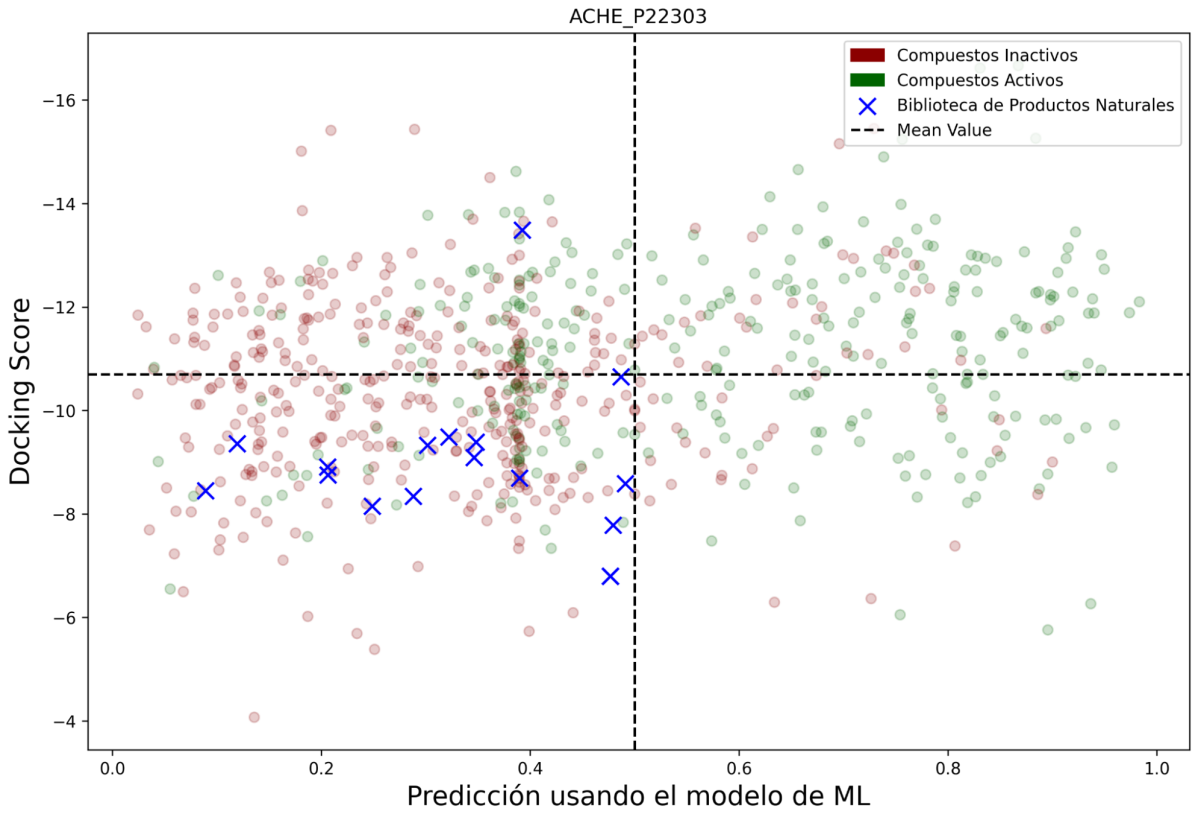


Figura 3.5: Resultados de Target Fisher del cribado de la de la biblioteca de productos naturales en Acetilcolinesterasa.

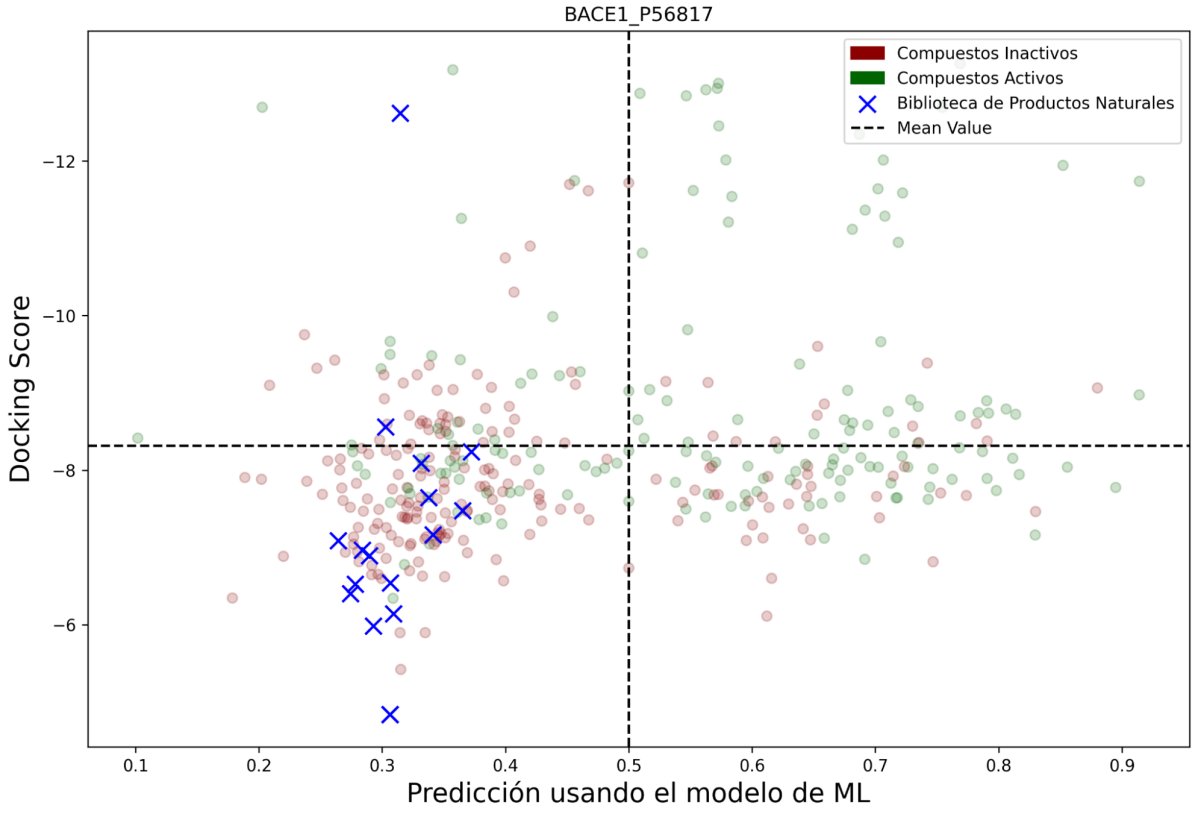


Figura 3.6: Resultados de Target Fisher del cribado de la biblioteca de productos naturales en Beta-secretasa 1.

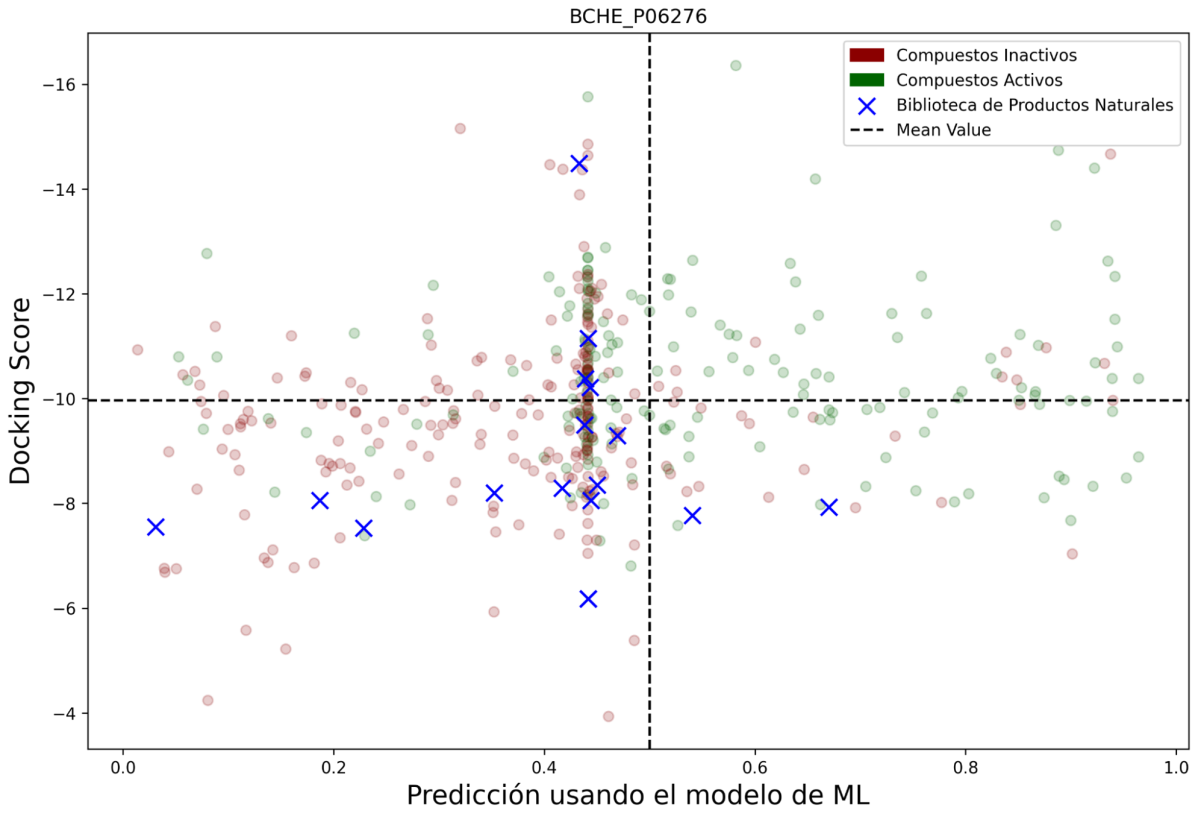


Figura 3.7: Resultados de Target Fisher del cribado de la biblioteca de productos naturales en Butirilcolinesterasa.

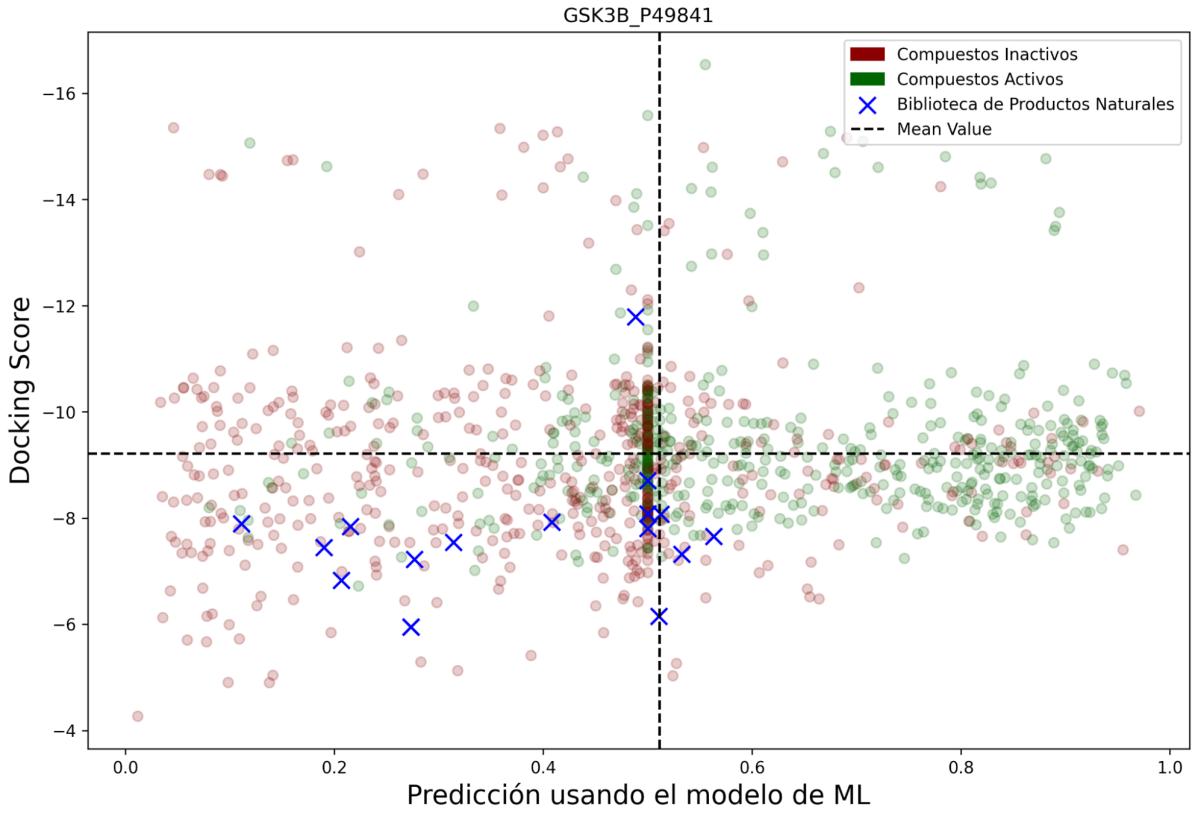


Figura 3.8: Resultados de Target Fisher del cribado de la de la biblioteca de productos naturales en GSK3-β

.

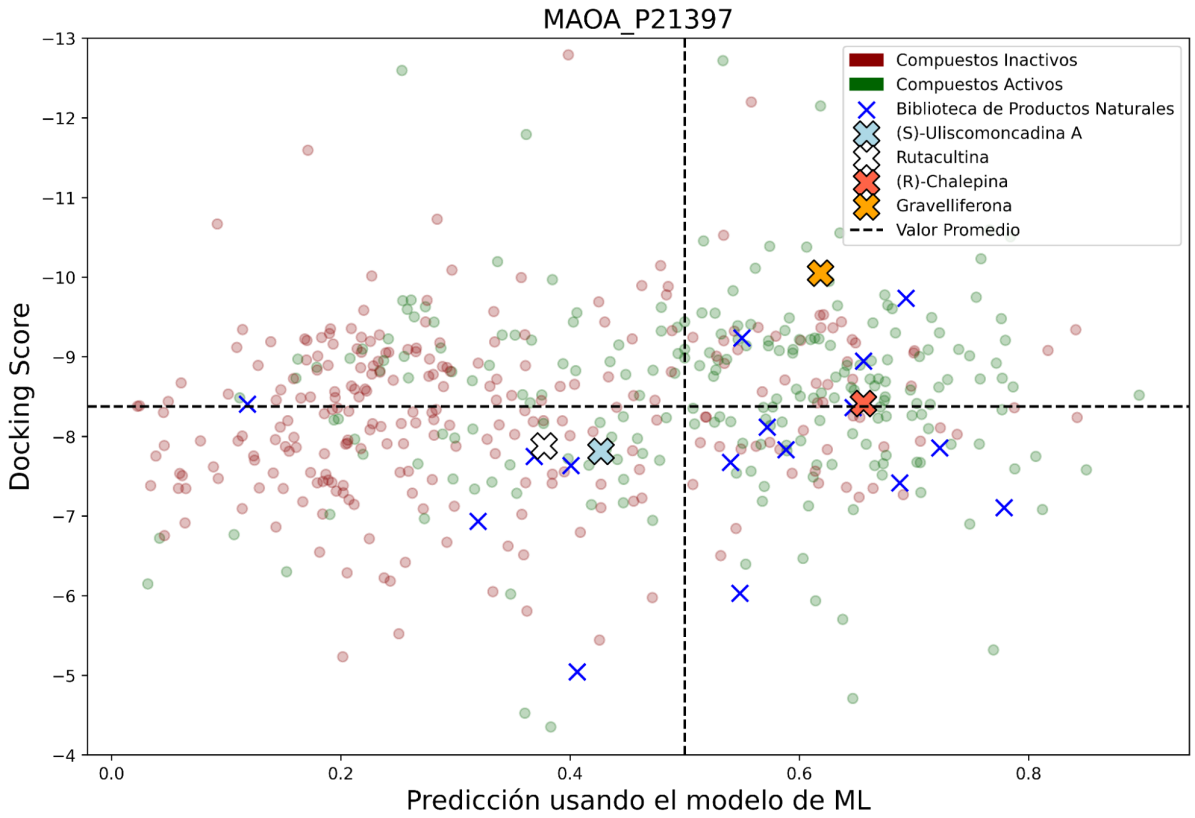


Figura 3.9: Resultados de Target Fisher del cribado de la de la biblioteca de productos naturales en Monoamino Oxidasa A.

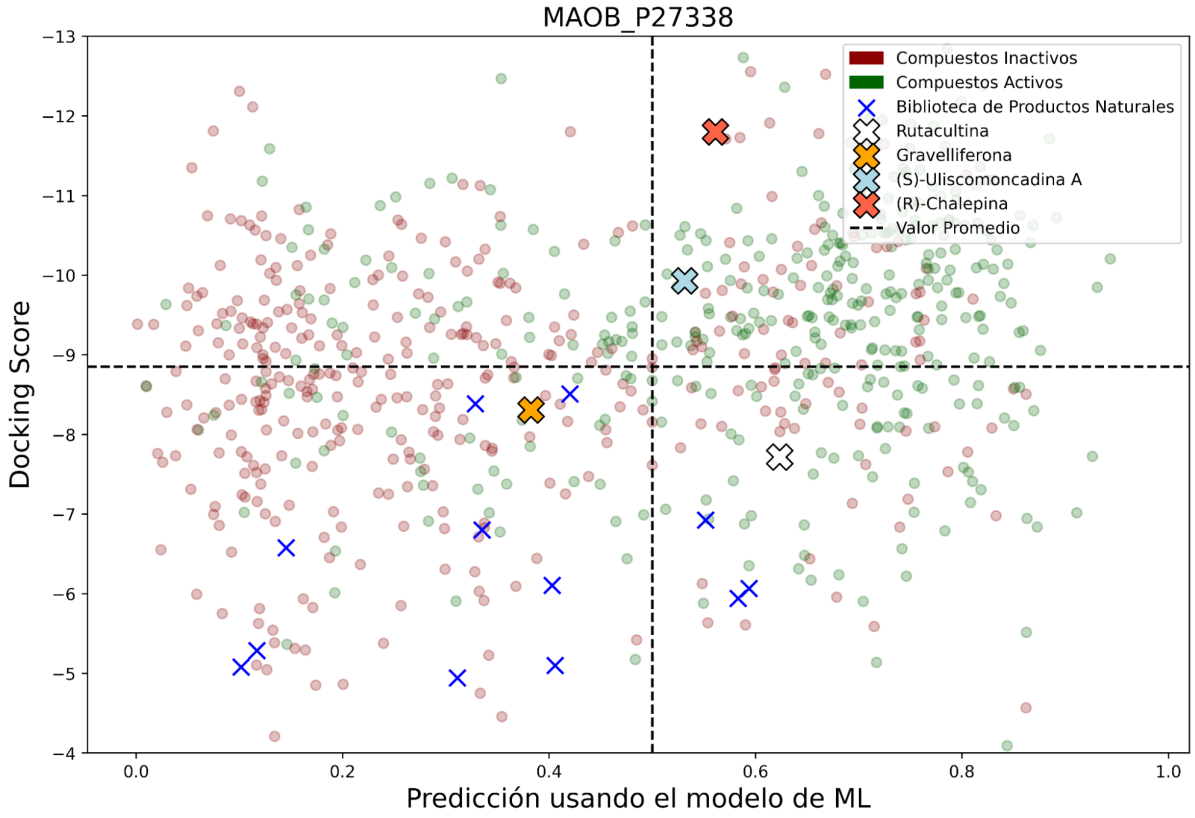


Figura 3.10: Resultados de Target Fisher del cribado de la biblioteca de productos naturales en Monoamino Oxidasa B.

inhibidor de MAO-A (figura 3.9), mientras que (*S*)-uliscomoncadina A y (*R*)-chalepina se destacaron en el cribado contra MAO-B (figura 3.10).

Producto de estas predicciones, se procedió al reaislamiento de las cuatro cumarinas a partir de *Helietta apiculata*, siguiendo el protocolo detallado en el capítulo 6.

3.4 Ensayos de actividad biológica

En la tabla 3.1 se muestran los resultados de los ensayos de actividad biológica de las cuatro cumarinas mencionadas y de algunos otros metabolitos secundarios que se encontraban disponibles.

Tabla 3.1: Resultados de actividad biológica. Los valores en porcentaje reflejan la disminución de la actividad de la proteína al ser tratada con el compuesto a una concentración de 10 µM.

Compuesto	**IC_{50} (µM)**					
	AChE	**BuChE**	**BACE1**	**GSK3β**	**MAO-A**	**MAO-B**
1	> 10	> 10 (35 %)	> 10			
2					> 50	**0,45 ± 0,03**
3				> 10	> 50	> 50 (49 %)
4	> 10	> 10	> 10	> 10	> 50 (20 %)	**8,68 ± 0,29**
5			> 10		> 50	**0,81 ± 0,08**
6	> 10	> 10		> 10	> 50 (22 %)	> 50 (31 %)
7	> 10	> 10 (21 %)	> 10			

Como se puede observar, se aprovechó para evaluar en un mayor número de ensayos en comparación con las predicciones positivas obtenidas por el Target Fisher. Sin embargo, los únicos compuestos activos fueron los predichos por el programa (*S*)-uliscomoncadina A y (*R*)-chalepina en MAO-B, sumando a la rutacultina cuya predicción había sido concluyente. Por su parte, la gravelliferona resultó ser un caso de falso positivo, ya que no es activa frente a MAO-A. Todos los demás ensayos confirmaron la predicción negativa

predicha por el programa previamente. En el caso de los resultados de leve actividad presentados entre paréntesis, estos podrían, dependiendo del caso, dar origen a nuevos proyectos de diversificación estructural racional de estos compuestos. En este sentido, cobra importancia una de las ventajas mencionadas en el capítulo 2 de las metodologías de predicción de blancos basadas en estructura, dado que las predicciones traen consigo el posible modo de unión de la molécula con el blanco molecular, permitiendo plantear posibles modificaciones en base a su interacción con residuos del sitio activo.

3.5 Estudio del modo de unión de los compuestos activos

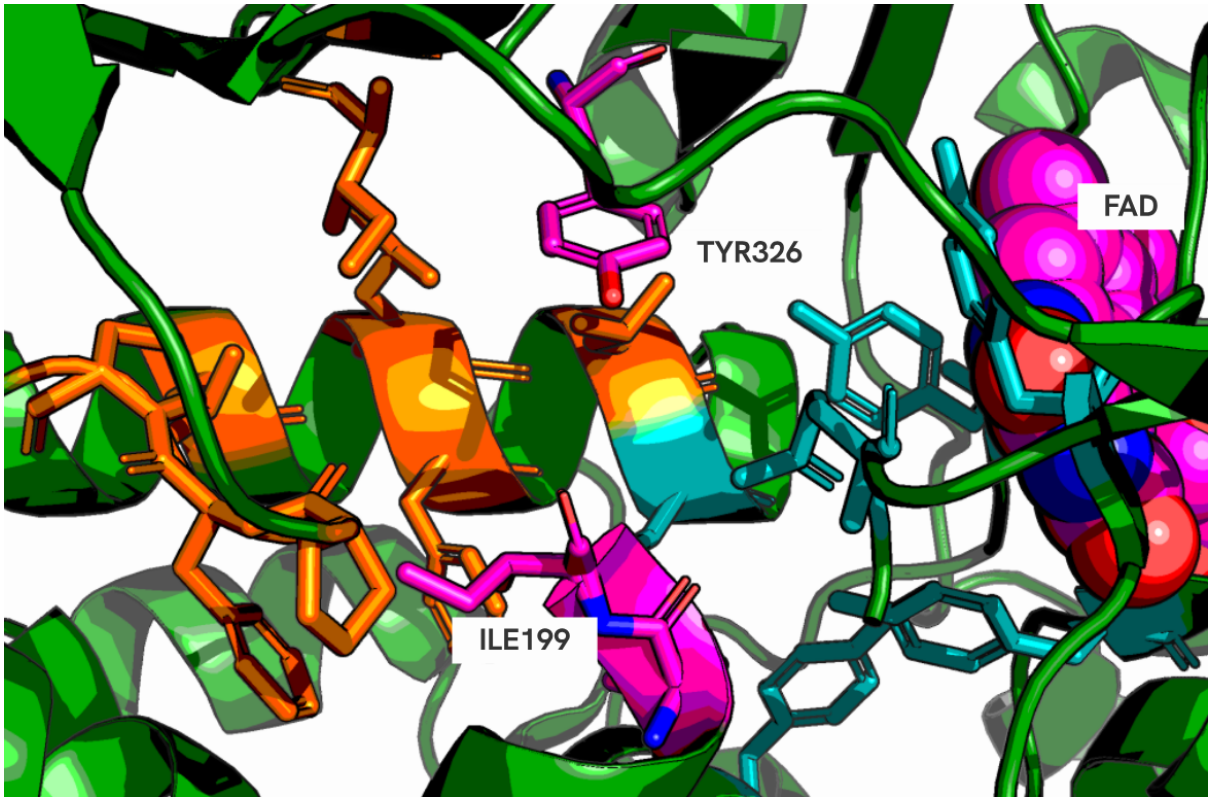


Figura 3.11: Representación del sitio activo de MAO-B. Los residuos no polares, representados con color naranja se encuentran separados de los polares, representados con color celeste, por los residuos ILE199 y TYR326.

Para estudiar en profundidad estas relaciones estructura-actividad, es esencial una exploración detallada de los modos de unión de estos compuestos y del sitio de unión. El bolsillo de unión de la MAO-B, como se muestra en la figura 3.11, es bipartito, con una cavidad de entrada altamente lipofílica y una cavidad de sustrato más pequeña e hidrofílica cerca del cofactor FAD. Los residuos Ile199 y Tyr326 actúan como guardianes, convirtiendo el sitio en una cavidad monopartita de aproximadamente 700 $Å^3$.(28) En contraste, la MAO-A presenta una cavidad monopartita más pequeña (∼400 $Å^3$). Como resultado, los ligandos más pequeños que ocupan únicamente la cavidad del sustrato en la MAO-B tienden a ser menos selectivos, mientras que los ligandos más grandes que interactúan con la cavidad de entrada, en general, no inhiben la MAO-A.

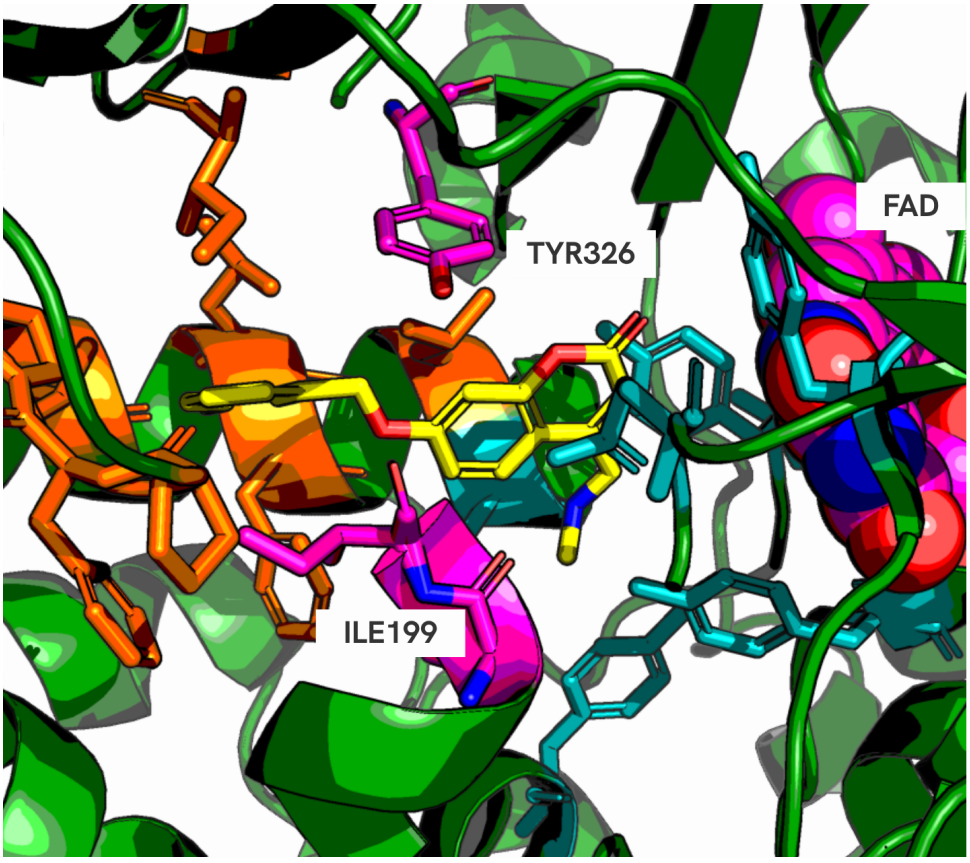


Figura 3.12: Modo de unión típicamente observado para cumarinas en MAO-B (PDB 2V61).

Como se mencionó previamente, la inhibición de monoamino oxidasas por parte de cumarinas es un campo ampliamente estudiado, existiendo incluso estructuras cristalográficas de la proteína unida a este tipo de compuestos (por ejemplo PDB IDs 2V60, 2V61, 7P4F y 7P4H). El típico modo de unión de las cumarinas reportadas, tal como se muestra en la figura 3.12, consiste en la orientación de la lactona hacia el cofactor FAD, quedando el resto de la cumarina apuntando hacia el lado contrario, ubicando el sustituyente no polar hacia el bolsillo lipofílico.

Por su parte, las cumarinas estudiadas en el presente trabajo podrían presentar un modo de unión distinto. Como se puede ver en la figura 3.13, de acuerdo a los resultados arrojados por Autodock Vina, la pose de mejor puntuación para todos los compuestos activos es aquella en donde la lactona se encuentra ubicada hacia el lado contrario al cofactor, siendo las interacciones con la cavidad hidrofílica reemplazadas por los grupos hidroxilo y éter del anillo alifático en los casos en donde esto fuera posible (chalepina y la uliscomoncadina A).

La razón detrás de este cambio en el modo de unión se debe a la presencia del grupo dimetil alilo, altamente lipofílico y ubicado en el carbono 3 del anillo lactónico. Dicho grupo se encuentra presente en todas las cumarinas estudiadas en este capítulo pero en ninguna de las reportadas anteriormente en bibliografía como ligandos. Además, esta característica probablemente explique la selectividad de estos compuestos hacia MAO-B, dado que pueden ocupar esta cavidad, ausente en MAO-A. Apoyando esta hipótesis se encuentra que la decursina, un análogo de la uliscomoncadina A, cuya estructura se muestra la figura 3.14 que carece del grupo dimetil alilo, presenta una inhibición mayor de

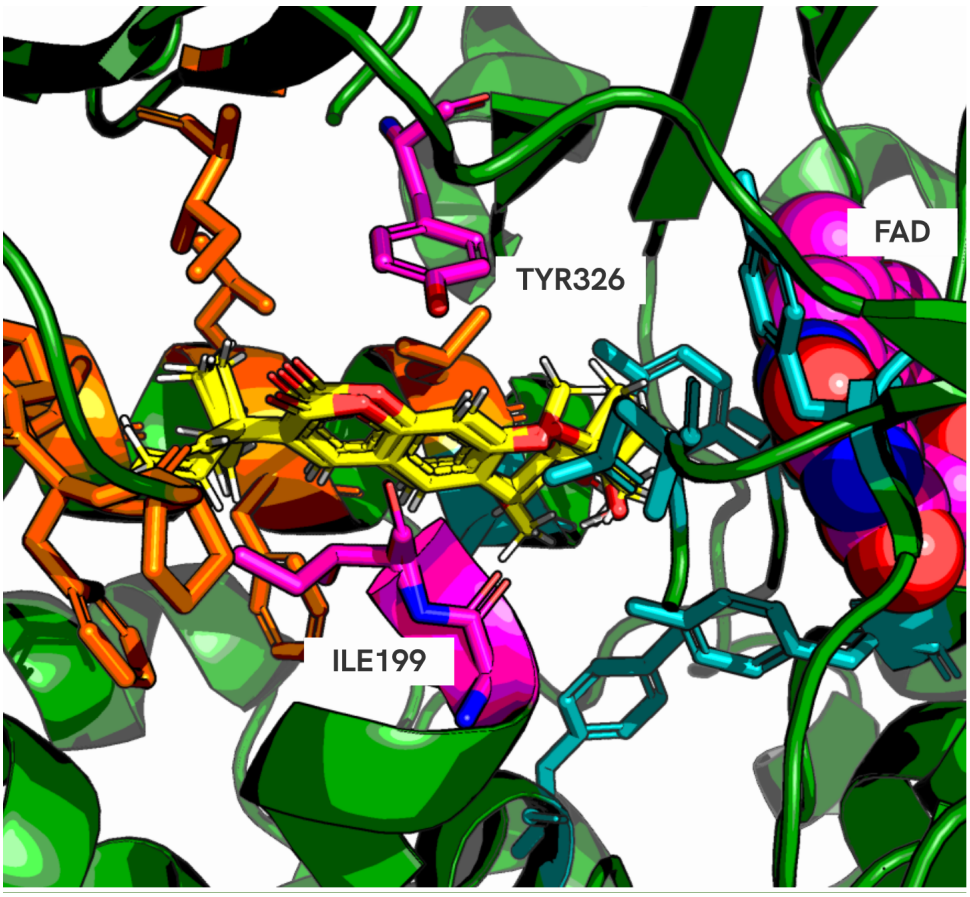


Figura 3.13: Modo de unión superpuesto de las cumarinas reportadas como activas en el presente trabajo.

MAO-A respecto de MAO-B.(29) Finalmente, la posibilidad de reemplazar la interacción original de la lactona por el tándem hidroxilo/éter parece estar correlacionada con la mayor actividad de los compuestos en los que esto fuera posible, como son la chalepina y la uliscomoncadina A.

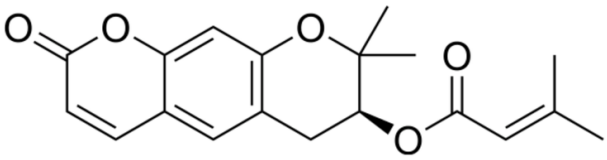


Figura 3.14: Estructura química de la decursina.

3.5.1. Simulaciones de dinámica molecular

Para el estudio en mayor profundidad de la interacción de estos compuestos con el blanco molecular, se procedió a realizar simulaciones de dinámica molecular. Las mismas permiten observar en detalle cómo los compuestos interactúan con el entorno proteico a nivel atómico en un contexto temporal, lo que ofrece una visión de la flexibilidad y el ajuste conformacional del ligando en el sitio activo de la proteína. De esta manera, es posible analizar la estabilidad del complejo proteína-ligando, estudiar las interacciones claves (como enlaces de hidrógeno, interacciones hidrofóbicas, y puentes salinos) y los cambios conformacionales tanto del ligando como de los residuos más cercanos. Además, estas simulaciones abren la puerta al cálculo de energías de interacción y perfiles de energía libre, proporcionando una comprensión detallada del posible modo de acción del ligando que justifique la selectividad hacia MAO-B.

Una manera de analizar la estabilidad del complejo formado es estudiar la variación del RMSD (Root Mean Square Deviation) a lo largo de la trayectoria de la simulación de dinámica molecular. El RMSD mide el desplazamiento promedio de los átomos en el complejo con respecto a una estructura de referencia inicial durante el transcurso de la simulación. Un valor de RMSD bajo y constante indica que el complejo se mantiene estable en una conformación bien definida, mientras que variaciones significativas o tendencias ascendentes en el RMSD sugieren cambios conformacionales o inestabilidad en la unión del ligando a la proteína. El análisis de la variación del RMSD permite evaluar si el ligando permanece adecuadamente acoplado al sitio activo de MAO-B o si tiende a separarse

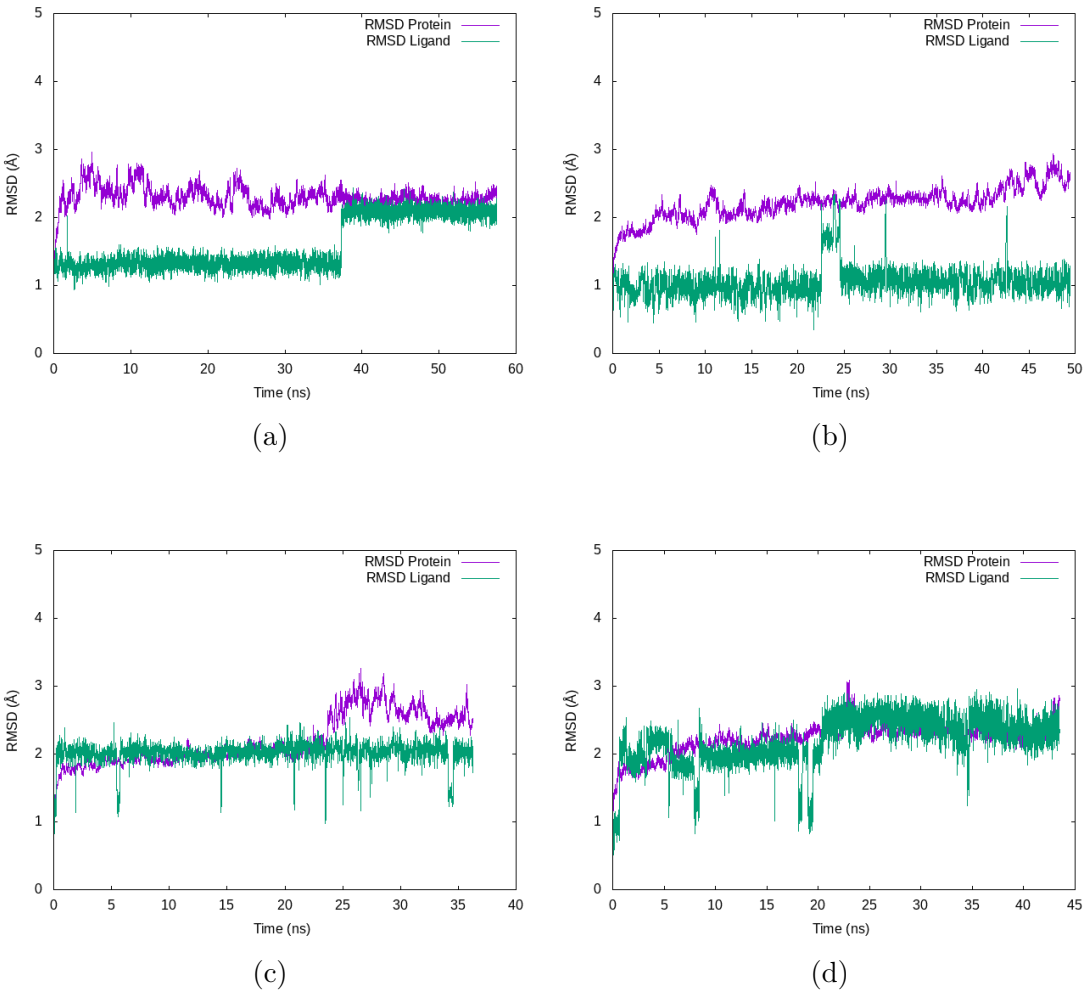


Figura 3.15: Gráficos de RMSD obtenidos durante la dinámica molecular de diferentes ligandos en MAO-B: (a) (S)-uliscomoncadina A, (b) (R)-chalepina, (c) rutacultina y (d) gravelliferona.

o a adoptar distintas conformaciones, lo cual es crítico para comprender la afinidad y selectividad del ligando hacia el blanco molecular.

Del gráfico de la (*S*)-uliscomoncadina A, figura 3.15a, se desprende un único salto antes de los 40 ns que corresponde al cambio conformacional del anillo alifático del compuesto, provocando un cambio en la interacción del grupo OH, permitiendo su interacción con el residuo Gln206. El resto de la estructura no sufre mayores alteraciones, indicando estabilidad en el modo de unión en los tiempos estudiados. En el caso de la (*R*)-chalepina, figura 3.15b, se observan dos saltos, que implican un pequeño cambio conformacional y

luego el restablecimiento a la posición original, manteniéndose estable durante el resto de la simulación. Las poses finales de estos dos compuestos se observan en la figura 3.16. Una observación interesante es que durante la simulación distintas moléculas de agua se incorporaron al sitio de unión, replicando la ubicación observada en las estructuras cristalinas disponibles en bibliografia.

El caso de la rutacultina es similar al de la chalepina, figura 3.15c, sufriendo su modo de unión algunas modificaciones a lo largo de la dinámica que luego fueron restituidas, llegando a un modo de unión similar al inicial pero dando una idea de la baja estabilidad del mismo.

Finalmente, el compuesto que sí sufrió grandes cambios a lo largo de la simulación fue la gravelliferona, figura 3.15d, particularmente por la cadena hidrocarbonada ubicada en el C-6, la cual provocó múltiples cambios en el modo de unión del compuesto, denotando la alta inestabilidad del mismo en el sitio activo, en correlación con lo visto experimentalmente.

3.5.2. Cálculos de energía de interacción

Finalmente, se calcularon las energías de unión (sin considerar efectos entrópicos) de las tres cumarinas activas, proporcionando una medida de la afinidad entre el ligando y la proteína, una menor energía de unión generalmente indica una interacción más fuerte y estable.

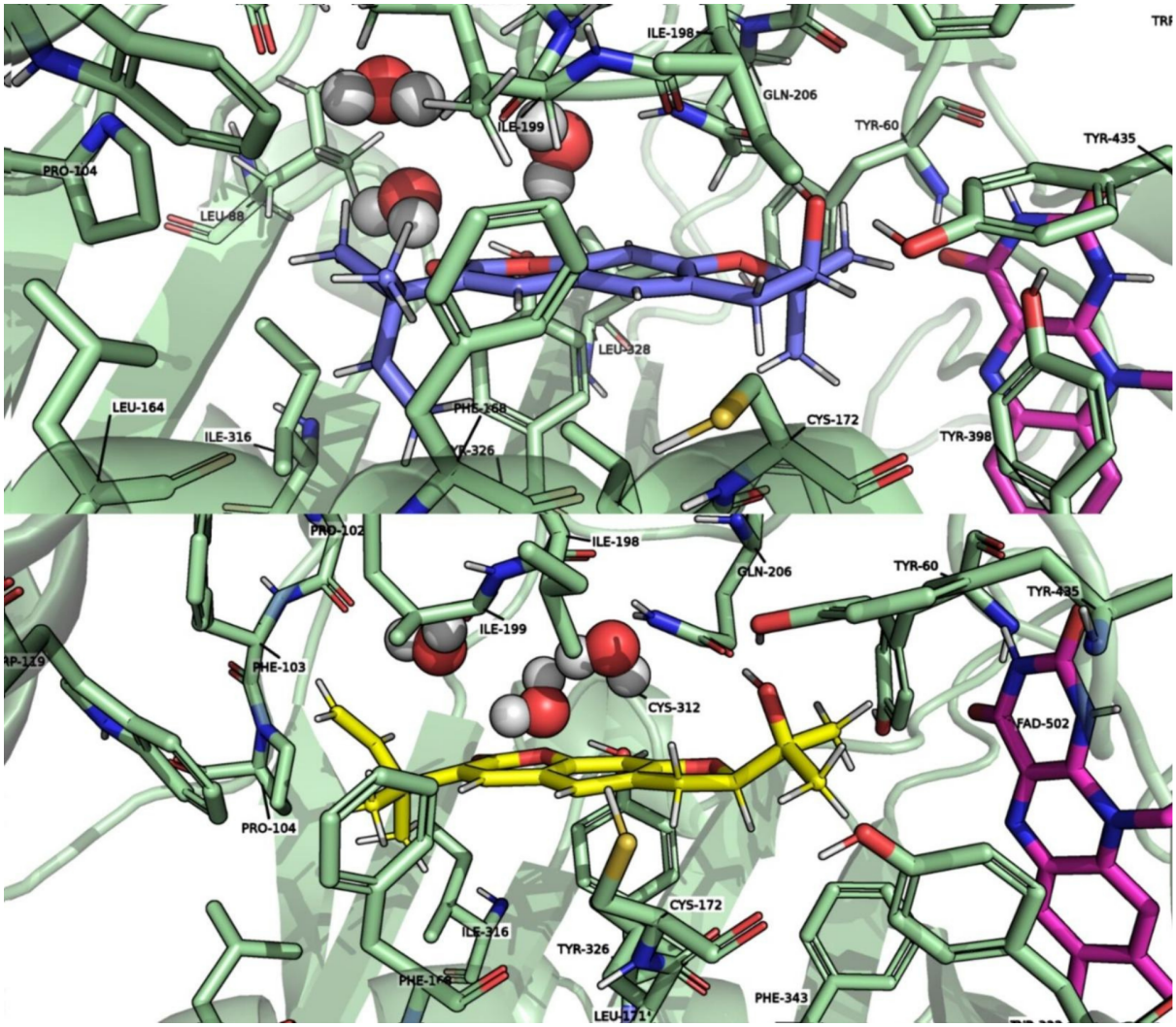


Figura 3.16: Modo de unión obtenido al final para la (S)-uliscomoncadina A (arriba) y la (R)-chalepina (abajo) de las simulaciones de dinámica molecular

Tabla 3.2: Valores de ΔH obtenidos a partir del cálculo de energía.

Compuesto	**ΔH (kcal/mol)**	**IC_{50} (µM)**
(R)-chalepina	-46 ± 3	0,45 ± 0,03
(S)-uliscomoncadina A	-49 ± 3	0,81 ± 0,08
Rutacultina	-34 ± 3	8,68 ± 0,29

Los resultados, mostrados en la tabla 3.2, presentaron una buena correlación con los valores experimentales: la cumarina con mayor valor de IC_{50}, es decir, la que necesita mayor concentración para inhibir MAO-B al 50 %, presentó el valor más alto de energía de unión, lo que indica una afinidad menor en comparación con las restantes. Esta correlación entre la energía de unión calculada y los valores de IC_{50} experimentales apoya la precisión de las simulaciones, permitiendo predecir la eficacia relativa de los ligandos y orientar el diseño de nuevas moléculas con mejores propiedades inhibitorias.

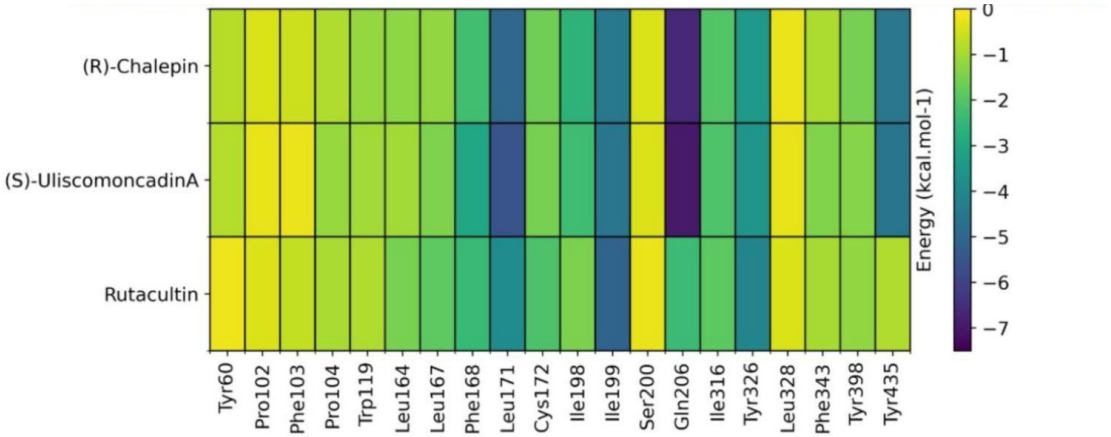


Figura 3.17: Descomposición de energía por residuo derivada del cálculo de MM-GBSA.

El análisis de la descomposición de energía por residuo (figura 3.17) muestra un patrón de interacción similar entre los 3 compuestos activos, con la (R)-chalepina y (S)-uliscomoncadina A mostrando interacciones significativas con los residuos Tyr435, Gln206 y el cofactor FAD, lo que podría explicar su mayor capacidad inhibitoria frente a la enzima.

3.6 Comparación con otras metodologías

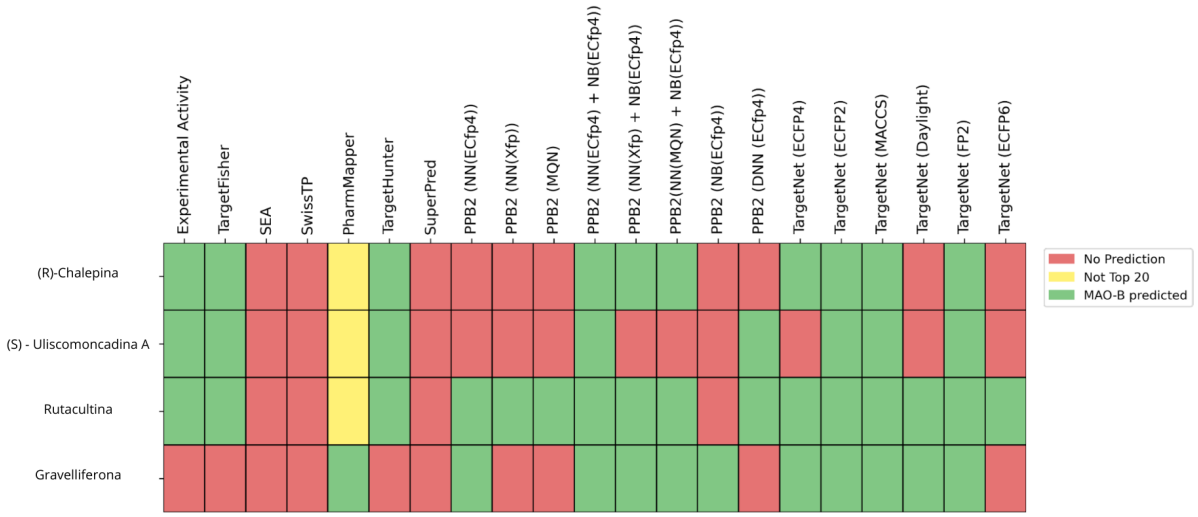


Figura 3.18: Predicciones de actividad de las cumarinas estudiadas en este capitulo mediante diversas metodologías de predicción de blancos moleculares.

Para evaluar el rendimiento de Target Fisher en comparación con otras herramientas de predicción de blancos, utilizamos tanto métodos basados en ligandos, como SEA(30), SwissTargetPrediction(31), TargetHunter(32), SuperPred(33), PPB2(34), TargetNet(35), como la herramienta basada en estructura PharmMapper(36), con el fin de predecir la actividad de los cuatro compuestos de cumarina probados. Dado que la inhibición de las monoaminooxidasas por cumarinas está bien documentada, esperábamos que la mayoría de los enfoques basados en ligandos identificaran las MAOs como posibles blancos. Sin embargo, como se muestra en la figura 3.18, solo TargetHunter y ciertos modelos de PPB2 y TargetNet identificaron con éxito la MAO-B como un blanco de estos compuestos. PharmMapper clasificó la MAO-B dentro de los 50 principales blancos predichos para todos los compuestos, pero la ubicó en el top 10 solo para la gravelliferona 4. Curiosamente, la rutacultina, la cumarina más simple entre las ensayadas, recibió predicciones positivas

más consistentes en todos los métodos, lo que resalta como la complejidad estructural puede dificultar la predicción a través de múltiples herramientas.

Por último, para evaluar la capacidad predictiva de Target Fisher con inhibidores menos convencionales de MAO-B, filtramos en la base de datos PubChem aquellos compuestos activos recientemente reportados con baja similitud estructural con otros inhibidores conocidos (puntuación de similitud de Tanimoto ≤ 0.4). Esta selección dio como resultado un total de 13 inhibidores. Esta comparación es particularmente desafiante, ya que la mayoría de los métodos alternativos dependen de servidores web, lo que limita evaluaciones más amplias, y muchos no proporcionan acceso a sus conjuntos de entrenamiento para una comparación justa. No obstante, como se muestra en la figura 3.19, el modelo específico de Target Fisher superó con éxito a todos los demás métodos en la predicción precisa de estos inhibidores estructuralmente únicos. Este resultado ilustra la robustez de Target Fisher para detectar actividad, incluso para compuestos fuera de los perfiles típicos de inhibidores de MAO-B.

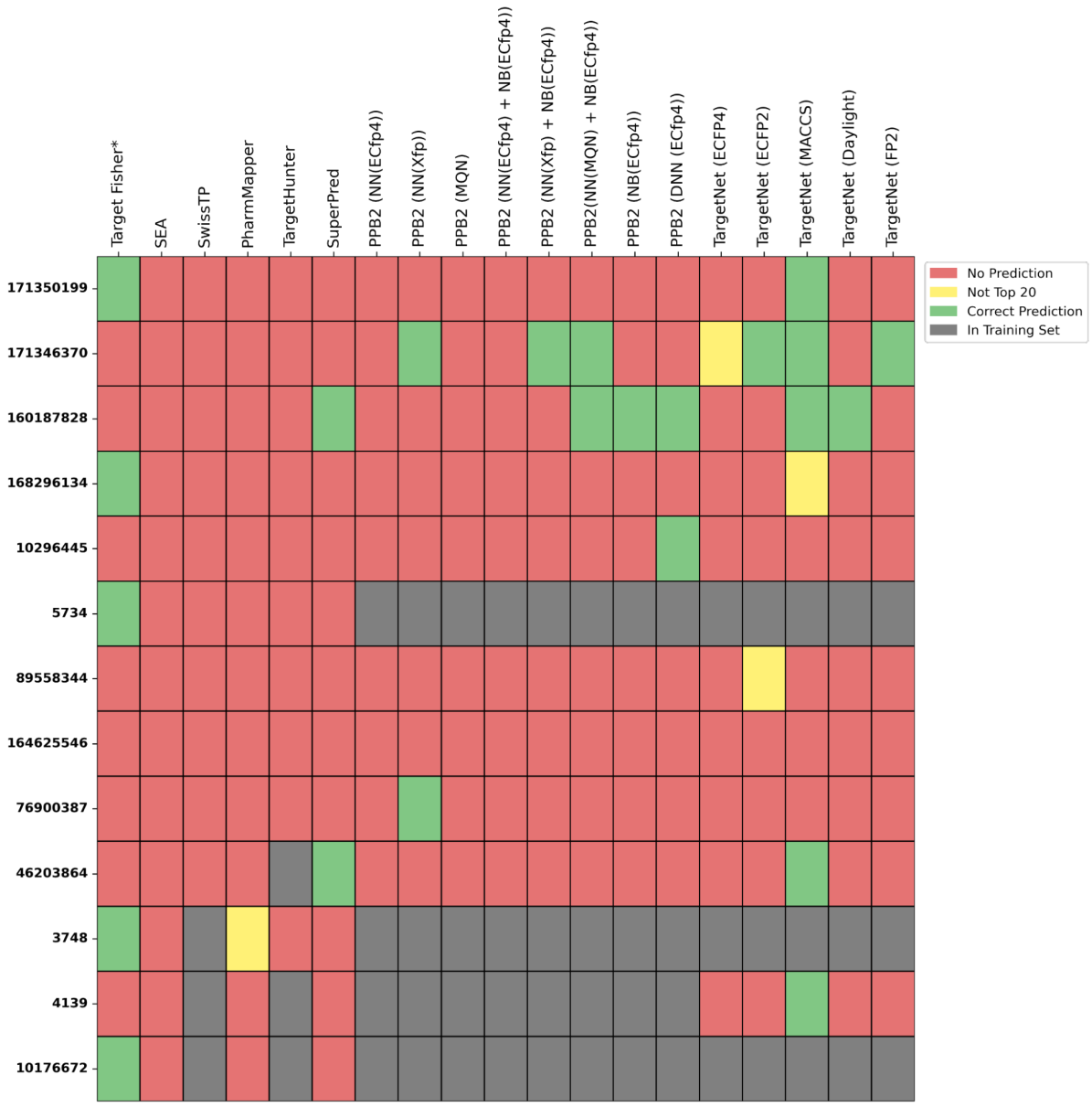


Figura 3.19: Evaluación de compuestos recientemente añadidos a PubChem, activos en MAO-B y con estructura novedosa respecto a los previamente ensayados, en distintas herramientas de predicción de blancos moleculares.

Bibliografía

[1] Davies, P.; Maloney, A. J. F. SELECTIVE LOSS OF CENTRAL CHOLINERGIC NEURONS IN ALZHEIMER'S DISEASE. *The Lancet* **1976**, *308*, 1403.

[2] Zhang, P.; Xu, S.; Zhu, Z.; Xu, J. Multi-Target Design Strategies for the Improved Treatment of Alzheimer's Disease. *Eur. J. Med. Chem.* **2019**, *176*, 228–247.

[3] Pope, C. N.; Brimijoin, S. Cholinesterases and the Fine Line between Poison and Remedy. *Biochem. Pharmacol.* **2018**, *153*, 205–216.

[4] Martins, M. M.; Branco, P. S.; Ferreira, L. M. Enhancing the Therapeutic Effect in Alzheimer's Disease Drugs: The Role of Polypharmacology and Cholinesterase Inhibitors. *ChemistrySelect* **2023**, *8*, e202300461.

[5] Qizilbash, N.; Birks, J.; Lopez Arrieta, J.; Lewington, S.; Szeto, S. WITHDRAWN: Tacrine for Alzheimer's Disease. *Cochrane Database Syst. Rev.* **2007**, CD000202.

[6] Liu, P.-P.; Xie, Y.; Meng, X.-Y.; Kang, J.-S. History and Progress of Hypotheses and Clinical Trials for Alzheimer's Disease. *Signal Transduct. Target. Ther.* **2019**, *4*, 29.

[7] Kandiah, N.; Pai, M.-C.; Senanarong, V.; Looi, I.; Ampil, E.; Park, K. W.; Karanam, A. K.; Christopher, S. Rivastigmine: the advantages of dual inhibition of acetylcholinesterase and butyrylcholinesterase and its role in subcortical vascular dementia and Parkinson's disease dementia. *Clinical Interventions in Aging* **2017**, *12*, 697–707.

[8] Kepp, K. P. Bioinorganic Chemistry of Alzheimer's Disease. *Chem. Rev.* **2012**, *112*, 5193–5239.

[9] Chen, X.-Q.; Chen, Y.; Liraz, O.; Moussa-Pacha, N. M.; Weinstock, M. Reconsideration of Anticholinesterase Therapeutic Strategies against Alzheimer's Disease. *ACS Chemical Neuroscience* **2018**, *9*, 1553–1563.

[10] Jones, D. N.; Raghanti, M. A. The Role of Monoamine Oxidase Enzymes in the Pathophysiology of Neurological Disorders. *J. Chem. Neuroanat.* **2021**, *114*, 101957.

[11] Binda, C.; Hubálek, F.; Li, M.; Edmondson, D. E.; Mattevi, A. Crystal Structure of Human Monoamine Oxidase B, a Drug Target Enzyme Monotopically Inserted into the Mitochondrial Outer Membrane. *FEBS Lett.* **2004**, *564*, 225–228.

[12] Goozee, K. G.; Shah, T. M.; Sohrabi, H. R.; Rainey-Smith, S. R.; Brown, B.; Verdile, G.; Martins, R. N. Examining the Potential Clinical Value of Curcumin in the Prevention and Diagnosis of Alzheimer's Disease. *Br. J. Nutr.* **2016**, *115*, 449–465.

[13] Griñán-Ferré, C.; Bellver-Sanchis, A.; Izquierdo, V.; Corpas, R.; Roig-Soriano, J.; Chillón, M.; Andres-Lacueva, C.; Somogyvári, M.; Sőti, C.; Sanfeliu, C.; Pallàs, M. The Pleiotropic Neuroprotective Effects of Resveratrol in Cognitive Decline and Alzheimer's Disease Pathology: From Antioxidant to Epigenetic Therapy. *Ageing Res. Rev.* **2021**, *67*, 101271.

[14] Ramsay, R. R. Molecular Aspects of Monoamine Oxidase B. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2016**, *69*, 81–89.

[15] Schedin-Weiss, S.; Inoue, M.; Hromadkova, L.; Teranishi, Y.; Yamamoto, N. G.; Wiehager, B.; Bogdanovic, N.; Winblad, B.; Sandebring-Matton, A.; Frykman, S.; Tjernberg, L. O. Monoamine Oxidase B Is Elevated in Alzheimer Disease Neurons, Is Associated with -Secretase and Regulates Neuronal Amyloid -Peptide Levels. *Alzheimers Res. Ther.* **2017**, *9*, 57.

[16] Giacobini, E.; Gold, G. Alzheimer Disease Therapy–Moving from Amyloid- to Tau. *Nat. Rev. Neurol.* **2013**, *9*, 677–686.

[17] Hooper, C.; Killick, R.; Lovestone, S. The GSK3 Hypothesis of Alzheimer's Disease. *J. Neurochem.* **2008**, *104*, 1433.

[18] Hardy, J. A.; Higgins, G. A. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in Pharmacological Sciences* **1992**, *13*, 383–388.

[19] Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297*, 353–356.

[20] Hubert, M.; Pikuła, J.; Konopka, T.; Jaźwiński, D.; Skrobot, M.; Skwierawska, O. Is It the Twilight of BACE1 Inhibitors? *Curr. Neuropharmacol.* **2021**, *19*, 54–80.

[21] Moussa-Pacha, N. M.; Abdin, S. M.; Omar, H. A.; Alniss, H.; Al-Tel, T. H. BACE1 Inhibitors: Current Status and Future Directions in Treating Alzheimer's Disease. *Med. Res. Rev.* **2020**, *40*, 339–384.

[22] Prati, F.; Bottegoni, G.; Bolognesi, M. L.; Cavalli, A. BACE-1 Inhibitors: From

Recent Single-Target Molecules to Multitarget Compounds for Alzheimer's Disease. *J. Med. Chem.* **2018**, *61*, 619–637.

[23] Argentinos, B. N. CIAR - Centro de Investigaciones Antonia Ramos. https://www.bosques.org.ar/ciar/, 2025; Accessed: 2025-02-24.

[24] Valdez, B.; Musikant, A. D.; Palermo, J. A. Chemical Modifications of 12-Methoxy-N-methylvoachalotine (12-MMV): Enhancing Efficacy against Phytopathogenic Fungi. *ACS Agricultural Science & Technology* **2023**, *4*, 102–109.

[25] Belén Valdez, M.; Bernal Giménez, D. M.; Fernández, L. R.; Musikant, A. D.; Ferri, G.; Sáenz, D.; Di Venosa, G.; Casas, A.; Avigliano, E.; Edreira, M. M.; Palermo, J. A. Antiparasitic Derivatives of the Furoquinoline Alkaloids Kokusaginine And Flindersiamine. *ChemMedChem* **2022**, *17*, e202100784.

[26] Pis Diez, C. M.; Céspedes, M.; Di Venosa, G. M.; Calvo, G.; Avigliano, E.; Casas, A. G.; Palermo, J. A. Synthesis and Cytotoxicity Evaluation of Olivacine-Indole Hybrids Tethered by Alkyl Linkers. *Nat. Prod. Res.* **2022**, *36*, 3657–3664.

[27] Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.

[28] Milczek, E. M.; Binda, C.; Rovida, S.; Mattevi, A.; Edmondson, D. E. The 'Gating' Residues Ile199 and Tyr326 in Human Monoamine Oxidase B Function in Substrate and Inhibitor Recognition. *FEBS J.* **2011**, *278*, 4860–4869.

[29] Baek, S. C.; Lee, H. W.; Ryu, H. W.; Kang, M.-G.; Park, D.; Kim, S. H.; Cho, M.-

L.; Oh, S.-R.; Kim, H. Selective Inhibition of Monoamine Oxidase A by Hispidol. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 584–588.

[30] Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nature biotechnology* **2007**, *25*, 197–206.

[31] Daina, A.; Michielin, O.; Zoete, V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic acids research* **2019**, *47*, W357–W364.

[32] Wang, L.; Ma, C.; Wipf, P.; Liu, H.; Su, W.; Xie, X.-Q. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *The AAPS journal* **2013**, *15*, 395–406.

[33] Gallo, K.; Goede, A.; Preissner, R.; Gohlke, B.-O. SuperPred 3.0: drug classification and target prediction—a machine learning approach. *Nucleic acids research* **2022**, *50*, W726–W731.

[34] Awale, M.; Reymond, J.-L. Polypharmacology browser PPB2: target prediction combining nearest neighbors with machine learning. *Journal of chemical information and modeling* **2018**, *59*, 10–17.

[35] Yao, Z.-J.; Dong, J.; Che, Y.-J.; Zhu, M.-F.; Wen, M.; Wang, N.-N.; Wang, S.; Lu, A.-P.; Cao, D.-S. TargetNet: a web service for predicting potential drug–target interaction profiling via multi-target SAR models. *Journal of computer-aided molecular design* **2016**, *30*, 413–424.

[36] Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic acids research* **2017**, *45*, W356–W360.

Capítulo 4

Cribado virtual multiblanco en una quimioteca sintética: identificación de inhibidores de GSK-3β

El laboratorio de neurofármacos del Instituto de Química Médica del CSIC (Madrid, España), responsable de los ensayos descritos en el capítulo anterior, cuenta con una quimioteca de 1444 moléculas pequeñas. Esta colección, fruto de más de tres décadas de trabajo en diversos proyectos, incluye una amplia variedad de familias químicas, así como intermediarios de síntesis y subproductos de reacción aislados y almacenados.

En este capítulo se describe el cribado virtual realizado sobre esta quimioteca utilizando la herramienta **Target Fisher**, considerando los seis blancos moleculares evaluados e introducidos en el capítulo anterior. Esto se realizó con el objetivo de probar la utilidad de la herramienta para el cribado de quimiotecas de moléculas pequeñas, permitiendo la búsqueda de nuevas actividades biológicas distintas a las originalmente planteadas en su síntesis. Como resultado de esta campaña se identificaron varios candidatos prometedores, los cuales fueron ensayados *in vitro* dando como resultado una serie de potentes inhibidores de la proteína **GSK-3**β, una diana clave en enfermedades neurodegenerativas.

Posteriormente, se llevó a cabo un análisis exhaustivo de las relaciones estructura-actividad (SAR) de dichos compuestos para entender los factores moleculares determinantes de la interacción con GSK-3β. Este análisis incluyó simulaciones de dinámica molecular para explorar el modo de interacción de los inhibidores más prometedores. Finalmente, se estableció una ruta alternativa de síntesis que permitió la generación de nuevos compuestos diseñados para explotar las relaciones estructura-actividad identificadas. Estos compuestos ampliaron la diversidad química disponible y mantuvieron la afinidad hacia GSK-3β, validando el enfoque integrado de cribado virtual y optimización sintética.

4.1 Análisis fisicoquímico de la quimioteca del grupo de neurofármacos

Un paso fundamental en la caracterización de una quimioteca es el análisis de sus propiedades fisicoquímicas, ya que estas determinan su viabilidad como candidatos a fármacos. En este caso, se evaluaron los compuestos de la quimioteca en estudio y se compararon con dos bases de datos de referencia: **DrugBank,**(1) que contiene moléculas aprobadas para uso médico, y **B3DB**,(2) que incluye compuestos clasificados según su capacidad para atravesar la barrera hematoencefálica (BHE), dividiéndolos en permeables (BBB^+) y no permeables (BBB^-).

En la figura 4.1 se muestra la comparación de las propiedades fisicoquímicas normalizadas de las bases de datos, donde se observa que los compuestos de la quimioteca analizada presentan un perfil muy similar al de los compuestos BBB^+. Esto es consistente con el hecho de que la quimioteca fue desarrollada por un grupo de neurofármacos, por lo que se esperaba que los compuestos tuvieran características compatibles con la permeabilidad a la BHE. Sin embargo, cabe destacar que muchos de estos compuestos fueron sintetizados hace más de 20 años, lo que resalta la importancia de su reevaluación en el contexto de nuevos modelos predictivos y criterios actuales de desarrollo de fármacos para el sistema nervioso central (SNC).

Al analizar las propiedades individuales, se observa que la gran mayoría de los compuestos en la quimioteca cumplen con las Reglas de Lipinski,(3) explicadas previamente en el capitulo 1.

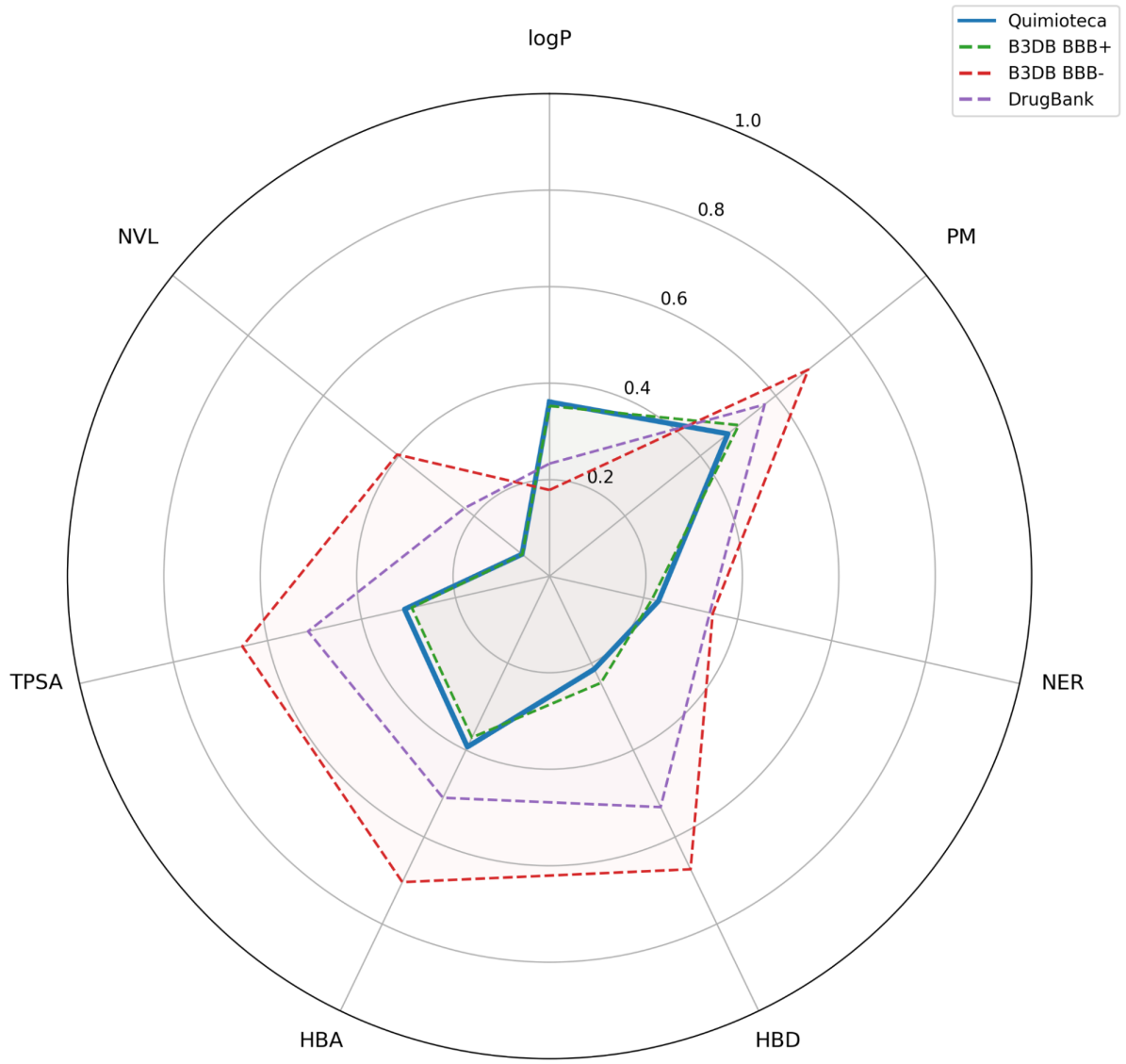


Figura 4.1: Gráfico de radar que compara las propiedades fisicoquímicas normalizadas de diferentes bases de datos de compuestos. La línea sólida representa los valores promedio de la quimioteca en estudio, mientras que las líneas discontinuas corresponden a otras bases de datos. Las propiedades representadas incluyen Peso Molecular (PM), Logaritmo del coeficiente de partición octanol-agua (LogP), Número de Donores de Enlace de Hidrógeno (HBD), Número de Aceptores de Enlace de Hidrógeno (HBA), Área de Superficie Polar Topológica (TPSA), Número de Violaciones de las Reglas de Lipinski (NVL) y Número de Enlaces Rotables (NER).

Estas propiedades en general se consideran indicadores importantes de la capacidad de una molécula para ser absorbida, distribuida y excretada eficientemente en el organismo. El cumplimiento de estas reglas sugiere que los compuestos de la quimioteca, en caso de demostrarse una actividad biológica relevante, tienen el potencial de convertirse en candidatos a fármacos.

4.2 Predicción de actividad biológica y selección de compuestos para evaluación experimental

Con el objetivo de identificar posibles compuestos bioactivos en el contexto de enfermedades neurodegenerativas, se realizaron predicciones de actividad biológica utilizando **Target Fisher**. En este caso, se utilizaron nuevamente los seis blancos moleculares de relevancia en enfermedades neurodegenerativas, para los cuales se dispone de ensayos: acetilcolinesterasa (AChE), butirilcolinesterasa (BuChE), monoamino oxidasa A (MAO-A), monoamino oxidasa B (MAO-B), β-secretasa 1 (BACE1) y glicógeno sintasa quinasa 3 beta (GSK-3β).

Al igual que en el capítulo anterior, para cada blanco molecular se generó un gráfico de distribución de **predicción vs. docking score** con cuatro cuadrantes (figuras 4.2–4.5), permitiendo visualizar la relación entre la probabilidad de interacción de los compuestos y su afinidad calculada por docking. En cada uno de estos gráficos se destacan los compuestos seleccionados para evaluación experimental.

A partir de estos resultados, se seleccionaron un conjunto de compuestos siguiendo

los siguientes criterios:

1. **Coincidencia entre predicción y docking score**: Se priorizaron aquellos compuestos que mostraban altas probabilidades de interacción con los blancos de interés y que, en algunos casos, presentaban además valores favorables en el docking.

2. **Disponibilidad del compuesto**: Se consideró la cantidad disponible de cada compuesto en la quimioteca.

3. **Representación dentro de una misma familia**: Cuando varios compuestos de la misma familia química cumplían con los criterios anteriores, se seleccionó uno de ellos para optimizar los recursos experimentales.

En el caso de **AChE** (figura 4.2), se seleccionaron los compuestos **8** y **9**, ya que ambos presentaron un *docking score* inferior a la media y una buena predicción de actividad. También se evaluaron los compuestos **10** y **11** debido a que fueron predichos como activos por el modelo de inteligencia artificial.

En el cribado contra **BuChE** (figura 4.4), se seleccionaron los compuestos **10**, **11** y **12**. Cabe destacar que, dado que AChE y BuChE suelen estudiarse en conjunto, todos los compuestos seleccionados para uno de los blancos fueron también evaluados en el otro. Además, otros compuestos como **13** fueron evaluados en BuChE al haber sido preparados inicialmente para otros blancos moleculares.

Para **BACE1** (figura 4.3), se seleccionaron los compuestos **12** y **13**. El compuesto **11**, ya disponible por su preparación para otros ensayos, también fue evaluado.

En el caso de las **monoamino oxidasas**, **14** y **15** fueron seleccionados para su evalua-

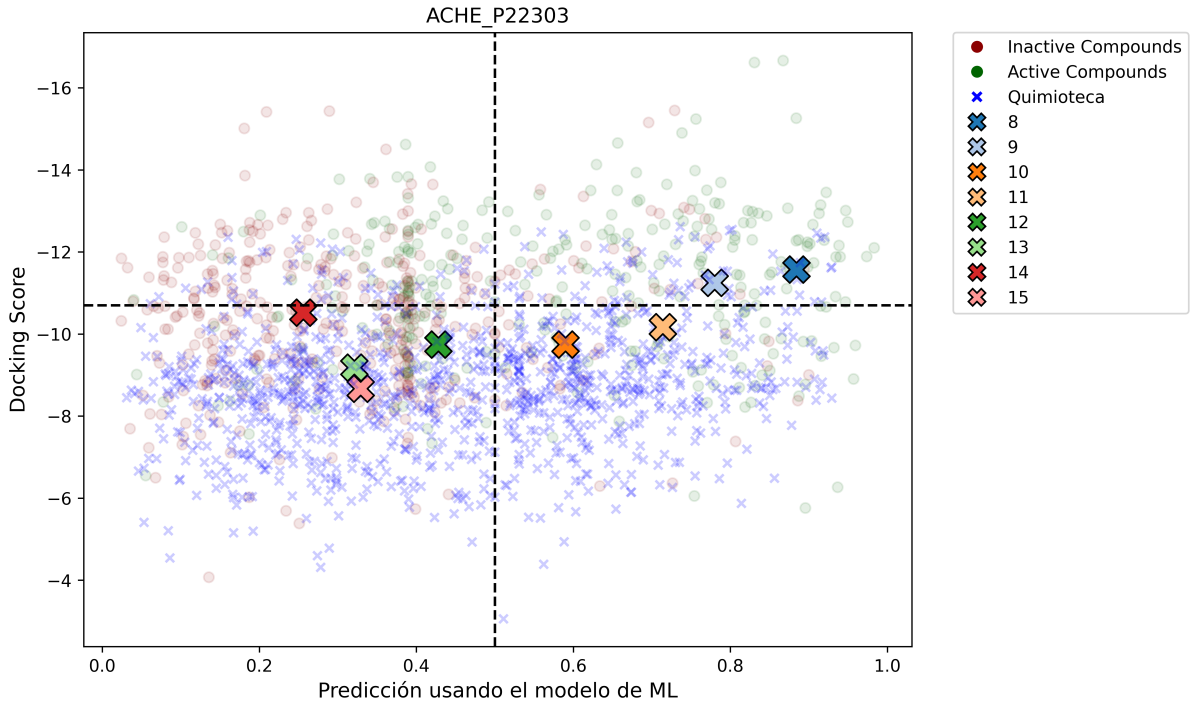


Figura 4.2: Resultados de Target Fisher del cribado de la quimioteca en Acetilcolinesterasa.

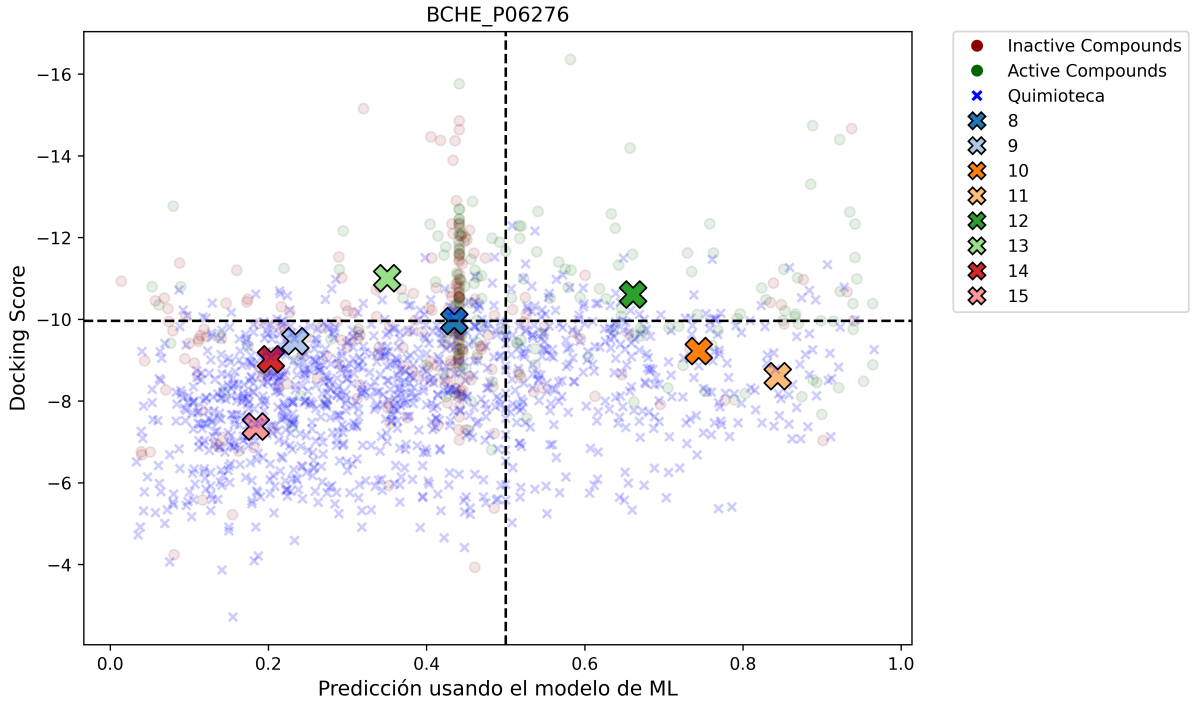


Figura 4.3: Resultados de Target Fisher del cribado de quimioteca en Beta-secretasa 1.

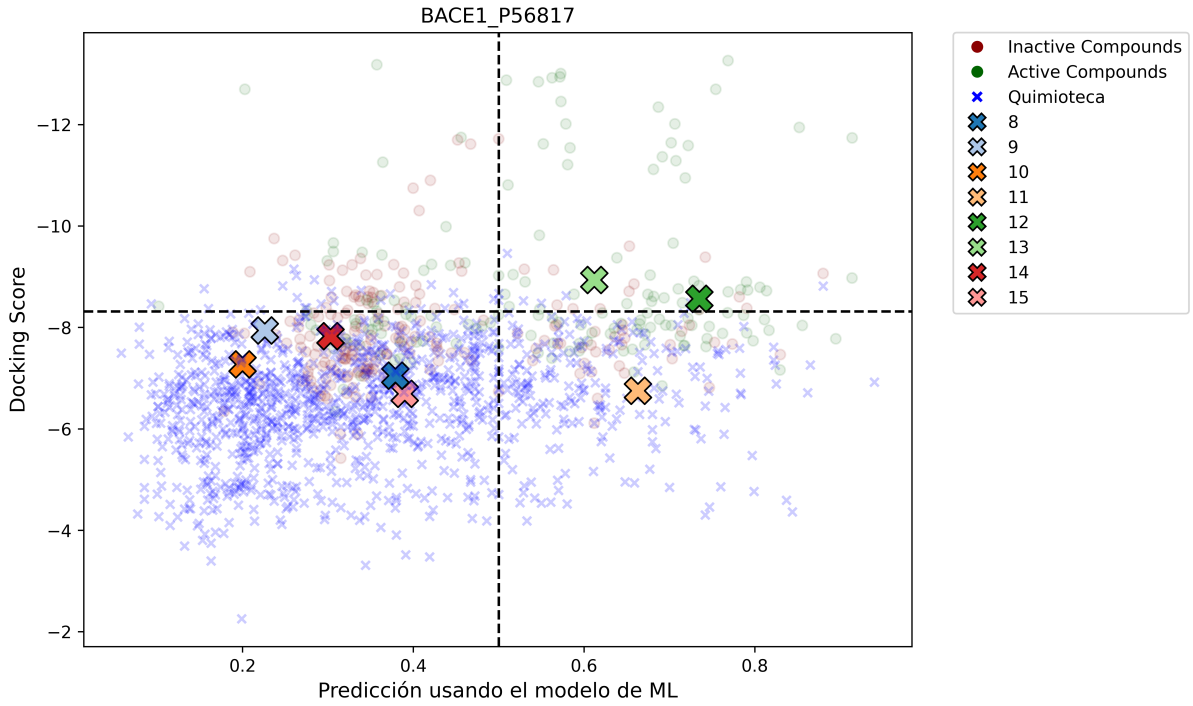


Figura 4.4: Resultados de Target Fisher del cribado de la quimioteca en Butirilcolinesterasa.

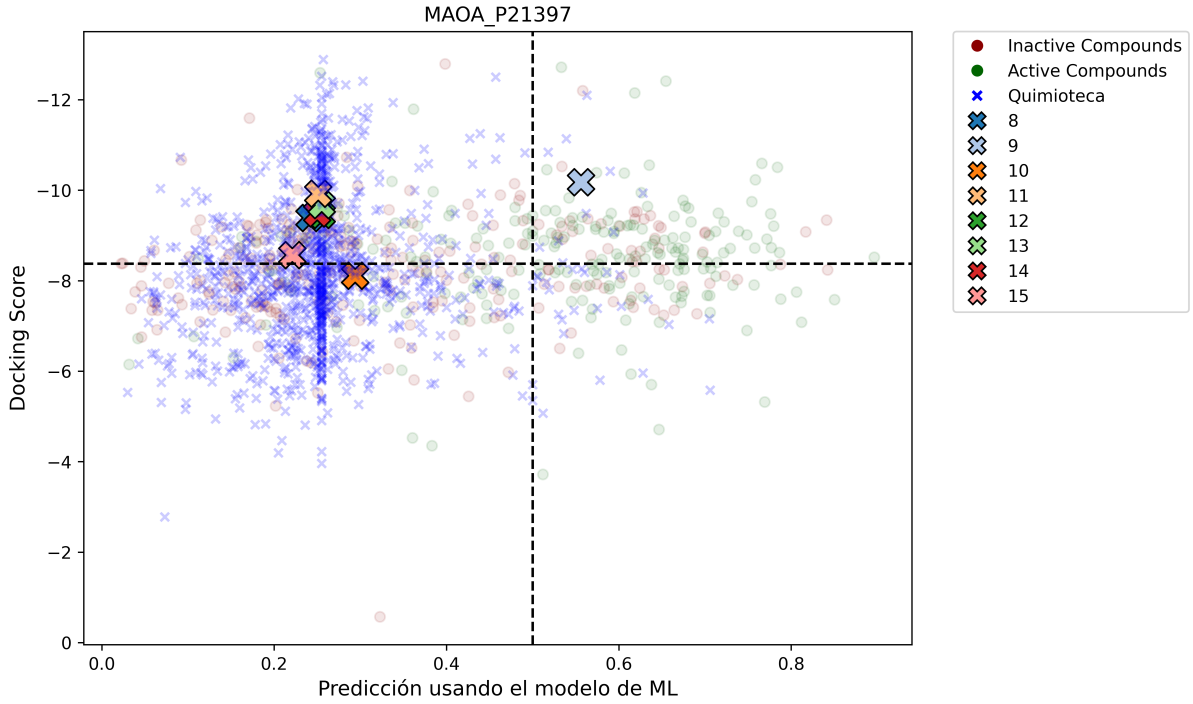


Figura 4.5: Resultados de Target Fisher del cribado de la quimioteca en GSK-3β.

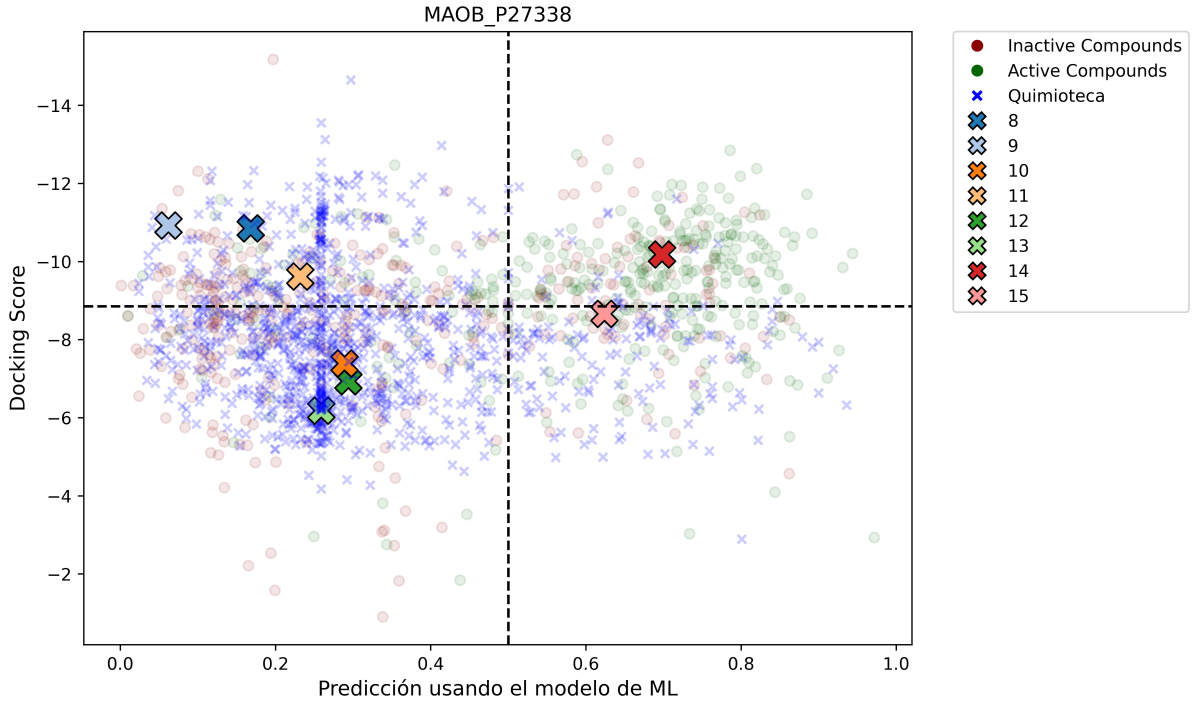


Figura 4.6: Resultados de Target Fisher del cribado de la quimioteca en Monoamino Oxidasa A.

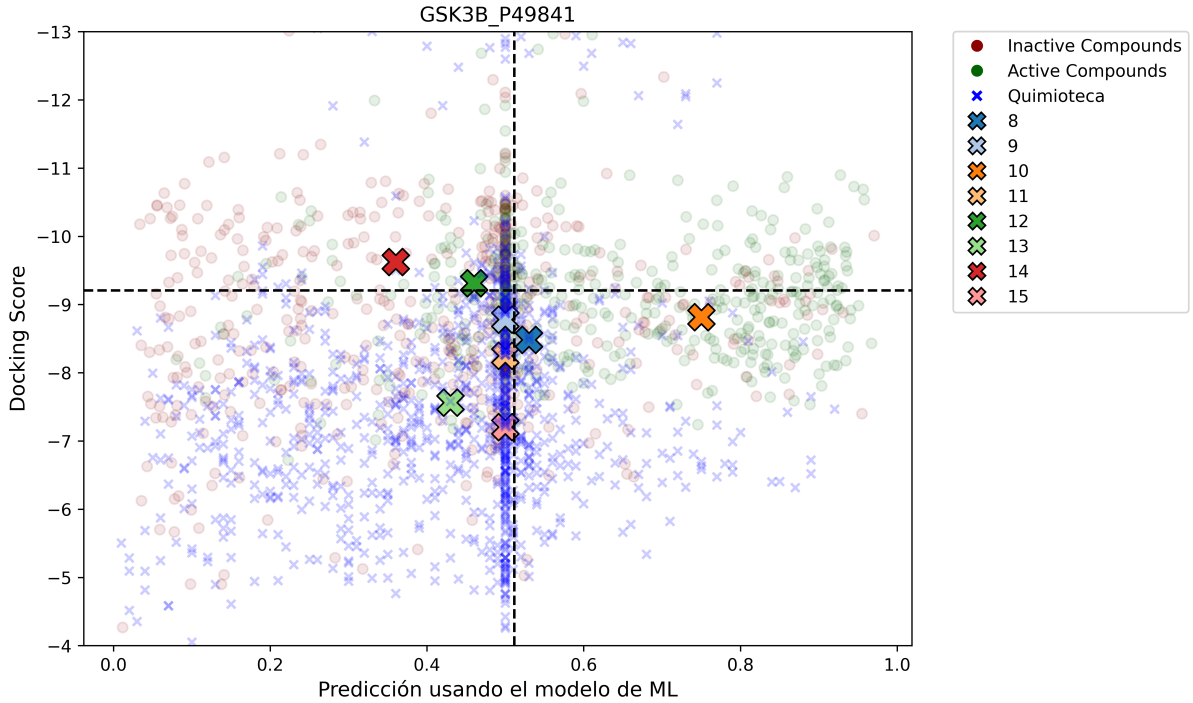


Figura 4.7: Resultados de Target Fisher del cribado de la quimioteca en Monoamino Oxidasa B.

ción en **MAO-B** (figura 4.7), mientras que el compuesto **9** fue evaluado frente a **MAO-A** (figura 4.6). Como ocurre con las colinesterasas, los compuestos seleccionados para una de las MAOs también fueron evaluados en la otra. Asimismo, los compuestos **10**, **12** y **13**, preparados para otros blancos, fueron incluidos en estos ensayos cruzados.

Finalmente, en el cribado frente a **GSK-3**β (figura 4.5), se seleccionó el compuesto **10** como candidato para evaluación experimental, a pesar de no presentar un buen puntaje de *docking*. Esta decisión se basó en su alta probabilidad de actividad predicha por el modelo de inteligencia artificial, la disponibilidad suficiente del compuesto para ensayos, y el hecho de que su estructura química se encontraba representada múltiples veces dentro de la quimioteca. Esto último resultaba especialmente interesante, ya que, en caso de confirmarse su actividad, podría sugerir una familia estructural con potencial frente a este blanco molecular.

En la Tabla 4.1 se presentan los resultados de actividad de los compuestos seleccionados. De todos ellos, solo el compuesto **10** mostró un IC_{50} menor al umbral establecido para GSK-3β. Por otro lado, el compuesto **13** mostró también una leve inhibición de BuChE, lo cual coincide con su buena interacción con la enzima (según el *docking*), aunque contradice la predicción del modelo.

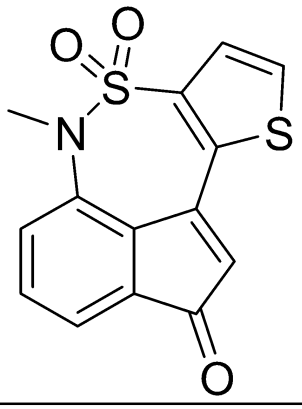
La estructura del compuesto activo, **10**, se muestra en la Tabla 4.2. Para preservar la confidencialidad de la quimioteca, solo se presenta la estructura de este compuesto y no la de los restantes que fueron evaluados. Se puede observar que **10** presenta un sistema aromático fusionado y grupos funcionales clave, como una **sulfonamida *N*-metilada y una cetona**, que podrían desempeñar un papel crucial en su actividad biológica como

Tabla 4.1: Resultados de actividad de los compuestos seleccionados.

Compuesto	AChE	BuChE	MAO-A	MAO-B	BACE1	GSK3β	Predicción
8	> 10	> 10 (21 %)					AChE
9			> 50	> 50			MAO-A
10*	> 10	> 10	> 50 (26 %)	> 50		7.91±0.28	GSK3β
11	> 10	> 10	> 50	> 50	> 10		BACE1, AChE
12*	> 10	> 10	> 50	> 50	> 10		BACE1
13	> 10	> 10 (26 %)	> 50	> 50	> 10	> 10	BACE1
14			> 50	> 50			MAO-B
15	> 10	> 10	> 50	> 50			

*Mal disuelto en DMSO

Tabla 4.2: Estructura y actividad del compuesto **10** en GSK-3β.

Compuesto	**GSK-3β IC_{50} (µM)**
	**7.91 ± 0.28**

inhibidor de la enzima.

Este compuesto fue sintetizado originalmente en el contexto de un proyecto llevado a cabo por el grupo del Dr. Vega, predecesor al de la Dra. María Isabel Rodríguez Franco, pero nunca llegó a ser publicado. Dicho proyecto, desarrollado entre los años 80 y 90, tenía como objetivo el diseño de derivados de la imipramina (figura 4.8) con menor toxicidad, manteniendo su potencial terapéutico como antidepresivos tricíclicos.(4; 5; 6; 7; 8) De esta manera, como se mostrará en la sección a continuación, en la quimioteca se encontraba toda una familia de compuestos sintetizados durante este proyecto, algunos de ellos publicados e inclusive con pruebas de actividad *in vivo*.

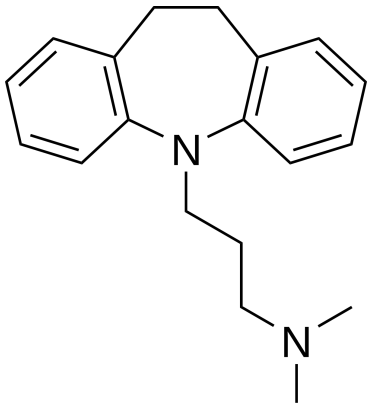


Figura 4.8: Estructura química de la imipramina.

4.3 Ensayos *in vitro* de una familia de benzotiazepinas presentes en la quimioteca

Para profundizar en los resultados obtenidos con **10** y explorar la posibilidad de que otros compuestos estructuralmente relacionados también presentaran actividad en **GSK-3**β, se realizó una búsqueda de moléculas en la quimioteca que compartieran el mismo esqueleto central. Para ello se realizó un análisis de similitud molecular utilizando **RDKit**, describiendo a las moléculas mediante **Morgan Fingerprints** (radio 2, 2048 bits) y seleccionando aquellas con mayor similitud estructural. Este procedimiento permitió recuperar 16 compuestos adicionales con el mismo esqueleto de **benzotiazepina** con diversas modificaciones en regiones clave de la molécula, figura 4.9. Al comparar estos compuestos con **10**, se observan variaciones en:

- El anillo de tiofeno, que en algunos casos fue reemplazado por un **benceno** o un **anillo imidazólico**.
- El carbono puente entre los anillos aromáticos, en algunos casos dio lugar a la

formación de un cuarto ciclo y en otros permitió la introducción de una cadena alquílica.

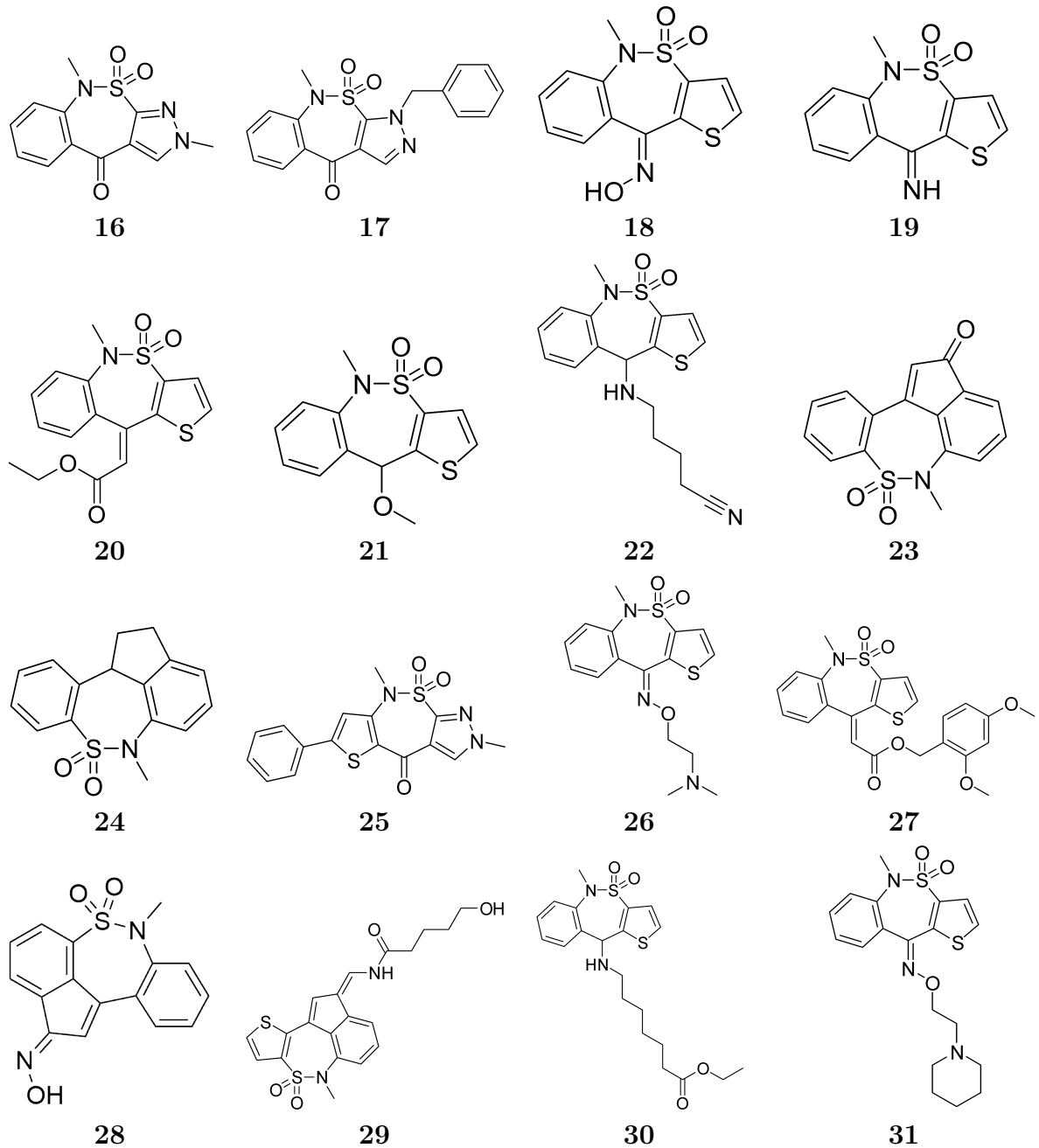


Figura 4.9: Estructuras de los compuestos análogos con esqueleto de benzotiazepina.

Los resultados de actividad, presentados en la Tabla 4.3, muestran que, además de **10**, dos compuestos adicionales, **28** y **23**, exhibieron actividad significativa frente a GSK-3β. En particular, **28** demostró una potencia superior, con un orden de magnitud menor en

Tabla 4.3: Actividad inhibitoria de los análogos de benzotiazepina frente a GSK-3β.

Compuesto	**GSK-3β IC$_{50}$ (µM)**	**Compuesto**	**GSK-3β IC$_{50}$ (µM)**
16	> 10	**24**	> 10
17	> 10	**25**	> 10 (21 %)
18	> 10	**26**	> 10
19	> 10	**27**	> 10
20	> 10	**28**	**0.055 ± 0.010**
21	> 10	**29**	> 10
22	> 10	**30**	> 10
23	**7.33 ± 0.16**	**31**	> 10

su IC50. Es importante destacar que ninguno de estos tres compuestos activos había sido publicado previamente, lo que sugiere que este núcleo estructural podría representar un punto de partida interesante para el diseño de futuros inhibidores de **GSK-3β**.

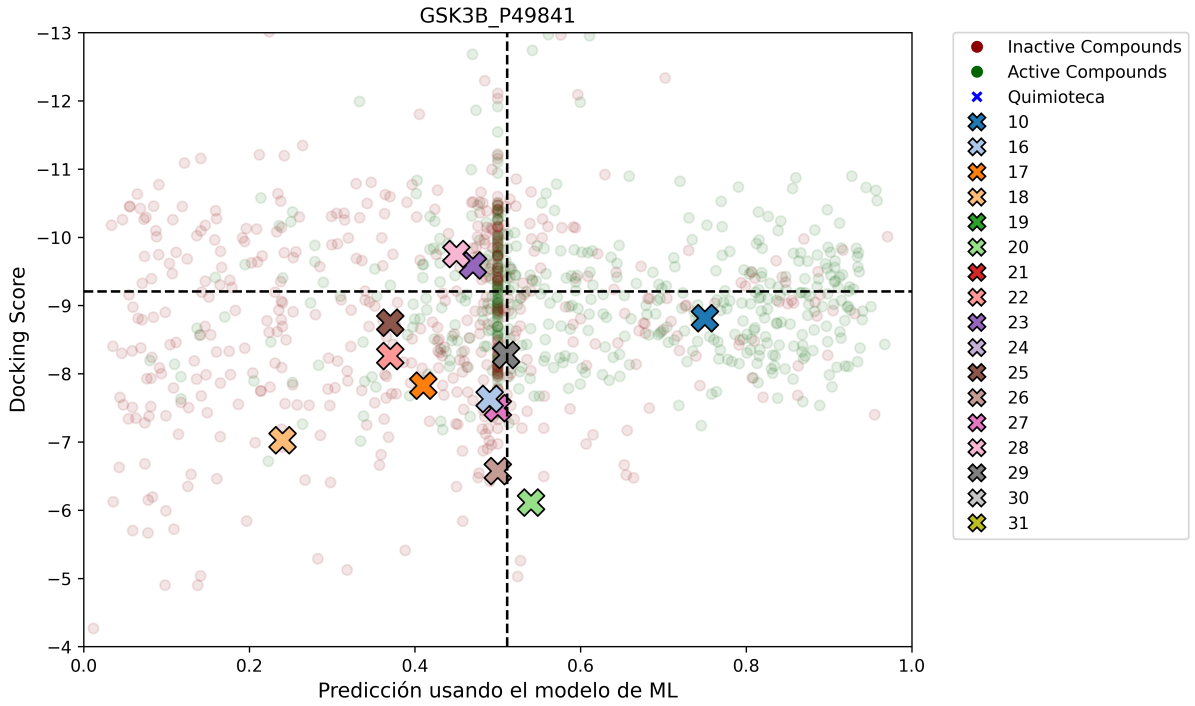


Figura 4.10: Distribución de los compuestos con estructura de benzotiazepina en el gráfico de predicción vs. docking score para GSK-3β.

En la figura 4.10 se observa que todos los compuestos con el esqueleto de benzotiacepina se agrupan en la misma región del espacio de predicción, lo que indica que Target Fisher los clasifica como similares en términos de su potencial de interacción con **GSK-**

3β. Sin embargo, los dos compuestos activos, **28** y **23**, se destacan principalmente por sus valores más favorables de docking score y no por una predicción particularmente alta del modelo de aprendizaje automático. Esto resalta el valor de considerar ambas métricas de forma integrada, ya que una señal clara por una de ellas puede ser suficiente para priorizar un compuesto en determinados contextos.

4.4 Relaciones estructura-actividad y análisis del modo de unión

Como se observa en los resultados, otros dos compuestos, **28** y **23**, presentaron actividad significativa frente a GSK-3β al igual que **10**. La posibilidad de evaluar todo este conjunto de moléculas relacionadas abre la puerta a establecer varias relaciones estructura-actividad.

4.4.1. Comparación entre 23 y 10

23 difiere de **10** en el reemplazo del anillo de tiofeno por un anillo de benceno. Este cambio implica un intercambio entre dos grupos que son considerados bioisósteros ya que han sido reportados como funcionalmente equivalentes en sistemas biológicos.(9; 10) Este hecho se puede fundamentar a partir del uso de la herramienta *SwissBioisostere*,(11) que permite analizar el efecto de un cambio en un grupo funcional buscándolo en distintas bases de datos y comparando cada caso. La búsqueda del reemplazo del grupo benceno, ya sea por el 2-tiofeno o el 3-tiofeno, arroja un total de 3700 resultados, indicando que es un reemplazo bastante utilizado (el mismo análisis pero utilizando pirrol arroja menos de

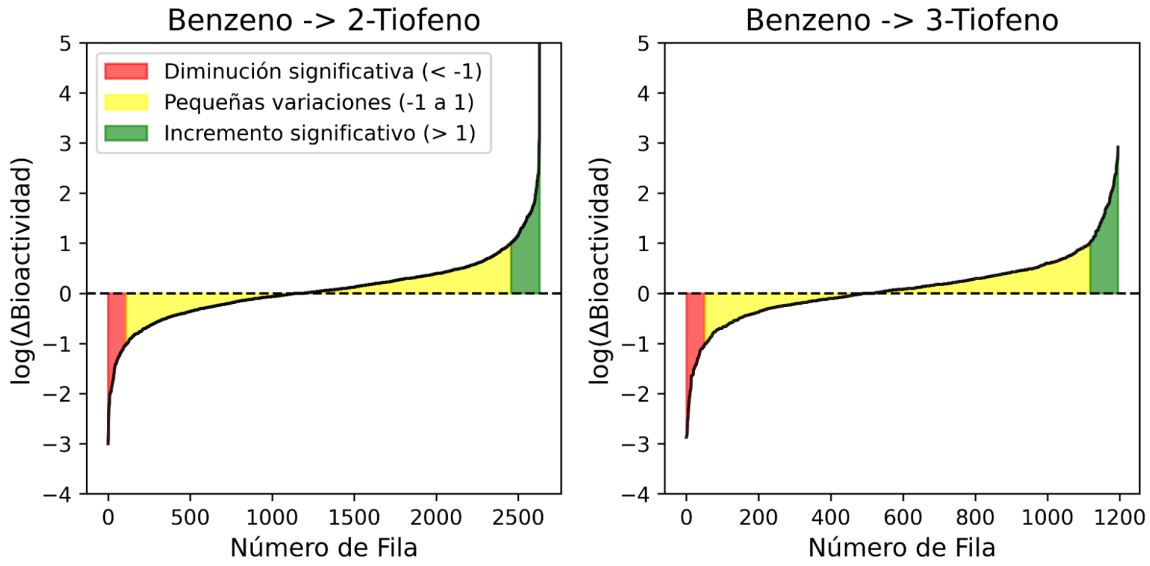


Figura 4.11: Efecto en la actividad biológica al cambiar benceno por tiofeno reportado en biliografia.

200 resultados), como se observa en la figura 4.11. Además, la gran mayoría de los casos reportados mostraron solo una leve variación en la actividad, teniendo el producto obtenido una potencia similar del mismo orden de magnitud al intercambiar ambos grupos. Por último, la variación en el área topológica superficial polar producto de este cambio es nula, y la variación promedio en la solubilidad, estimada utilizando logP, es de -0.13, valor también bajo si consideramos que el cambio por pirrol disminuye este valor hasta -1.

4.4.2. Modificación de 23 a 28

28 resulta de la formación de una oxima a partir de **23**. Esta modificación incrementa la actividad en dos órdenes de magnitud, lo que sugiere que la presencia de este grupo mejora significativamente las interacciones moleculares responsables de la actividad biológica.

4.4.3. Otras relaciones estructura-actividad (SAR) identificadas

Al analizar el resto de los compuestos, se identificó que las siguientes modificaciones en la región de la ciclopentenona (o su oxima en **28**) afectan negativamente la actividad:

1. La inserción de grupos voluminosos, como en los compuestos **29**, **30**, **31**, conduce a la pérdida de actividad.

2. El reemplazo del anillo de ciclopentenona por un éster (**20**) también resulta en pérdida de actividad.

3. La eliminación del carbonilo alfa-beta insaturado (**24**) reduce la actividad significativamente.

4. Sustitución completa del anillo por un grupo carbonilo, imina u oxima (**16**, **17**, **18**, **19**, **21**, **23**, **25**) provoca la pérdida total de actividad.

Del análisis anterior se desprende la alta relevancia de un grupo hidrofílico, ya sea como receptor (carbonilo) o también dador de puentes de hidrógeno (oxima) en una región específica de la molécula. La ausencia de un grupo de estas características lleva a la pérdida de la actividad (**26**), efecto que también se observa cuando estos grupos se ubican ligeramente hacia el interior de la molécula, con mayor impedimento estérico.

Por último, la quimioteca estudiada no permite estudiar relaciones estructura-actividad relacionadas a los otros anillos aromáticos de la molécula, producto de la poca variabilidad de los mismos entre los compuestos disponibles.

4.5 Puesta a punto de la síntesis de los compuestos de la serie benzotiazepínica

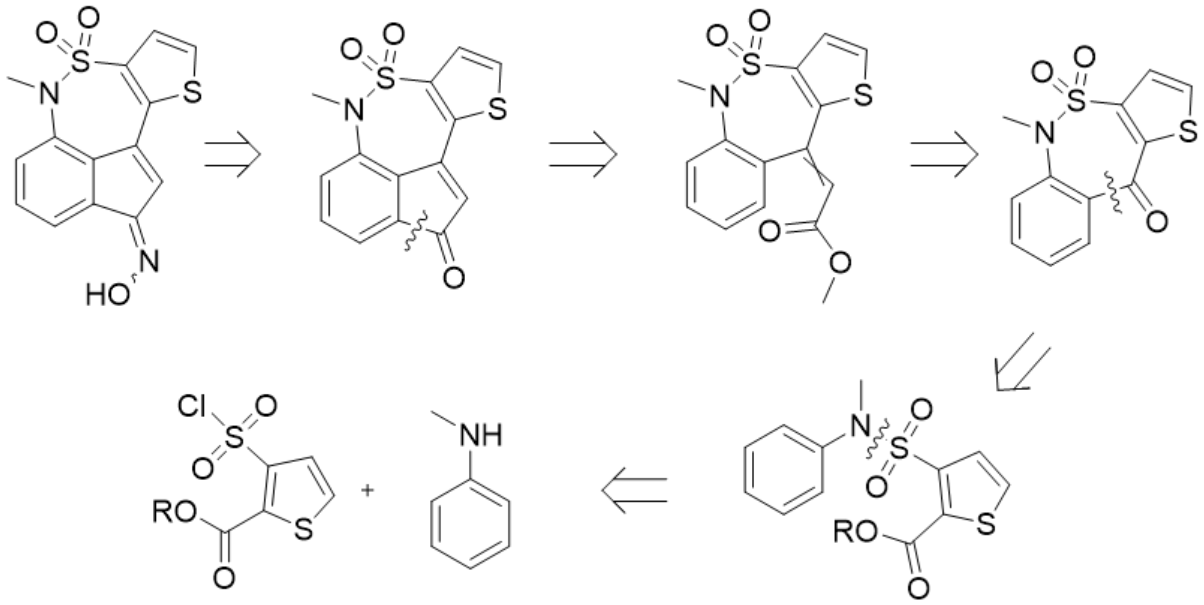
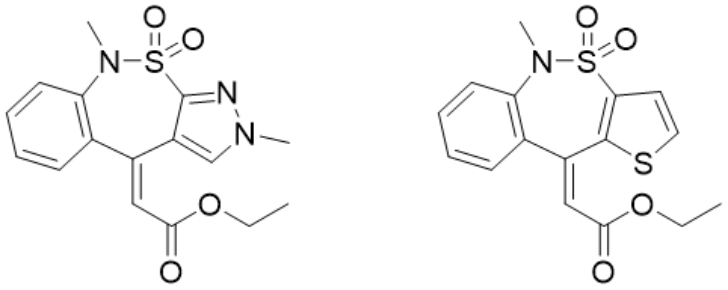


Figura 4.12: Análisis retrosintético de los compuestos de la serie benzotiazepínica.

Con el objetivo de aumentar la cantidad de derivados y profundizar en los estudios de estructura-actividad, se evaluó la posibilidad de sintetizar nuevos análogos. Para ello se planteó una retrosíntesis posible, basada en las **rutas utilizadas previamente** para la preparación de este tipo de compuestos y la **disponibilidad de intermediarios** en cantidades adecuadas.

Un posible análisis retrosintético se presenta en la figura 4.12. La oxima puede obtenerse a partir de una cetona. A continuación se propone la desconexión de la ciclopentenona, obteniendo un éster, disponible con dos heterociclos distintos en la quimioteca en cantidades suficientes para llevar a cabo varios pasos de reacción. Igualmente, este intermediario puede obtenerse a partir de la formación de un doble enlace a partir de un carbonilo,

también disponible. La formación de este tipo de compuestos ya había sido reportada, a partir de anilinas metiladas y un 2-éster,3-clorosulfonil heterocíclo comercial.(5)



(a) Intermediario 1 (b) Intermediario 2

Figura 4.13: Intermediarios disponibles en la quimioteca para la síntesis de nuevos derivados.

El primer paso para la obtención de análogos a partir de estos intermediarios consiste en una acilación de Friedel-Crafts que permita la formación de la ciclopentenona. Dicha reacción no suele ser sencilla cuando se parte de un grupo éster, y por lo general el mismo debe ser convertido a un cloruro de ácido.

La secuencia de reacciones planteada a partir de los intermediarios que se observan la figura 4.13 se muestra en la figura 4.14. En primer lugar se realiza la hidrólisis del grupo éster en medio básico (KOH, MeOH) y posterior acidificación. El ácido carboxílico obtenido fue convertido en el correspondiente cloruro de ácido utilizando $SOCl_2$. La acilación mencionada se realiza finalmente utilizando $AlCl_3$ como catalizador.

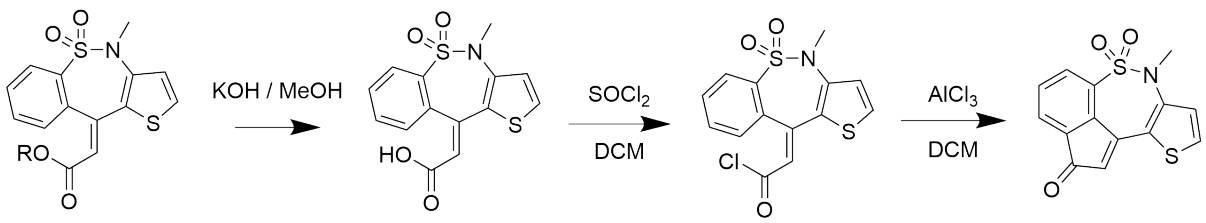


Figura 4.14: Esquema de síntesis para la formación de la ciclopentenona.

Una ruta alternativa para la formación del anillo en un paso directamente desde el

éster es mediante el uso de un superácido como el trifluorometansulfónico, figura 4.15 .(12) Finalmente, la oxima se obtiene a partir de hidroxilamina a reflujo en medio básico.

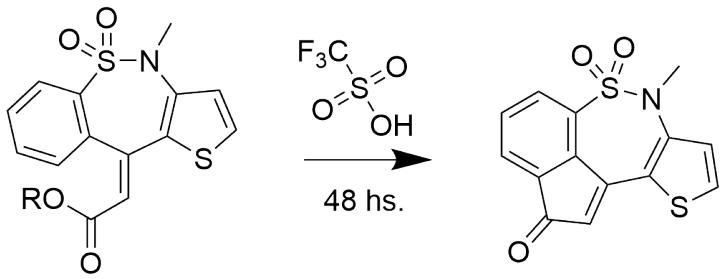
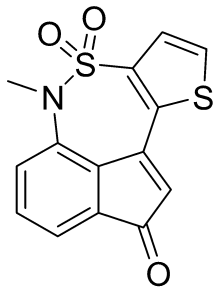
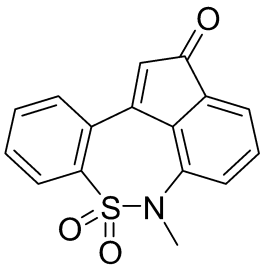
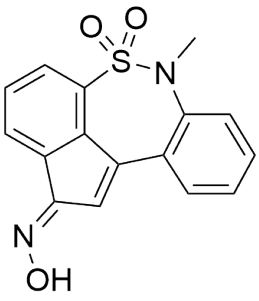
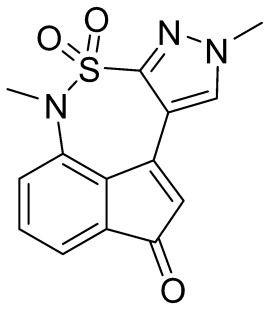


Figura 4.15: Ruta alternativa para la síntesis de los derivados de benzotiazepina.

Se probaron ambas rutas sintéticas experimentalmente, permitiendo la obtención del análogo heterocíclicos con anillo de pirazol, **32** y la síntesis de **23** en mayor cantidad. No fue posible sintetizar las oximas derivadas de **10** y **32** debido a falta de tiempo durante mi estancia en Madrid.

Las estructuras de dichos compuestos junto con su actividad biológica se presentan en la Tabla 4.4. Se observa que la diferencia de dos órdenes de magnitud entre las ciclopentenonas y las oximas se sigue manteniendo, con una afinidad superior por la enzima en el caso de los derivados con tiofeno.

Tabla 4.4: Resumen de estructuras y actividades de los compuestos activos frente a GSK-3β.

			
10	**23**	**28**	**32**
Mayor a 7,91 µM	7,33 µM	0,055 µM	8,92 µM

4.6 Estudio del modo de unión

La elucidación del modo de interacción de estos compuestos con el sitio activo de **GSK-3**β resulta clave para interpretar sus diferencias en actividad biológica y guiar el diseño racional de nuevos derivados. El modo de unión, es decir, la orientación y las interacciones específicas que establece cada molécula dentro del sitio catalítico, determina en gran medida su afinidad y selectividad. Tal como se desarrolló en el Capítulo 3, este tipo de información puede obtenerse mediante el uso combinado de *docking molecular* y simulaciones de dinámica molecular.

Los resultados del *docking* molecular para los compuestos carbonílicos **10**, **23** y **32** revelaron tres poses posibles para cada uno, todas con orientaciones generales similares. En los tres casos, se observa que el grupo sulfonamida se orienta hacia la región del solvente, mientras que el grupo carbonilo se dirige hacia el interior del sitio de unión, en proximidad al residuo ASP133. Considerando la pose de mejor puntuación, presentada en la figura 4.16, la orientación general de la pose más representativa es idéntica para **10** y **32**; sin embargo, en el caso de **23**, se observa una diferencia notable: el anillo aromático proveniente de la anilina se orienta hacia el interior de la cavidad, a diferencia de lo que ocurre en los otros dos compuestos, donde dicho anillo queda proyectado hacia el exterior, hecho que viene acompañado de una reorientación del grupo carbonilo.

Las simulaciones de dinámica molecular realizadas para estos tres compuestos evidenciaron que el modo de unión propuesto es estable tanto para **10** como para **32**. En cambio, en el caso de **23**, el compuesto no logra mantenerse dentro del sitio activo, siendo

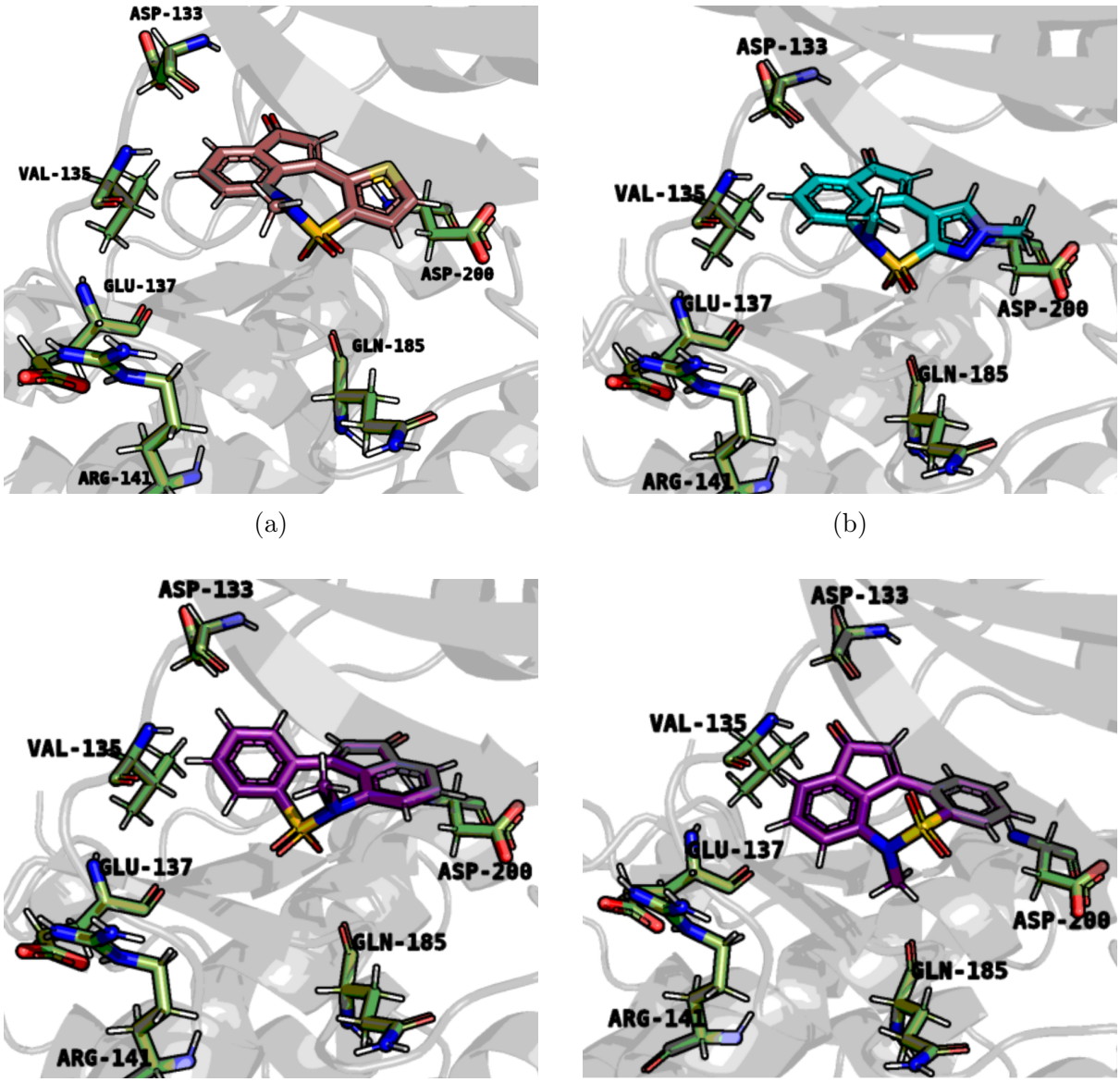


Figura 4.16: Poses de docking de los compuestos **10**, **32** y **23** en el sitio activo de GSK-3β. Pose (a) corresponde a **10**, pose (b) a **32**, pose (c) es la primera y pose (d) es la tercera de **23**.

expulsado de la cavidad durante la simulación, como se observa en los gráficos de RMSD presentados en el Capítulo 6. No obstante, la pose 3 de **23** —que presenta una orientación similar a la observada para los otros dos compuestos— difiere en menos de un 4 % en la puntuación de docking respecto a la pose 1. Al utilizar esta última como punto de partida para la simulación de dinámica molecular, se observó una mayor estabilidad, lo que sugiere que el modo de acción más plausible de **23** corresponde a esta orientación alternativa.

Con respecto a las oximas, se estudió también el posible modo de unión de **28** asi como también las derivadas de **10** y **23**, **33** y **34**. El modo de unión encontrado en las primeras poses de las tres es equivalente y similar a los compuestos carbonílicos precursores, como se observa en la figura 4.17. En la figura se pueden observar las poses obtenidas, todas ellas estables en las simulaciones de dinámica molecular.

4.6.1. Análisis de las energías de interacción ΔH entre los compuestos y la proteína

Utilizando las dinámicas anteriores, se evaluaron las energías de interacción entre los 6 compuestos activos y la proteína. Por un lado, el ΔH global mostró, excepto en el caso de **33**, donde las oximas alcanzaron menores energías de unión que los compuestos carbonílicos, en concordancia con los resultados experimentales.

Por otro lado, la descomposición de estas energías permitió analizar estas interacciones en mayor profundidad. En primer lugar, los residuos **ILE 62**, **VAL 70** y **LEU 188** actúan como puntos de anclaje comunes para todos los compuestos, constituyendo un núcleo

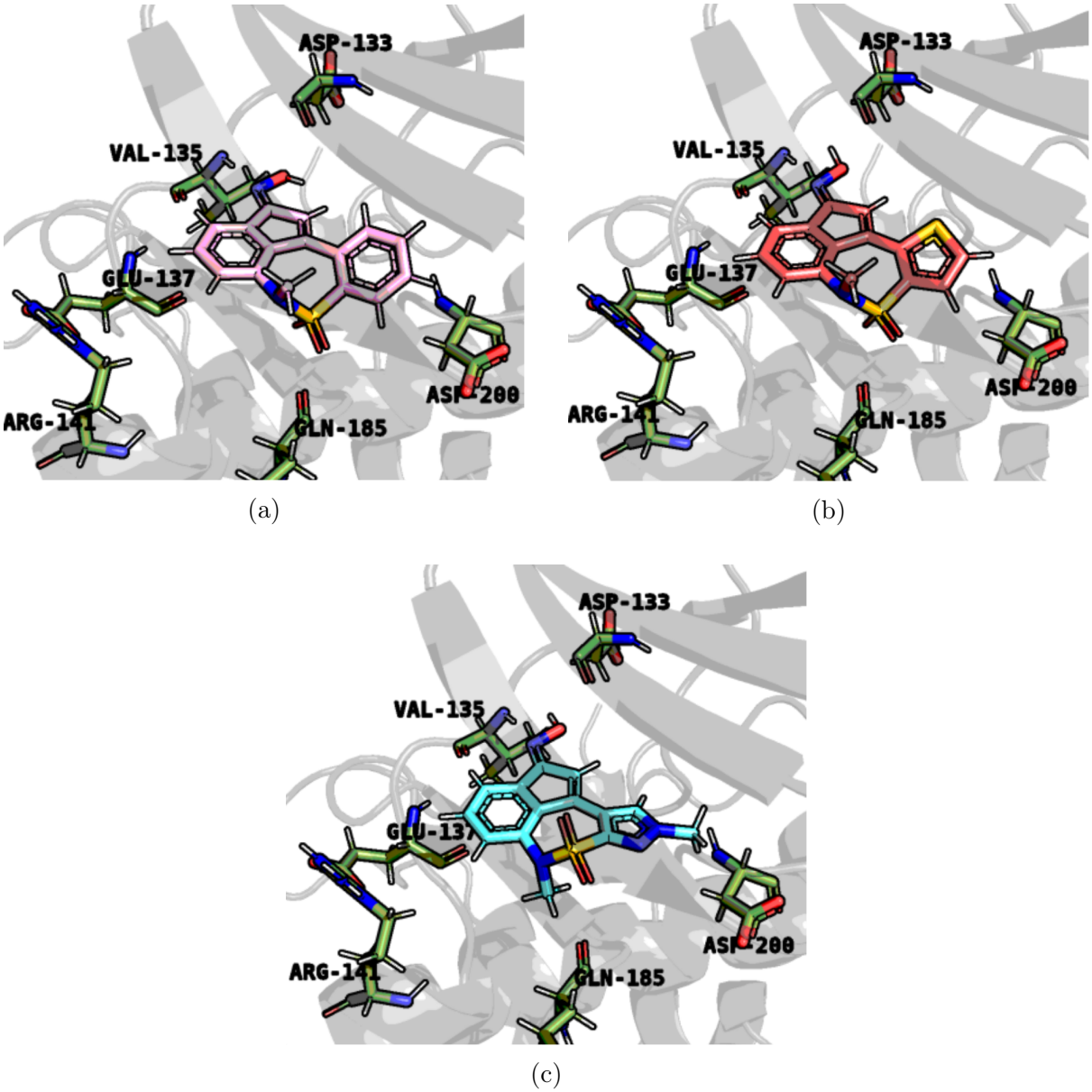


Figura 4.17: Poses de docking de los compuestos **28** (a), **33** (b) y **34** (c) en el sitio activo de GSK-3β.

Tabla 4.5: Energías de interacción y actividades biológicas de los compuestos activos.

Compuesto	ΔH **(kcal/mol)**	$\mathbf{IC_{50}}$ **(µM)**
10	-22 ± 2	$< 7{,}91$
23	-18 ± 2	7.33
32	-32 ± 3	8.92
28	-27 ± 2	0.055
34	-24 ± 2	
33	-28 ± 3	

hidrofóbico esencial para la afinidad de unión. Sin embargo, los análogos con oximas tienden a presentar interacciones puente de hidrógeno con los residuos **TYR 134**, **VAL 135**, **ASP 133**, prácticamente ausentes en los compuestos carbonílicos y que podrían justificar su mayor afinidad experimental.

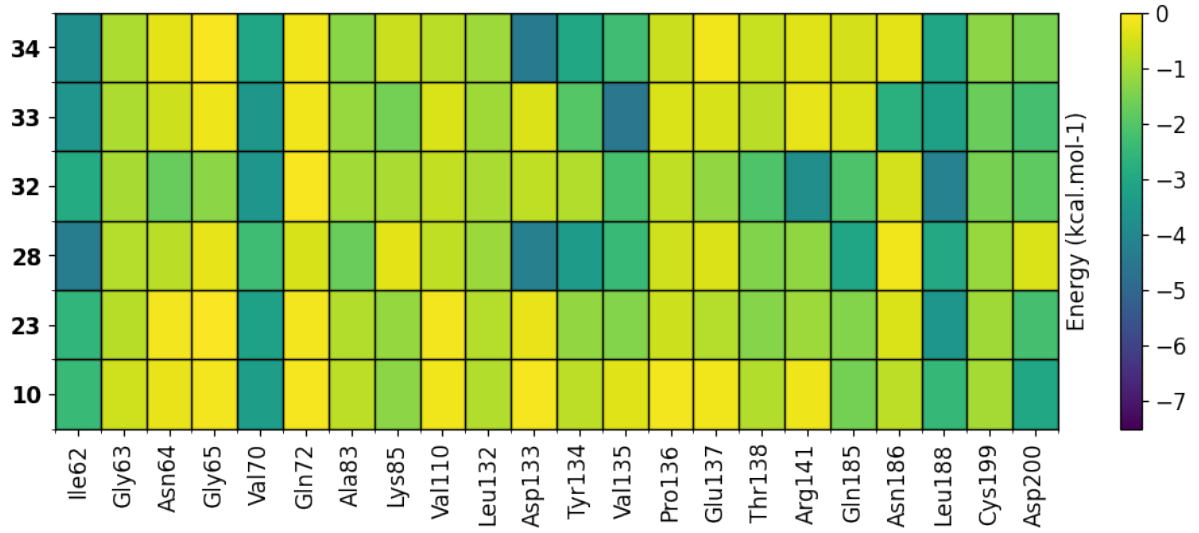


Figura 4.18: Descomposición de las energías de interacción por residuo para los compuestos activos.

Lo desarrollado en este capítulo sienta las bases para continuar y profundizar en el estudio de esta serie de compuestos. Por un lado, se cuenta ahora con una comprensión más precisa del sitio de unión a GSK-3β, gracias al análisis estructural llevado a cabo mediante docking molecular y simulaciones de dinámica molecular. Por otro, las rutas sintéticas empleadas para obtener estos compuestos han sido mayormente optimizadas, lo que permite pensar en futuras modificaciones estructurales. En particular, se plantea como estrategias posibles la variación del heterociclo o la sustitución del grupo oxima por otros grupos funcionales que mantengan las interacciones clave con la proteína y conserven o incluso mejoren la actividad biológica observada. Estas alternativas pueden ser empleadas en futuros ensayos biológicos, ampliando el estudio hacia la evaluación de

toxicidad y la selectividad frente a otras quinasas relevantes.

Bibliografía

[1] Knox, C. *et al.* DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Res.* **2024**, *52*, D1265–D1275.

[2] Meng, F.; Xi, Y.; Huang, J.; Ayers, P. W. A Curated Diverse Molecular Database of Blood-Brain Barrier Permeability with Chemical Descriptors. *Sci. Data* **2021**, *8*, 289.

[3] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

[4] Díaz, J. A.; Vega, S.; Expósito, M. A.; Mateo, C. C. S.; Darias, V. Synthesis and Antidepressant Evaluation of New Hetero[2,1]Benzothiazepine Derivatives. *Arch. Pharm. (Weinheim)* **1996**, *329*, 352–360.

[5] Vega, S.; Diaz, J. A. Synthesis of Thieno[3,4-c] and Thieno[3,2-c][2,1]Benzothiazepines. *J. Heterocycl. Chem.* **1993**, *30*, 1509–1512.

[6] Vega, S.; Díaz, J. A. Synthesis of New Hetero[2,1]Benzothiazepine Derivatives. *J. Heterocycl. Chem.* **1997**, *34*, 1191–1194.

[7] Arranz, E.; Díaz, J. A.; Morante, E.; Pérez, C.; Vega, S. Preparation and Regiochemical Assignments of New Pyrazolo[3,4-c][2,1]Benzodiazepines. *J. Heterocycl. Chem.* **1996**, *33*, 151–156.

[8] Vega, S.; Díaz, J. A.; Darias, V.; Mateo, C. C. S. Synthesis of New Thieno and Pyrazolo[2,1]Benzothiazepine Derivatives with Potential Antidepressant Properties. *J. Heterocycl. Chem.* **2000**, *37*, 389–393.

[9] Alalawy, A. I.; Alatawi, K.; Alenazi, N. A.; Qarah, A. F.; Alatawi, O. M.; Alnoman, R. B.; Alharbi, A.; El-Metwaly, N. M. Synthesis, Molecular Modeling, and Anticancer Activity of New Thiophene and Thiophene-Pyrazole Analogues Incorporating Benzene-Sulfonamide Moiety as Carbonic Anhydrase Isozymes (CA-IX and CA-XII). *J. Mol. Struct.* **2024**, *1295*, 136609.

[10] Flood, A.; Trujillo, C.; Sanchez-Sanz, G.; Kelly, B.; Muguruza, C.; Callado, L. F.; Rozas, I. Thiophene/Thiazole-Benzene Replacement on Guanidine Derivatives Targeting 2-Adrenoceptors. *Eur. J. Med. Chem.* **2017**, *138*, 38–50.

[11] Cuozzo, A.; Daina, A.; Perez, M. A. S.; Michielin, O.; Zoete, V. SwissBioisostere 2021: Updated Structural, Bioactivity and Physicochemical Data Delivered by a Reshaped Web Interface. *Nucleic Acids Res.* **2022**, *50*, D1382–D1390.

[12] Wilsdorf, M.; Leichnitz, D.; Reissig, H.-U. Trifluoromethanesulfonic Acid Catalyzed Friedel–Crafts Alkylations of 1,2,4-Trimethoxybenzene with Aldehydes or Benzylic Alcohols. *Org. Lett.* **2013**, *15*, 2494–2497.

Capítulo 5

Síntesis de derivados de triptamina inspirados en alcaloides naturales

Los metabolitos secundarios aislados de organismos marinos continúan siendo una fuente prolífica de nuevos compuestos bioactivos.(1) Si bien las vías biosintéticas activas y los precursores en estos organismos se asemejan a los presentes en organismos terrestres, los productos naturales marinos suelen ser muy diferentes, con esqueletos carbonados variados y grupos funcionales raros en metabolitos de plantas. Como ejemplo de grupo funcional poco frecuente en metabolitos de plantas se puede mencionar los halógenos; en los organismos marinos es bastante frecuente encontrar compuestos halogenados, sobre todo bromados, mientras que éstos son extremadamente raros entre los productos naturales de plantas. Esto se debe a la presencia de enzimas específicas (sobre todo bromoperoxidasas) muy activas en los organismos marinos. Muchos metabolitos secundarios de origen marino suelen tener actividad biológica, y esto se debe a que fueron biosintetizados para cumplir roles de tipo ecológico, generalmente como antipredatorios (antifeedants) o antiincrustantes (antifouling), o como antibióticos o antivirales naturales para el organismo productor que generalmente no posee sistema inmune, defensa física o posibilidad de escape ante un predador.

Más allá de la actividad biológica que puedan cumplir para el organismo marino, muchos de estos metabolitos poseen interesantes actividades farmacológicas. Esto, sumado a la novedad estructural, vuelve a los metabolitos marinos buenos candidatos para el screening biomédico. Sin embargo, al día de hoy la cantidad de productos naturales marinos que han llegado al mercado es escasa. El problema no es la falta de actividad, sino la poca sustentabilidad para el aislamiento en cantidades (a nivel de gramo) necesarias para atravesar con éxito las diferentes etapas del desarrollo de un fármaco. Por lo tanto, al

día de hoy, los productos naturales marinos son considerados más como una fuente de inspiración para el desarrollo de drogas terapéuticas que como un recurso natural en sí mismo.

A lo largo de los años, nuestro grupo se ha dedicado al aislamiento y elucidación de nuevos productos naturales, con un enfoque particular en organismos marinos. En ese contexto, en un artículo reciente fue publicado el aislamiento, elucidación estructural y estudios preliminares de actividad biológica de una serie de nueve alcaloides bromopirrólicos, las aspidostomidas A-H, a partir del briozoo *Aspidostoma giganteum*, colectado en el Golfo de San Jorge.(2)

Como se observa en la figura 5.1, las aspidostomidas presentan estructuras de dibromotirosina o de bromotriptofano formando amidas lineales o lactamas tipo pirrolocetopiperazinas con un ácido bromopirrolcarboxílico, mientras que la aspidazida A es una acil-azida asimétrica formada mediante una unión *N*-*N* de dos lactamas pirrolocetopiperazínicas con diferente sustitución en el pirrol. Precisamente, el grupo bromopirrolcarboxílico es el fragmento estructural común en todos los compuestos identificados.

Los anillos pirrolicos halogenados, indoles halogenados, y los anillos de dibromotirosina son fragmentos comunes en una gran variedad de productos naturales marinos bioactivos y por lo tanto buenos bloques de construcción para la preparación de compuestos con actividad biológica o farmacológica.(3) En particular dos de las aspidostomidas (aspidostomida E, IC_{50}: 7.8 μM y la G, IC_{50}: 15 μM), mostraron en ensayos preliminares citotoxicidad frente a la línea celular 786-O de carcinoma renal humano.

La **Aspidostomida G** en particular presenta una estructura molecular sencilla, en

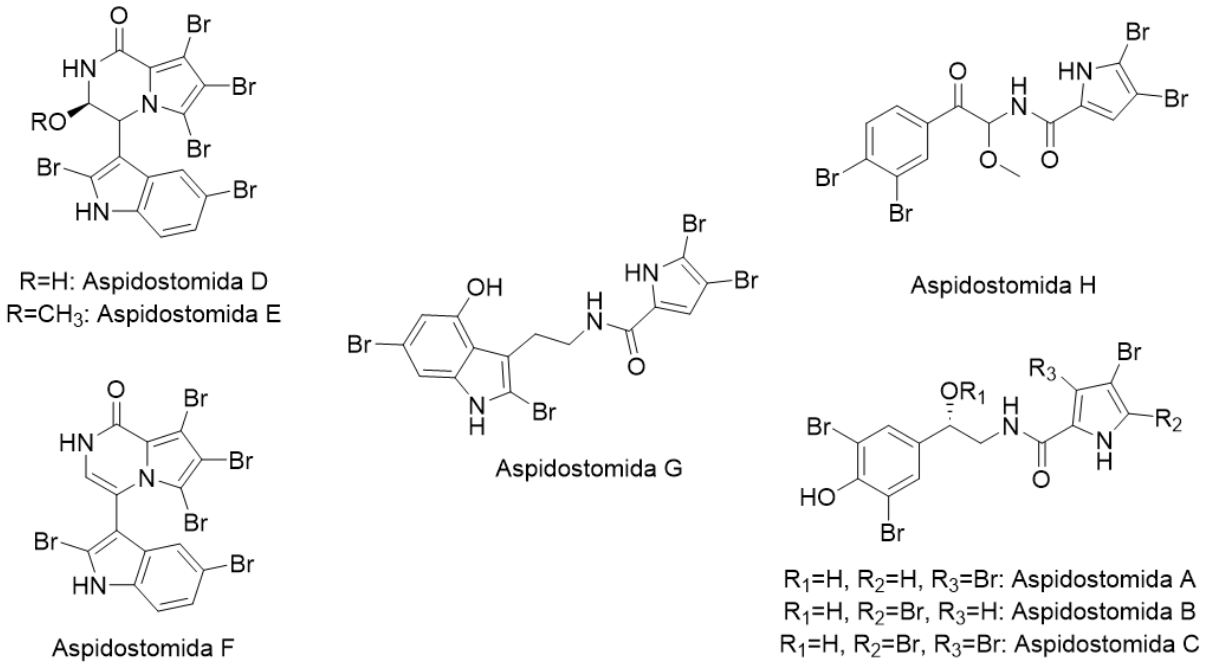


Figura 5.1: Estructuras químicas de las aspidostomidas A–H y la aspidazida A, aisladas del briozoo *Aspidostoma giganteum*.

la cual se distinguen dos fragmentos estructurales: una triptamina sustituída y un ácido pirrolcarboxíllico, unidos por un enlace amida. Esto permite plantear de manera simple la preparación de análogos, trabajando en la sustitución de ambos fragmentos estructurales por separado.

Como se mencionó previamente, la aspidostomida G presentó actividad frente a la línea tumoral renal 786-O. Los tumores renales constituyen un 3,5 % de todos los tipos de cáncer, su prognosis es pobre, y su incidencia se ha incrementado en los últimos años. La mayor parte de los tumores renales, sobre todo los carcinomas de células claras (85 % de ellos, entre ellos 786-O), sobreexpresan la enzima CK2 (casein quinasa 2), la cual es el blanco molecular típico, no sólo para tumores renales, sino también de pulmón.(4) La CK2 es una enzima capaz de fosforilar una gran cantidad de sustratos, y se la ha asociado con una variedad de enfermedades además del cáncer, tales como la fibrosis quística, la

esclerosis múltiple y la hipertrofia cardíaca.

Por otro lado, en uno de los primeros artículos del grupo se aislaron otros dos alcaloides derivados del triptofano, las **chondriamidas A y B**, a partir del alga *Chondria artropurpurea*, recolectada en Miramar, Provincia de Buenos Aires.(5) En este caso, los compuestos presentan una estructura bis-indólica, con ambos heterociclos unidos por un grupo amida y dobles enlaces intermedios (Figura 5.2). Ambos compuestos mostraron en su momento leve actividad antifúngica frente a *Aspergillus oryzae* y *Trichophyton mentagrophytes*.

Junto con estos compuestos, del mismo extracto también se reportó el aislamiento de varios productos de foto-oxidación de los dobles enlaces. Inclusive, se comprobó que la simple exposición a la luz natural en solución metanólica comenzaba los procesos de oxidación. El aislamiento posterior por otro grupo, a partir de la misma especie, de un isómero de las chondriamidas con el doble enlace de la enamina en configuración *Z*, reforzó esta teoría acerca de la labilidad fotoquímica de esta familia de compuestos.(6)

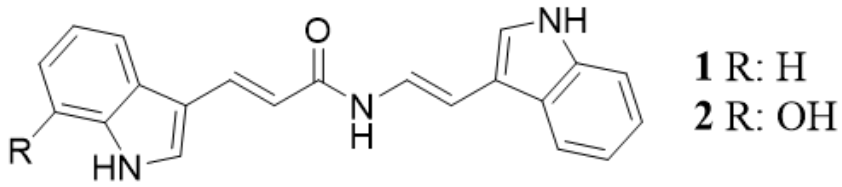


Figura 5.2: Estructuras químicas de las chondriamidas A (1) y B (2)

Estos compuestos, la aspidostomida G y las chondriamidas, son de especial interés en este trabajo por dos razones: son productos naturales autóctonos cuya actividad biológica no ha sido estudiada en profundidad y presentan estructuras base relativamente sencillas a las que se puede llegar mediante síntesis de pocos pasos. Esto facilita, incluso sin necesidad

de realizar su síntesis total, la obtención de análogos inspirados en los productos naturales originales, que pueden ser altamente diversificados y utilizados para explorar su potencial actividad biológica.

El objetivo original de esta parte del trabajo fue: por un lado, la síntesis de análogos inspirados en la aspidostomida G y las chondriamidas; y por otro, la implementación de estrategias de diseño racional enfocadas en la quinasa CK2 como blanco molecular modelo. Para ello, se trabajó inicialmente en la optimización de reacciones clave para la síntesis de análogos de aspidostomida G. En paralelo, se construyó una quimioteca virtual de análogos potencialmente sintetizables a partir de los precursores disponibles en el laboratorio, la cual fue evaluada frente a múltiples blancos moleculares mediante las metodologías computacionales desarrolladas a lo largo de esta tesis. Los compuestos sintetizados fueron evaluados *in vitro* en diversos ensayos de actividad biológica, entre los que se incluyen ensayos de actividad antifúngica.

5.1 Síntesis de análogos de Aspidostomida G

En esta sección se describe la puesta a punto de distintas rutas de síntesis para la obtención de análogos del alcaloide natural.

5.1.1. Formación de amidas utilizando triptaminas comerciales

Realizando un análisis retrosintético de la estructura base de esta molécula, se puede ver que la misma contiene dos heterociclos unidos por una amida. Dicho grupo funcional puede formarse mediante el acoplamiento de una triptamina y un ácido pirrol carboxíli-

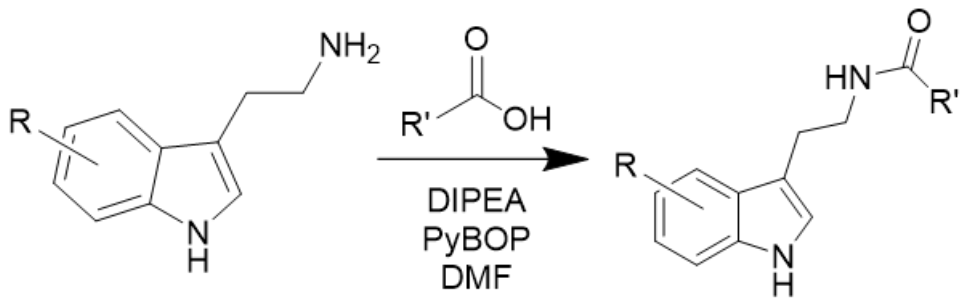


Figura 5.3: Reacción de acoplamiento entre una amina y un ácido carboxílico utilizando PYBOP como agente de acoplamiento.

co, como se observa en la figura 5.3. Si bien no existe una gran variedad de triptaminas sustituídas disponibles comercialmente, o su costo en general es bastante alto, en el laboratorio se contaba con 5-benciloxi triptamina, que se usó para poner a punto esta primera reacción de acoplamiento con otros ácidos heterocíclicos comerciales.

5.1.2. Diversificación estructural del anillo indólico mediante la síntesis de triptaminas

Con el fin de aumentar las posibilidades de diversificación estructural, se puso a punto la síntesis de triptaminas mediante la ruta sintética que se observa en la figura 5.4A. En primer lugar, se lleva a cabo la formilación en C-3 de un indol sustituido mediante una reacción de Vilsmeier-Haack.(7) Luego, mediante una reacción de Henry,(8) el aldehído resultante es convertido en un nitrovinil indol, que finalmente es reducido a la triptamina correspondiente.

Las reacciones mencionadas fueron puestas a punto sin inconvenientes con la excepción del último paso, que inicialmente estaba planteado usando $LiAlH_4$ en THF, se dió bajos rendimientos. Sorprendentemente, la reducción de grupos nitro mediante hidruros está reportada como de alta dificultad, mientras que resulta ser una de las más sencillas al

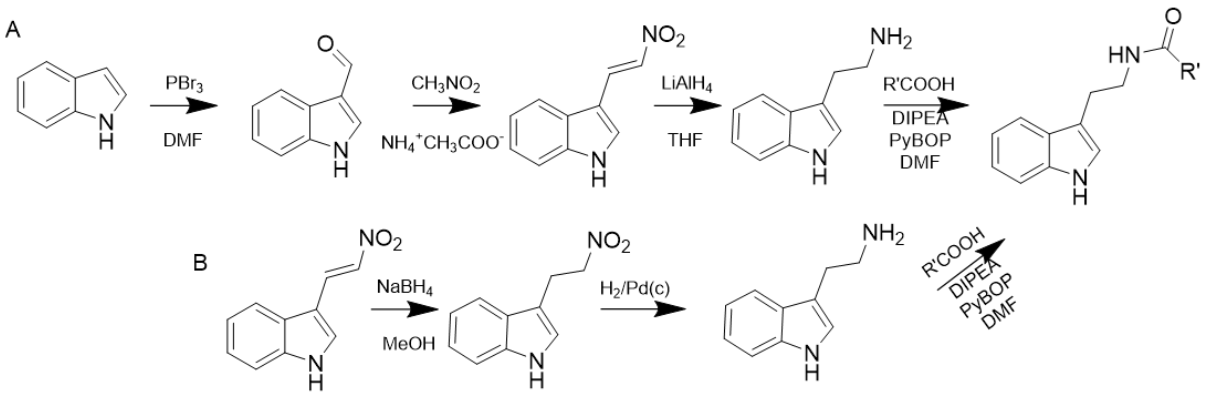


Figura 5.4: (A) Ruta sintética inicial desarrollada para la obtención de triptaminas sustituidas.

emplear H_2 y Pd(C).(9; 10) Por este motivo, se planteó una ruta alternativa en dos etapas para la reducción, que consiste en aprovechar la capacidad del $NaBH_4$ de reducir dobles enlaces conjugados(11) y luego la reducción del grupo NO_2 con H_2/Pd(c), como se observa en la figura 5.4B. Finalmente, al igual que en la ruta sintética anterior, la amida se forma a partir del acoplamiento mediado por PyBOP entre la triptamina y el correspondiente ácido carboxílico.

5.1.3. Síntesis *one pot* de aspidostomidas a partir de la síntesis de indol de Fisher

Una alternativa para la síntesis de aspidostomidas consiste en emplear la reacción de síntesis de indoles de Fischer, a partir de una fenilhidrazina y el dietilacetal del 4-aminobutiraldehído. La exploración de esta vía se vio motivada por las dificultades asociadas a la purificación de las triptaminas obtenidas mediante rutas reductivas, así como por las limitaciones derivadas del uso de agentes reductores, los cuales impiden la incorporación de grupos funcionales sensibles al proceso, tales como nitros o grupos carbonilo o carboxilo. Asimismo, se ha observado que la reducción puede provocar la dehalogenación

del anillo indólico halogenado (sobre todo cuando está bromado) en determinadas posiciones —particularmente en la posición 6—, lo cual resulta especialmente problemático en el caso de la aspidostomida G, cuyo esqueleto presenta un sustituyente precisamente en esa posición.(12) Estos inconvenientes impulsaron la búsqueda de rutas sintéticas alternativas más compatibles con la diversidad estructural deseada.

Dicha síntesis puede realizarse mediante dos estrategias, como se observa en la figura 5.5: por un lado, puede sintetizarse primero la triptamina mediante la síntesis de Fisher y luego formar el enlace amida con PyBOP (vía 1); o bien puede formarse primero esta amida, para luego desproteger el aldehído en medio ácido y formar luego el anillo indólico directamente a partir de la fenilhidrazina (vía 2).

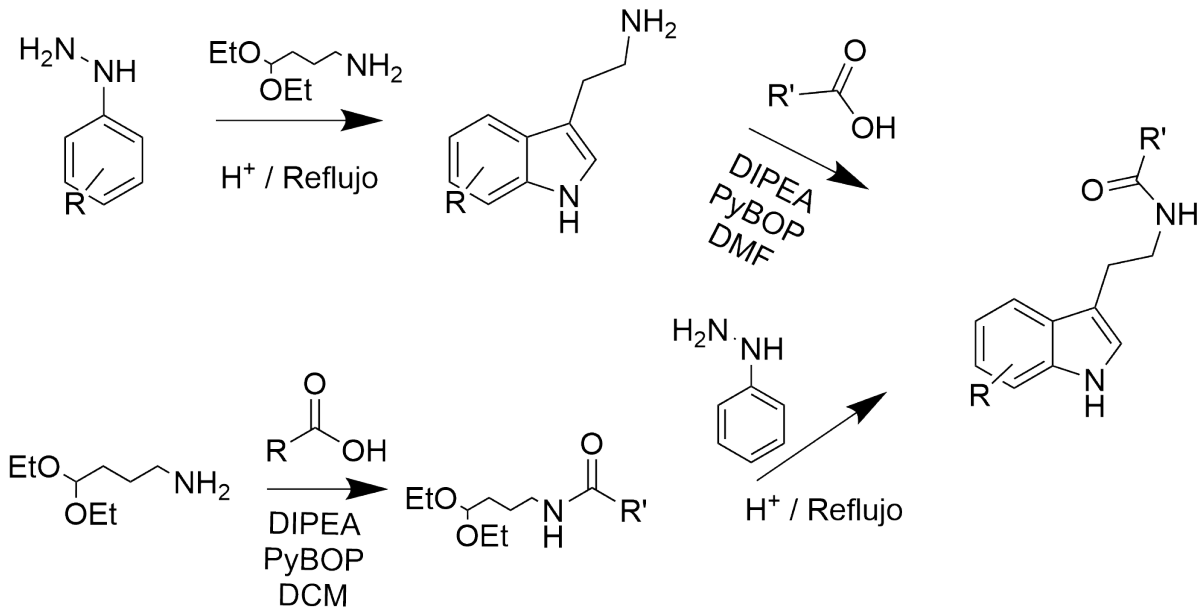


Figura 5.5: Estrategias sintéticas alternativas para la obtención de esqueletos tipo aspidostomida mediante la síntesis de indoles de Fischer.

Ambas estrategias fueron probadas satisfactoriamente, realizando la reacción de Fisher tanto con H_2SO_4 como con $ZnCl_2$, siendo este último catalizador el de mejores resultados. Además, se logró optimizar la vía 2 de manera que, al utilizar diclorometano en lugar

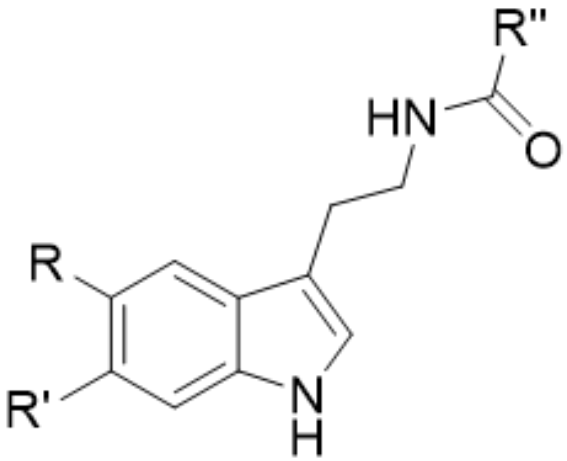
de DMF como solvente en la etapa de formación de la amida, fue posible realizar la transformación completa en un mismo balón de reacción ("*one-pot*"). En la práctica, se lleva a cabo el acoplamiento entre el grupo amino del dietilacetal del 4-aminobutiraldehído y el ácido en diclorometano, se monitorea el avance por TLC, y una vez finalizada la reacción, se evapora el solvente. Luego se añade el ácido y se calienta a reflujo para inducir la formación del indol mediante la reacción de Fisher, obteniendo así la aspidostomida sin necesidad de aislar intermediarios. Esta estrategia simplifica considerablemente el procedimiento y facilita la obtención de compuestos por esta vía.

Cabe destacar que esta ruta ofrece un enfoque versátil que puede ser aprovechado para la síntesis de nuevas triptaminas, permitiendo así el acceso a una mayor diversidad de derivados en condiciones no reductivas. No obstante, una desventaja a considerar es que, en el caso de utilizar hidrazinas *m*-sustituidas, es probable la formación de mezclas de triptaminas isómeras, lo cual puede dificultar la purificación de los productos finales y disminuir el rendimiento.

En la Figura 5.6 se presentan los análogos de aspidostomida sintetizados durante la puesta a punto de las reacciones anteriores.

5.1.4. Reemplazo isostérico del grupo amida

Otra estrategia para la síntesis de análogos de la aspidostomida G es mediante el reemplazo isostérico del grupo amida por un anillo de triazol. Es bien conocido que un anillo triazólico es un bioisóstero de un grupo amida, ya que presenta tamaño, polaridad y capacidad similar de interacción, con la diferencia de que el grupo triazol no es



Cmp.	R	R'	R"
35	H	H	2-Pirrol
36	H	H	2-Furano
37	H	H	3-Furano
38	H	H	2-Tiofeno
39	H	H	3-Tiofeno
40	H	H	2-Py
41	Br	H	2-Pirrol
42	Br	H	2-Furano
43	Br	H	3-Furano
44	Br	H	2-Tiofeno
45	Br	H	3-Tiofeno
46	BzO	H	2-Pirrol
47	BzO	H	2-Furano
48	BzO	H	3-Furano
49	BzO	H	2-Tiofeno
50	BzO	H	3-Tiofeno
51	H	H	2-Pirrol
52	H	H	2-Furano
53	H	H	2-Tiofeno
54	H	Cl	2-Pirrol
55	H	H	etil Indol

Figura 5.6: Análogos de aspidostomida sintetizados.

hidrolizable.(13) Por estos motivos, este reemplazo es muy utilizado en química medicinal.

La ruta clásica para la síntesis de triazoles es mediante una reacción click por lo cual se requiere la presencia de un alquino y de una azida, y ambos grupos funcionales pueden estar ubicados en el anillo indólico o en el otro heterociclo. En particular, el presente trabajo se centró en la preparación de alquinos indólicos, dado que se contaba con una cantidad considerable y variada de bromuros bencílicos, muy fácilmente convertibles en azidas.

En primer lugar se intentó obtener este tipo de indoles mediante la reacción de Bestmann-Ohira (Figura 5.7). Dicha reacción permite obtener alquinos a partir de aldehídos, permitiendo acoplar este paso a la reacción de Vilsemeier-Haack planteada anteriormente:

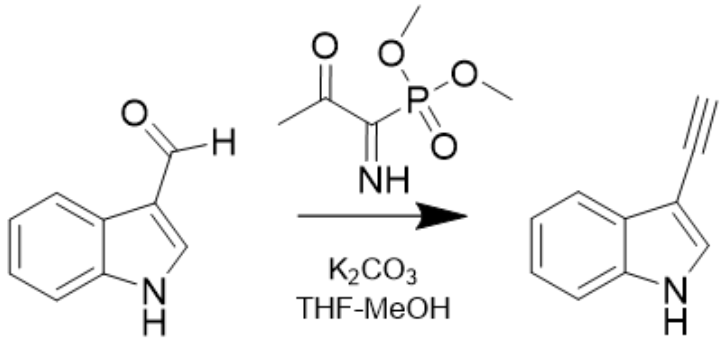


Figura 5.7: Obtención de alquinos a partir de la reacción de Bestmann-Ohira

Lamentablemente, si bien dicha reacción ha mostrado cierta robustez a distintos sustituyentes en el aldehído, los rendimientos disminuyen cuando el aldehído posee H acídicos, como es el caso del indol o pirrol.

De forma alternativa, se planteó llegar hasta el alquino indólico mediante la secuencia de reacciones observada en la figura 5.8, en la que el paso clave es una reacción de Sonogashira. En primer lugar se lleva a cabo la iodación del indol en C-3, para luego proceder a la *N*-metilación del indol. Dicho precursor iodado es sometido luego a una reacción de Sonogashira con trimetilsililacetileno para acoplar de esta forma el alquino. A continuación, una reacción de desililación con reactivos fluorados permite obtener el alquino indólico deseado. En principio se agregó el paso de metilación del indol para poner a punto la reacción siguiente, y eliminar el NH indólico como un factor posiblemente problemático en dicho proceso.



Figura 5.8: Secuencia sintética alternativa para la obtención de 3 alquino indoles

Finalmente, la formación del triazol se realizó mediante una clásica reacción tipo "*click*" con la azida seleccionada, utilizando ascrobato de sodio y sulfato cúprico en DMF, obteniendo el compuesto **56**, figura 5.9.

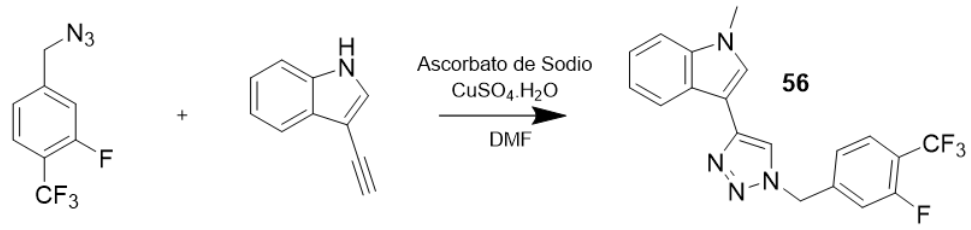


Figura 5.9: Reacción click para la formación del triazol

5.2 Diseño racional de análogos con potencial actividad sobre CK2

Como se mencionó en la introducción, debido a la sobreexpresión de la enzima CK2 en la línea celular tumoral en la que la aspidostomida G presentó actividad, esta proteína se seleccionó como blanco molecular. La búsqueda de análogos que podrían interactuar con la enzima se realizó enfocándose en su estructura, con miras a obtener la mejor interacción posible con el sitio activo, utilizando tanto *docking* como dinámica molecular.

5.2.1. Estudio de la posible interacción entre la aspidostomida G y CK2

En primer lugar, se analizaron las estructuras cristalinas de CK2 disponibles en el Protein Data Bank (PDB). La relevancia de esta enzima en diversos procesos biológicos, junto con su implicación en múltiples patologías, ha despertado un creciente interés en la

comunidad científica, lo cual se ve reflejado en la gran cantidad de estructuras cristalinas depositadas, cada una con distintos ligandos co-cristalizados. Se seleccionaron, entonces, aquellas estructuras que presentaban las mejores métricas de calidad y cuyos ligandos eran estructuralmente diversos, con el objetivo de capturar la flexibilidad conformacional del sitio activo.

Paralelamente, se compiló un conjunto de compuestos con actividad experimental frente a CK2 a partir de la base de datos ChEMBL. Se consideraron como activos aquellos compuestos con valores de IC_{50} menores a 100 nM, y como inactivos aquellos con IC_{50} mayores a 10 000 nM. A fin de ampliar el set de entrenamiento, se incorporaron también *decoys* generados mediante el servidor DUDE-Z, los cuales presentan propiedades fisicoquímicas similares a las de los compuestos activos, pero se asume que carecen de actividad. El conjunto final estuvo compuesto por 1821 moléculas: 501 activos, 666 inactivos y 654 decoys.

Con este conjunto, se buscó construir un modelo capaz de distinguir compuestos activos de inactivos frente a CK2 mediante docking molecular. Para ello, se evaluaron distintos programas, ya mencionados en capítulos anteriores: AutoDock Vina, Smina, Vinardo, AutoDock4, FRED y Hybrid. Los cuatro primeros son de licencia libre, mientras que FRED y Hybrid forman parte de la suite comercial de OpenEye, cuya licencia académica fue provista al grupo de trabajo. El mejor desempeño se obtuvo con Hybrid, que alcanzó un valor de ROC AUC de 0.828. Este rendimiento superior puede atribuirse a que Hybrid permite evaluar simultáneamente los ligandos sobre múltiples estructuras cristalinas y en lugar de explorar libremente el espacio conformacional, inicia el proceso de acoplamiento

alineando el ligando candidato sobre la posición de un ligando co-cristalizado en el sitio.

5.2.2. Estudios por dinámica molecular de la interacción entre la aspidostomida G y CK2

A partir de las estructuras cristalinas seleccionadas y del modelo generado con Hybrid, se llevaron a cabo simulaciones de dinámica molecular para un conjunto representativo de compuestos activos e inactivos. Se seleccionaron 30 moléculas y se utilizaron como punto de partida las poses de docking más favorables obtenidas con Hybrid. Las simulaciones se realizaron sobre la estructura cristalina considerada más adecuada, y en todos los casos se observó que los compuestos permanecieron estables en el sitio activo durante el transcurso de la simulación. Los cálculos de energía de interacción revelaron cierta diferenciación entre activos e inactivos, como se muestra en la figura 5.10, lo que sugiere que esta métrica podría ser útil como criterio de priorización en etapas finales del proceso de selección de candidatos.

En particular, se estudió la interacción de la **aspidostomida G** con CK2 mediante una simulación de dinámica molecular. Luego de **45 ns**, la pose del compuesto se mantuvo estable en el sitio activo. La energía de interacción, calculada promediando los últimos 5 ns de la simulación, fue de -44.3 kcal/mol, un valor comparable al observado para los compuestos activos del conjunto de validación y en línea con la hipótesis inicial de que este sea efectivamente el blanco molecular del producto natural.

De este análisis pueden extraerse varias conclusiones relevantes. En primer lugar, se observa que la conservación de una pose estable durante la simulación no garantiza, por

sí sola, que una molécula se comporte como inhibidor efectivo de la enzima, como quedó evidenciado en varios de los compuestos inactivos utilizados durante la validación de la metodología. Aún resta determinar si la extensión de las simulaciones a tiempos más prolongados permitiría observar la disociación del complejo proteína-ligando en los casos inactivos; sin embargo, dicha exploración excede los objetivos del presente trabajo.

Por otro lado, como se observa en la figura 5.10, se identificó una cierta separación entre las energías de interacción calculadas para compuestos activos e inactivos, lo que sugiere que esta métrica podría constituir un criterio útil para la priorización de candidatos. No obstante, es importante destacar que el cálculo de estas energías conlleva un costo computacional considerable, por lo que su aplicación se propone como una herramienta complementaria a emplearse en etapas avanzadas del proceso de selección, más que como un filtro inicial.

5.2.3. Armado de una base de datos de todos los análogos sintetizables de aspidostomida G

Una vez desarrollado el protocolo anterior se procedió a generar *in silico* una base de datos que contuviera todos los análogos de aspidostomida G sintetizables a partir de los precursores accesibles por el grupo. Como se mencionó en la introducción, la aspidostomida G presenta en su estructura dos fragmentos principales: una triptamina sustituida y un ácido pirrol carboxílico. Utilizando las reacciones propuestas en el plan de trabajo en conjunto con bloques de construcción disponibles en el laboratorio, y aquellos compuestos potencialmente adquiribles, se construyeron todas las triptaminas y ácidos carboxílicos

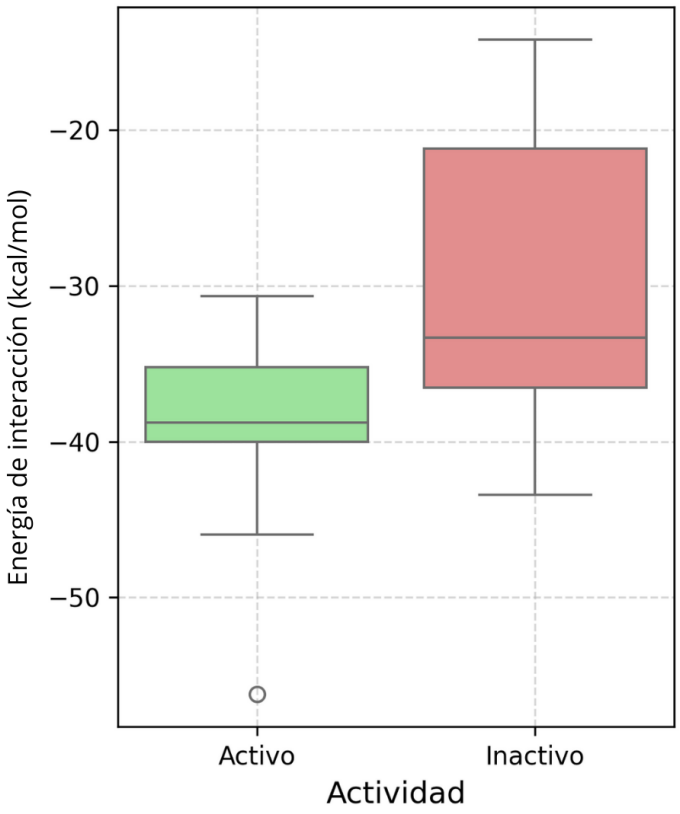


Figura 5.10: Distribución de las energías de interacción calculadas a partir de simulaciones de dinámica molecular para un conjunto de compuestos activos e inactivos.

aromáticos factibles de ser sintetizados. Finalmente, mediante el software de química combinatoria SmiLib(14) se generaron todas las combinaciones posibles de estos fragmentos, obteniendo una quimioteca virtual con 3591 compuestos potencialmente sintetizables.

5.2.4. Estudios de docking molecular

La quimioteca virtual de análogos de aspidostomida G fue evaluada mediante cribado en CK2 utilizando Hybrid, que es el programa de docking que mejores resultados había mostrado. Sin embargo, los resultados, que se muestran en la figura 5.11, no fueron muy alentadores puesto que los análogos obtuvieron unas puntuaciones con mayor semejanza al conjunto de compuestos inactivos, con tan solo 46 compuestos superando a la mediana de

la distribución de activos, y ninguno de ellos con una puntuación demasiado prometedora.

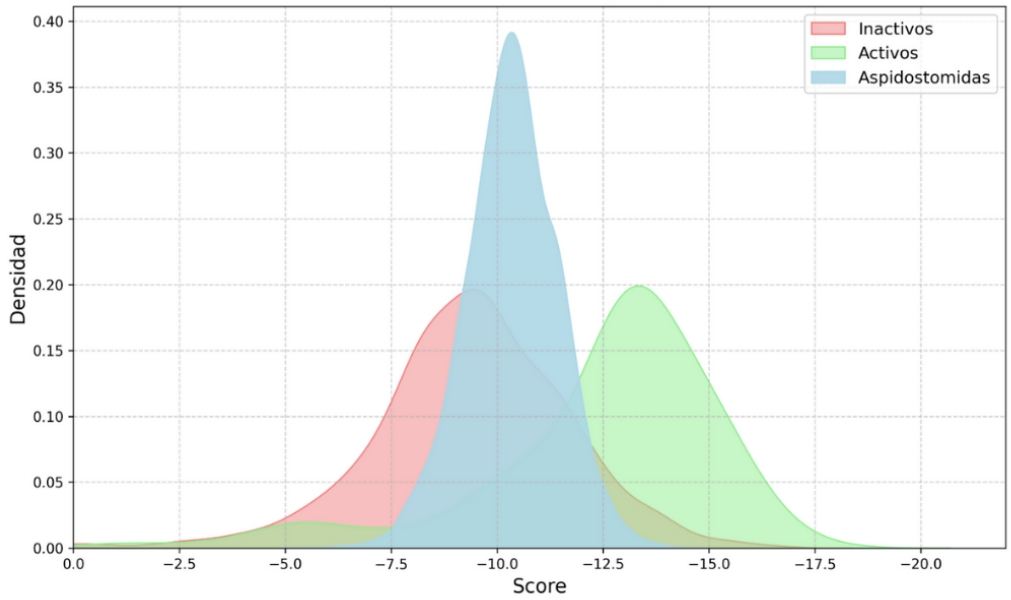


Figura 5.11: Distribución de los valores de docking obtenidos con Hybrid para compuestos activos, inactivos y la quimioteca virtual de aspidostomidas.

Otra estrategia, derivada de lo desarrollado en el capítulo 2, consistió en evaluar los compuestos mediante un modelo de inteligencia artificial entrenado específicamente para CK2. Esta aproximación replica la lógica de Target Fisher, utilizando perfiles de interacción derivados del docking como entradas para un clasificador. Para ello, se seleccionaron los 200 compuestos con mejor puntuación obtenida mediante Hybrid, y se calcularon sus huellas de interacción. Estas fingerprints energéticas fueron luego evaluadas en un modelo preliminar de machine learning específico para esta enzima, anterior al modelo final presentado en el capítulo 2. Sin embargo, como se observa en la figura 5.12, los resultados obtenidos no fueron promisorios.

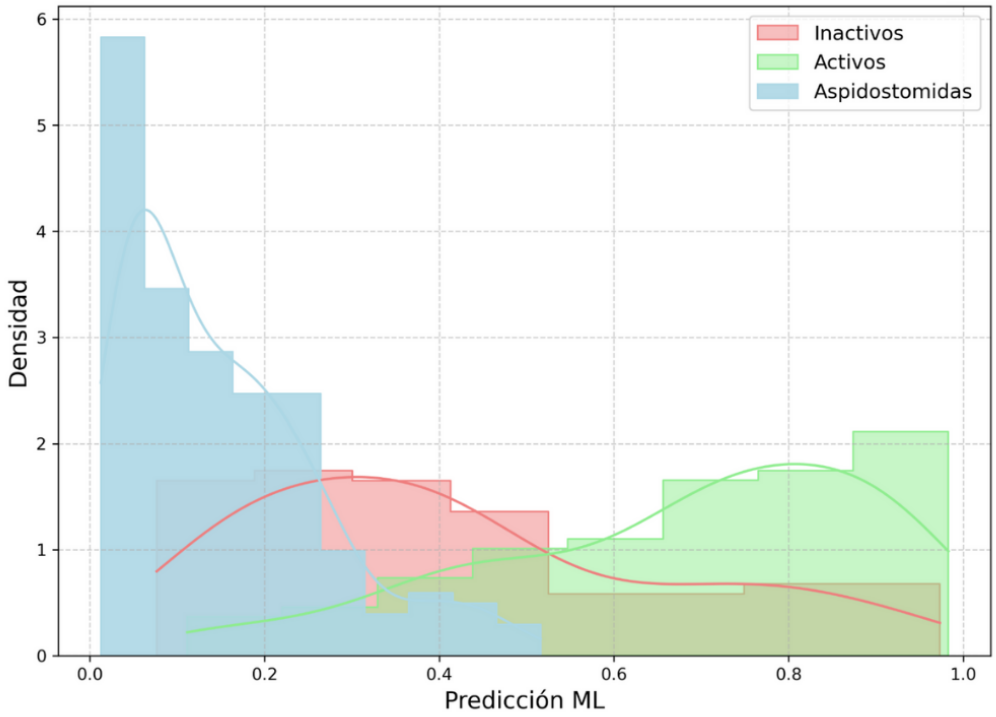


Figura 5.12: Distribución de probabilidades de predicción obtenidas mediante un modelo preliminar de machine learning entrenado para CK2, aplicadas a las fingerprints energéticas derivadas del docking de los 200 compuestos mejor posicionados.

5.3 Exploración de la actividad biológica

El diseño racional de derivados enfocados en CK2 sufrió de diversas dificultades. En primer lugar, como se observa en el apartado anterior, los resultados del *screening* virtual no fueron prometedores. Si bien, tanto las simulaciones de dinámica molecular como los modelos de docking e inteligencia artificial mostraron una alta capacidad clasificatoria, resulta muy difícil encontrar compuestos objetivo a partir de los resultados obtenidos. Esto se debe a que, como se puede observar en las figuras 5.11 y 5.12, prácticamente no se observan compuestos con valor de predicción mayor a 0.5 o valores de puntuación de docking que se asemejen a los de los compuestos activos. A este hecho se le sumó que no

se contaba con el ensayo experimental sobre la enzima, y su obtención y puesta a punto no estaba garantizada por motivos económicos. Finalmente, la síntesis de derivados fue interrumpida por la pandemia de COVID-19, lo que implicó una pausa forzada de casi dos años de actividad experimental. En particular, durante ese tiempo se avanzó en el desarrollo del Target Fisher y en su aplicación al estudio de compuestos con potencial actividad sobre blancos relevantes para enfermedades neurodegenerativas, descriptos en capítulos anteriores.

No obstante, a medida que se fueron obteniendo los análogos, estos compuestos fueron igualmente evaluados en distintos ensayos fenotípicos y dirigidos, aprovechando las colaboraciones vigentes en el grupo y la disponibilidad de modelos biológicos.

Por un lado, en el marco de una colaboración con el grupo del Dr. Martín Edreira (IQUIBICEN, FCEN.UBA), la Dra. Lucía Fernández evaluó algunos de los compuestos sintetizados frente a *Trypanosoma cruzi*, el parásito causante de la enfermedad de Chagas. Esta enfermedad, endémica en América Latina, representa un importante problema de salud pública, especialmente en zonas rurales, y se encuentra actualmente desatendida en términos de desarrollo de nuevos tratamientos. Los ensayos se realizaron sobre el estadio intracelular amastigote del parásito, que representa la forma clínica más relevante durante la infección en humanos. Los resultados, que se observan en la figura 5.13, muestran que los compuestos presentaron valores de IC_{50} en torno a 20 μM, menor al control negativo y en el orden de magnitud del benznidazol (Bz), uno de los fármacos utilizados para tratar la enfermedad, cuyo IC_{50} en este ensayo es 10 μM. Si bien esta actividad no puede considerarse suficiente para considerarlos *hits*, los resultados son alentadores ya

que sugieren que existe una actividad biológica específica que podría mejorar con otros compuestos.

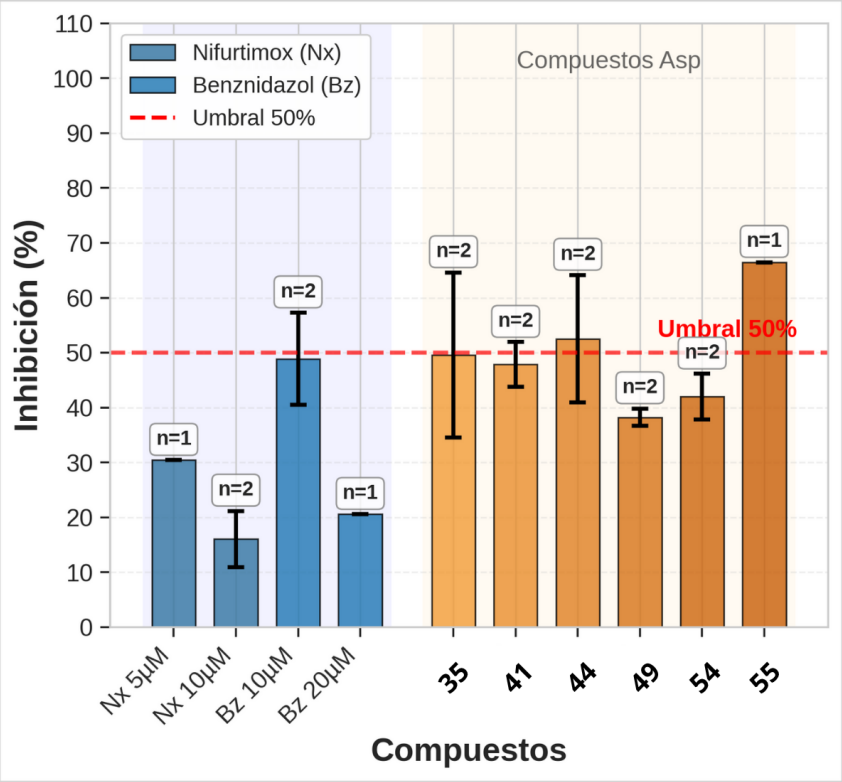


Figura 5.13: Actividad de compuestos seleccionados frente a *Trypanosoma cruzi* en su forma intracelular amastigote. Se muestran los valores de IC_{50} comparados con el control negativo y los fármacos de referencia benznidazol (Bz) y Nifurtimox (Nx).

De forma complementaria, y como parte de los objetivos del proyecto de la unidad ejecutora en UMYMFOR, en el instituto se puso a punto una plataforma de *screening* que permitió evaluar de manera sistemática compuestos con potencial actividad frente a hongos fitopatógenos. Esta plataforma posibilitó la evaluación de algunos de los derivados sintetizados frente a tres cepas de hongos de gran relevancia para cultivos de interés económico en nuestro país: *Fusarium solani*, *Fusarium graminearum* y *Fusarium tucumaniae*. Estas cepas representan una amenaza significativa para diversos cultivos de relevancia económica, por lo que la identificación de nuevos compuestos con actividad antifúngica constituye una línea estratégica de investigación.

Tabla 5.1: Resultados de los ensayos de actividad antifúngica frente a cepas de *Fusarium*. Se muestran las concentraciones inhibitorias mínimas (MIC) para cada compuesto, expresadas en μg/ml y μM.

Muestra	*F. solani*		*F. graminearum*		*F. tucumaniae*	
	μg/ml	μM	μg/ml	μM	μg/ml	μM
Pyraclostrobin	3.12	8.04	1.56	4.02	0.78	2
35	50	197.5	100	395	50	197.5
36	–	–	–	–	–	–
40	50	150.5	50	150.5	100	301
41	–	–	–	–	–	–
42	100	300	50	150	100	300
43	50	143	50	143	50	143
54	25	86	100	347.5	100	347.5
55	25	75	–	–	–	–

Los resultados, presentados en la tabla 5.1, muestran que, en general, los derivados ensayados no exhibieron una actividad antifúngica destacada, registrándose concentraciones inhibitorias al menos un orden de magnitud superiores a las del control positivo. Sin embargo, cabe señalar que el compuesto 20, un análogo estructural de las chondriamidas, se destacó dentro del panel evaluado al mostrar la mejor respuesta antifúngica relativa, lo que sugiere que modificaciones adicionales sobre esta estructura podrían conducir a derivados con mayor eficacia.

Finalmente, como parte de la colaboración con el grupo de la Dra. María Isabel Rodríguez-Franco del Instituto de Química Médica (IQM), CSIC Madrid —colaboración ya mencionada a lo largo de esta tesis—, algunos de los compuestos sintetizados fueron evaluados frente a los blancos moleculares disponibles en dicho laboratorio. Los resultados, presentados en la Tabla 5.2, no evidenciaron una actividad significativa en ninguno de los blancos ensayados. No obstante, se observó una inhibición residual sobre la enzima monoamino oxidasa B (MAO-B) y sobre la butirilcolinesterasa (BuChE) por parte de

uno de los análogos, lo cual podría guiar la preparación de análogos adicionales con otras modificaciones estructurales orientadas a mejorar la afinidad hacia estas dianas.

Comp	**MAO-A**	**MAO-B**	**GSK3β**	**AChE**	**BuChE**
41	> 50	> 50 (42 %)	—	> 10	> 10
42	> 50	> 50 (26 %)	—	> 10	> 10 (26 %)
44	> 50	> 50 (47 %)	> 10	> 10	> 10 (38 %)
46	> 50	> 50 (50 %)	> 10	> 10	> 10
48	> 50	> 50 (43 %)	> 10	> 10	> 10

Tabla 5.2: Actividad biológica de compuestos Asp sobre distintas enzimas. Todos los valores corresponden a concentraciones inhibitorias 50 (IC_{50}) expresadas en µM.

Al momento de redacción de esta tesis, esta línea de investigación permanece abierta, con una reorientación hacia el diseño de inhibidores de la enzima MAO-B. Esta decisión se basa en que se trata de uno de los pocos modelos validados experimentalmente, y permite aprovechar la experiencia adquirida en su inhibición a lo largo del Capítulo 3, donde algunos compuestos mostraron una leve interacción y se cuenta además con el ensayo biológico disponible.

Los próximos pasos se enfocarán en la reconstrucción de la quimioteca virtual, incorporando tanto los nuevos reactivos adquiridos en el laboratorio como los análogos posibles mediante la ruta sintética desarrollada para la obtención de derivados triazólicos. Esta nueva biblioteca será evaluada en el modelo correspondiente dentro de Target Fisher, y los compuestos más prometedores serán seleccionados en función de su viabilidad sintética, con el objetivo de validar las hipótesis computacionales antes de avanzar hacia objetivos sintéticos de mayor complejidad.

5.4 Diseño de nuevos inhibidores de MAO-B

Como se discutió en la sección anterior, ninguno de los compuestos derivados de aspidostomida evaluados presentó una actividad biológica sobresaliente que justificara una profundización inmediata en su estudio. Sin embargo, a lo largo del proyecto se optimizaron diversas rutas sintéticas que permiten el acceso a miles de análogos estructurales. Además, en esta Tesis se describió el desarrollo de una metodología para la predicción de actividad biológica, y se identificaron compuestos cuyas interacciones con MAO-B y GSK-3β fueron estudiadas en profundidad.

Si bien Target Fisher fue concebido como una herramienta de screening virtual *inverso*, cada uno de los modelos incorporados puede emplearse también para realizar un screening convencional, es decir, el cribado de miles de compuestos sobre un único blanco molecular.

Dado que se cuenta con un ensayo experimental disponible, un modelo computacional validado, y que algunos compuestos demostraron una inhibición leve de la enzima MAO-B (tabla 5.2), se decidió reorientar el proyecto hacia el diseño racional de nuevos inhibidores de dicha enzima.

En primer lugar, se realizó un cribado virtual de la quimioteca de aspidostomidas contra MAO-B, utilizando el modelo correspondiente de Target Fisher. Los resultados se resumen en la figura 5.14, donde puede observarse que la mayoría de los compuestos presentan puntuaciones de docking en torno al promedio del modelo. No obstante, en el cuadrante superior derecho se identificaron 639 compuestos (18 % del total) con un alto puntaje de afinidad predicha.

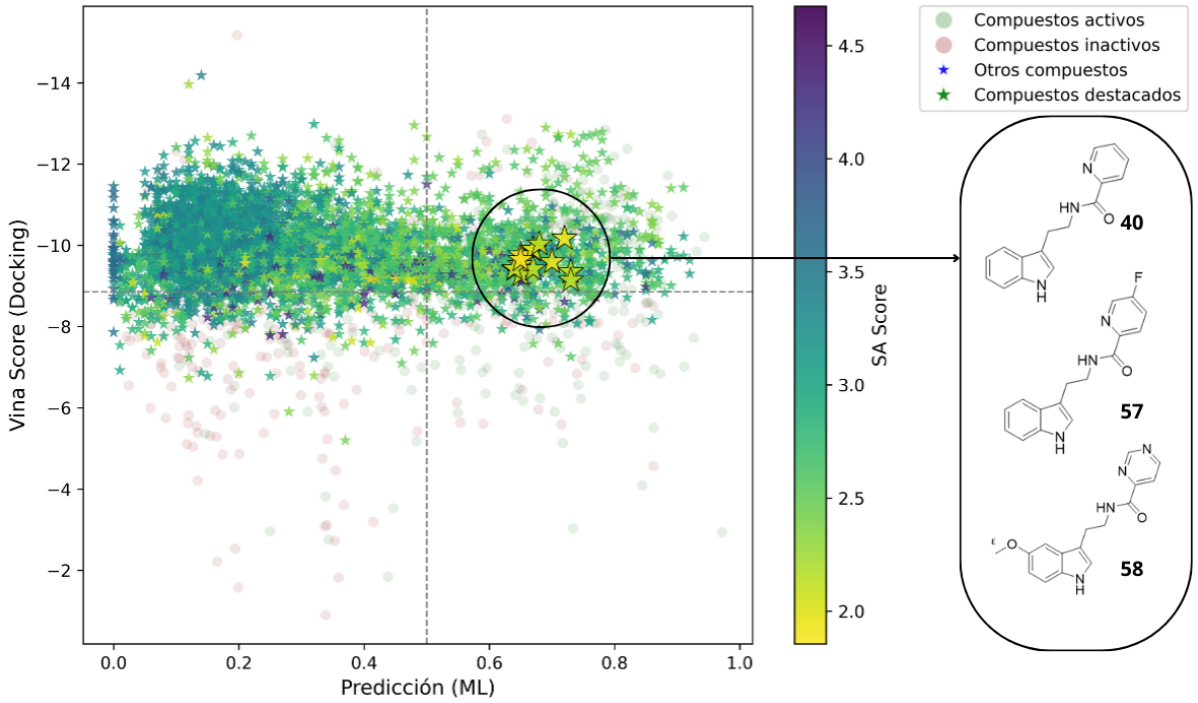


Figura 5.14: Cribado virtual de una quimioteca de aspidostomidas frente al blanco molecular MAO-B utilizando el modelo disponible en Target Fisher.

Para validar experimentalmente la hipótesis de inhibición de MAO-B de forma eficiente, se priorizó este subconjunto de compuestos mediante un criterio de facilidad sintética. Se utilizó para ello el SA score, implementado en la librería RDKit (15), el cual evalúa la facilidad de síntesis de una molécula en una escala de 1 a 10, siendo 1 el valor que indica mayor simplicidad sintética.

Finalmente, se observó que varios de los compuestos que cumplen las características anteriores contienen un anillo de piridina como sustituyente. Actualmente, se está llevando a cabo la síntesis de una pequeña quimioteca de análogos que incluye los análogos de aspidostomidas posibles a partir de triptaminas y ácidos carbixilicos que contengan un anillo de piridina y sean comerciales. Una vez obtenida esta serie de análogos, los mismos serán evaluados frente a la enzima y se decidirán los pasos a seguir dependiendo de los resultados. Así, este trabajo abre una nueva línea dentro del proyecto, centrada en la

validación funcional de las predicciones realizadas mediante la metodología desarrollada en esta Tesis.

Bibliografía

[1] Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2019**, *36*, 122–173.

[2] Patiño C, L. P.; Muniain, C.; Knott, M. E.; Puricelli, L.; Palermo, J. A. Bromopyrrole Alkaloids Isolated from the Patagonian Bryozoan *Aspidostoma Giganteum*. *J. Nat. Prod.* **2014**, *77*, 1170–1178.

[3] Rane, R.; Sahu, N.; Shah, C.; Karpoormath, R. Marine Bromopyrrole Alkaloids: Synthesis and Diverse Medicinal Applications. *Curr. Top. Med. Chem.* **2013**, *14*, 253–272.

[4] *Renal Cell Carcinoma: Molecular Targets and Clinical Applications*, second edition ed.; Springer, 2011.

[5] Palermo, J. A.; Flower, P. B.; Seldes, A. M. Chondriamides A and B, New Indolic Metabolites from the Red Alga Chondria Sp. *Tetrahedron Lett.* **1992**, *33*, 3097–3100.

[6] A new indole derivative from the red alga Chondria atropurpurea. Isolation, structure determination, and anthelmintic activity. *PubMed* **1998**, *61*, 1688–90.

[7] Author, A.; Author, B. The Vilsmeier-Haack Reaction. *Journal Name* **2000**, *10*, 100–120.

[8] Author, C.; Author, D. Henry Reaction: Recent Developments. *Chemistry Journal* **2005**, *15*, 200–250.

[9] House, H. O. *Modern Synthetic Reactions*, 2nd ed.; 1972.

[10] Brown, H. C. In *Advances in Organometallic Chemistry*; Stone, F. G. A., West, R., Eds.; Academic Press, 1973; Vol. 11; pp 1–20.

[11] Kadin, S. B. Reduction of Conjugated Double Bonds with Sodium Borohydride. *J. Org. Chem.* **1966**, *31*, 620–622.

[12] Scott Wiens, P.; Johnson, J. L.; Gribble, G. W. Concerning the Preparation of 6-Bromotryptamine. *Tetrahedron* **2021**, *85*, 132055.

[13] Bonandi, E.; Christodoulou, M. S.; Fumagalli, G.; Perdicchia, D.; Rastelli, G.; Passarella, D. The 1,2,3-Triazole Ring as a Bioisostere in Medicinal Chemistry. *Drug Discov. Today* **2017**, *22*, 1572–1581.

[14] Schüller, A.; Hähnke, V.; Schneider, G. SmiLib v2.0: A Java-Based Tool for Rapid Combinatorial Library Enumeration. *QSAR & Combinatorial Science* **2007**, *26*, 407–410.

[15] Landrum, G.; others RDKit: Open-source cheminformatics. 2006–; https://github.com/rdkit/rdkit, Accessed: 2025-05-08.

Capítulo 6

Materiales y métodos

6.1 Métodos computacionales - Generalidades

6.1.1. Docking molecular

6.1.1.1. Preparación de blancos moleculares y protocolos de acoplamiento molecular

Las estructuras experimentales se descargaron del Protein Data Bank (PDB) (1). La preparación general del blanco consistió en la eliminación de moléculas de agua, la adición de protones y la asignación de cargas parciales, utilizando UCSF Chimera 1.16 (2), salvo en los casos en que se indica lo contrario. El sitio de unión fue definido en torno al bolsillo reportado para el ligando co-cristalizado o con base en información bibliográfica.

A continuación, se detallan los protocolos específicos de acoplamiento molecular según el programa utilizado:

1. **AutoDock Vina 1.2.0** (3):
   - El archivo `PDBQT` del blanco fue generado con AutoDock Tools (4).
   - La caja de acoplamiento fue definida manualmente para cubrir el sitio de unión conocido.
   - Se emplearon un tamaño de grilla de `0.375` y un valor de *exhaustiveness* de `15`.
   - En algunos casos, se utilizaron funciones de puntuación alternativas disponibles en la versión extendida de Vina, como las de AutoDock4 y Smina.

2. **DOCK 6.9** (5):

   - El blanco molecular fue preparado con UCSF Chimera, eliminando moléculas de agua, añadiendo hidrógenos y asignando cargas AMBER.
   - El sitio de unión fue identificado mediante el método de esferas, utilizando los comandos `sphgen` y `sphere_selector`.
   - Se construyó una rejilla de puntuación con el comando `grid`, configurada para emplear la función Grid Score.
   - El acoplamiento se realizó en modo flexible, permitiendo explorar distintas conformaciones de los ligandos

3. **HYBRID y FRED (OpenEye)** (6; 7):

   - La preparación del blanco molecular se realizó con el asistente `Make Receptor` incluido en la suite OpenEye (8).
   - Para ambos programas se utilizaron sus parámetros por defecto.

6.1.1.2. Preparación de ligandos

Las estructuras tridimensionales de los compuestos evaluados experimentalmente para cada blanco molecular fueron obtenidas desde las bases de datos PubChem (9) y ChEMBL (10). La división entre conjuntos de compuestos activos e inactivos se realizó utilizando un valor de corte en la actividad (IC_{50}), que varió según el blanco molecular y el experimento a realizar.

Las estructuras tridimensionales fueron generadas y minimizadas utilizando el software Balloon (11; 12), el cual fue integrado a la plataforma Target Fisher por tratarse de una herramienta de código abierto. No obstante, para las pruebas realizadas con los programas HYBRID y FRED se empleó el software OMEGA (13), incluido en la suite de OpenEye, por su mayor precisión y velocidad en la generación de conformaciones.

Las cargas de los compuestos fueron asignadas utilizando Open Babel(**?**) (obabel) a un pH fisiológico de 7.4, salvo en los casos en que se utilizaron FRED o HYBRID, donde las cargas y estados de protonación fueron definidos utilizando la herramienta FixpKa, también incluida en el paquete de OpenEye.

Finalmente, los archivos `PDBQT` necesarios para los acoplamientos con AutoDock Vina fueron generados utilizando Meeko (14), una herramienta desarrollada para facilitar la preparación automática de ligandos compatible con dicho motor de acoplamiento.

6.1.1.3. Obtención de vectores de interacción

Los vectores de interacción proteína-ligando fueron obtenidos a partir de una versión modificada de la herramienta OBEnergy, disponible en la caja de herramientas química Open Babel. Dicha modificación, llamada MIEF (15), permite el cálculo de la energía de interacción entre el ligando y la proteína utilizando distintos campos de fuerza de mecánica molecular. Para todos los resultados mostrados en esta Tesis, se utilizó MMFF94, pero también se cuenta con la posibilidad de usar GAFF.

6.1.2. Construcción de modelos de aprendizaje automático

La información contenida en el vector de interacción para cada compuesto activo o inactivo fue utilizada para entrenar un modelo de clasificación binaria utilizando Máquinas de Vectores de Soporte (SVM), un algoritmo incluido en la librería scikit-learn (16). Dado que el vector original contenía la energía de interacción del ligando con todos los residuos de la proteína, fue necesario realizar un primer paso de preprocesamiento que consistió en mantener solo los residuos presentes en el bolsillo de unión para reducir la relación ruido/señal en los datos. Este paso fue realizado utilizando un script en PyMOL desarrollado internamente, el cual seleccionó solo los residuos que estaban a 5.0 Å o menos de distancia de cualquier molécula acoplada. El conjunto de datos filtrado fue luego dividido aleatoriamente en dos subconjuntos: el conjunto de entrenamiento (80 %) y el conjunto de prueba (20 %).

Los hiperparámetros C y γ fueron ajustados para optimizar el rendimiento del algoritmo SVM, utilizando GridSearchCV con validación cruzada de 5 pliegues y un kernel RBF. Los hiperparámetros óptimos fueron seleccionados en función del rendimiento promedio en el conjunto de prueba utilizando el área bajo la curva característica de operación del receptor (ROC AUC); también se tomaron en cuenta otras métricas como F1-score, precisión y exactitud.

Con el fin de entender la influencia de los datos en la frontera entre ambas clases en el rendimiento del modelo, se estudió el uso de una ventana de tamaño fijo entre ambas clases. Este umbral se definió como una diferencia del 5 % o 10 % en el valor de logIC50

del blanco. Los compuestos con una actividad dentro de la ventana fueron excluidos del conjunto de datos. Además, también se quiso entender el impacto de aplicar un filtro de similitud entre los compuestos en el conjunto de datos. Esto se realizó agrupándolos usando un coeficiente de Tanimoto, y se seleccionó un compuesto de cada clúster para el conjunto final de datos. El número de clústeres fue controlado seleccionando un umbral de similitud que llevara a subconjuntos activos e inactivos de tamaño similar, con el fin de evitar un desbalance entre ambas clases.

Este procesamiento de los datos llevó a la obtención de 5 conjuntos de datos para cada blanco: dos que incluyen tanto el filtrado por similitud con diferentes tamaños de ventana, y tres en los que solo se modifica el tamaño de la ventana, sin filtrar por similitud estructural.

Como paso final, el mejor modelo para cada blanco fue seleccionado y se utilizaron importancias por permutación para calcular el impacto relativo de cada característica en el rendimiento del modelo. Los modelos fueron re-entrenados, reduciendo el número de características a la mitad e incrementándolas hasta obtener un rendimiento similar al del modelo original.

6.1.3. Simulaciones de dinámica molecular

Las simulaciones de dinámica molecular (MD) fueron realizadas sobre los complejos formados por los compuestos activos validados experimentalmente. Cada complejo fue inmerso en una caja de agua modelo TIP3P con un margen de 15 Å desde el complejo hasta el borde de la caja. El blanco molecular fue descrito utilizando el campo de fuerza

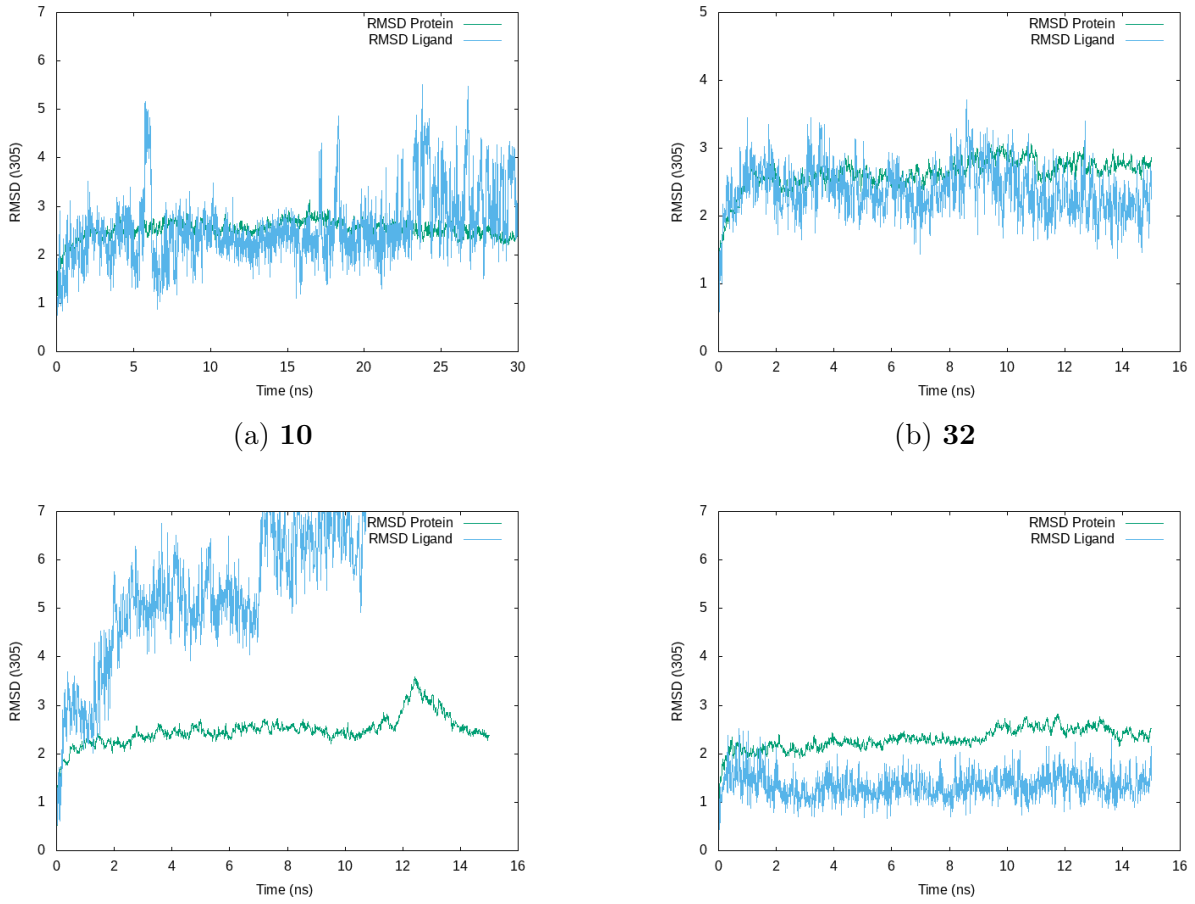
Amber14SB (17), mientras que las topologías de los ligandos fueron generadas mediante el campo de fuerza GAFF (Generalized Amber Force Field) (18), con cargas derivadas mediante el método RESP. Las cargas RESP fueron calculadas utilizando el módulo *Antechamber* y el software Gaussian16 (19), siendo *Leap* y *Antechamber* parte del paquete AmberTools (20).

Todas las simulaciones de dinámica molecular fueron llevadas a cabo utilizando el software NAMD 2.13 (21). Se estableció un corte de 12 Å para las interacciones de van der Waals y las fuerzas electrostáticas de largo alcance se calcularon mediante el método de suma de mallas de Ewald (Particle Mesh Ewald), con una resolución de grilla de 1.0 Å. Las contribuciones 1–4 fueron multiplicadas por un factor de 0.83 para respetar los requerimientos del campo de fuerza AMBER.

El protocolo de minimización consistió en una serie de etapas (incluyendo hidrógenos, moléculas de agua, cadenas laterales de residuos y el sistema completo) con un máximo de 1.000.000 de pasos por etapa. Posteriormente, se realizó un calentamiento gradual del sistema desde 0 hasta 310 K. Las simulaciones de equilibración y producción se realizaron a temperatura constante (310 K), empleando dinámica de Langevin con un coeficiente de amortiguación de 5 ps^{-1}, mientras que la presión fue mantenida constante en 1 atm mediante el método de pistón de Nose-Hoover Langevin, con un período de decaimiento de 200 fs y una constante de tiempo de amortiguación de 100 fs. Se utilizó un paso de integración de 1 fs, y los enlaces que involucran átomos de hidrógeno en moléculas de agua fueron restringidos utilizando el algoritmo SHAKE.

6.1.4. Simulaciones de dinámica molecular de los compuestos 10, 23, 32, 32, 34 y 33.

Las simulaciones de dinámica molecular se hicieron de acuerdo a lo indicado en la sección 6.1.3. A continuación se muestran los gráficos de RMSD obtenidos en dichas simulaciones.



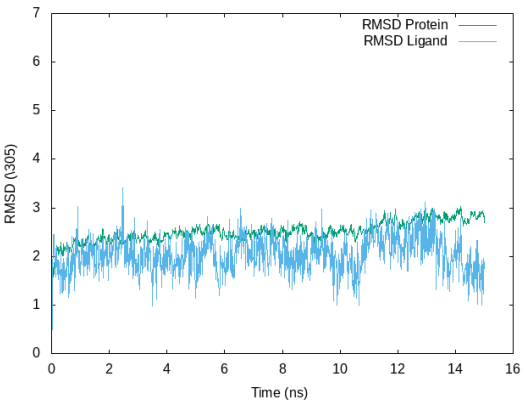
(a) **10**

(b) **32**

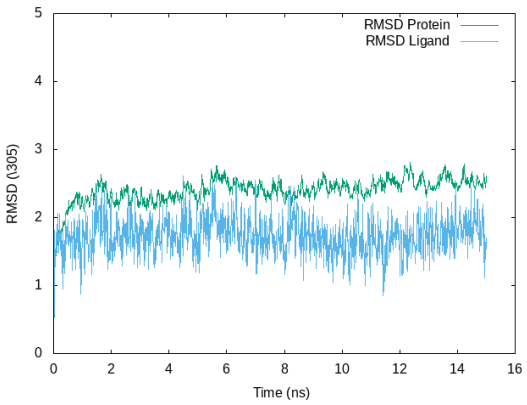
(c) **23 - pose 1**

(d) **23 - pose 3**

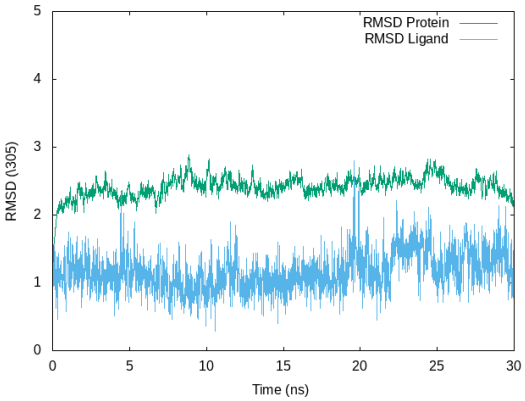
Figura 6.1: Gráficos de RMSD obtenidos durante las simulaciones de dinámica molecular de los compuestos **10**, **23** y **32**.



(a) **32**



(b) **34**



(c) **33**

Figura 6.2: Gráficos de RMSD obtenidos durante las simulaciones de dinámica molecular de los compuestos **32**, **34** y **33**.

6.2 Métodos experimentales - Generalidades

6.2.1. Equipos y condiciones generales

Todos los solventes utilizados para cromatografía o síntesis fueron previamente destilados. Los espectros de RMN fueron medidos utilizando los siguientes equipos: *Bruker AC-200* (200.13 MHz), *Fourier 300* (300 MHz) y *Bruker Avance III* (500.13 MHz). Se tomó como referencia interna la señal residual del solvente no deuterado. Todos los espectros de RMN 2D (COSY, HSQC-DEPT, HMBC, NOESY) fueron medidos utilizando secuencias de pulsos estándar.

Los espectros *ESI-HRMS* fueron registrados en un espectrómetro *Bruker micrOTOF-Q II*.

Las separaciones y análisis por CCD se realizaron en placas de sílica gel *Merck 60* F_{254}, y las cromatografías en columna se hicieron en todos los casos con *Merck Sílicagel* (230–400 mesh). Las placas de CCD fueron observadas bajo luz UV (254 nm) y reveladas con una solución de vainillina al 2 % en H_2SO_4/EtOH.

Las reacciones de hidrogenación se realizaron utilizando el equipo Parr 3911 disponible en el Departamento de Química Orgánica.

Los reactivos utilizados en las síntesis fueron adquiridos a través de proveedores comerciales, incluyendo *Sigma-Aldrich*, *BLDPharm* y *Aaron Chemicals LLC*, sin necesidad de purificación adicional. La separación de los compuestos se llevó a cabo mediante cromatografía en columna convencional utilizando sílica gel, o bien empleando el sistema automatizado *Biotage Isolera Selekt*, según correspondiera.

6.2.2. Aislamiento, purificación e identificación de cumarinas

La corteza de canela de venado *Helietta apiculata* Benth fue recolectada en el Centro de Investigaciones Antonia Ramos (CIAR) en Campo Ramón, provincia de Misiones, Argentina. La corteza (800 g) fue macerada en etanol durante una semana a temperatura ambiente. Posteriormente, el extracto fue filtrado y el solvente eliminado a presión reducida para obtener el extracto etanólico crudo (CV, 38.4 g).

Para separar los alcaloides, el extracto CV fue tratado con una solución de ácido clorhídrico hasta alcanzar un pH = 3 y luego se extrajo con acetato de etilo (CV1, 12.8 g). Esta fase orgánica se llevó a sequedad y luego se particionó entre éter de petróleo y una solución acuosa de metanol al 90 %, obteniéndose los subextractos lipofílico (CV2, 2.6 g) y polar (CV3, 9.0 g).

Tres gramos de CV3 fueron sometidos a cromatografía líquida de vacío utilizando mezclas de diclorometano y acetato de etilo de polaridad creciente como eluyente. Seis fracciones (CV3.1–CV3.6) fueron reunidas de acuerdo con sus perfiles cromatográficos en capa fina. La fracción CV3.5 fue fraccionada mediante cromatografía flash en columna con un gradiente de diclorometano : acetona. Este proceso permitió obtener uliscomoncadina A (1, 28.6 mg), chalepina (2, 67.9 mg) y algunas mezclas de compuestos. Estas últimas fueron purificadas mediante cromatografía líquida de vacío en fase reversa con metanol:agua (9:1) como eluyente, lo cual permitió obtener rutacultina pura (3, 6.3 mg). Gravelliferona (4, 13.7 mg) fue obtenida luego de permeación de CV3.2 a través de Sephadex con metanol como solvente.

Todas las estructuras fueron identificadas mediante espectroscopía de RMN, corroborando la información con reportes previos de la literatura (22; 23).

6.2.3. Medición y simulación de Dicrosismo Circular Electrónico (ECD)

La medición de ECD se realizó en un espectrómetro de CD Jasco (J-815, Jasco Inc. Japón) con una cubeta de cuarzo de 2 mm de camino óptico. Las soluciones se prepararon en metanol (Aberkon, grado espectroscópico) a 1 mM para chalepina y uliscomoncadina A. Se llevó a cabo una búsqueda conformacional con MMFF94 disponible en la versión 12.5.139 de Maestro de Schrödinger. Los confórmeros de baja energía generados dentro de la ventana de energía de 5 kcal/mol, fueron sometidos a una optimización geométrica utilizando el método DFT con el nivel de teoría B3LYP/6-31G. Para asegurar que estos confórmeros fueran mínimos de energía potencial, se realizaron cálculos de frecuencia al mismo nivel de teoría.

Los cálculos de un solo punto de los confórmeros de baja energía se realizaron a nivel B3LYP/6-31G(d,p). Los efectos del solvente fueron considerados utilizando un modelo de continuidad polarizable (PCM), con MeOH como solvente. Nuevamente, se hizo una selección de confórmeros de baja energía, manteniendo aquellos confórmeros dentro de 2 kcal/mol. Para *(R)*-chalepina, se mantuvieron 27 confórmeros, mientras que para *(S)*-uliscomoncadina A, se seleccionaron 28 confórmeros.

Estos confórmeros de baja energía fueron la entrada para los cálculos TD-DFT (24), que se realizaron a diferentes niveles de teoría para *(R)*-chalepina, de modo que se pudiera

seleccionar la mejor combinación de funcional y base, en este caso CAMB3LYP (curva verde). Debido a las similitudes estructurales entre ambos compuestos, este funcional también se utilizó en el caso de *(S)*-uliscomoncadina A.

El ECD calculado se generó con Gaussview (24), a partir de los archivos de salida de Gaussian, después de ponderar los espectros ECD de cada conformador seleccionado por su coeficiente de Boltzmann, tomando en cuenta las energías de los cálculos de un solo punto a los niveles de teoría correspondientes.

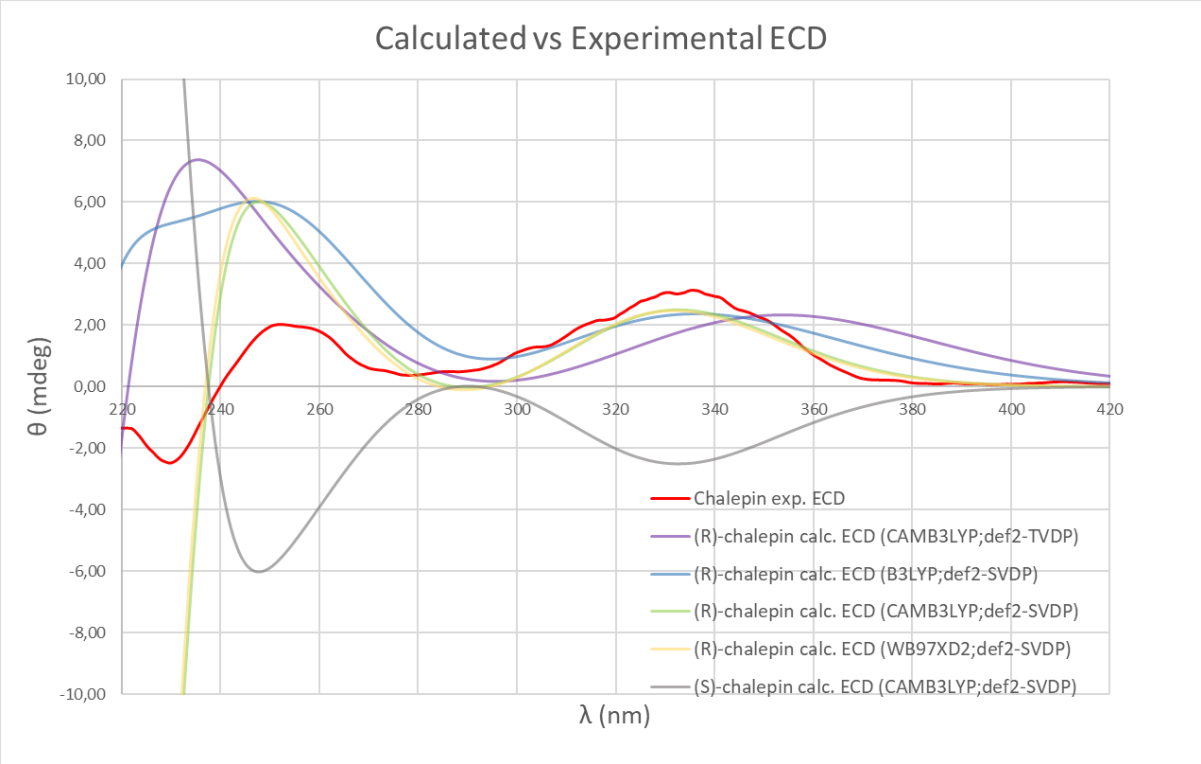


Figura 6.3: ECD del compuesto **2**.

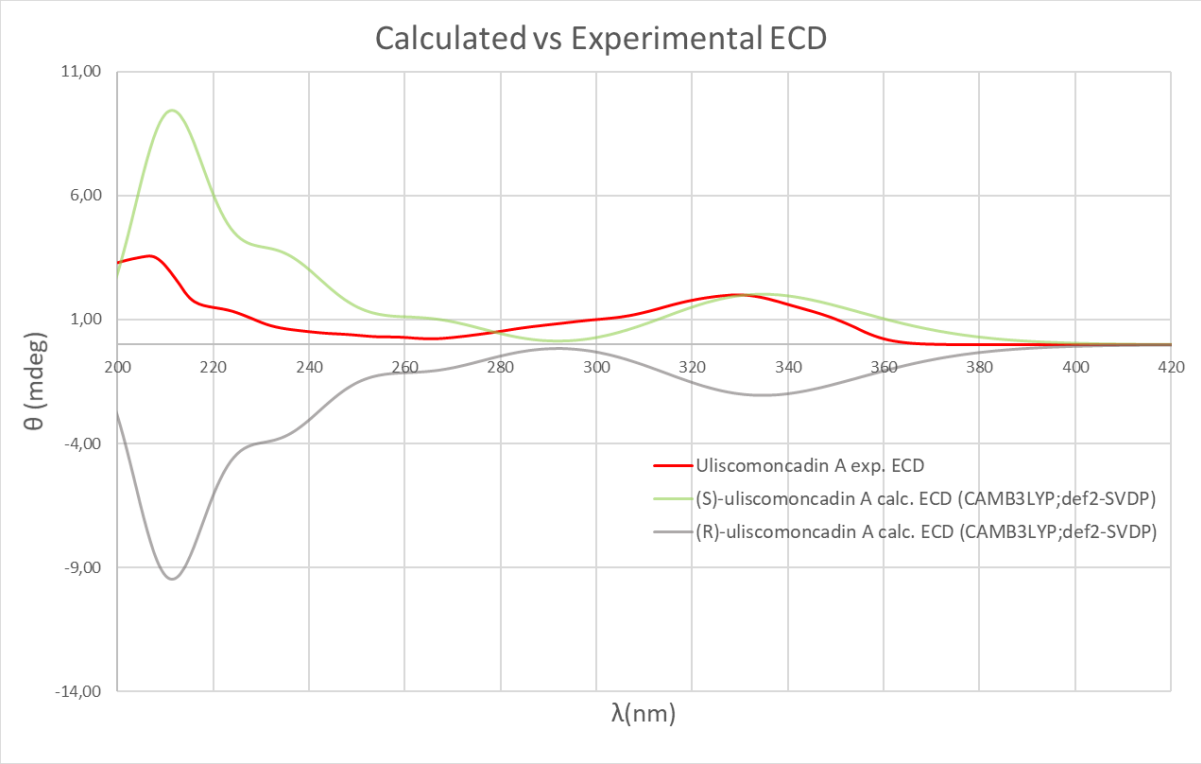


Figura 6.4: ECD del compuesto **6**.

6.2.3.1. *(R)*-Chalepina (2)

^{1}H NMR (500 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 7.48 (s, 1H, H-4), 7.19 (s, 1H, H-5), 6.71 (s, 1H, H-8), 6.16 (dd, J = 17.3, 10.7 Hz, 1H, H-10), 5.06–5.09 (señales superpuestas, 2H, H-11a,b), 1.47 (s, 6H, H-12,13), 3.12–3.27 (señales superpuestas, H-14a,b), 4.71 (dd, J = 9.3, 8.4 Hz, 1H, H-15), 1.36 (s, 3H, H-17), 1.23 (s, 3H, H-18). ^{13}C NMR (126 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 160.5 (C-2), 131.1 (C-3), 138.2 (C-4), 113.3 (C-4a), 123.4 (C-5), 124.7 (C-6), 162.4 (C-7), 97.3 (C-8), 154.8 (C-8a), 40.5 (C-9), 145.8 (C-10), 112.3 (C-11), 26.3 (C-12,13,17), 29.8 (C-14), 91.1 (C-15), 71.9 (C-16), 24.4 (C-18). $[\alpha]_{D}^{25} = -6{,}3$ (c 0.78 in $CHCl_3$)

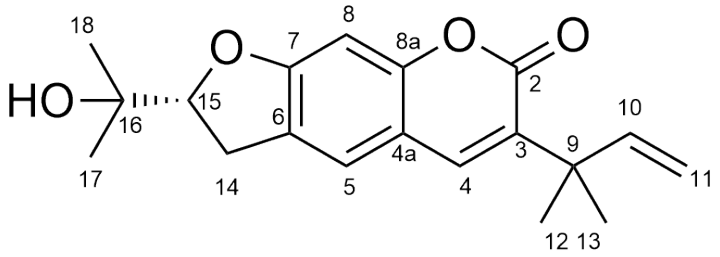


Figura 6.5: Estructura del compuesto **2**.

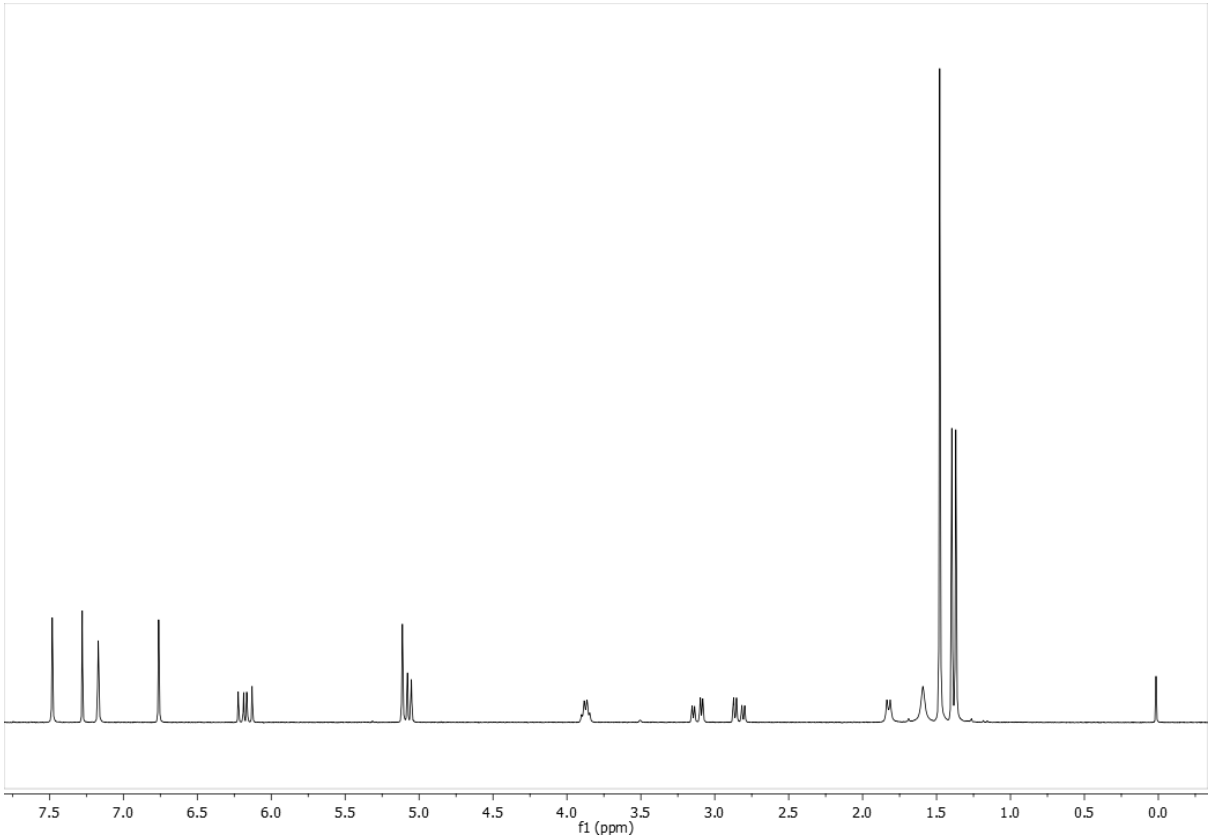


Figura 6.6: ^{1}H NMR del compuesto **2**.

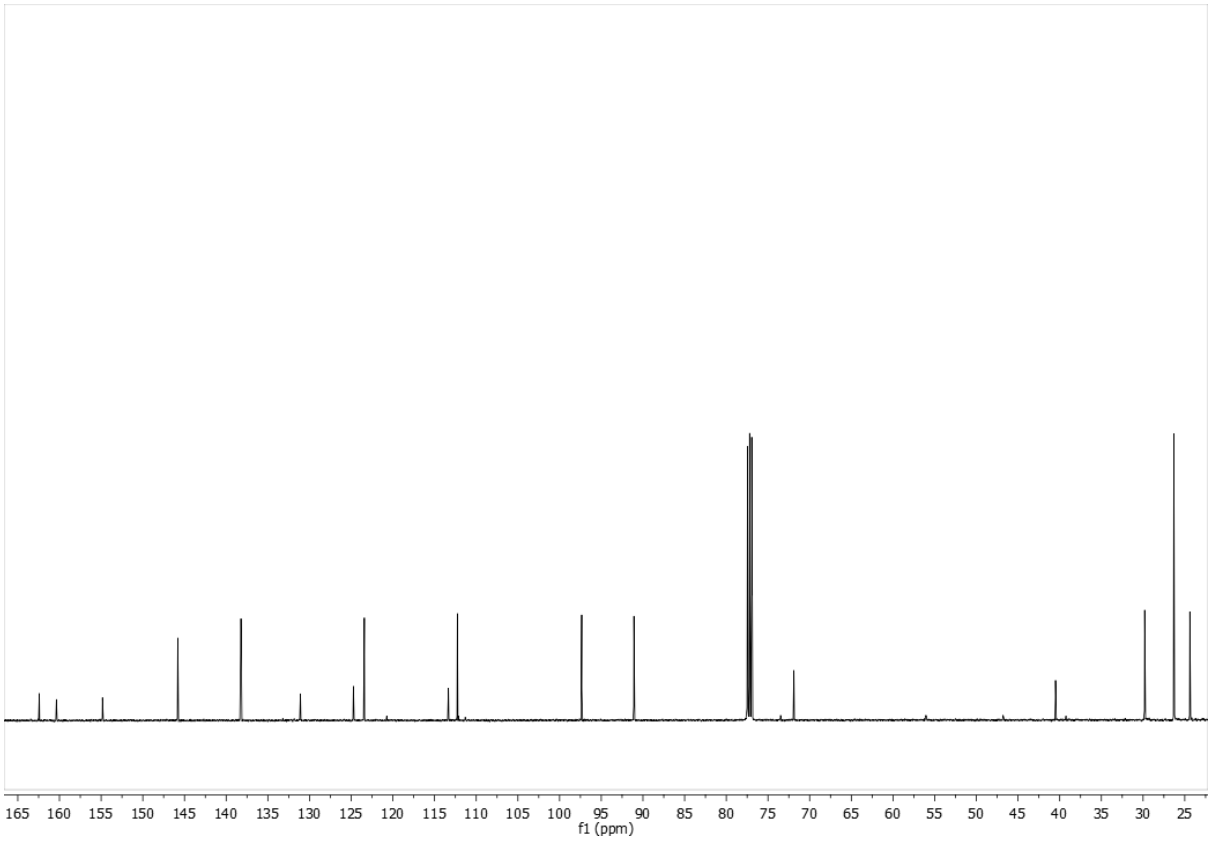


Figura 6.7: ^{13}C NMR del compuesto **2**.

6.2.3.2. Gravelliferona (3)

^{1}H NMR (500 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 7.55 (s, 1H, H-4), 7.17 (s, 1H, H-5), 7.02 (s, 1H, H-8), 6.16 (dd, J = 17, 11 Hz, 1H, H-10), 5.08 (señales superpuestas, 2H, H-11a,b), 1.45 (s, 6H, H-12,13), 3.37 (d, J = 7 Hz, 2H, H-14), 5.32 (dddd, J = 7, 6, 3, 2 Hz, 2H, H-15), 1.75 (s, 3H, H-17), 1.79 (s, 3H, H-18). ^{13}C NMR (126 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 161.3 (C-2), 131.2 (C-3), 138.8 (C-4), 112.8 (C-4a), 128.3 (C-5), 125.5 (C-6), 157.7 (C-7), 102.6 (C-8), 153.4 (C-8a), 40.5 (C-9), 145.8 (C-10), 112.8 (C-11), 26.5 (C-12,13), 18.6 (C-14), 121.5 (C-15), 134.9 (C-16), 26.0 (C-17), 18.0 (C-18).

Referencia (25): 1H NMR (300 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 7.54 (s, 1H, H-4), 7.19 (s, 1H, H-5), 6.94 (s, 1H, H-8), 6.18 (dd, J = 10.8, 16.8 Hz, 1H, H-10), 5.08 (m, 2H, H-11a,b), 1.45 (s, 6H, H-12,13), 3.39 (d, J = 7.2 Hz, 2H, H-14), 5.32 (t, 2H, H-15), 1.78 (s, 3H, H-17), 1.81 (s, 3H, H-18).

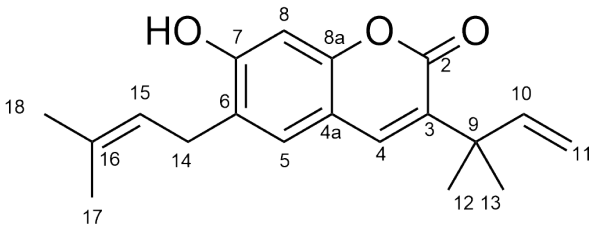


Figura 6.8: Estructura del compuesto 3.

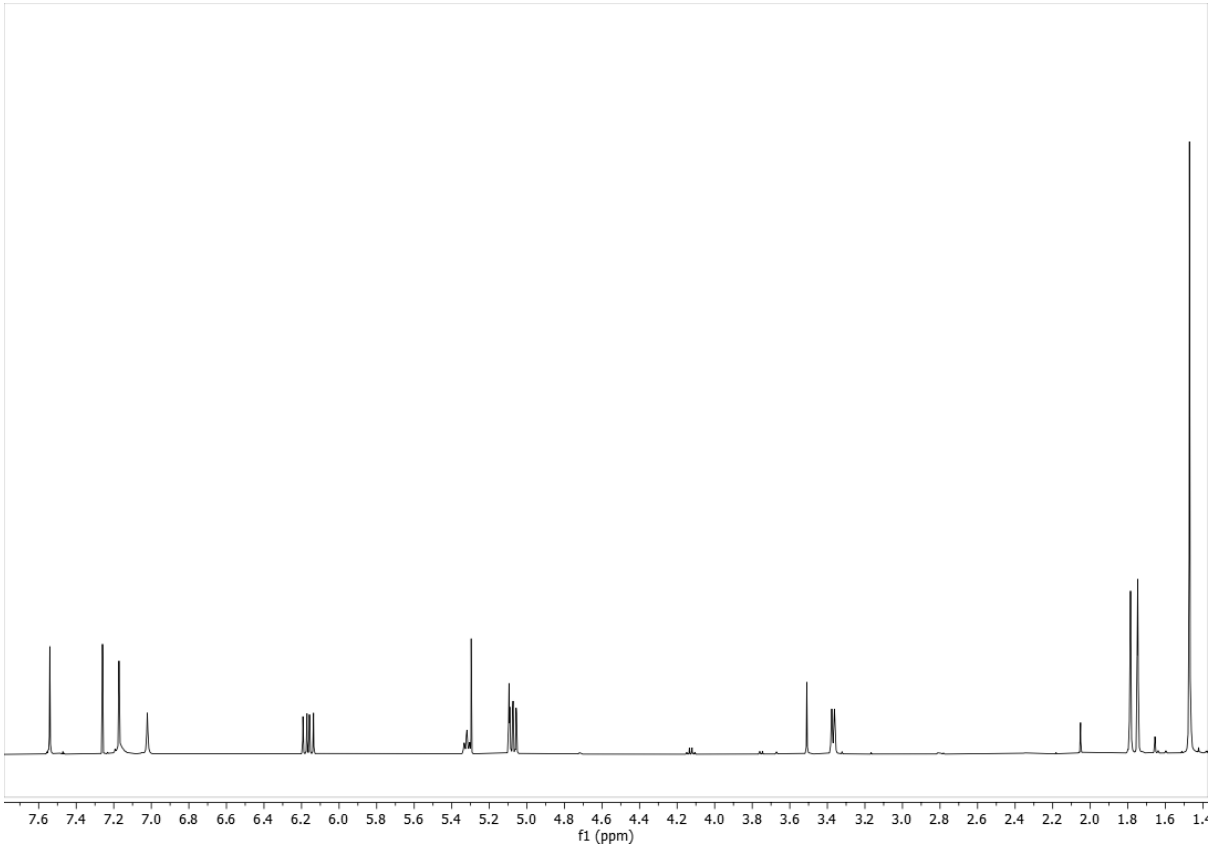


Figura 6.9: ^{1}H NMR del compuesto 3.

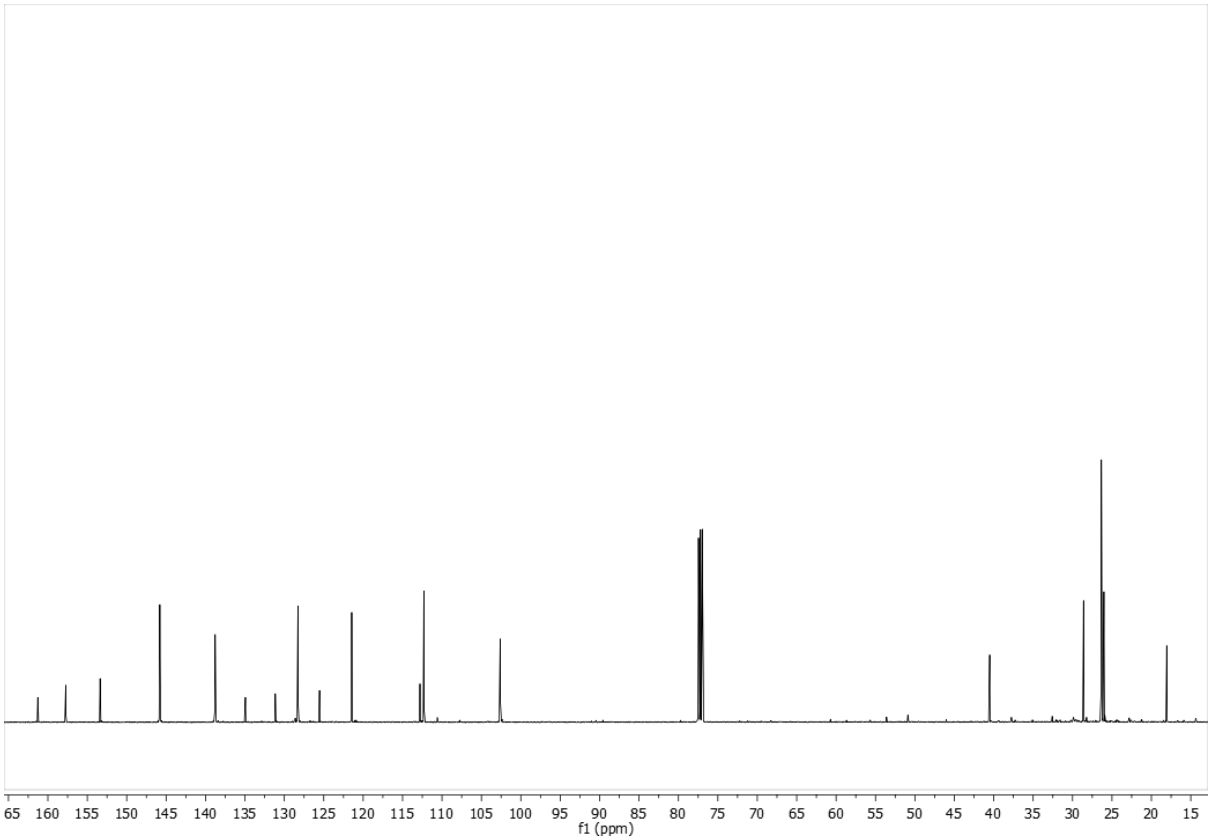


Figura 6.10: ^{13}C NMR del compuesto 3.

6.2.3.3. Rutacultina (4)

^{1}H NMR (200 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 7.51 (s, 1H, H-4), 6.85 (s, 1H, H-5), 6.81 (s, 1H, H-8), 6.20 (m, 1H, H-10), 5.09 (señales superpuestas, 2H, H-11a,b), 1.48 (s, 6H, H-12,13), 3.93 (s, OCH_3), 3.91 (s, OCH_3).

Referencia (26): ^{1}H NMR (Cl_3CD) δ = 7.49 (H-4), 6.85 (H-5), 6.81 (H-8), 6.20 (c, H-10), 5.06 and 5.04 (señales superpuestas, 2H, H-11a,b), 1.50 (s, 6H, H-12,13), 3.92 (s, OCH_3), 3.89 (s, OCH_3).

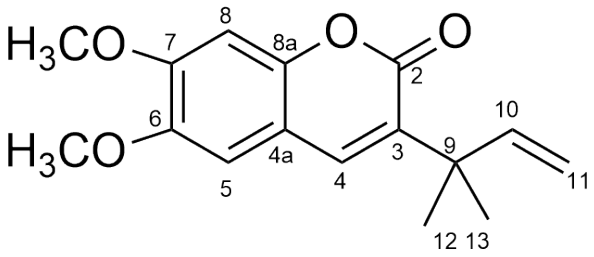


Figura 6.11: Estructura del compuesto **4**.

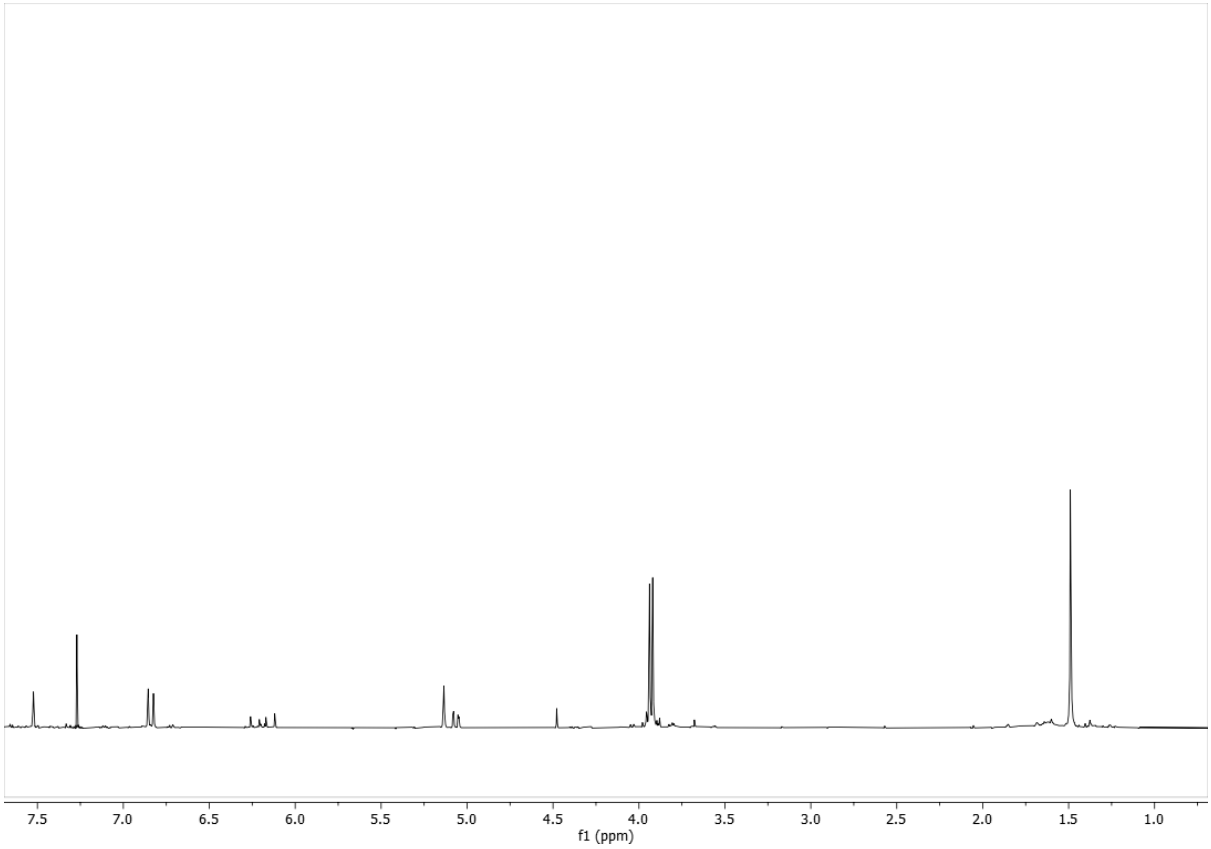


Figura 6.12: ^{1}H NMR del compuesto **4**.

6.2.3.4. *(S)*-Uliscomoncadina A (5)

(1):1H NMR (500 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 7.47 (s, 1H, H-4), 7.16 (s, 1H, H-5), 6.75 (s, 1H, H-8), 6.16 (dd, J = 17, 11 Hz, 1H, H-10), 5.09 (señales superpuestas, 2H, H-11a,b), 1.48 (s, 6H, H-12,13), 3.10 (ddd, J = 17, 5, 1 Hz, 1H, H-14a), 2.82 (dd, J = 17, 5 Hz, 1H, H-14b), 3.86 (q, J = 5 Hz, 1H, H-15), 1.35 (s, 3H, H-17), 1.37 (s, 3H, H-18). ^{13}C NMR (126 MHz, $CDCl_3$, 25°C, TMS, ppm) δ = 160.3 (C-2), 132.1 (C-3), 137.6 (C-4), 113.6 (C-4a), 128.9 (C-5), 116.2 (C-6), 155.6 (C-7), 104.1 (C-8), 153.5 (C-8a), 40.6 (C-9), 145.6 (C-10), 112.3 (C-11), 26.3 (C-12,13), 30.9 (C-14), 69.5 (C-15), 78.1 (C-16), 25.1 (C-17), 22.2 (C-18). (27) $[\alpha]_D^{25}$ = +28,2 (c 0.38 in MeOH).

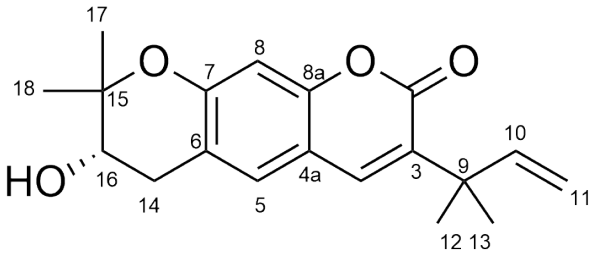


Figura 6.13: Estructura del compuesto **5**.

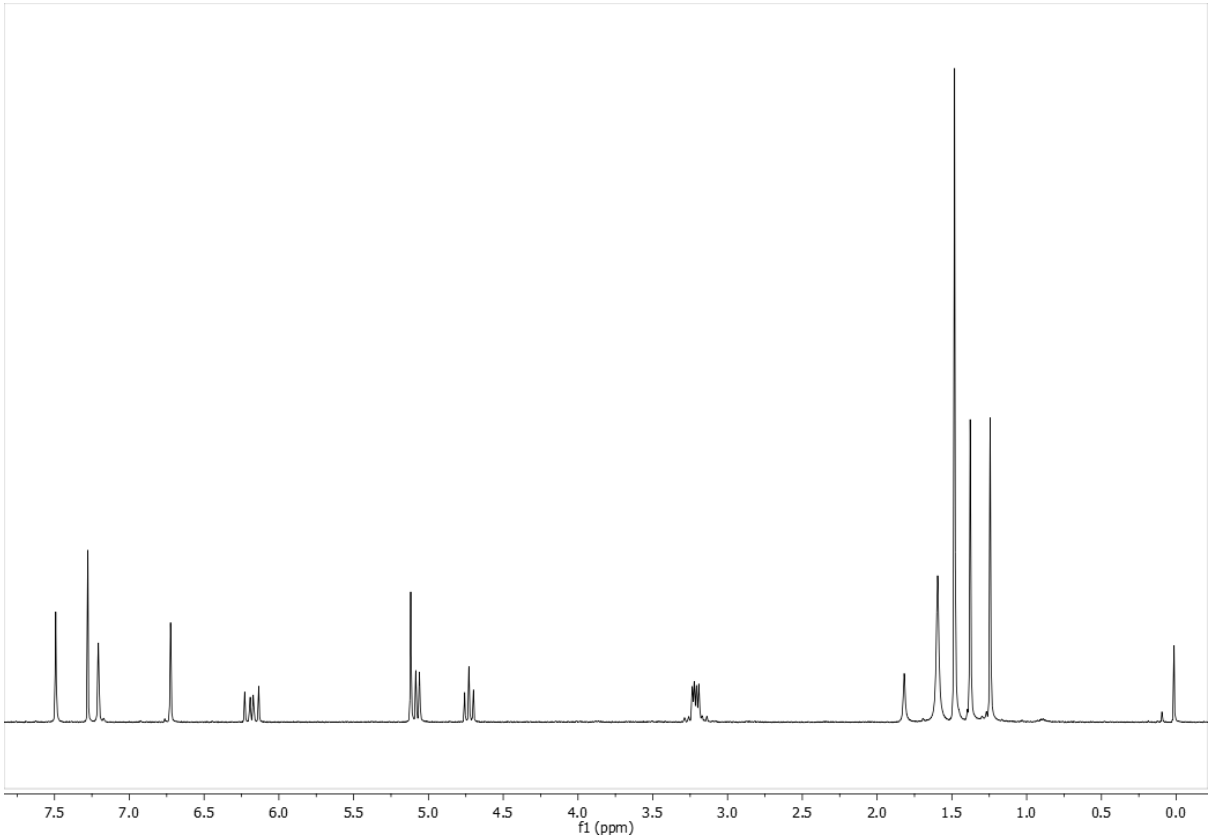


Figura 6.14: ^{1}H NMR del compuesto **5**.

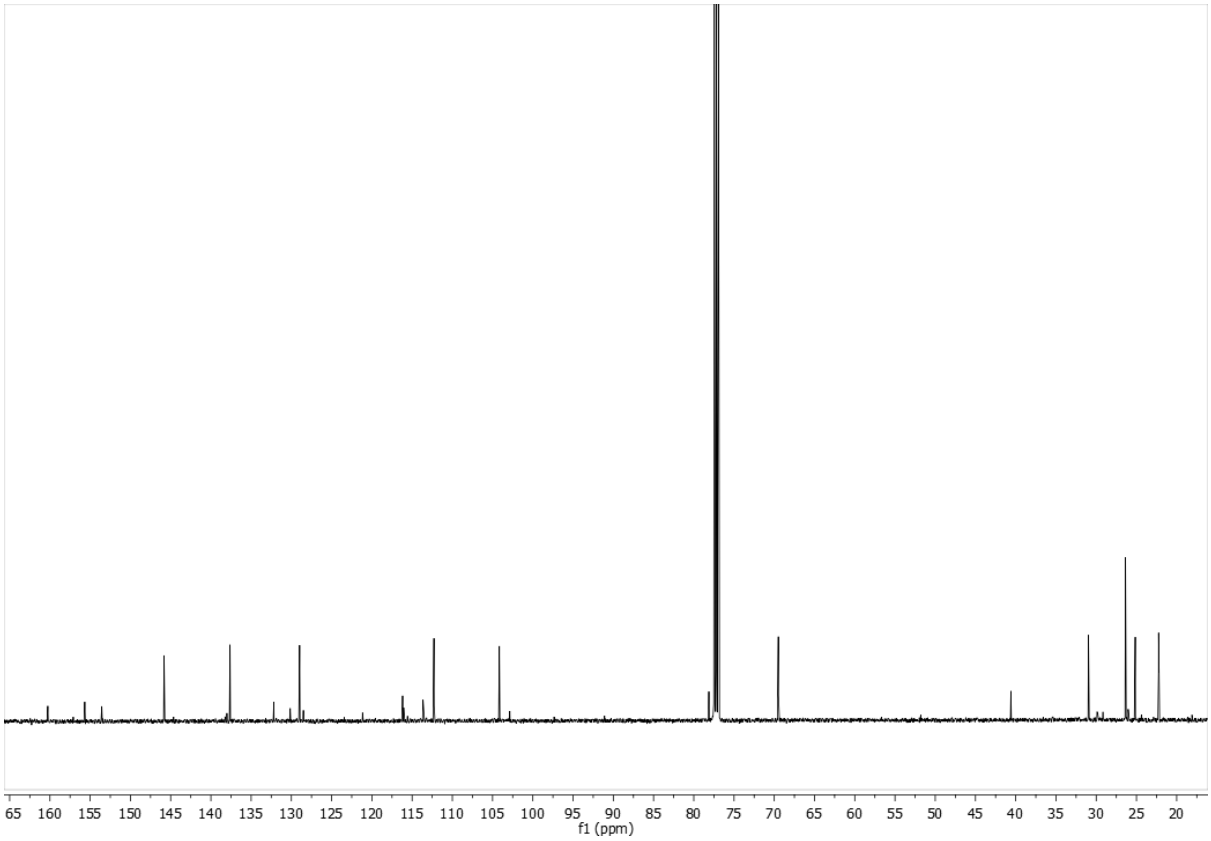


Figura 6.15: ^{13}C NMR del compuesto **5**.

6.2.4. Protocolos de síntesis

6.2.4.1. Síntesis del compuesto 32

Se disolvieron 100 mg del éster de partida en 5 mL de metanol, y se añadió un exceso de hidróxido de potasio (KOH, 10 equiv.) disuelto en 1 mL de agua. La mezcla se agitó a temperatura ambiente (22–25 °C) durante 2 horas. Finalizado este tiempo, se evaporó el metanol bajo presión reducida utilizando un evaporador rotatorio. El residuo acuoso se acidificó con HCl 1 M y se extrajo con acetato de etilo (3 × 10 mL). Dado que el compuesto obtenido presentaba baja solubilidad en la fase orgánica, fue necesario utilizar un volumen mayor del habitual para asegurar una extracción completa. La fase orgánica se secó sobre sulfato de sodio anhidro, se filtró y se concentró a presión reducida

El residuo seco se disolvió en 5 mL de diclorometano (DCM) y se le añadió cloruro de tionilo ($SOCl_2$, 3 equiv.). La reacción se dejó bajo agitación a temperatura ambiente durante 2 horas, y el progreso se monitoreó mediante cromatografía en capa fina (TLC), observándose la desaparición del ácido carboxílico intermedio y la aparición de un nuevo producto.

Sin purificación intermedia, se añadió directamente cloruro de aluminio ($AlCl_3$, 2 equiv.) al mismo balón. Se observó un cambio de color pronunciado hacia negro, característico de la formación del producto deseado. Para el análisis por TLC, se realizó una microextracción con NaOH 1 M (2 mL), lo que permitió eliminar el cloruro de ácido residual.

El producto final se aisló como un sólido blanco, con un rendimiento de 86 mg (80 %

con respecto al éster de partida).

6.2.4.2. Síntesis de 32 mediante activación superácida del éster precursor

En un balón seco bajo atmósfera inerte, se añadieron 50 mg del éster de partida y se disolvió en 0.4 mL de ácido trifluorometansulfónico. La mezcla de reacción se mantuvo bajo agitación a temperatura ambiente durante 48 horas, sin necesidad de calentamiento adicional. Luego de comprobar por CCD que el reactivo inicial desapareció por completo, se agrega acetato de etilo a la mezcla y se lleva a una ampolla de decantación. Se lava tres veces con agua destilada y la fase orgánica es llevada a sequedad. El producto final se aisló como un sólido blanco, con un rendimiento de 40 mg (80 %).

6.2.4.3. Compuesto 32

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.97 (s, 1H, H-18), 7.48 (dd, J = 6,8, 1.4 Hz, 1H, H-13), 7.41 (dd, J = 7,8, 6.8 Hz, 1H, H-12), 7.36 (dd, J = 7,8, 1.4 Hz, 1H, H-11), 6.04 (s, 1H, H-16), 4.08 (s, 3H, H-21), 3.26 (s, 3H, H-10). ^{13}C NMR (125 MHz, $CDCl_3$) δ 195.23 (C-15), 150.21 (C-6), 149.53 (C-4), 136.18 (C-3), 136.01 (C-2), 133.51 (C-14), 132.77 (C-18), 132.09 (C-12), 133.19 (C-11), 121.98 (C-13), 119.99 (C-16), 113.35 (C-5), 40.64 (C-21), 40.36 (C-10).

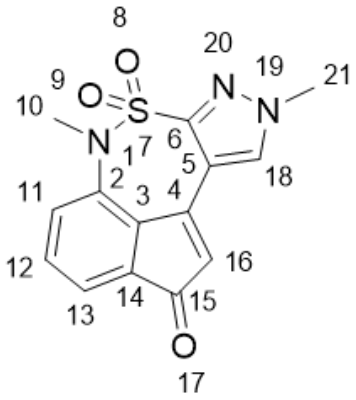


Figura 6.16: Estructura del compuesto **32**.

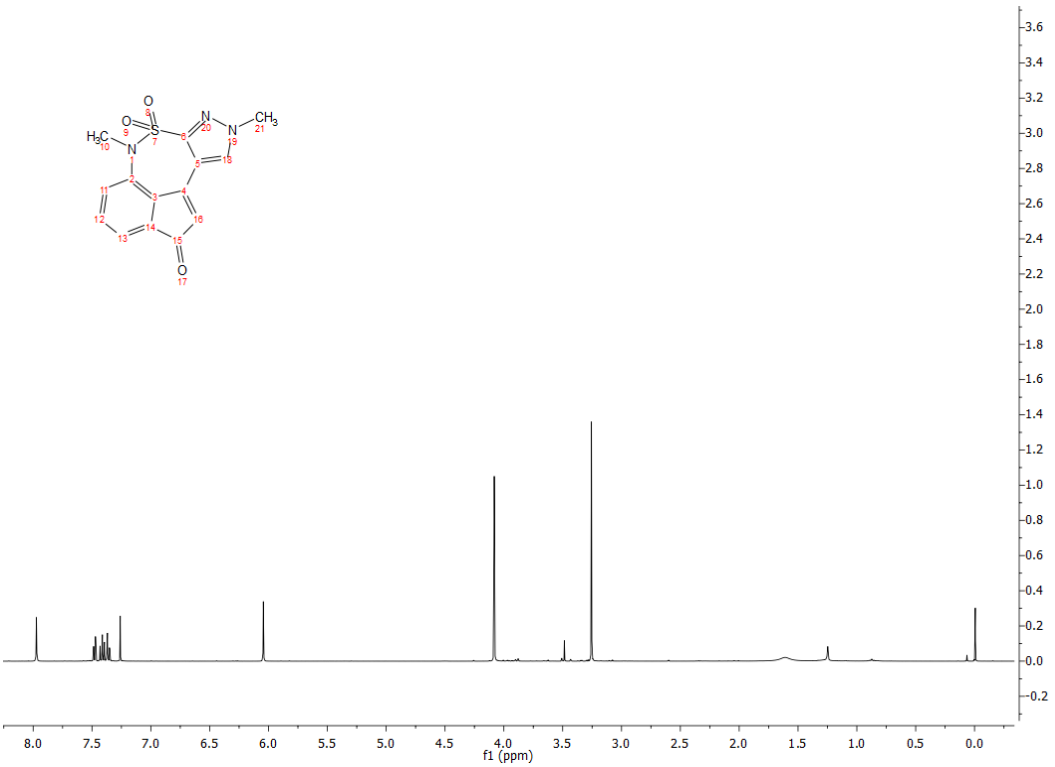


Figura 6.17: ^{1}H NMR del compuesto **32**.

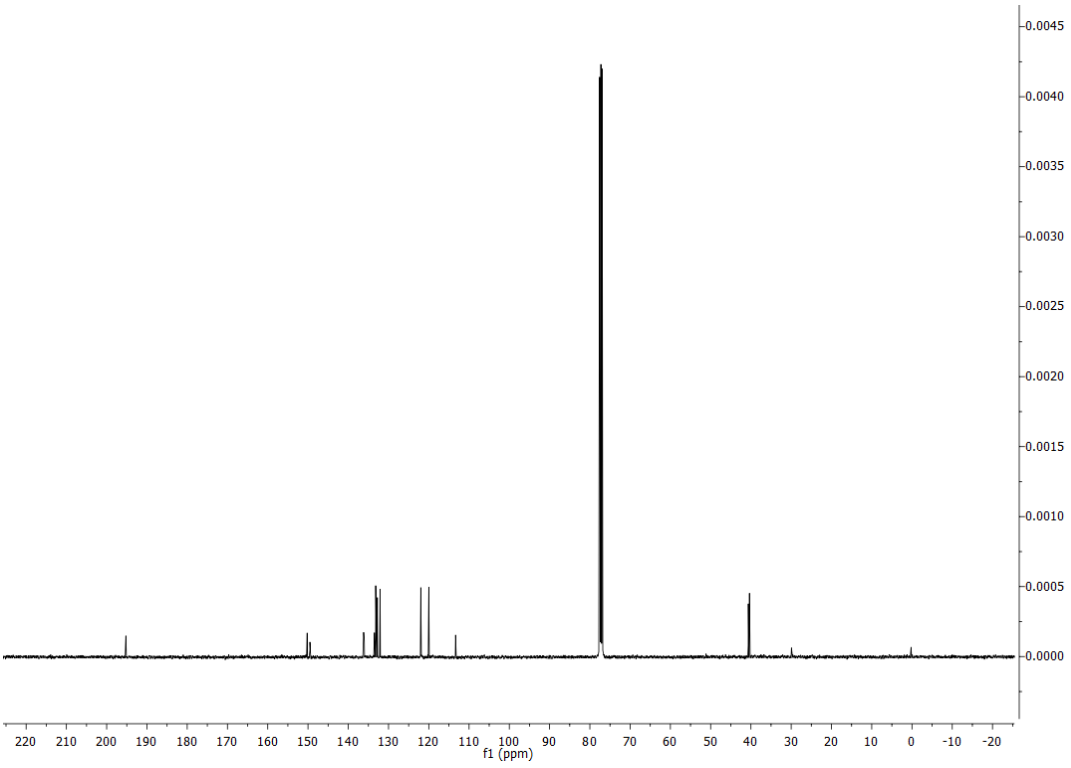


Figura 6.18: ^{13}C NMR del compuesto **32**.

6.2.5. Síntesis clásica de triptaminas

6.2.5.1. Síntesis del derivado formilado mediante reacción de Vilsmeier–Haack

En un balón se colocaron 1.1 equivalentes de N,N-dimetilformamida (DMF) y se enfrió en un baño de hielo. Bajo agitación, se añadió lentamente tribromuro de fósforo (PBr_3) en cantidad estequiométrica, manteniendo la temperatura baja durante 30 minutos. En algunos casos, se observó la formación de una mezcla viscosa; en tales situaciones, se añadió diclorometano (CH_2Cl_2) para facilitar la agitación.

Transcurrido ese tiempo, se añadió gota a gota indol (1.0 equivalente) disuelto en CH_2Cl_2 (o alternativamente en DMF), y se dejó la reacción bajo agitación a temperatura ambiente durante 3 horas. La reacción se monitorea por cromatografía en capa fina (TLC, éter de petróleo : acetato de etilo (1:1)) para observar la desaparición del indol de partida.

Una vez completada la reacción, se adicionó acetato de sodio (NaOAc) 1 M para cortar la reacción. La mezcla se mantuvo bajo agitación a temperatura ambiente hasta la aparición de un precipitado blanco a amarillento y la decoloración de la fase líquida. El precipitado obtenido se eliminó por filtración y se lavó con agua para eliminar el exceso de DMF. Dependiendo del sustituyente del indol puede que no se observe precipitación, en esos casos la mezcla se calienta a 100°C por 10 minutos y luego se extrae con acetato de etilo. En todos los casos se obtiene el compuesto buscado con rendimientos de moderados a altos.

6.2.5.2. Síntesis del derivado nitrovinílico mediante reacción de Henry

En un balón se colocó el derivado indol-aldehído (1.0 equivalente) junto con una pequeña cantidad de acetato de amonio (p.ej., una punta de espátula). A continuación, se añadió nitrometano en exceso, el cual actuó como solvente de la reacción. La mezcla se calentó a reflujo durante 1 hora, manteniendo la temperatura del baño de calentamiento en aproximadamente 90 ºC.

Durante el transcurso de la reacción, se observó la formación de un sólido amarillo, indicativo del avance del proceso. Una vez completada la conversión, la mezcla se dejó enfriar a temperatura ambiente y se adicionó agua destilada para cortar la reacción. Posteriormente, se realizó una extracción del producto orgánico utilizando acetato de etilo. Una vez llegado a sequedad se obtiene el producto como un aceite amarillo con rendimientos superiores al 90 %.

6.2.5.3. Reducción de nitro olefinas a triptaminas mediante $LiAlH_4$

Bajo atmósfera inerte de nitrógeno, se preparó una suspensión agitada de hidruro de litio y aluminio ($LiAlH_4$) (6 equivalentes) en tetrahidrofurano (THF, volumen igual al del reactivo) en un balón de fondo redondo. A esta suspensión se añadió gota a gota una solución de nitro olefina (1 equivalente) en THF (10 mL por mmol de sustrato) a 0 ºC.

La mezcla se dejó alcanzar la temperatura ambiente y se agitó durante 36 horas. Finalizado el tiempo de reacción, se procedió a la eliminación del exceso de $LiAlH_4$ mediante la adición lenta de agua destilada hasta que cesó la efervescencia. La mezcla resultante se diluyó con éter etílico y luego se añadió una solución acuosa saturada de sal de Rochelle

(tartrato sódico potásico), manteniendo la agitación durante 24 horas adicionales.

Las fases se separaron y la fase orgánica se extrajo con HCl acuoso 1 M. La fase acuosa ácida se alcalinizó con KOH acuoso 3 M hasta pH básico, y se extrajo nuevamente con éter etílico. Las fracciones orgánicas se reunieron, se secaron sobre sulfato de sodio anhidro, se filtraron y concentraron *in vacuo* para obtener la triptamina deseada, que no requirió purificación adicional.

6.2.5.4. Formación de amidas mediante acoplamiento con pyBOP

El ácido carboxílico correspondiente (1 equivalente), (PyBOP) (1 equivalente) y la trietilamina (TEA) (3 equivalentes) se disolvieron en DMF anhidro bajo atmósfera de argón y se agitaron durante 15 minutos a temperatura ambiente. Posteriormente, se añadió la amina primaria adecuada (2 equivalentes) en una sola porción, y la reacción se dejó agitar durante 18 horas a temperatura ambiente. Luego, la mezcla se vertió en 75 mL de acetato de etilo (EtOAc) y 25 mL de agua destilada. La capa acuosa se desechó y la capa orgánica se lavó dos veces con agua (25 mL) y una vez con agua salada (25 mL). La capa orgánica se secó con sulfato de magnesio y se evaporó bajo presión reducida. Los sólidos crudos se purificaron por cromatografía en columna (5–20 % EtOAc/Hexanos) para obtener los productos puros.

6.2.5.5. Obtención de triptaminas a partir de la hidrogenación de nitro etil indoles

El líquido amarillo obtenido en el paso anterior fue sometido a una reducción catalítica con hidrógeno (H_2) sobre Pd/C. Para ello, se colocó el líquido amarillo en un hidrogenador

y se utilizó metanol como solvente y se dejó reaccionar durante 8 horas.

Una vez completada la reacción, se filtró la mezcla a través de un lecho de Celite para eliminar el catalizador. El filtrado obtenido contenía la triptamina, que se utilizó sin necesidad de una purificación adicional.

6.2.5.6. Formación de Aspidostomidas *one pot* mediante acoplamiento con PyBOP y síntesis de indoles de Fischer

El protocolo comienza con el acoplamiento utilizando PyBOP, siguiendo el procedimiento previamente descrito, pero con diclorometano (DCM) como solvente y 4-amino butiraldehído dietil acetal como la amina. Una vez completado el acoplamiento, se evapora el solvente y, en el mismo balón, se lleva a cabo la reacción de Fischer agregando ácido sulfúrico al 4 % y cloruro de zinc ($ZnCl_2$), dejando reaccionar a reflujo durante 2 horas.

Al finalizar la reacción, el crudo obtenido se extrae con agua y acetato de etilo. El producto final se obtiene con alta pureza.

6.2.5.7. Iodación de indoles en C-3

El indol fue yodado utilizando yodo molecular (1 equivalente) y KOH (2.5 equivalentes) en DMF, a temperatura ambiente durante 3 horas. La reacción se detuvo mediante la adición de una solución saturada de bisulfito de sodio, se ajustó el pH a 3 con HCl y se extrajo con acetato de etilo. El producto se obtuvo como un sólido marrón, con altos rendimientos.

6.2.5.8. *N*-Metilación de 3-yodoindol

A una solución de KOH (3 equivalentes) en DMF se le agregó 3-yodoindol y se dejó reaccionar durante 30 minutos. Posteriormente, se añadió yoduro de metilo (1.1 equivalentes) y se continuó la reacción durante 1 hora. El progreso fue monitoreado por TLC (éter de petróleo:acetato de etilo 9:1), observando la desaparición del compuesto de partida.

6.2.5.9. Reacción de Sonogashira

A una solución de 3-yodo-*N*-metilindol en DMF anhidro, bajo atmósfera de argón, se agregaron CuI (0.3 equivalentes) y $Pd(PPh_3)_2Cl_2$ (0.15 equivalentes), previamente secados al vacío durante 1 hora. Luego se añadió trietilamina (1.2 equivalentes) y trimetilsiletinilo (1.5 equivalentes), dejando reaccionar durante 3 horas. El progreso se controló por TLC (éter de petróleo:acetato de etilo 9:1). El producto fue tratado posteriormente con fluoruro de tetrabutilamonio en THF durante 12 horas, obteniéndose el derivado 3-alquinoindol con rendimientos moderados.

6.2.5.10. Formación de azidas a partir de halogenuros de alquilo

A una solución del halogenuro de alquilo en DMF se le adicionó un exceso de azida de sodio (NaN_3), y se dejó reaccionar durante 3 horas, monitoreando por TLC (diclorometano : éter de petróleo 1:1). Una vez finalizada la reacción, se lavó con solución salina saturada (NaCl) y agua (3 veces cada una).

6.2.5.11. Formación de triazoles a partir de una reacción "click"

A una solución del alquino (6.24 mg, 1.0 equivalente) en DMF se le añadió ascorbato de sodio (0.25 equivalentes) y $CuSO_4 \cdot 5H_2O$ (0.5 equivalentes), dejándose reaccionar durante 30 minutos. Luego se incorporó la azida correspondiente (8.25 mg, 1.1 equivalentes), y se continuó la reacción durante 3 horas. El seguimiento se realizó por TLC (éter de petróleo:acetato de etilo 9:1). Una vez finalizada, se realizaron tres lavados con solución salina saturada y agua. Debido a la baja masa obtenida, el producto crudo se purificó por cromatografía en placa, aislándose 3.23 mg (39 %) de un aceite incoloro.

6.2.6. Información estructural de los compuestos sintetizados

6.2.6.1. Compuesto 35

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.3, 6.2.5.4, rendimiento: 12 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.65 (d, J = 6,9 Hz, 1H, H-4), 7.39 (d, J = 8,2 Hz, 1H, H-7), 7.23 (t, J = 7,1 Hz, 1H, H-6), 7.15 (t, J = 7,1 Hz, 1H, H-5), 7.07 (d, J = 2,3 Hz, 1H, H-2), 6.90 (m, 1H, H-5′), 6.35 (m, 1H, H-3′), 6.19 (m, 1H, H-4′), 3.76 (q, J = 6,6 Hz, 2H, H-β), 3.07 (t, J = 6,4 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 161.1 (Cγ), 136.8 (C7a), 127.7 (C3a), 126.3 (C2′), 122.7 (C6), 122.5 (C2), 121.4 (C5′), 120.0 (C5), 119.2 (C4), 113.5 (C7), 111.6 (C3), 110.1 (C4′), 108.8 (C3′), 40.2 (Cβ), 26.1 (Cα). HRESIMS calculado $[C_{15}H_{16}N_3O]^+$ 254,1288 obtenido 254,1261

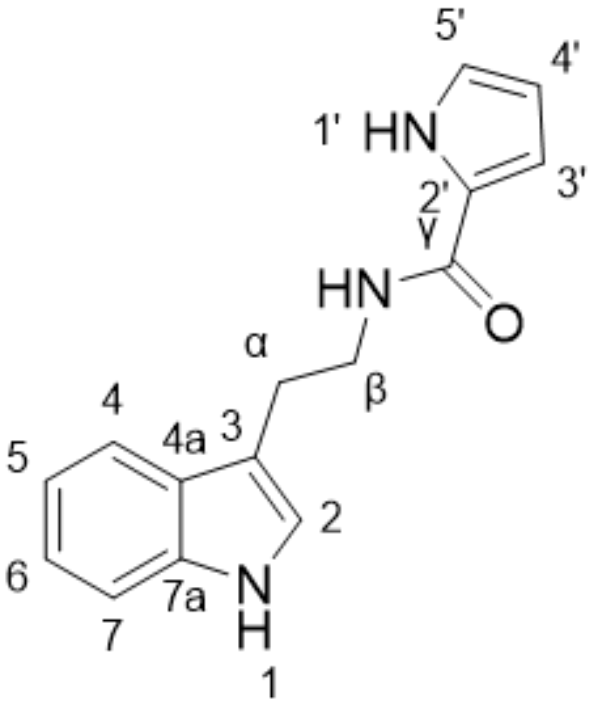


Figura 6.19: Estructura del compuesto 35.

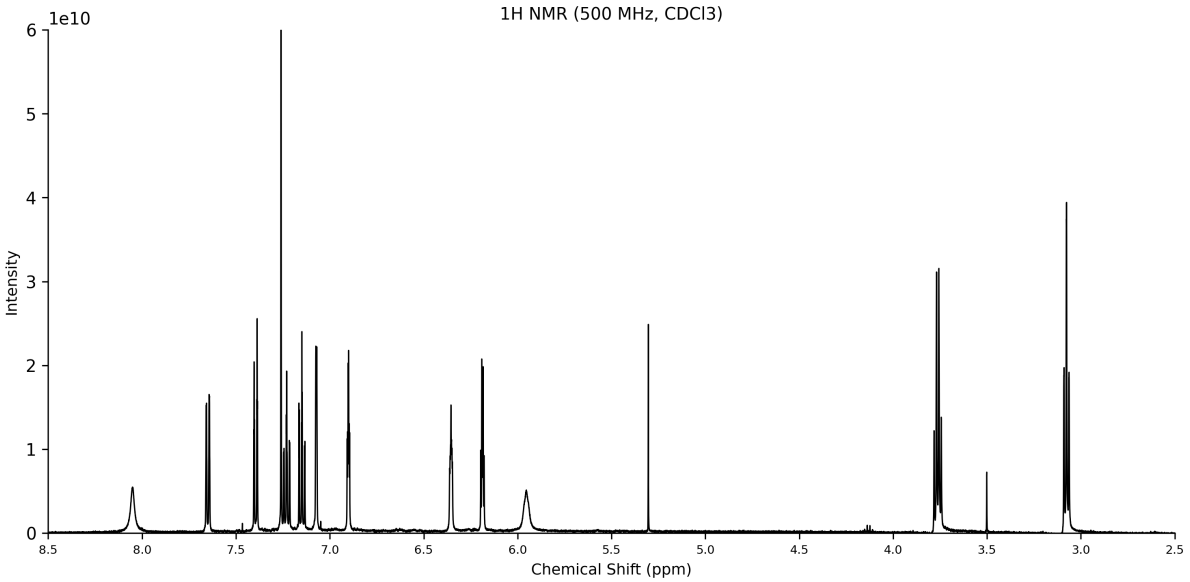


Figura 6.20: ^{1}H NMR del compuesto 35.

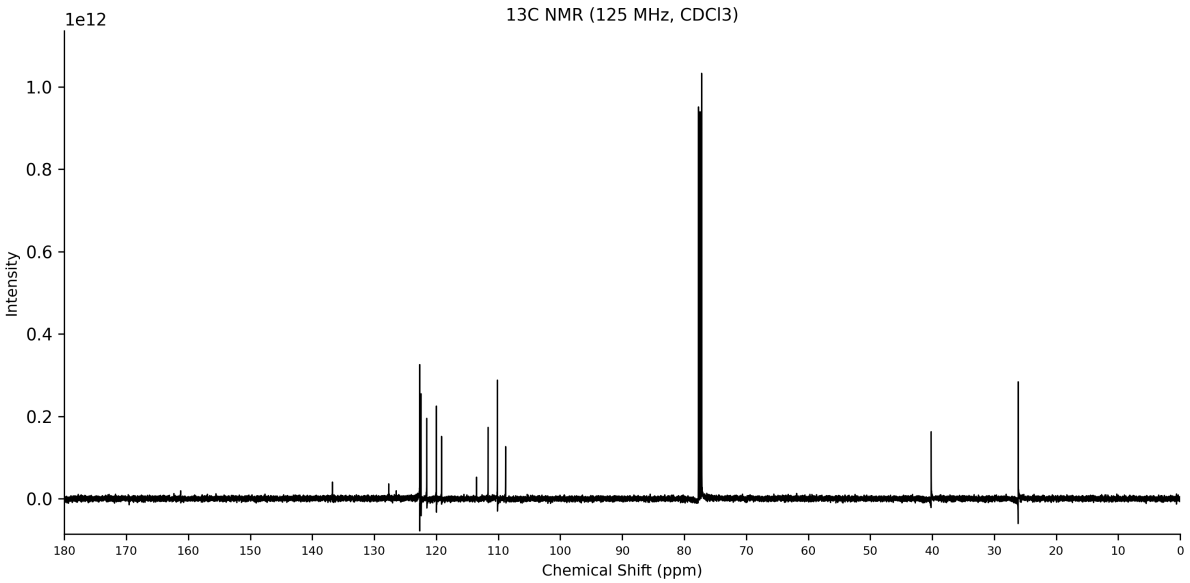


Figura 6.21: ^{13}C NMR del compuesto 35.

‘

6.2.6.2. Compuesto 36

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.3, 6.2.5.4, rendimiento: 40 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.65 (d, J = 7,9 Hz, 1H, H-4), 7.39 (d, J = 7,9 Hz, 1H, H-7), 7.35 (s, 1H, H-5′), 7.28–7.18 (t, 1H, H-6), 7.17–7.11 (t, 1H, H-5), 7.09 (m, 1H, H-3′), 6.47 (m, 1H, H-4′), 3.77 (q, J = 6,8 Hz, 2H, H-β), 3.10 (s, 1H, H-2), 3.08 (t, J = 6,8 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 158.6 (Cγ), 148.3 (C2′), 143.9 (C5′), 136.6 (C7a), 127.4 (C3a), 122.4 (C6), 122.2 (C2), 119.7 (C5), 118.9 (C4), 114.1 (C3′), 113.1 (C3), 112.2 (C4′), 111.4 (C7), 39.5 (Cβ), 25.7 (Cα). HRESIMS calculado $[C_{15}H_{15}N_2O_2]^+$ 255,1128 obtenido 255,1118

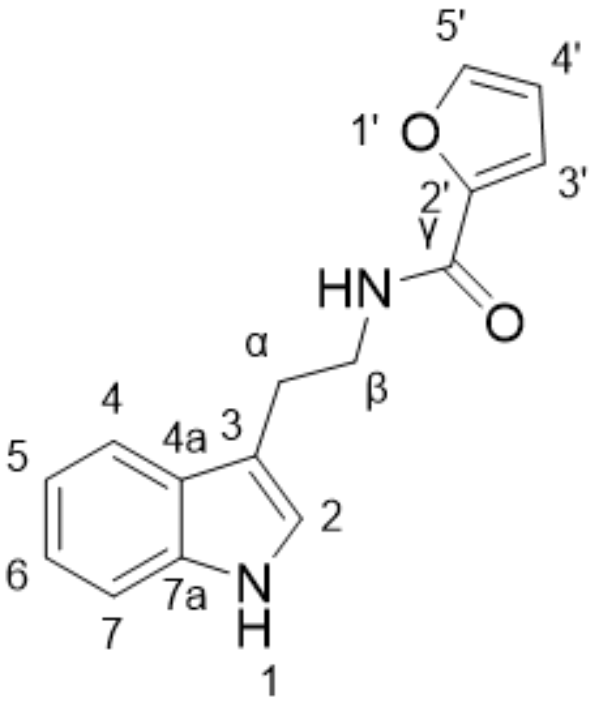


Figura 6.22: Estructura del compuesto 36.

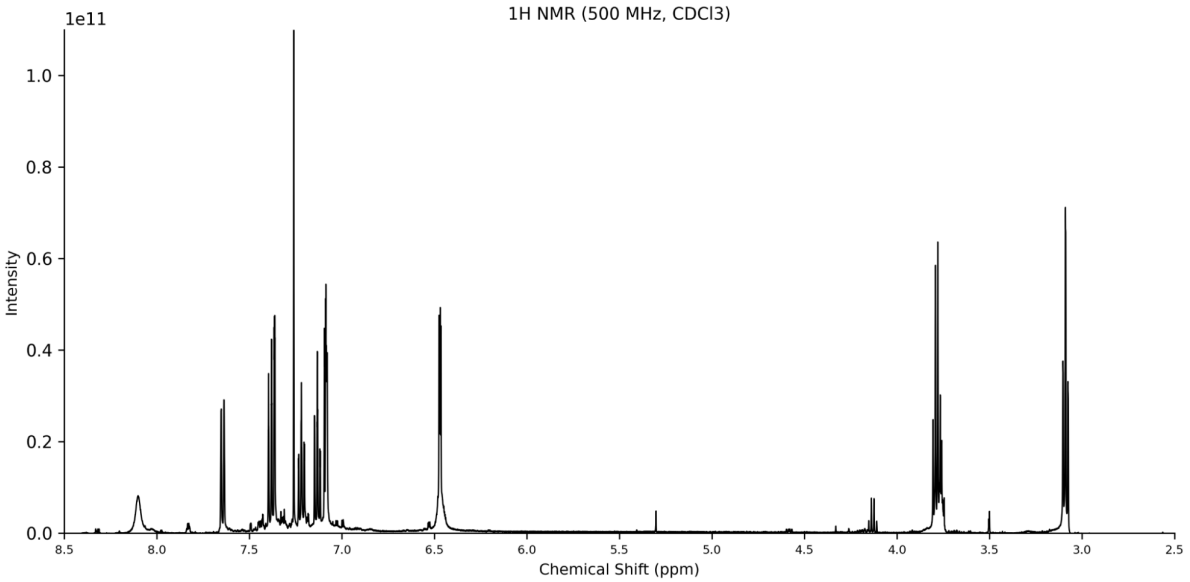


Figura 6.23: ^{1}H NMR del compuesto 36.

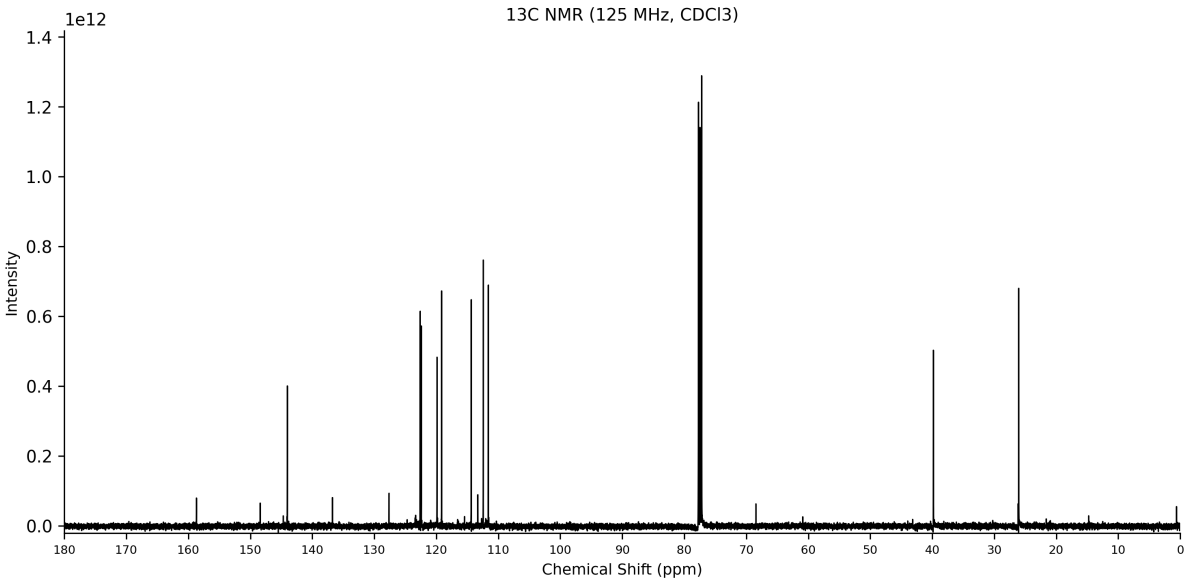


Figura 6.24: ^{13}C NMR del compuesto 36.

6.2.6.3. Compuesto 37

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.3, 6.2.5.4, rendimiento: 35 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.64 (d, J = 8 Hz, 1H, H-4), 7.4 (dd, J = 1,6, 0.9 Hz, 1H, H-2′), 7.4 (t, J = 1,6 Hz, 1H, H-5′), 7.4 (d, J = 8,2 Hz, 1H, H-7), 7.23 (td, J = 8,2, 1.2 Hz, 1H, H-6), 7.15 (td, J = 8,0, 1.0 Hz, 1H, H-5), 6.47 (m, 1H, H-4′), 3.75 (q, J = 6,7 Hz, 2H, H-β). ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.7 (Cγ), 144.7 (C2′), 143.8 (C5′), 136.6 (C7a), 127.5 (C3a), 122.9 (C6), 122.5 (C3′), 122.2 (C5), 119.8 (C2), 118.9 (C4), 113.3 (C3), 111.5 (C7), 108.3 (C4′), 40.0 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{15}H_{15}N_2O_2]^+$ 255,1128 obtenido 255,1148

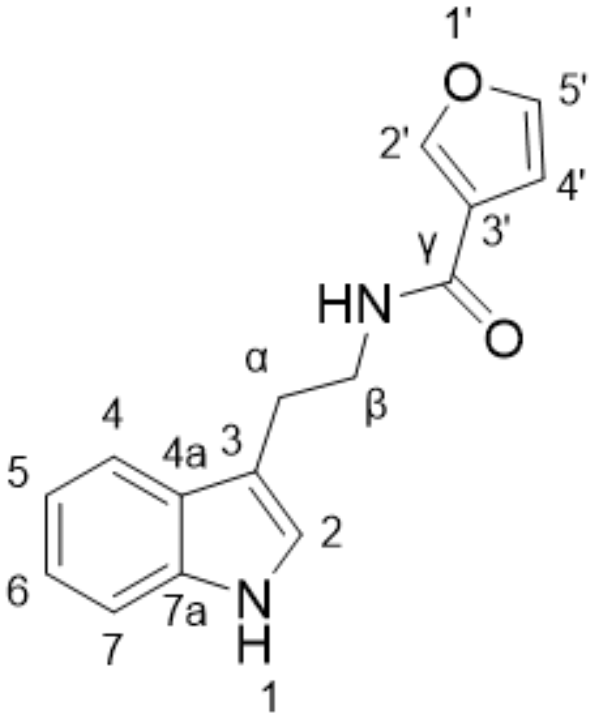


Figura 6.25: Estructura del compuesto 37.

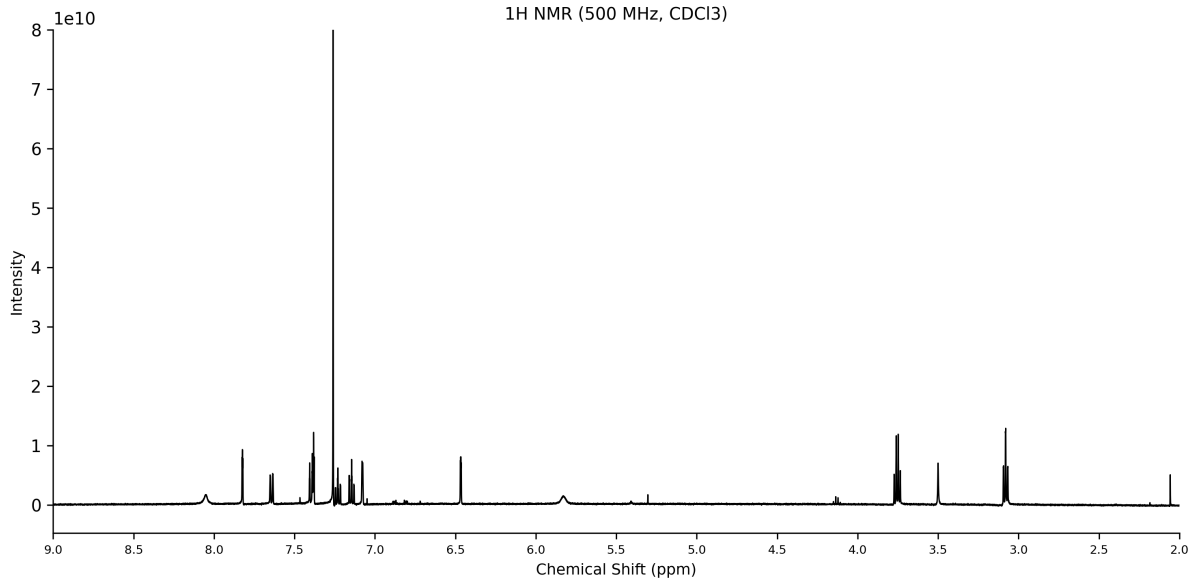


Figura 6.26: ^{1}H NMR del compuesto 37.

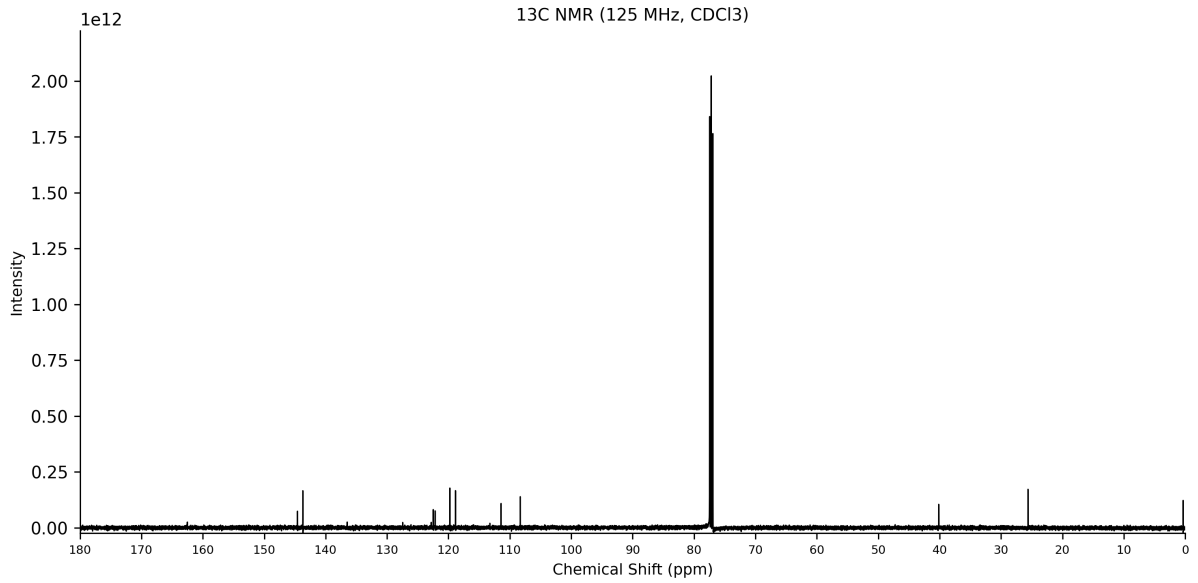


Figura 6.27: ^{13}C NMR del compuesto 37.

6.2.6.4. Compuesto 38

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.5, 6.2.5.4, rendimiento: 40 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.73 (dd, J = 2,9, 1.4 Hz, 1H), 7.65 (d, J = 7,9 Hz, 1H, H-4), 7.39 (dt, J = 8,1, 0.9 Hz, 1H, H-7), 7.31–7.22 (m, 1H), 7.31–7.22 (m, 1H), 7.23 (td, J = 8,0, 1.2 Hz, 1H, H-6), 7.14 (t, J = 8,0, 0.9 Hz, 1H, H-5), 3.77 (q, 2H, H-β). ^{13}C NMR (125 MHz, $CDCl_3$) δ 163.2 (Cγ), 137.8 (C3′), 136.6 (C7a), 128.1, 127.5 (C3a), 126.5, 126.1, 122.5 (C6), 122.3 (C2), 119.8 (C5), 118.9 (C4), 113.2 (C3), 111.5 (C7), 40.2 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{15}H_{15}N_2OS]^+$ 271,0900 obtenido 271,0897

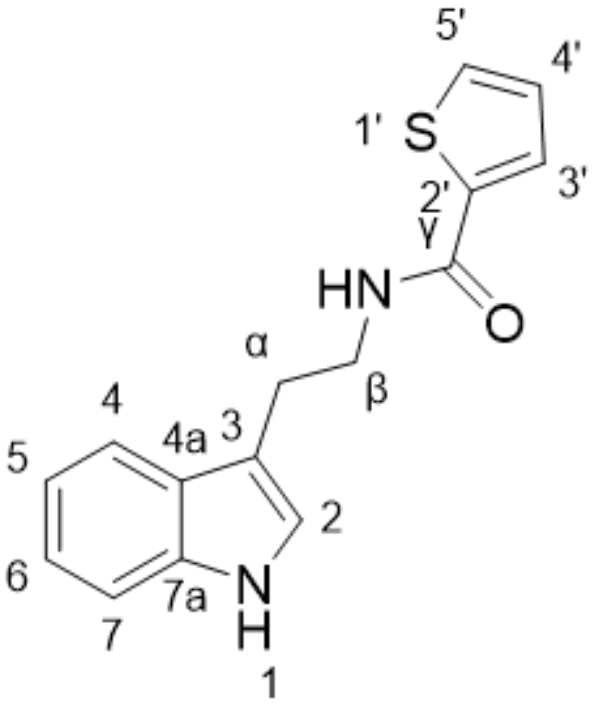


Figura 6.28: Estructura del compuesto 38.

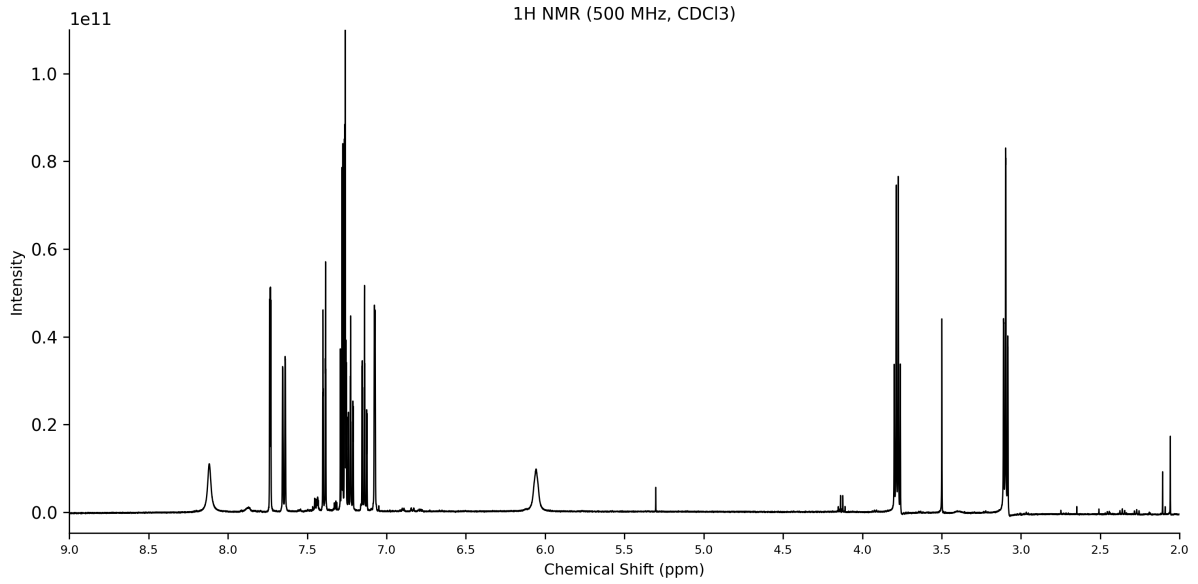


Figura 6.29: ^{1}H NMR del compuesto 38.

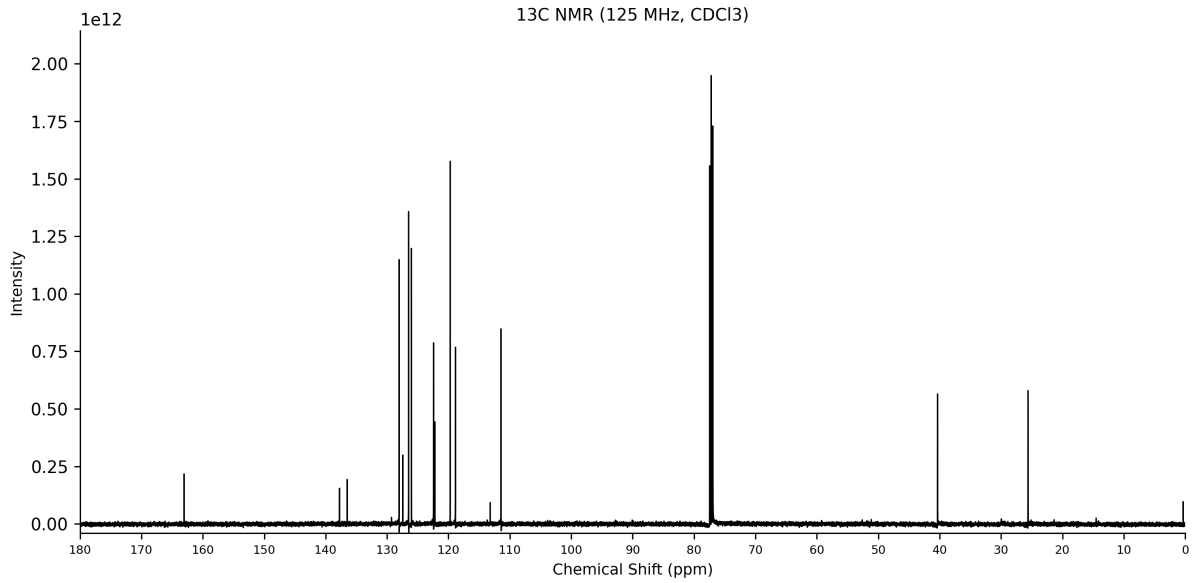


Figura 6.30: ^{13}C NMR del compuesto 38.

6.2.6.5. Compuesto 40

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.5, 6.2.5.4, rendimiento: 35 %. ^{1}H RMN (500 MHz, $CDCl_3$) δ 7.74 (dd, J = 2,9, 1.5 Hz, 1H, H-2'), 7.45 (d, J = 7,5 Hz, 2H, H-4"), 7.38 (t, J = 7,5 Hz, 1H, H-2"), 7.32 (t, J = 7,5 Hz, 1H, H-3"), 7.28 (m, H-4',5'), 7.28 (d, J = 8,8 Hz, 1H, H-7), 7.26 (m, H-4',5'), 7.14 (d, J = 2,4 Hz, 1H, H-4), 7.05 (d, J = 2,4 Hz, 1H, H-2), 6.96 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 5.00 (s, 2H, H-ε), 3.75 (q, J = 6,3 Hz, 2H, H-β), 3.04 (t, J = 6,5 Hz, 2H, H-α). ^{13}C RMN (125 MHz, $CDCl_3$) δ 163.1 (Cγ), 153.5 (C-5), 137.8 (C-3'), 137.7 (C-1"), 131.8 (C-7a), 128.6 (C-2"), 128.2 (C-3"), 128.0 (C-3a), 128.0 (C-2'), 127.8 (C-4"), 126.6 (C-4',5'), 126.1 (C-4',5'), 123.1 (C-2), 113.4 (C-6), 113.1 (C-3), 112.2 (C-7), 102.2 (C-4), 71.0 (Cε), 40.3 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{16}H_{16}N_3O]^+$ 266,1288 obtenido 266,1282

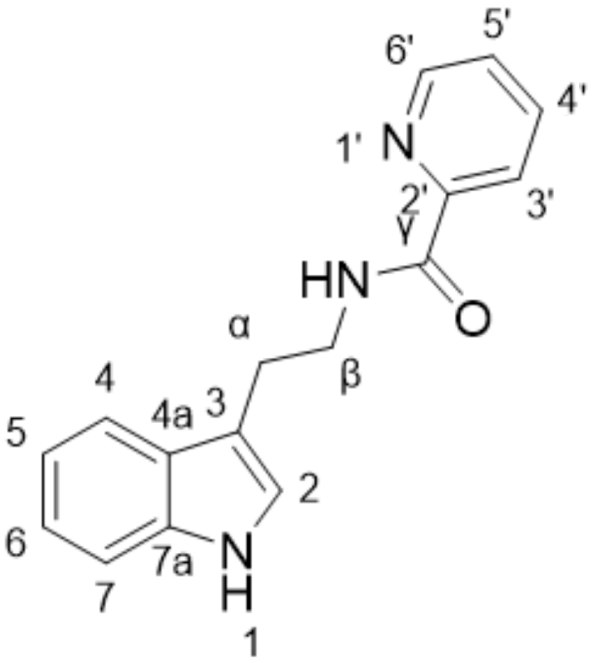


Figura 6.31: Estructura del compuesto 40.

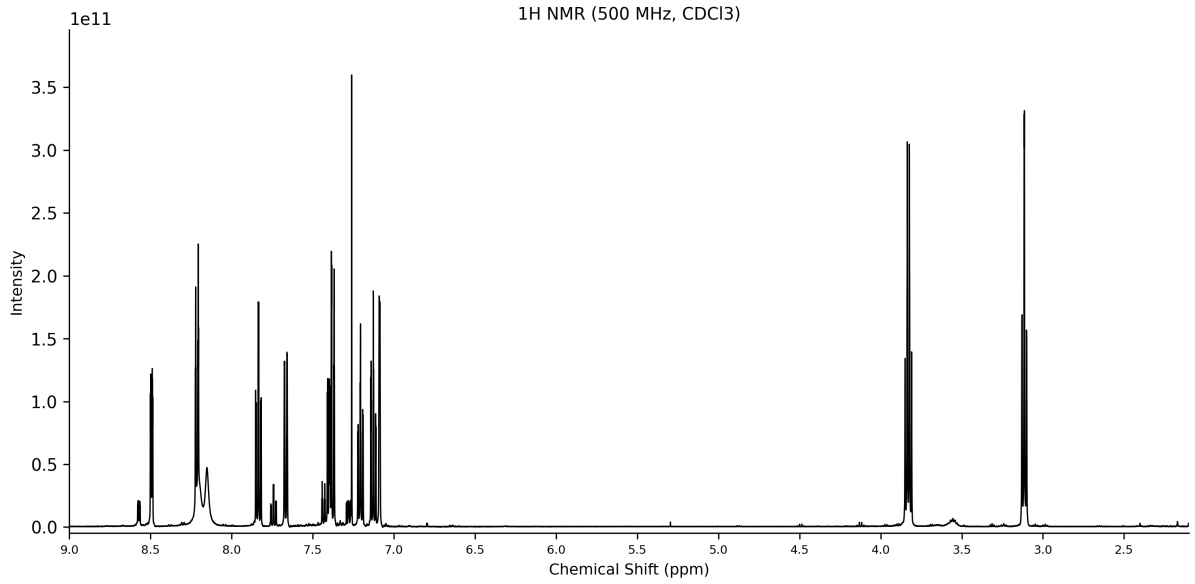


Figura 6.32: ^{1}H NMR del compuesto 40.

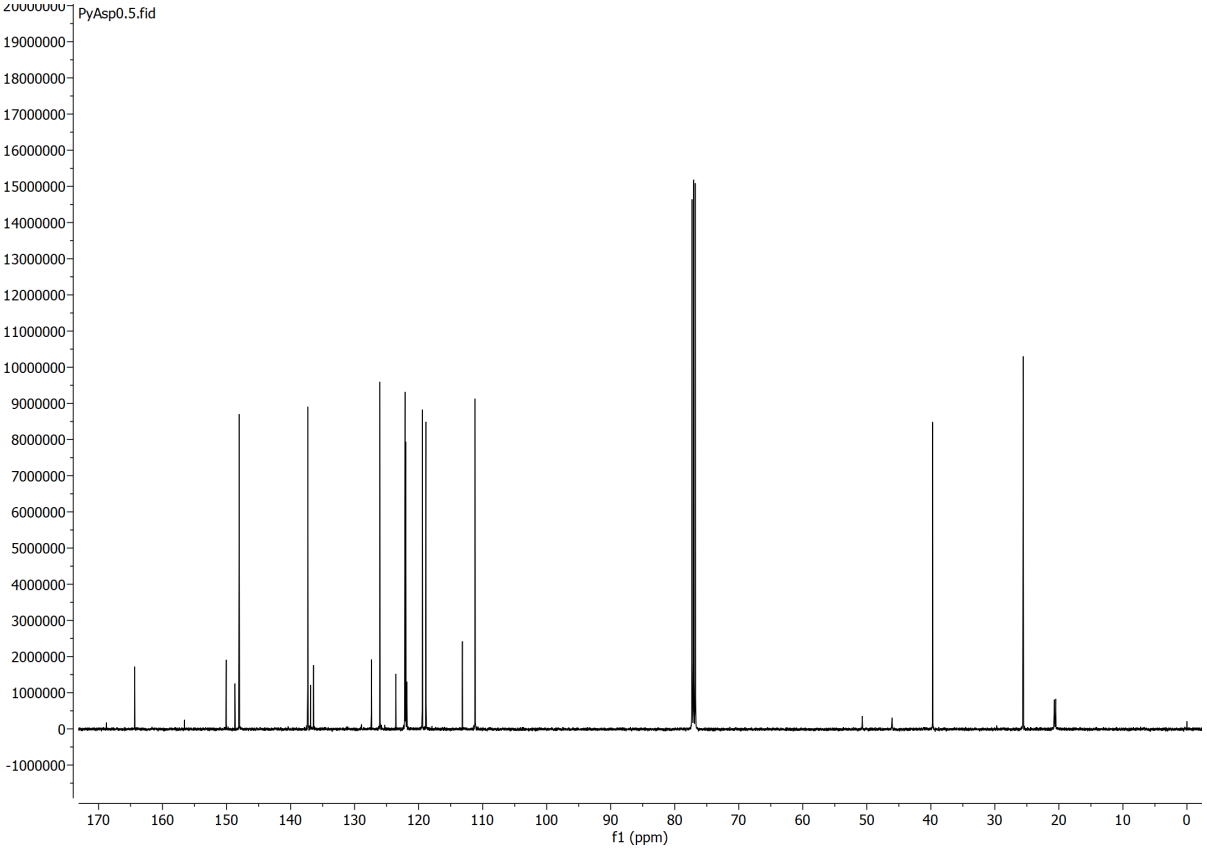


Figura 6.33: ^{13}C NMR del compuesto 40.

6.2.6.6. Compuesto 41

Aceite marrón obtenido mediante la reacción 6.2.5.6, rendimiento: 80 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (dd, J = 8,6, 1.4 Hz, 1H, H-6), 7.07 (d, J = 2,4 Hz, 1H, H-2), 7.76 (d, J = 1,8 Hz, 1H, H-4), 6.91 (td, J = 2,7, 1.3 Hz, 1H, H-5′), 7.25 (d, J = 8,6 Hz, 1H, H-7), 6.40 (ddd, J = 3,9, 2.5, 1.3 Hz, 1H, H-3′), 6.20 (dt, J = 3,8, 2.6 Hz, 1H, H-4′), 3.75 (q, J = 6,8 Hz, 2H, H-β), 3.01 (t, J = 6,8 Hz, 2H, H-α).^{13}C NMR (125 MHz, $CDCl_3$) δ 161.1 (Cγ), 135.1 (C7a), 129.3 (C3a), 126.2 (C2′), 113.1 (C5), 113.1 (C3), 121.6 (C4), 123.5 (C2), 125.3 (C6), 121.4 (C5′), 112.9 (C7), 110.0 (C3′), 108.7 (C4′), 39.7 (Cβ), 25.6 (Cα). HRESIMS calculado $[C_{15}H_{15}BrN_3O]^+$ 332,0393 obtenido 332,0384

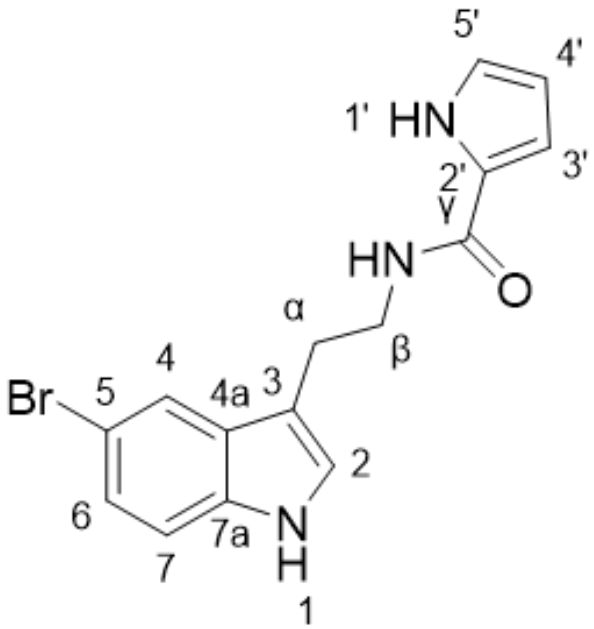


Figura 6.34: Estructura del compuesto 41.

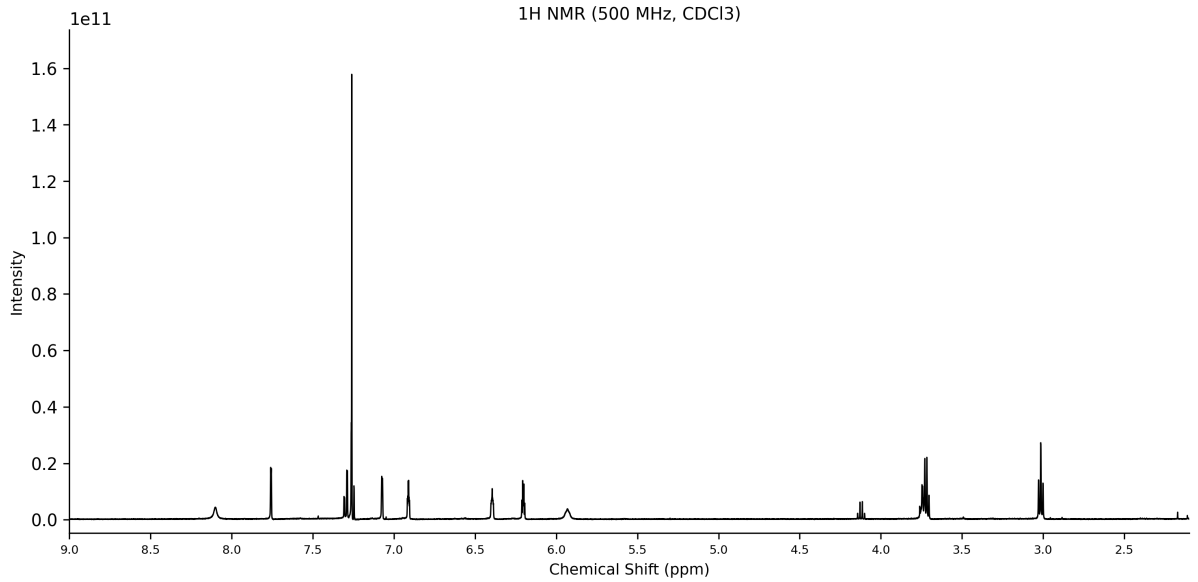


Figura 6.35: ^{1}H NMR del compuesto 41.

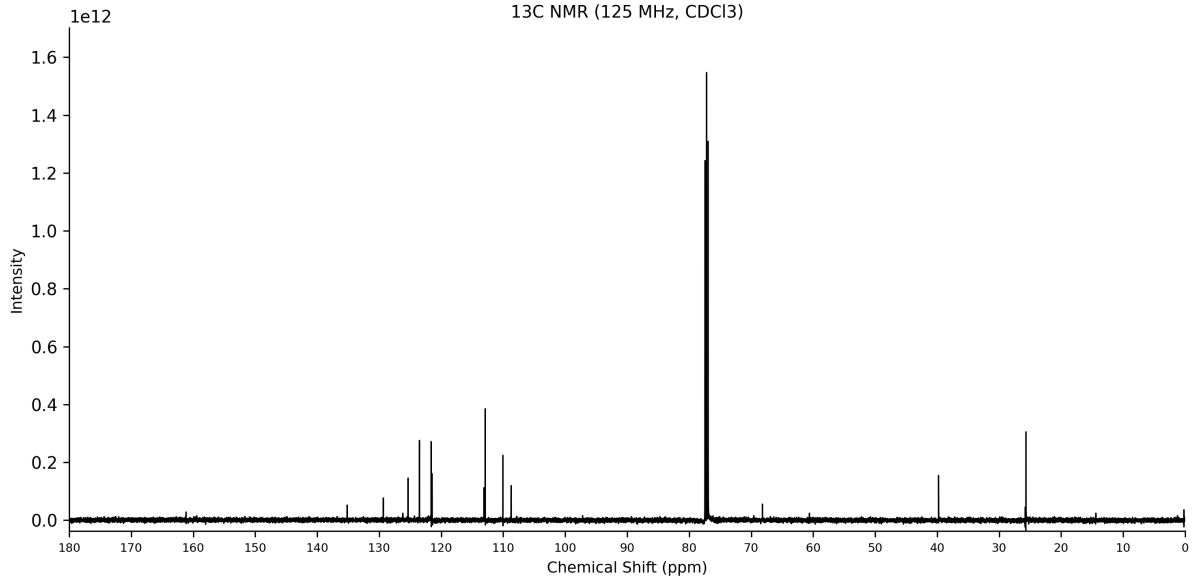


Figura 6.36: ^{13}C NMR del compuesto 41.

6.2.6.7. Compuesto 42

Aceite marrón obtenido mediante la reacción 6.2.5.6, rendimiento: 75 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.74 (d, H-4), 7.39 (dd, J = 1,7, 0.8 Hz, 1H, H-5′), 7.30 (dd, H-6), 7.25 (d, H-7), 7.10 (dd, J = 0,8 Hz, H-3′), 7.03 (dd, H-2), 6.48 (dd, J = 3,5, 1.8 Hz, 1H, H-4′), 3.74 (q, J = 6,8 Hz, 2H, H-β). ^{13}C NMR (125 MHz, $CDCl_3$) δ 158.6 (Cγ), 148.1 (C2′), 144.0 (C5′), 135.2 (C7a), 129.3 (C3a), 125.2 (C6), 123.5 (C2), 121.5 (C4), 114.3 (C3′), 112.9 (C5), 112.9 (C7), 112.8 (C3), 112.2 (C4′), 39.6 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{15}H_{14}BrN_2O_2]^+$ 333,0233 obtenido 333,0225

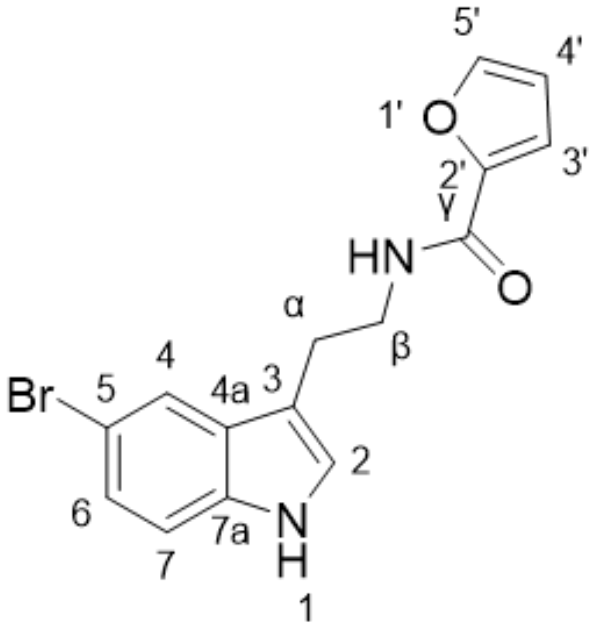


Figura 6.37: Estructura del compuesto 42.

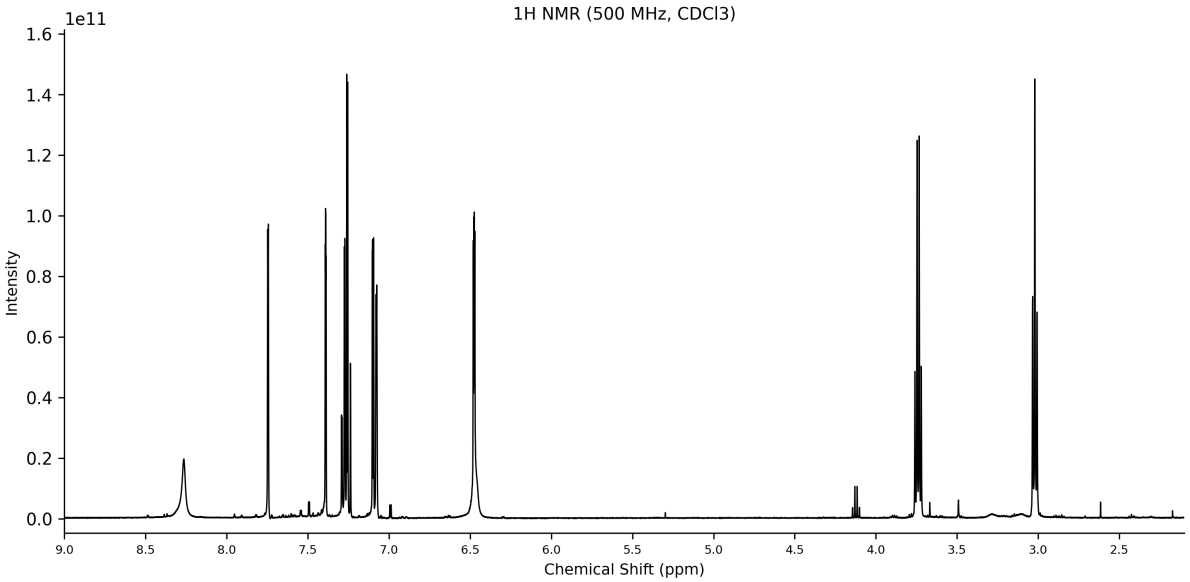


Figura 6.38: ^{1}H NMR del compuesto 42.

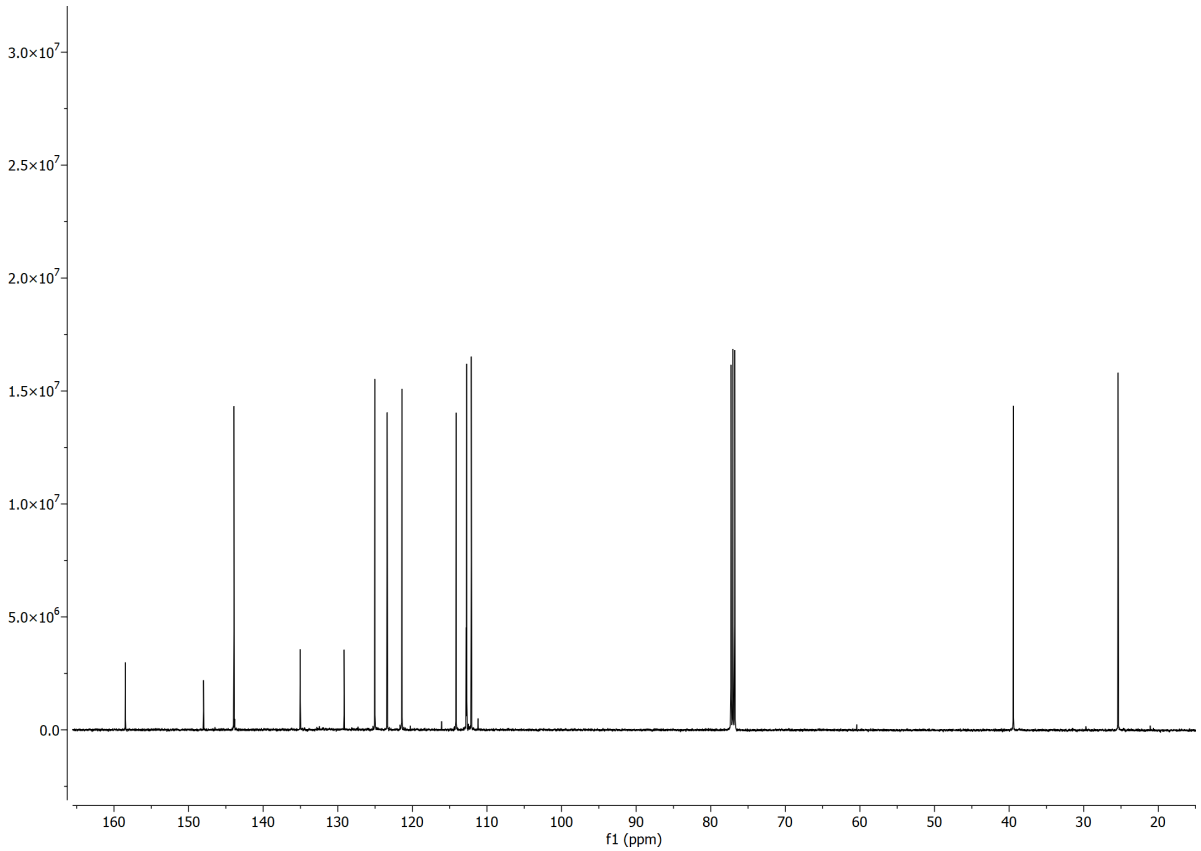


Figura 6.39: ^{13}C NMR del compuesto 42.

6.2.6.8. Compuesto 43

Aceite marrón obtenido mediante la reacción 6.2.5.6, rendimiento: 82 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.86 (dd, J = 1,6, 0.9 Hz, 1H, H-2′), 7.72 (d, J = 1,9 Hz, 1H, H-4), 7.39 (t, J = 1,7 Hz, 1H, H-5′), 7.27 (dd, J = 8,0, 6.1, 1.9 Hz, 1H, H-6), 7.23 (d, J = 8,6 Hz, 1H, H-7), 6.51 (dd, J = 1,9, 0.9 Hz, 1H, H-4′), 3.69 (q, J = 6,8 Hz, 2H, H-β), 2.99 (t, J = 6,8 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.9 (Cγ), 144.8 (C2′), 143.9 (C5′), 135.1 (C7a), 129.3 (C3a), 125.2 (C6), 123.5 (C2), 122.7 (C3′), 121.4 (C4), 113.0 (C7), 112.9 (C7), 112.8 (C3), 108.3 (C4′), 40.1 (Cβ), 25.3 (Cα). HRESIMS calculado $[C_{15}H_{14}BrN_2O_2]^+$ 333,0233 obtenido 333,0234

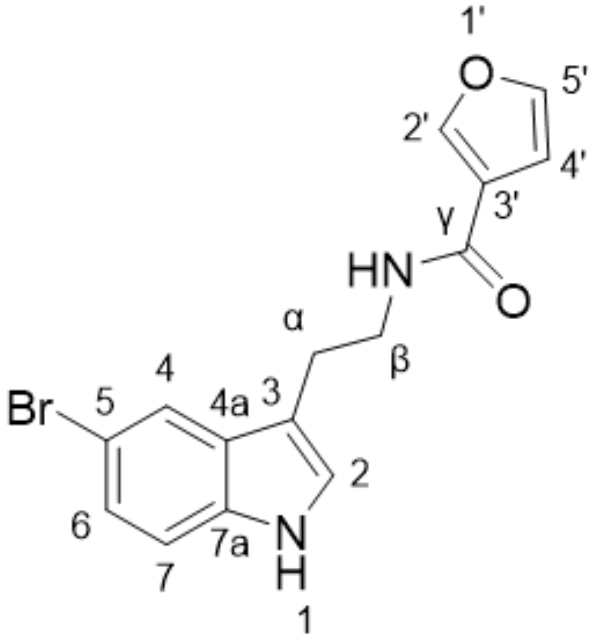


Figura 6.40: Estructura del compuesto 43.

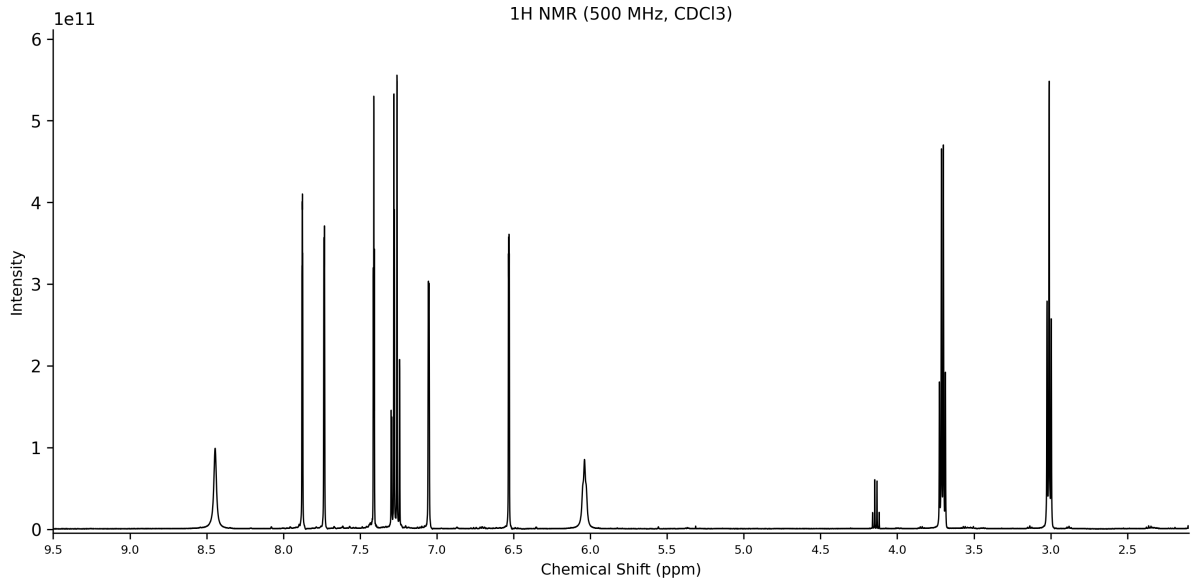


Figura 6.41: ^{1}H NMR del compuesto 43.

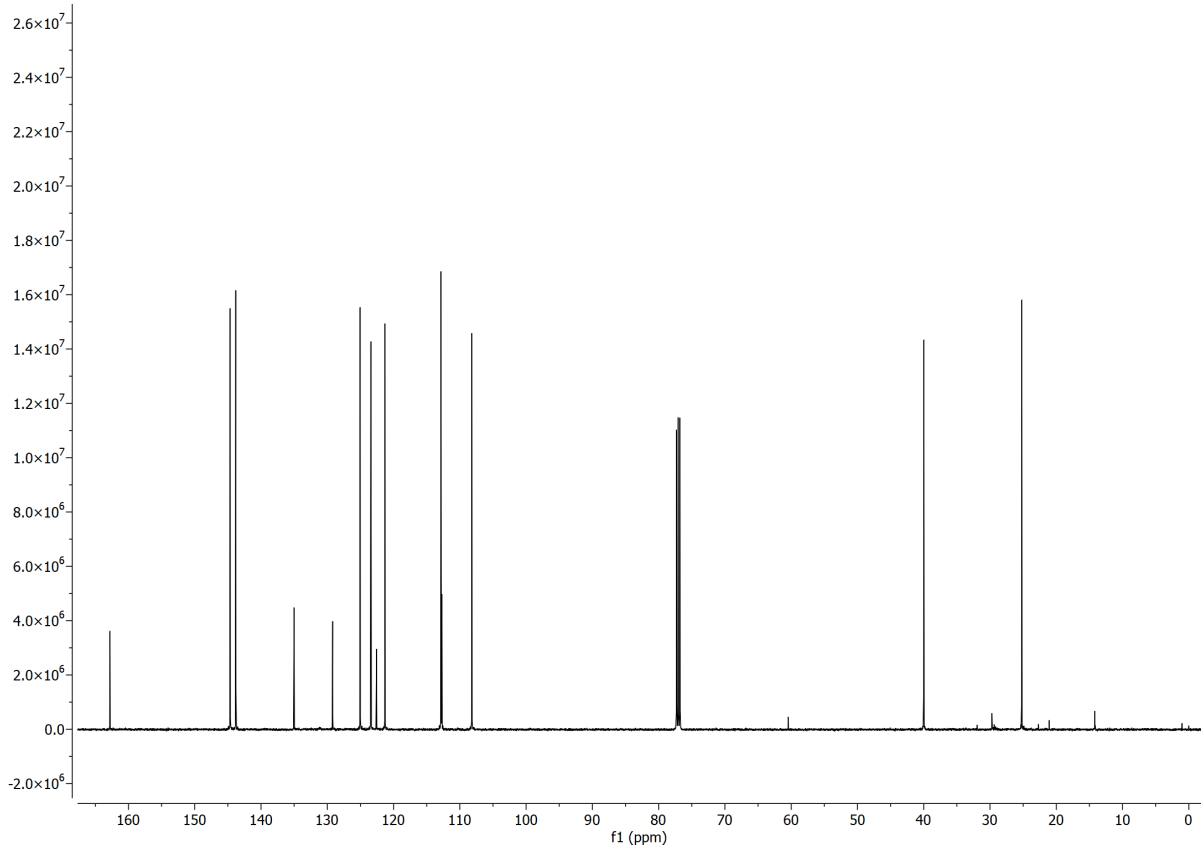


Figura 6.42: ^{13}C NMR del compuesto 43.

6.2.6.9. Compuesto 44

Aceite marrón obtenido mediante la reacción 6.2.5.6, rendimiento: 78 %.^{1}H NMR (500 MHz, $CDCl_3$) δ 7.76 (d, 1H, H-4), 7.45 (dd, J = 4,9, 1.1 Hz, 1H, H-5′), 7.38 (dd, J = 3,7, 1.1 Hz, 1H, H-3′), 7.26 (d, 1H, H-7), 3.75 (q, J = 7,2 Hz, 2H, H-β), 3.04 (t, J = 7,2 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.0 (Cγ), 139.1 (C2′), 135.1 (C7a), 129.9 (C5′), 129.3 (C3a), 128.1 (C3′), 127.8 (C4′), 125.3 (C6), 123.5 (C2), 121.6 (C4), 113.1 (C5), 113.0 (C3), 112.9 (C7), 40.5 (Cβ), 25.4 (Cα). HRESIMS calculado $[C_{15}H_{14}BrN_2OS]^+$ 349,0005 obtenido 349,0004

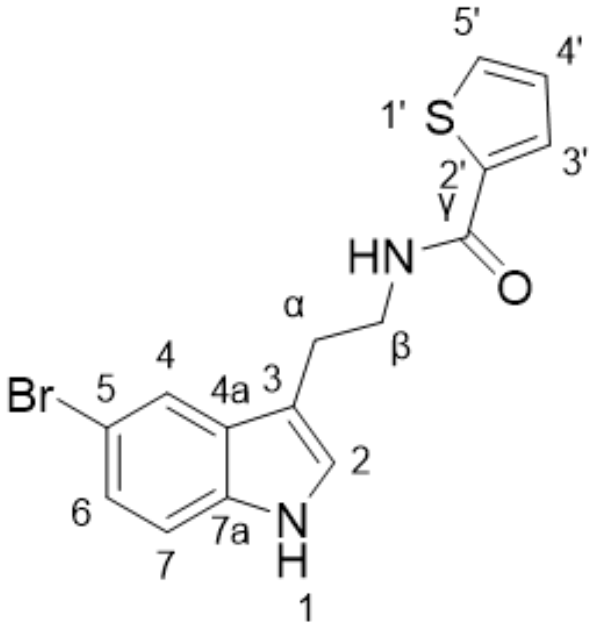


Figura 6.43: Estructura del compuesto 44.

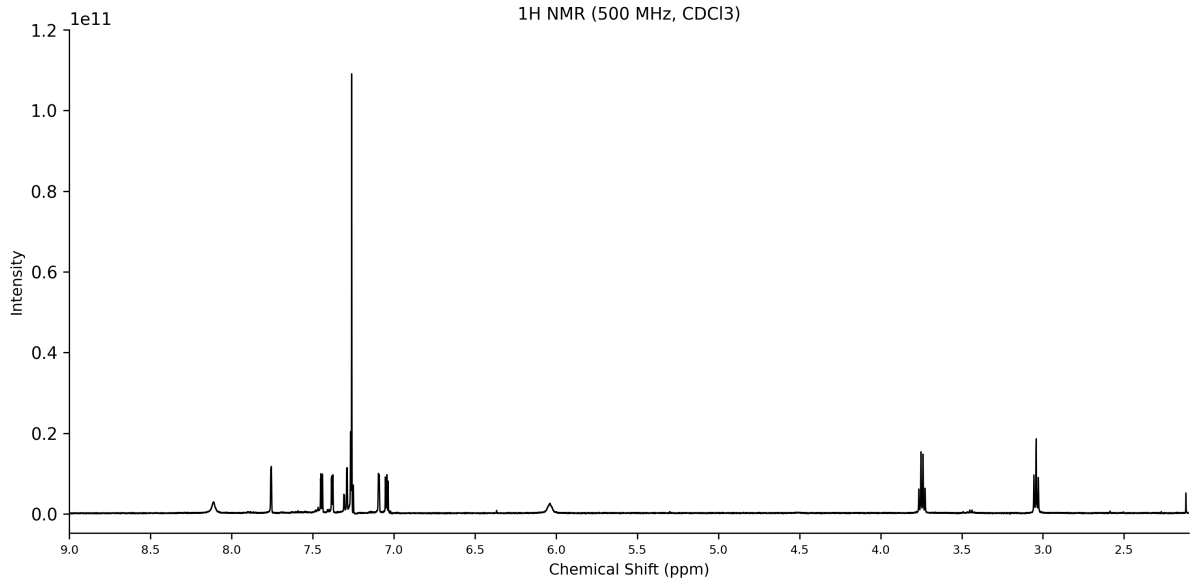


Figura 6.44: ^{1}H NMR del compuesto 44.

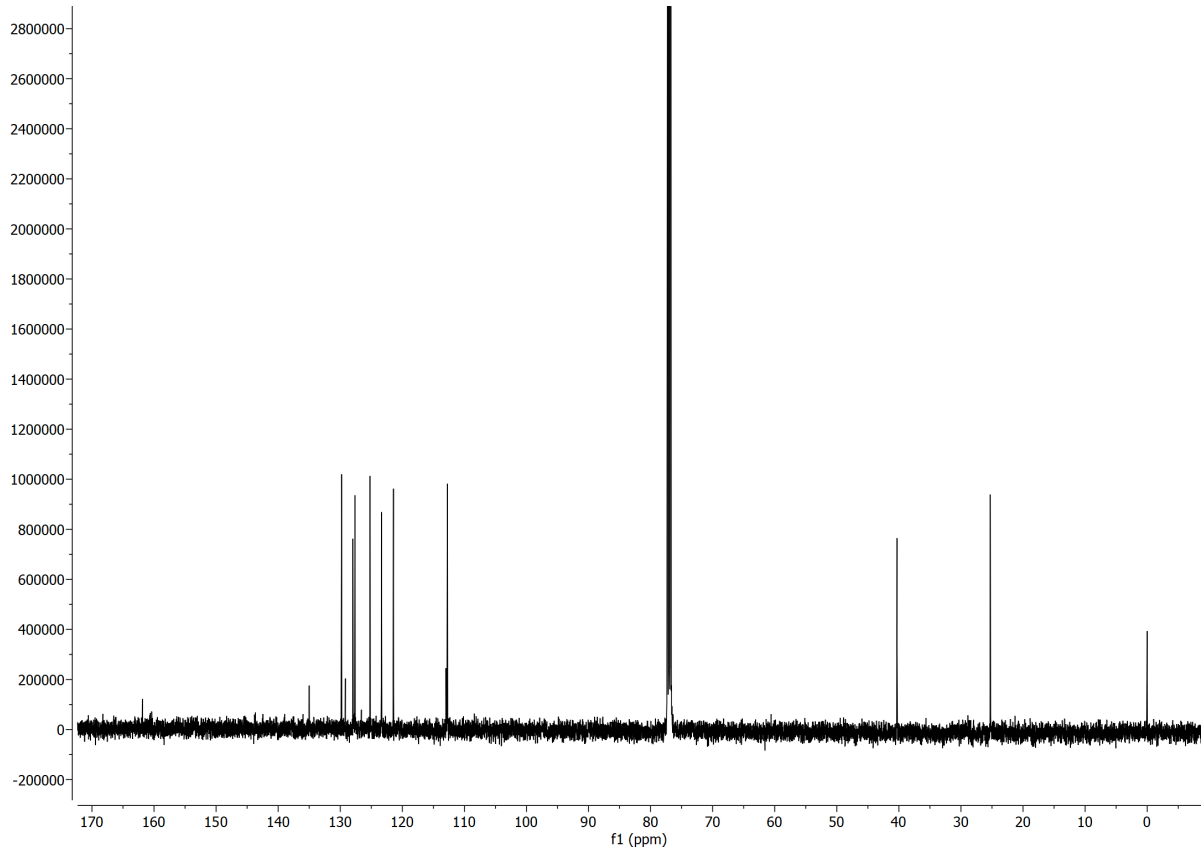


Figura 6.45: ^{13}C NMR del compuesto 44.

6.2.6.10. Compuesto 45

Aceite marrón obtenido mediante la reacción 6.2.5.6, rendimiento: 72 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.77 (dd, J = 2,8, 1.7 Hz, 1H, H-4), 7.73 (s, 1H, H-2′), 7.26 – 7.31 (m, 3H, H-2′, 6, 5′), 7.26 – 7.31 (m, 3H, H-2′, 6, 5′), 7.26 – 7.31 (m, 3H, H-2′, 6, 5′), 7.24 (d, 1H, H-7), 3.72 (q, 2H, H-β), 3.02 (t, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 163.2 (Cγ), 137.5 (C3′), 135.0 (C7a), 129.2 (C3a), 128.1 (C4), 126.6 (C2′, 6, 5′), 125.9 (C2′, 6, 5′), 125.1 (C2′, 6, 5′), 123.4 (C2), 121.4 (C2′), 112.9 (C5), 112.9 (C3), 112.8 (C7), 40.2 (Cβ), 25.2 (Cα).

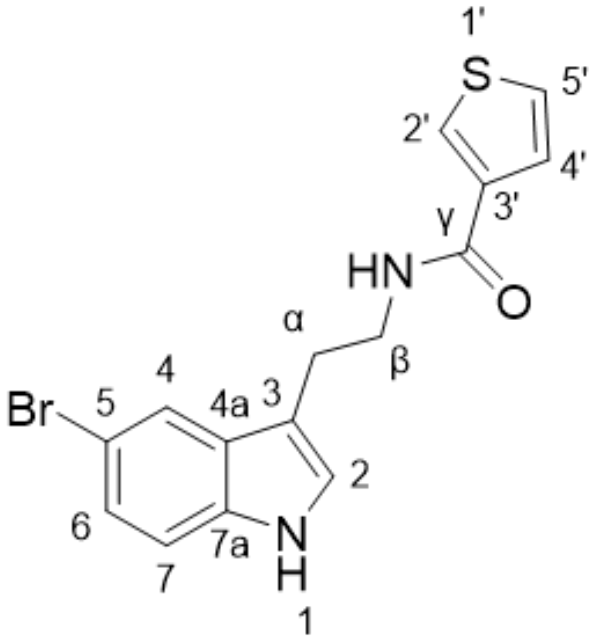


Figura 6.46: Estructura del compuesto 45.

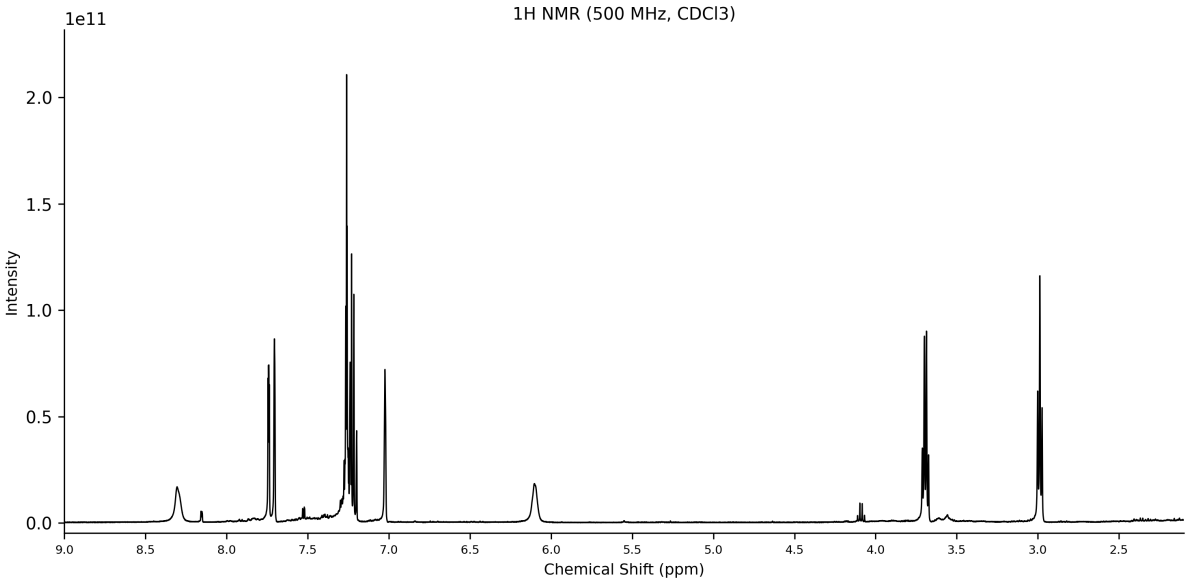


Figura 6.47: ^{1}H NMR del compuesto 45.

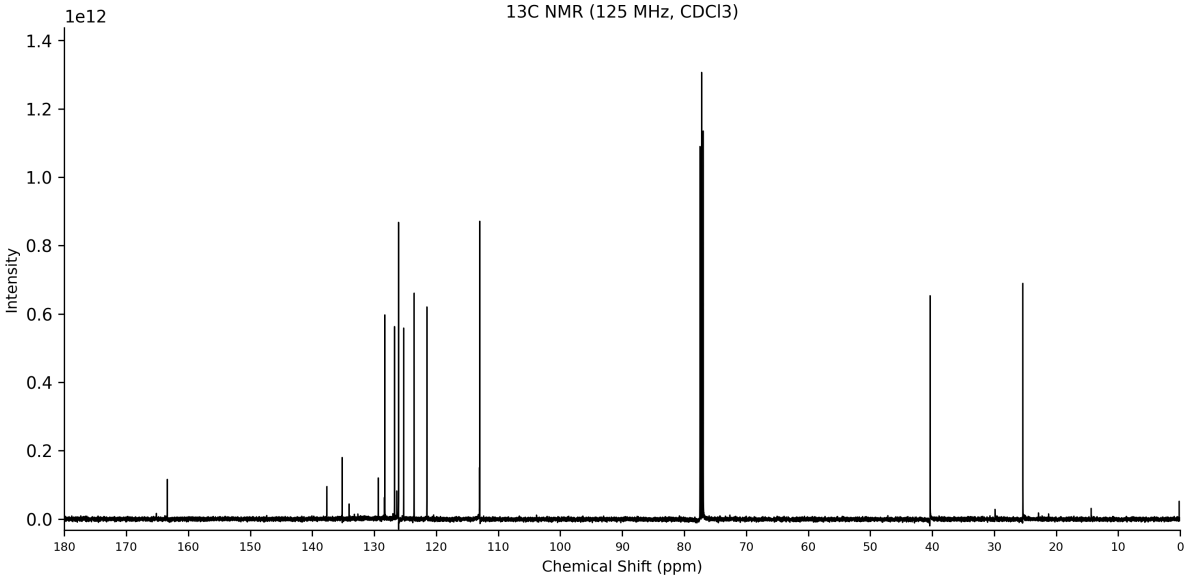


Figura 6.48: ^{13}C NMR del compuesto 45.

6.2.6.11. Compuesto 46

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 40 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 (d, J = 7,1 Hz, 2H, H-2″), 7.37 (t, J = 7,76 Hz, 1H, H-3″), 7.31 (t, J = 7,14 Hz, 1H, H-4″), 7.27 (d, 1H, H-7), 7.14 (d, J = 2,4 Hz, 1H, H-4), 7.02 (d, J = 2,4 Hz, 1H, H-2), 6.95 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 6.89 (q, J = 2,3 Hz, 1H, H-5′), 6.37 (s, 1H, H-3′), 6.18 (q, J = 2,9 Hz, 1H, H-4′), 5.01 (s, 2H, H-ε), 3.72 (q, J = 6,4 Hz, 2H, H-β), 3.01 (t, J = 6,6 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 161.3 (Cγ), 153.5 (C5), 137.7 (C1″), 131.8 (C7a), 128.6 (C3″), 127.9 (C4″), 127.9 (C3a), 127.8 (C2″), 126.2 (C2′), 123.1 (C2), 121.6 (C5′), 113.4 (C6), 113.1 (C3), 112.2 (C7), 109.9 (C4′), 108.9 (C3′), 102.2 (C4), 71.0 (Cε), 40.0 (Cβ), 25.7 (Cα). HRESIMS calculado $[C_{22}H_{22}N_3O_2]^+$ 360,1707 obtenido 360,1726

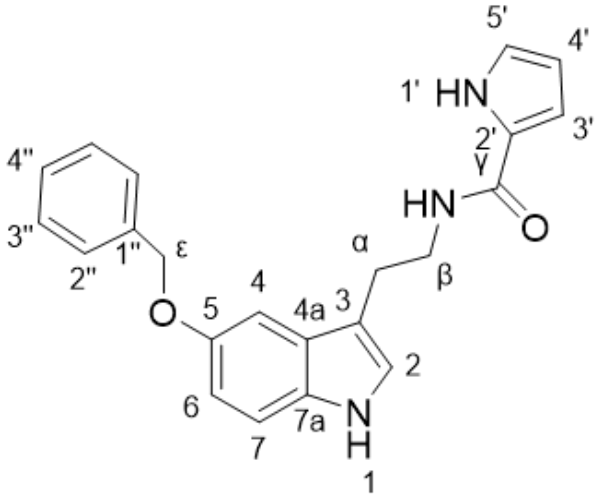


Figura 6.49: Estructura del compuesto 46.

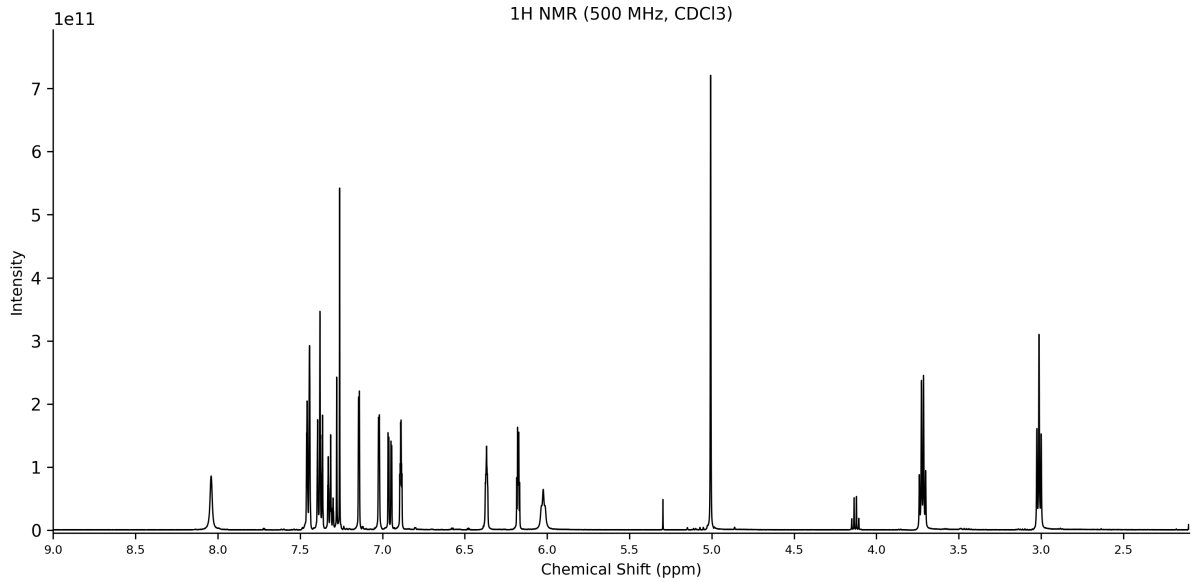


Figura 6.50: ^{1}H NMR del compuesto 46.

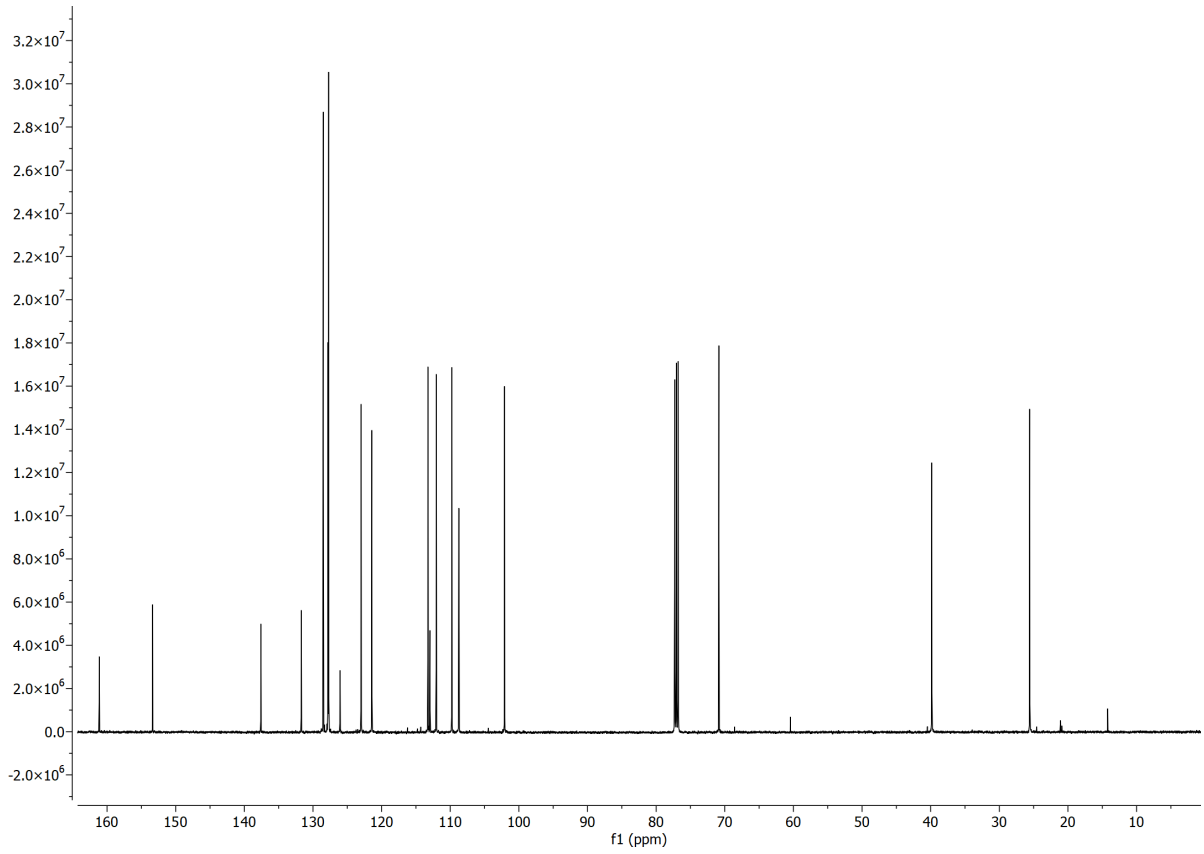


Figura 6.51: ^{13}C NMR del compuesto 46.

6.2.6.12. Compuesto 47

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 80 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.46 (d, J = 7,5 Hz, 2H, H-4″), 7.38 (t, J = 7,5 Hz, 1H, H-2″), 7.36 (dd, J = 1,8, 0.8 Hz, 1H, H-5′), 7.31 (t, J = 7,5 Hz, 1H, H-3″), 7.28 (d, J = 8,8 Hz, 1H, H-7), 7.15 (d, J = 2,4 Hz, 1H, H-4), 7.10 (dd, J = 3,45, 0.73 Hz, 1H, H-3′), 7.06 (d, J = 2,4 Hz, 1H, H-2), 6.96 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 6.46 (dd, J = 3,5, 1.8 Hz, 1H, H-4′), 5.05 (s, 2H, H-ε), 3.75 (q, J = 6,7 Hz, 2H, H-β), 3.03 (t, J = 6,7 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 158.5 (Cγ), 153.5 (C5), 148.3 (C2′), 143.9 (C5′), 137.7 (C1″), 131.9 (C7a), 128.7 (C2″), 128.0 (C3″), 127.9 (C3a), 127.8 (C4″), 123.0 (C2), 114.2 (C3′), 113.4 (C6), 113.0 (C3), 112.2 (C4′), 112.1 (C7), 102.3 (C4), 71.0 (Cε), 39.6 (Cβ), 25.7 (Cα). HRESIMS calculado $[C_{22}H_{21}N_2O_3]^+$ 361,1547 obtenido 361,1546

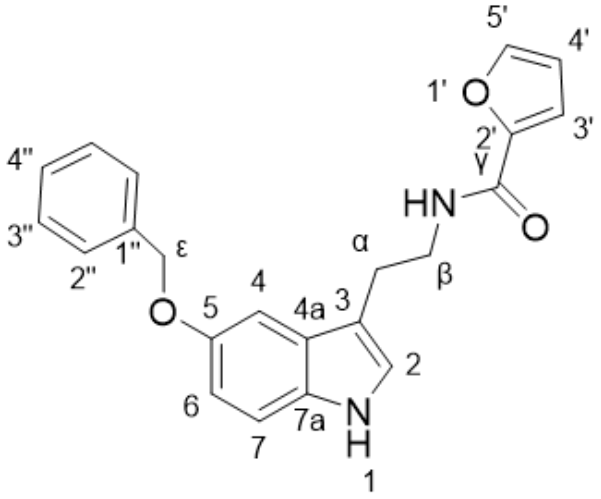


Figura 6.52: Estructura del compuesto 47.

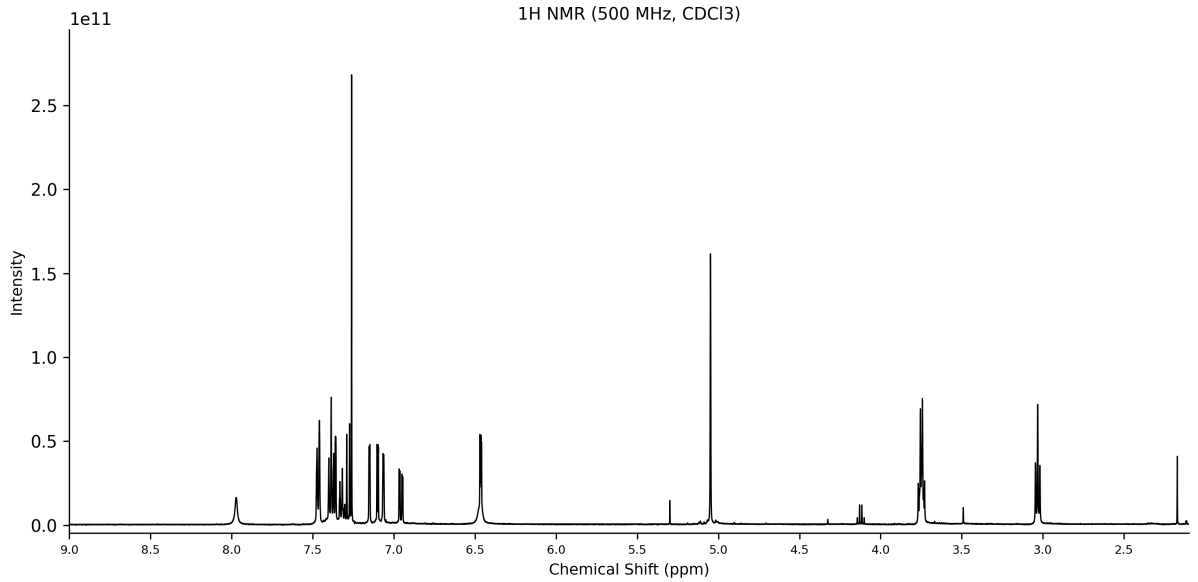


Figura 6.53: ^{1}H NMR del compuesto 47.

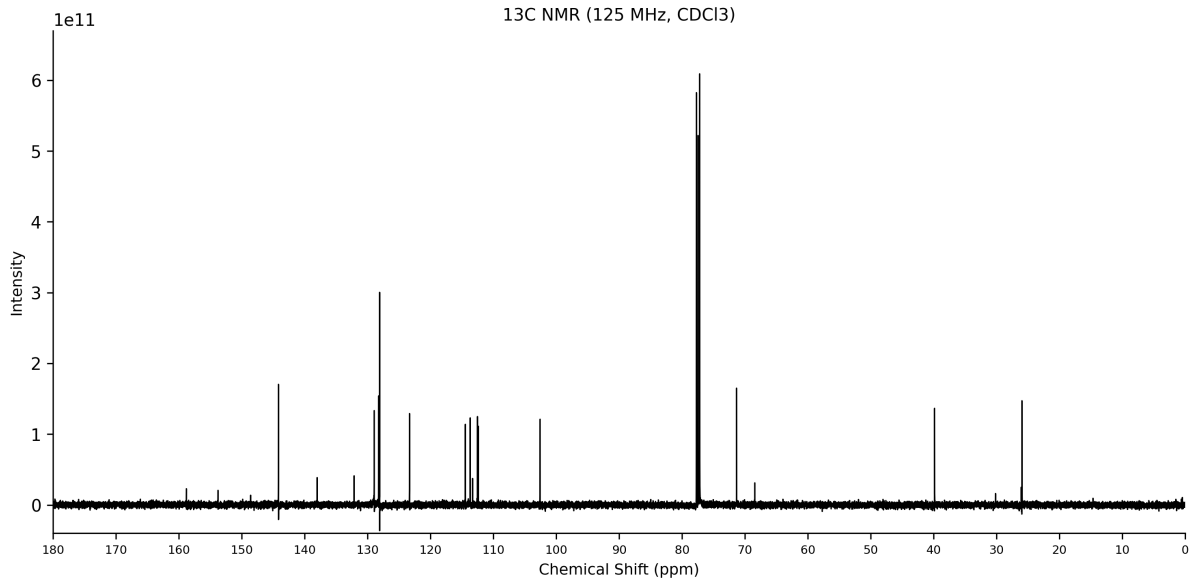


Figura 6.54: ^{13}C NMR del compuesto 47.

6.2.6.13. Compuesto 48

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 81 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.84 (dd, J = 1,5, 0.9 Hz, 1H, H-2′), 7.45 (d, J = 7,5 Hz, 2H, H-4″), 7.38 (t, J = 7,5 Hz, 1H, H-2″), 7.36 (m, 1H, H-5′), 7.32 (t, J = 7,5 Hz, 1H, H-3″), 7.27 (d, J = 8,7 Hz, 1H, H-7), 7.14 (d, J = 2,4 Hz, 1H, H-4), 7.02 (d, J = 2,4 Hz, 1H, H-2), 6.95 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 6.48 (dd, J = 1,9, 0.9 Hz, 1H, H-4′), 5.03 (s, 2H, H-ε), 3.71 (q, J = 6,5 Hz, 2H, H-β), 3.01 (t, J = 6,5 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.7 (Cγ), 153.5 (C5), 144.7 (C2′), 143.8 (C5′), 137.6 (C1″), 131.9 (C7a), 128.7 (C2″), 128.0 (C3″), 127.9 (C3a), 127.8 (C4″), 123.1 (C2), 122.8 (C3′), 113.3 (C6), 112.9 (C3), 112.2 (C7), 108.4 (C4′), 102.2 (C4), 71.0 (Cε), 40.1 (Cβ), 25.4 (Cα). HRESIMS calculado $[C_{22}H_{21}N_2O_3]^+$ 361,1547 obtenido 361,1544

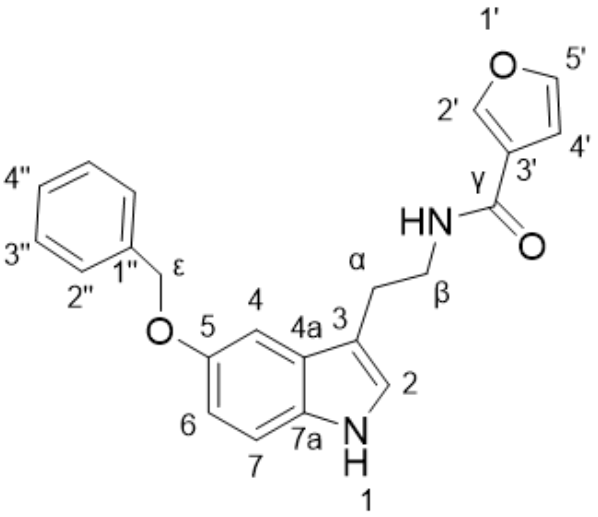


Figura 6.55: Estructura del compuesto 48.

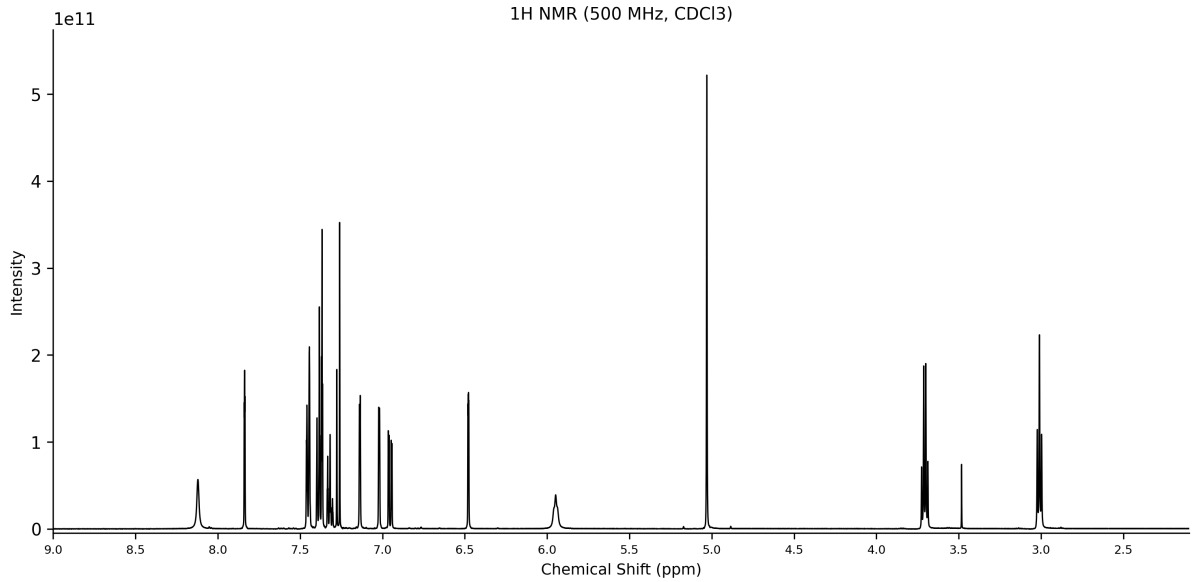


Figura 6.56: ^{1}H NMR del compuesto 48.

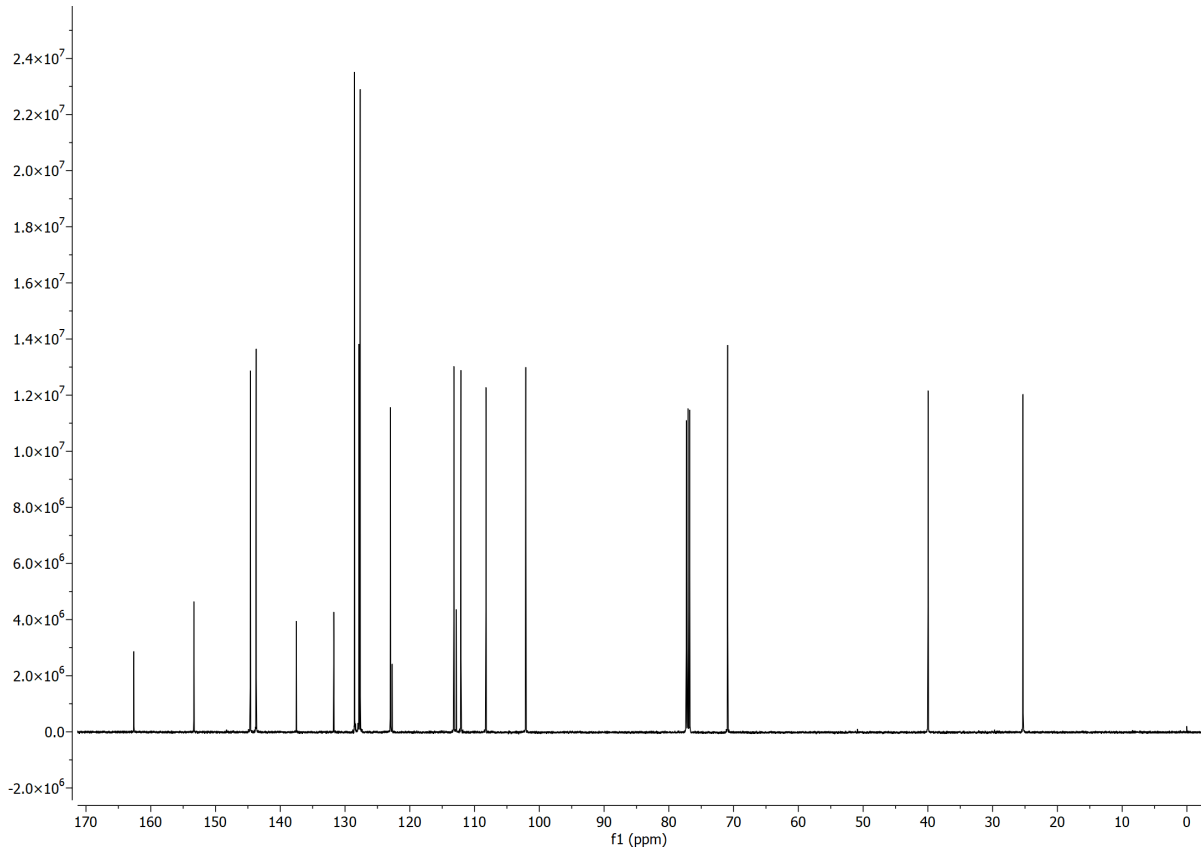


Figura 6.57: ^{13}C NMR del compuesto 48.

6.2.6.14. Compuesto 49

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 80 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 (d, J = 7,6 Hz, H-1″), 7.42 (dd, H-5′), 7.38 (t, J = 7,6 Hz, H-2″), 7.34 (dd, H-3′), 7.31 (t, J = 7,3 Hz, H-3″), 7.28 (d, J = 8,8 Hz, 1H, H-7), 7.14 (d, J = 2,4 Hz, 1H, H-4), 7.05 (d, J = 2,4 Hz, 1H, H-2), 7.01 (dd, J = 5,0, 3.7 Hz, 1H, H-4′), 6.95 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 5.00 (s, 2H, H-ε), 3.74 (q, J = 6,4 Hz, 2H, H-β), 3.04 (t, J = 6,4 Hz, 2H, H-α).^{13}C NMR (125 MHz, $CDCl_3$) δ 161.9 (Cγ), 153.4 (C5), 139.1 (C2′), 137.5 (Caromatic), 131.7 (C7a), 129.8 (C5′), 128.5 (C2″), 127.9 (C3′), 127.8 (C3″), 127.8 (C3a), 127.7 (C1″), 127.6 (C4′), 123.0 (C2), 113.3 (C6), 112.8 (C3), 112.1 (C7), 102.0 (C4), 70.9 (Cε), 40.4 (Cβ), 25.4 (Cα). HRESIMS calculado $[C_{22}H_{21}N_2O_2S]^+$ 377,1318 obtenido 377,1318

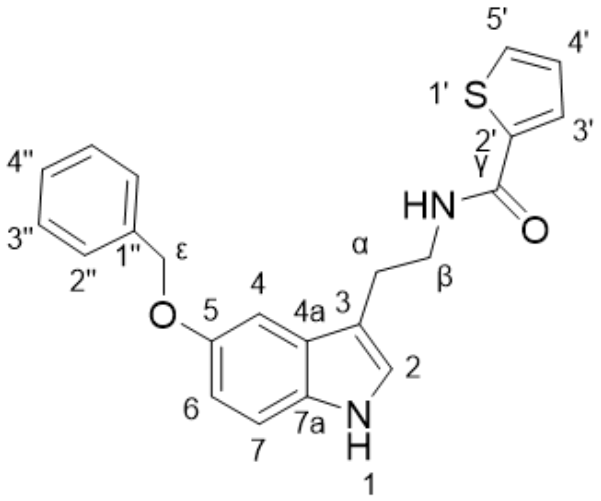


Figura 6.58: Estructura del compuesto 49.

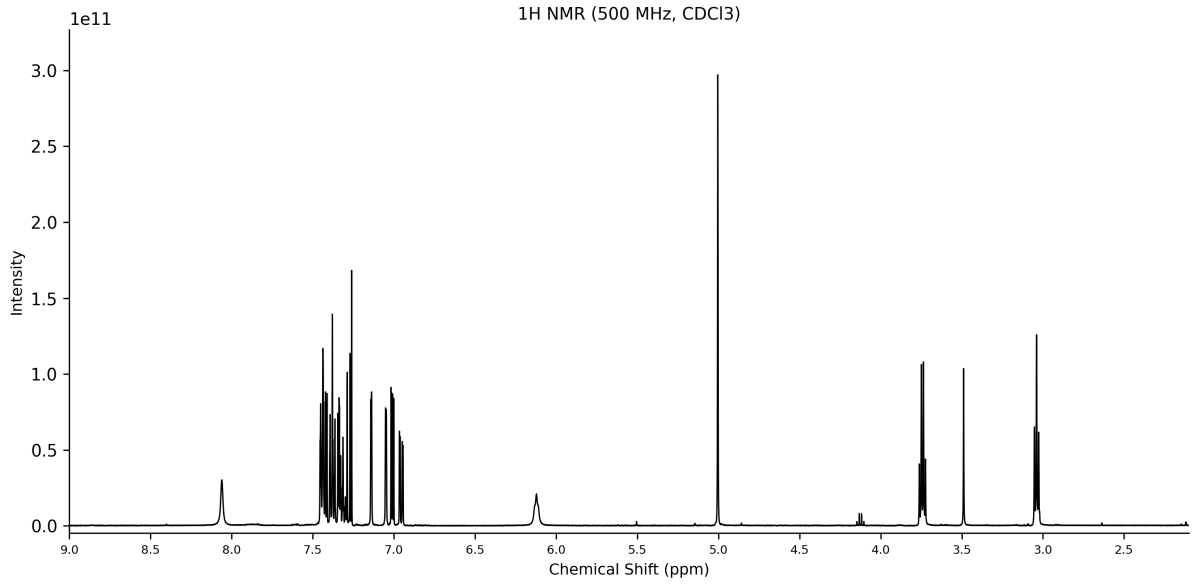


Figura 6.59: ^{1}H NMR del compuesto 49.

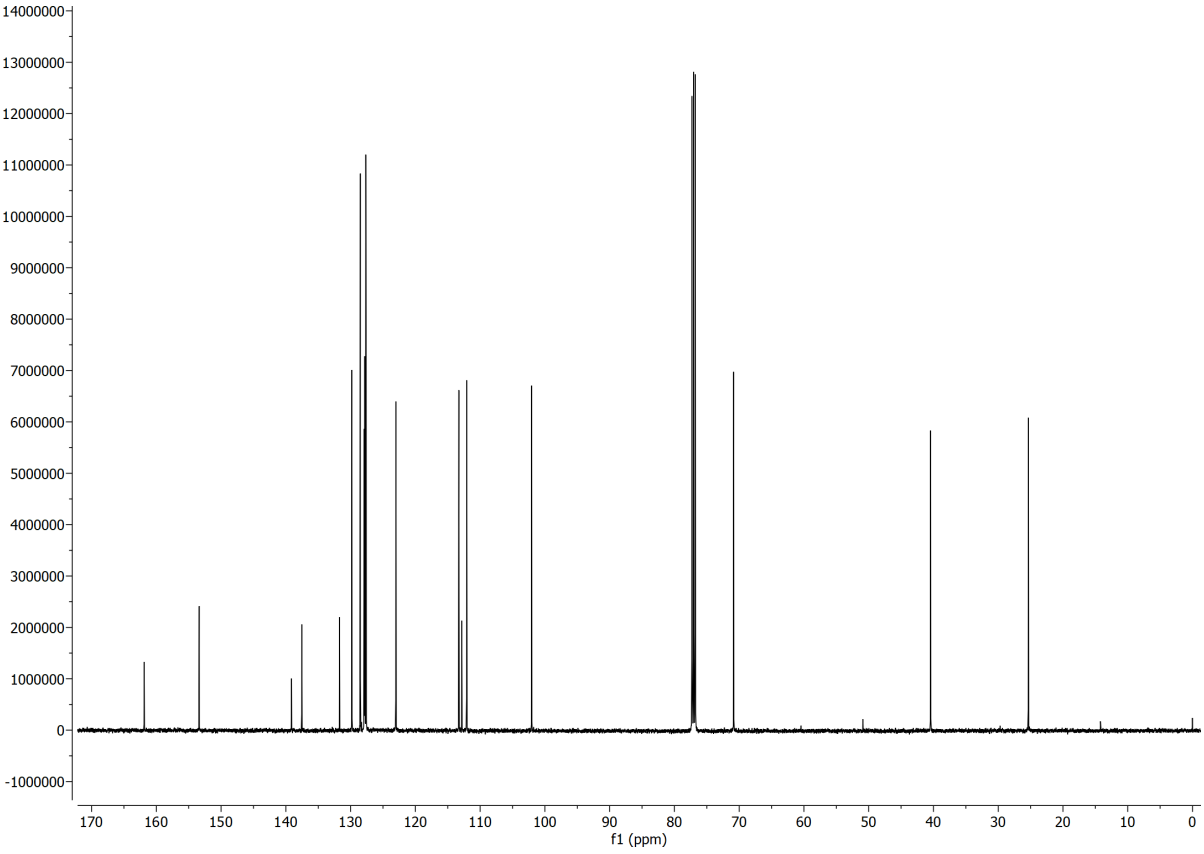


Figura 6.60: ^{13}C NMR del compuesto 49.

6.2.6.15. Compuesto 50

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 65 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.74 (dd, J = 1,5, 2.9 Hz, H-2′), 7.45 (d, J = 7,5 Hz, 2H, H-4″), 7.38 (t, J = 7,5 Hz, 1H, H-2″), 7.32 (t, J = 7,5 Hz, 1H, H-3″), 7.28 (m, H-4′, 5′), 7.28 (d, J = 8,8 Hz, 1H, H-7), 7.26 (m, H-4′, 5′), 7.14 (d, J = 2,4 Hz, 1H, H-4), 7.05 (d, J = 2,4 Hz, 1H, H-2), 6.96 (dd, J = 8,8, 2.4 Hz, 1H, H-6), 5.00 (s, 2H, H-ε), 3.75 (q, J = 6,3 Hz, 2H, H-β), 3.04 (t, J = 6,5 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 163.1 (Cγ), 153.5 (C5), 137.8 (C3′), 137.7 (C1″), 131.8 (C7a), 128.7 (C2″), 128.2 (C3″), 128.0 (C3a), 128.0 (C2′), 127.8 (C4″), 126.6 (C4′, 5′), 126.1 (C4′, 5′), 123.1 (C2), 113.4 (C6), 113.1 (C3), 112.2 (C7), 102.2 (C4), 71.0 (Cε), 40.3 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{22}H_{21}N_2O_2S]^+$ 377,1318 obtenido 377,1316

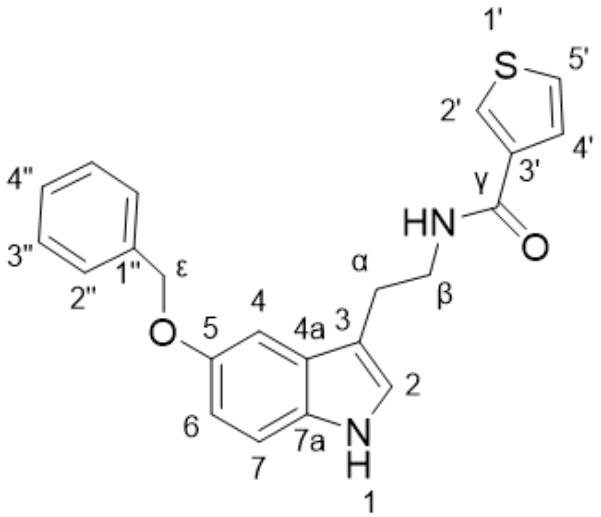


Figura 6.61: Estructura del compuesto 49.

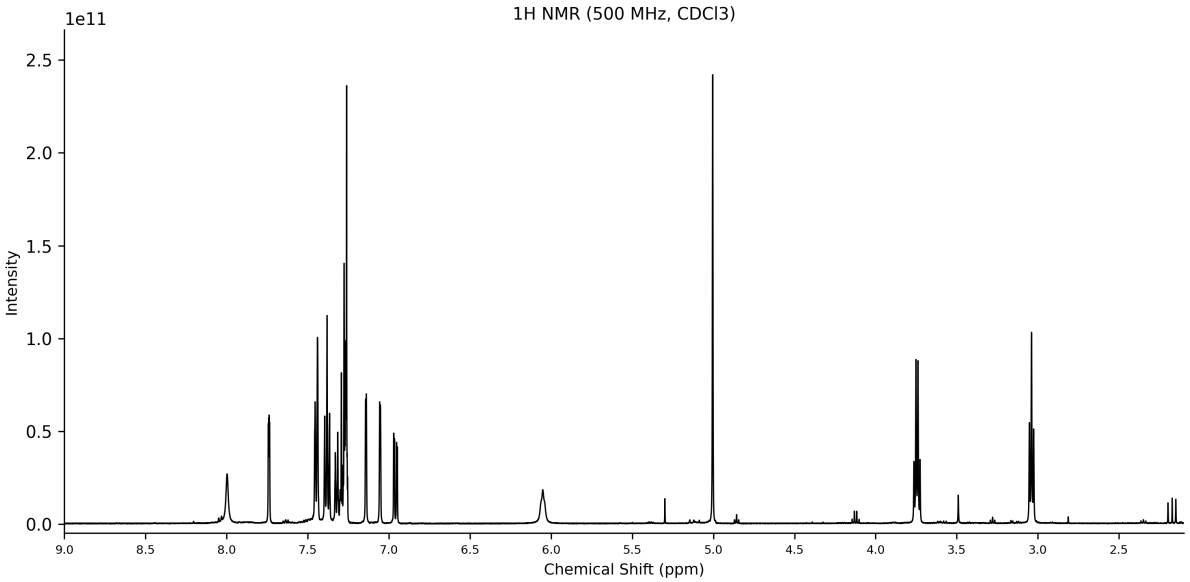


Figura 6.62: ^{1}H NMR del compuesto 49.

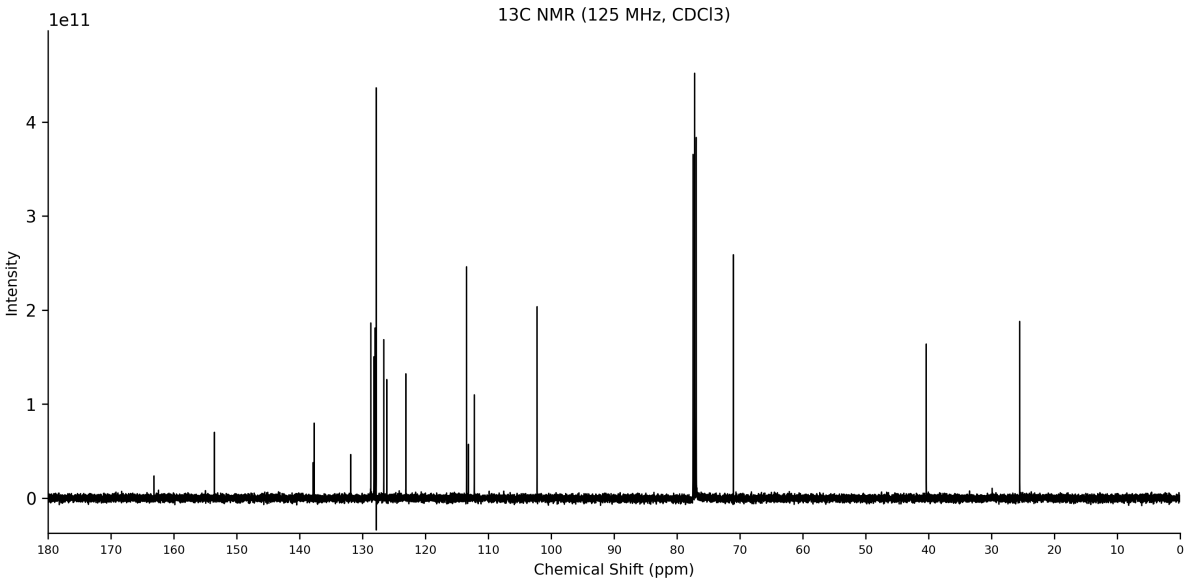


Figura 6.63: ^{13}C NMR del compuesto 49.

6.2.6.16. Compuesto 52

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.5, 6.2.5.4, rendimiento: 38 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.36 (dd, J = 1,8, 0.8 Hz, 1H, H-5′), 7.27 (d, J = 8,8 Hz, 1H, H-7), 7.09 (dd, J = 3,5, 0.8 Hz, 1H, H-3′), 7.07 – 7.03 (m, 2H, H-2), 7.07 – 7.03 (m, 2H, H-4), 6.87 (dd, J = 8,8, 2.5 Hz, 1H, H-6), 6.47 (dd, J = 3,5, 1.8 Hz, 1H, H-4′), 3.82 (s, 3H, H-ε), 3.76 (q, J = 6,7 Hz, 2H, H-β), 3.04 (t, J = 6,7 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 158.4 (Cγ), 154.1 (C5), 148.1 (C2′), 143.7 (C5′), 131.5 (C7a), 127.7 (C3a), 122.8 (C2), 114.0 (C3′), 112.8 (C3), 112.6 (C6), 112.1 (C4′), 112.0 (C7), 100.4 (C4), 55.8 (Cε), 39.5 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{16}H_{17}N_2O_3]^+$ 285,1234 obtenido 285,1217

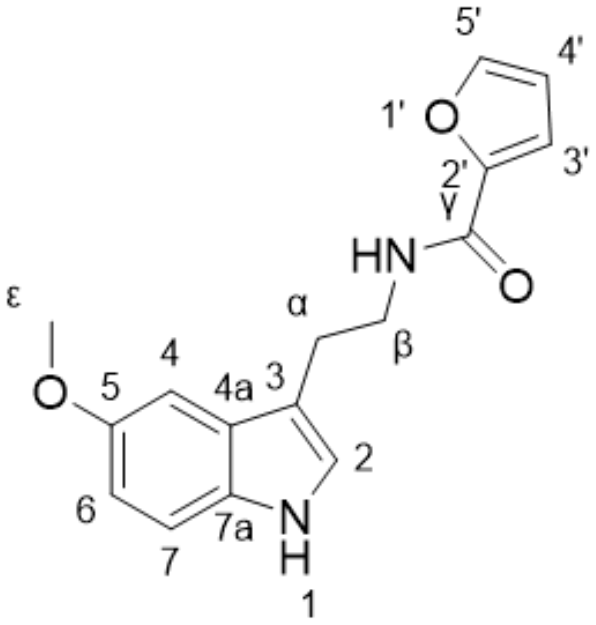


Figura 6.64: Estructura del compuesto 52.

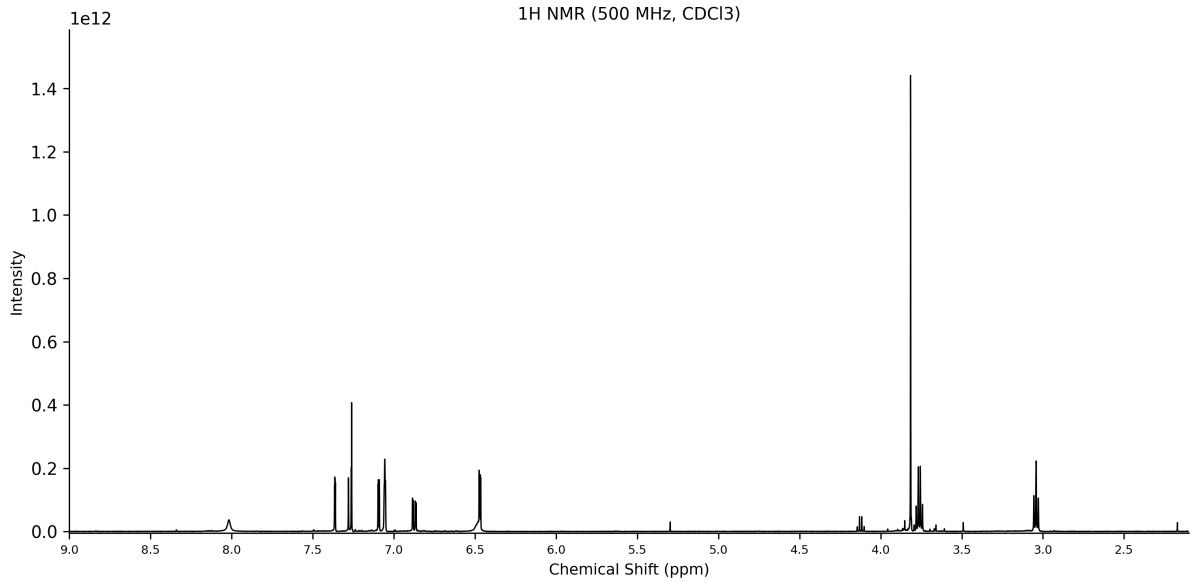


Figura 6.65: ^{1}H NMR del compuesto 52.

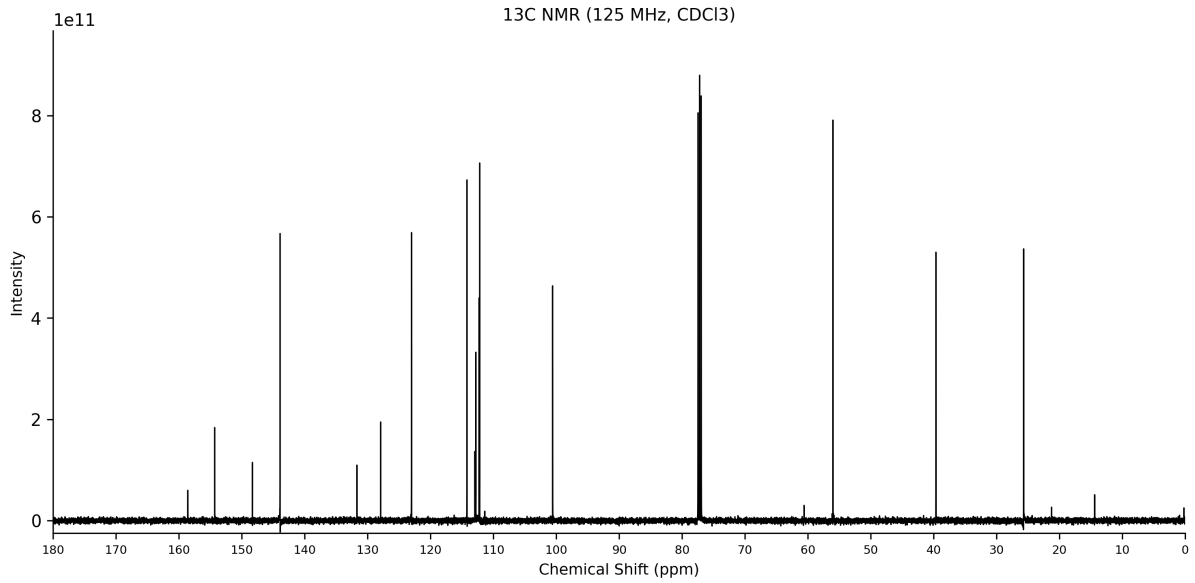


Figura 6.66: ^{13}C NMR del compuesto 52.

6.2.6.17. Compuesto 53

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.5, 6.2.5.4, rendimiento: 39 %.^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 (dd, J = 1 Hz, 5.0, 1.2 Hz, H-5′), 7.36 (dd, J = 1 Hz, 3.7, 1.2 Hz, H-3′), 7.30 (d, J = 1 Hz, 8.8 Hz, H-7), 7.06 (d, J = 1 Hz, 2.7 Hz, H-2), 7.06 (d, J = 1 Hz, 2.7 Hz, H-4), 7.05 (dd, J = 1 Hz, 4.8, 3.8 Hz, H-4′), 6.90 (dd, J = 1 Hz, 8.8, 2.4 Hz, H-6), 3.81 (s, 3H, H-ε), 3.76 (q, J = 6,7 Hz, 2H, H-β), 3.08 (t, J = 6,7 Hz, 2H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.0 (Cγ), 154.3 (C5), 139.3 (C2′), 131.6 (C7a), 129.9 (C5′), 128.0 (C3′), 127.9 (C3a), 127.7 (C4′), 123.0 (C2), 112.9 (C3), 112.9 (C6), 112.2 (C7), 100.5 (C4), 55.9 (Cε), 40.5 (Cβ), 25.5 (Cα). HRESIMS calculado $[C_{16}H_{17}N_2O_2S]^+$ 301,1005 obtenido 301,0995

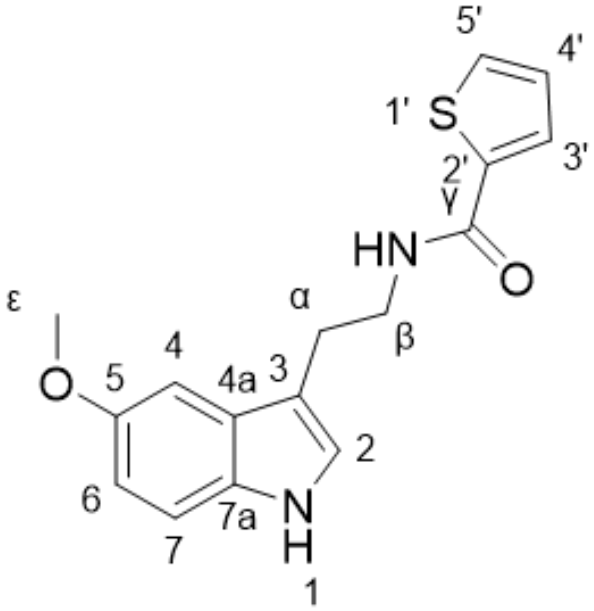


Figura 6.67: Estructura del compuesto 53.

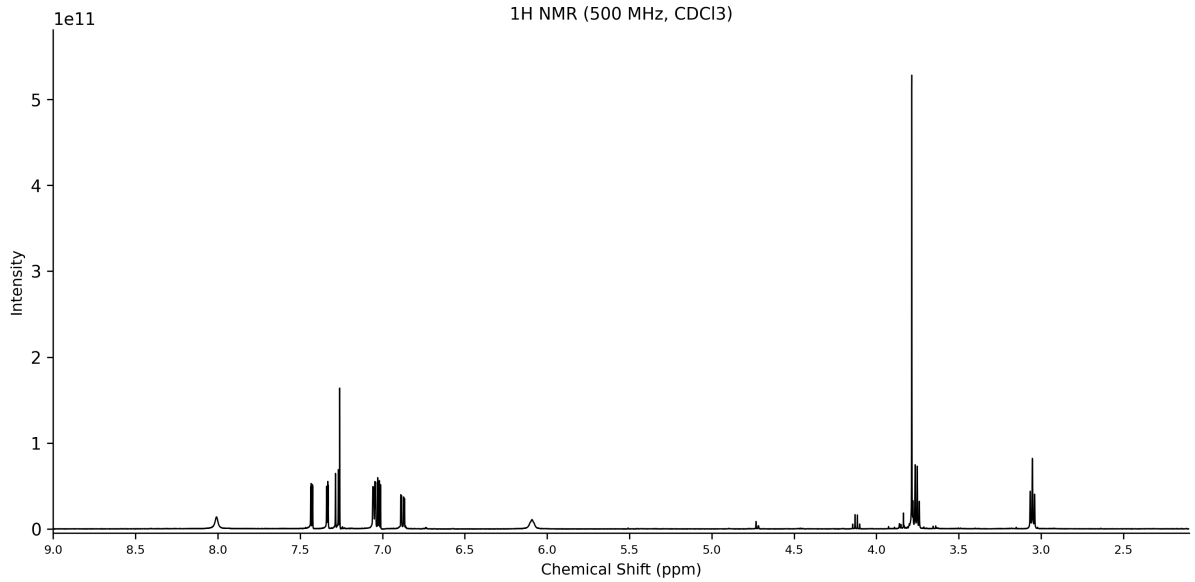


Figura 6.68: ^{1}H NMR del compuesto 53.

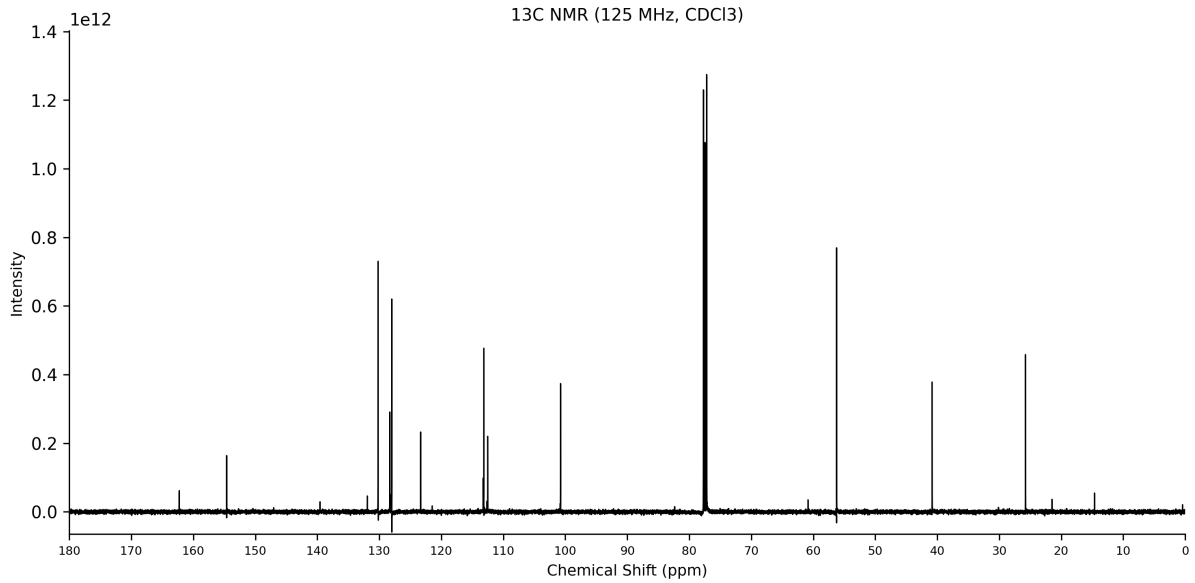


Figura 6.69: ^{13}C NMR del compuesto 53.

6.2.6.18. Compuesto 54

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.1, 6.2.5.2, 6.2.5.3, 6.2.5.4, rendimiento: 5 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.53 (d, J = 1 Hz, 8.5 Hz, H-4), 7.37 (d, J = 1 Hz, 1.8 Hz, H-7), 7.10 (dd, J = 1 Hz, 8.5, 1.8 Hz, H-5), 7.05 (d, J = 1 Hz, 2.3 Hz, H-2), 6.93 – 6.89 (m, J = 1 Hz, H, H-5′), 6.38 (s, J = 1 Hz, H, H-3′), 6.20 (d, J = 1 Hz, 3.5 Hz, H-4′), 3.72 (q, J = 2 Hz, 6.8 Hz, H-β), 3.03 (t, J = 2 Hz, 6.8 Hz, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) $_3$) δ 161.2 (Cγ), 136.9 (C7a), 128.4 (C3a), 126.1 (C2′), 126.1 (C6), 122.8 (C2), 121.5 (C5′), 120.5 (C5), 119.8 (C4), 113.5 (C3), 111.4 (C7), 110.0 (C4′), 108.7 (C3′), 39.8 (Cβ), 25.7 (Cα). HRESIMS calculado $[C_{15}H_{15}ClN_3O]^+$ 288,0898 obtenido 288,0899

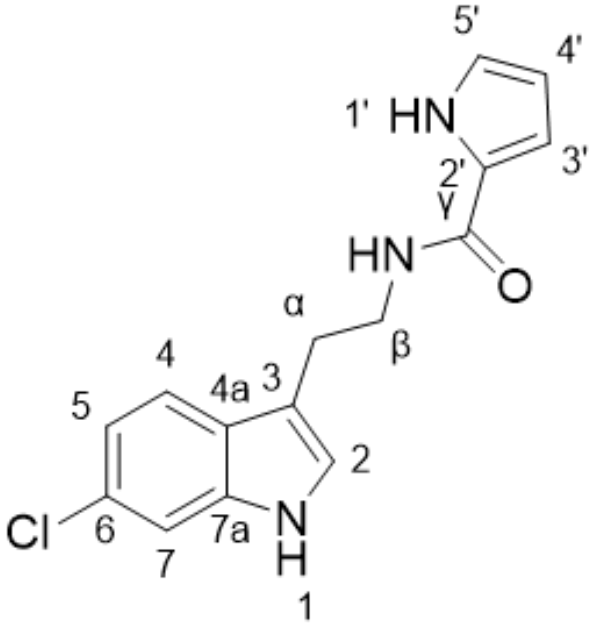


Figura 6.70: Estructura del compuesto 54.

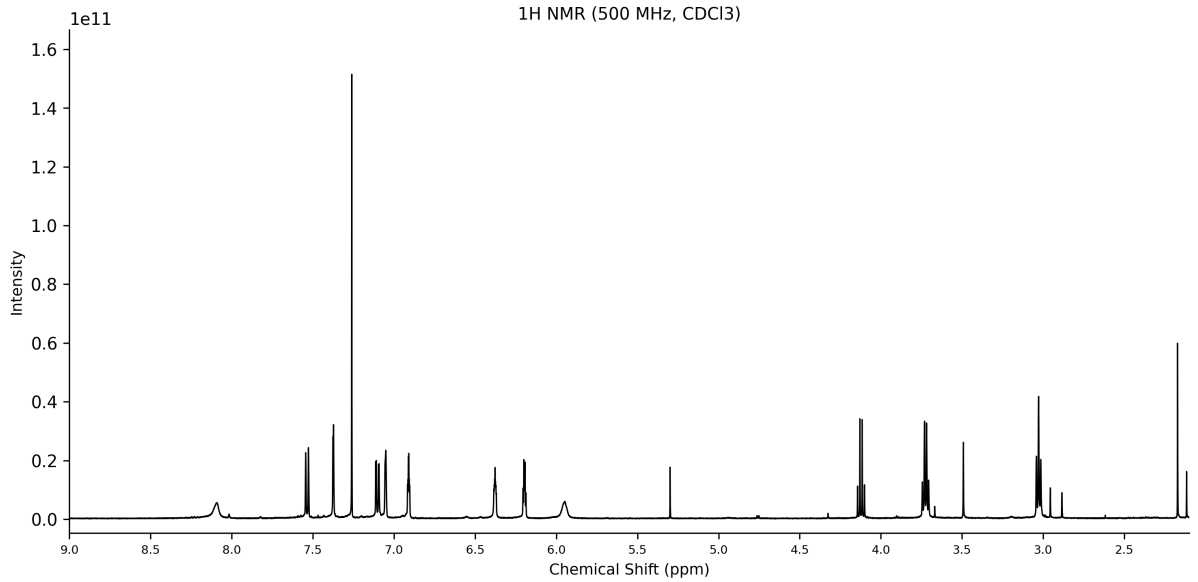


Figura 6.71: ^{1}H NMR del compuesto 54.

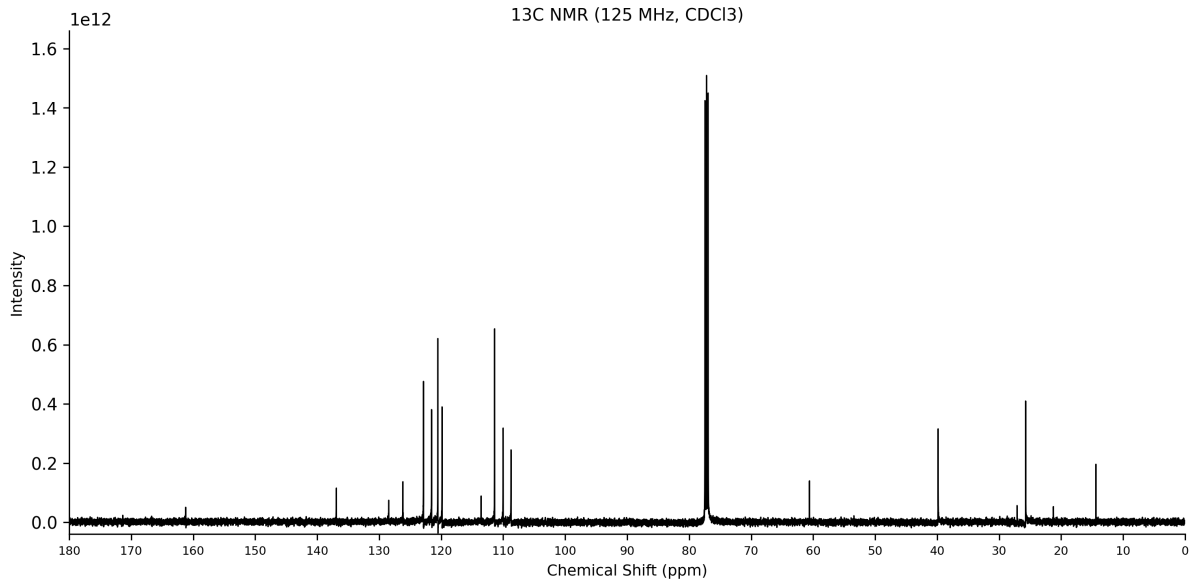


Figura 6.72: ^{13}C NMR del compuesto 54.

6.2.6.19. Compuesto 55

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 81 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.56 (d, J = 7,9 Hz, 1H, H-4′), 7.51 (d, J = 7,9 Hz, 1H, H-4), 7.35 (dd, J = 8,2, 4.9 Hz, 2H, H-7,7′), 7.25 – 7.16 (m, 2H, H-6), 7.25 – 7.16 (m, 2H, H-6′), 7.14 - 7.06 (m, $J = m$ Hz, 2H, H-5), 7.14 - 7.06 (m, $J = m$ Hz, 2H, H-5′), 6.89 (d, J = 2,3 Hz, 1H, H-2′), 6.56 (d, J = 2,3 Hz, 1H, H-2), 3.52 (q, J = 6,9 Hz, 2H, H-β), 3.07 (t, J = 7,1 Hz, 2H, H-α′), 2.82 (t, J = 6,6 Hz, 2H, H-α), 2.51 (t, J = 7,2 Hz, 2H, H-β′). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.7 (Cγ), 136.5 (C7a, 7a′), 127.4 (C3a), 127.3 (C3a′), 122.3 (C6), 122.2 (C2), 122.1 (C6′), 122.1 (C2′), 119.6 (C5), 119.5 (C5′), 119.0 (C4′), 118.9 (C4), 115.0 (C3), 112.9 (C3′), 111.3 (C7,7′), 39.4 (Cβ), 37.5 (Cβ′), 25.2 (Cα), 21.5 (Cα′).

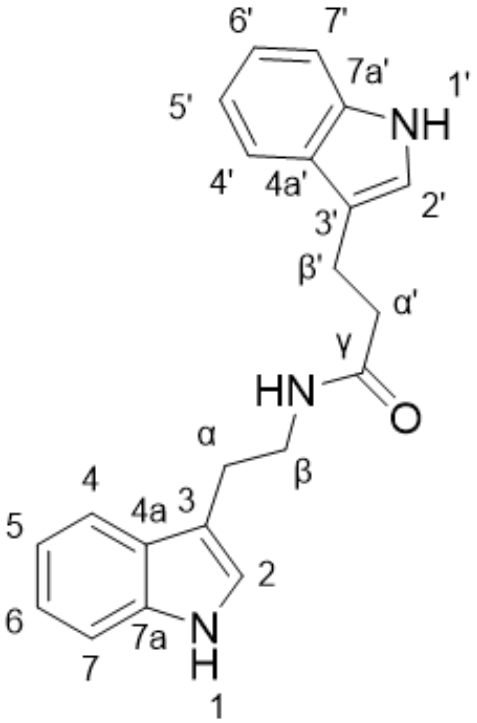


Figura 6.73: Estructura del compuesto 55.

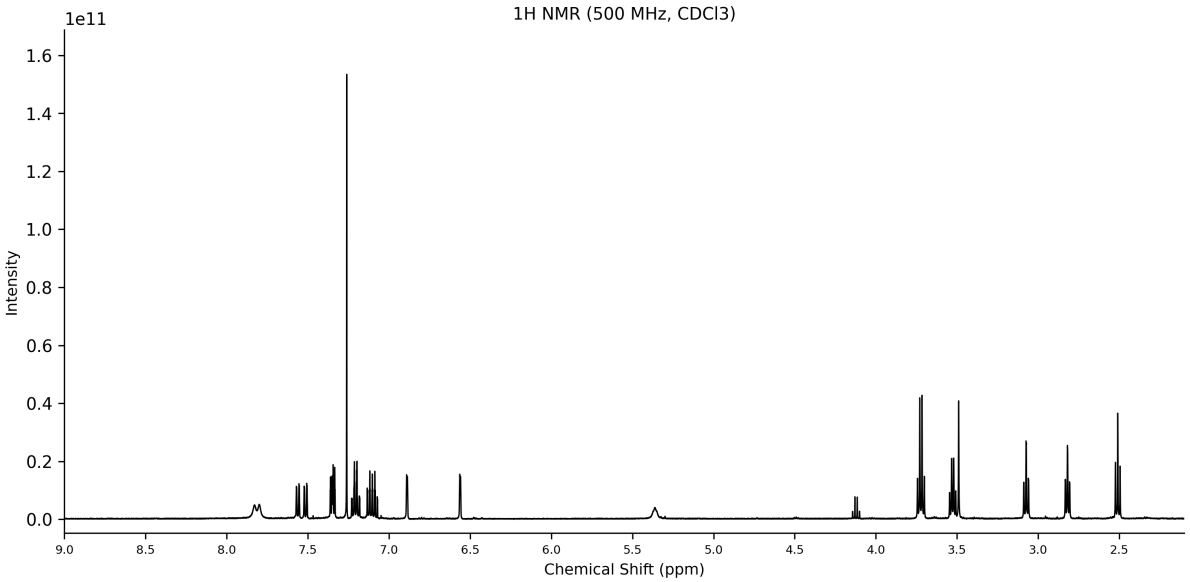


Figura 6.74: ^{1}H NMR del compuesto 55.

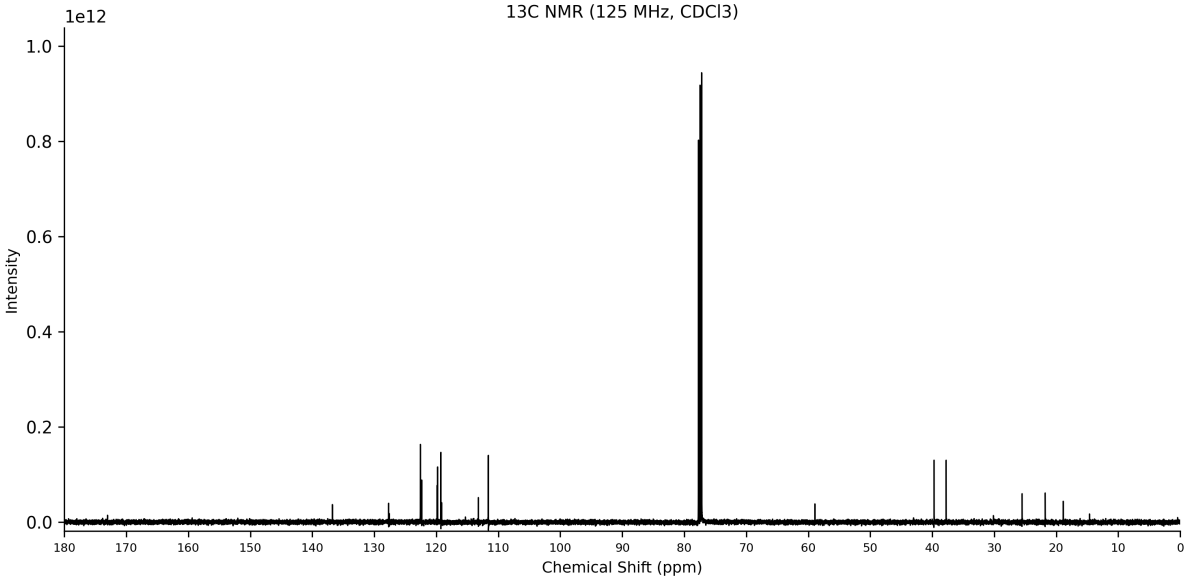


Figura 6.75: ^{13}C NMR del compuesto 55.

6.2.6.20. Compuesto 56

Aceite incoloro obtenido mediante la secuencia de reacciones 6.2.5.7, 6.2.5.2, 6.2.5.8, 6.2.5.9, rendimiento: 15 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (d, J = 7,9 Hz, 1H, H-4), 7.71 (d, J = 4,1 Hz, 1H, H-11), 7.62 (s, 2H, H-2, 14), 7.51 (dd, J = 8,9, 4,3 Hz, 1H, H-13), 7.40 – 7.22 (m, 3H, H-6), 7.40 – 7.22 (m, 3H, H-5), 7.40 – 7.22 (m, 3H, H-15), 7.37 (d, J = 8,5 Hz, 1H, H-7), 5.63 (s, 2H, H-12), 3.65 (s, 3H, H-1a). ^{13}C NMR (125 MHz, $CDCl_3$) δ 160.8 (CCF), 158.8 (CCF), 145.0 (C10), 137.3 (C7a), 133.7 (C18), 129.3 (CCF_3), 127.4 (C2, 14), 126.8 (C13), 125.6 (C3a), 122.4 (C6), 120.3 (C5), 119.9 (C4), 118.1 (C11), 118.1 (C15), 109.8 (C7), 105.8 (C3), 53.0 (C12), 33.2 (C1a).

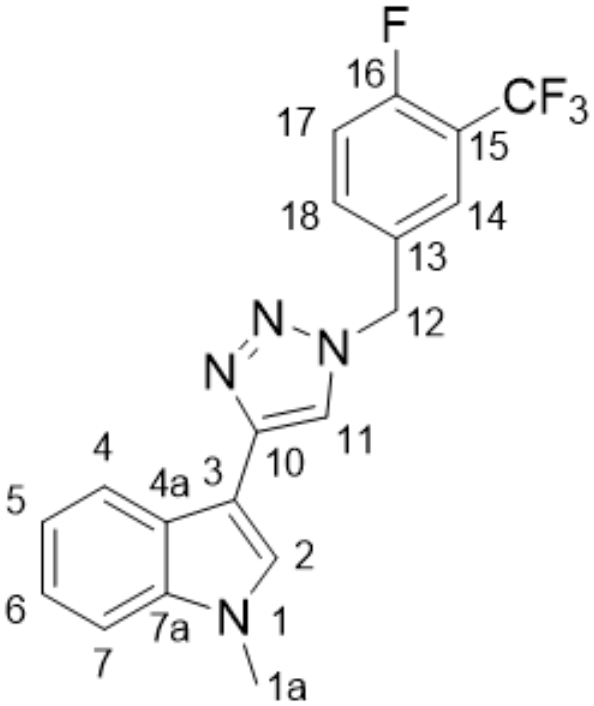


Figura 6.76: Estructura del compuesto 56.

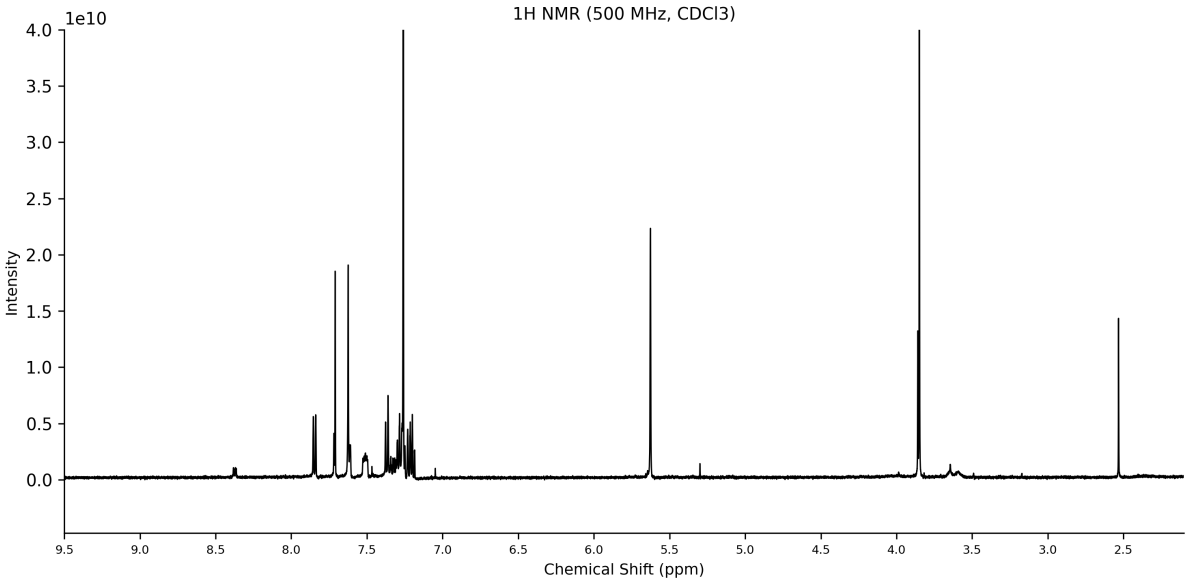


Figura 6.77: ^{1}H NMR del compuesto 56.

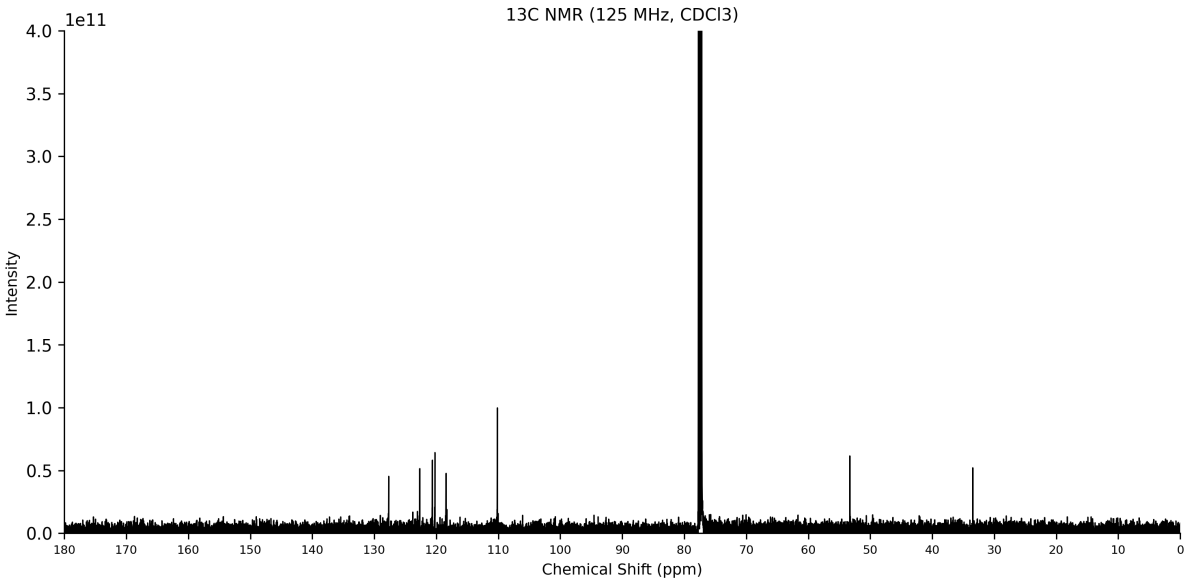


Figura 6.78: ^{13}C NMR del compuesto 56.

6.2.6.21. Compuesto 57

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 52 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.32 (d, $J = 1$ Hz, 2.8H, H-6′), 8.32 (d, $J = 1$ Hz, 2.8H, H-6′), 8.25 (dd, $J = 1$ Hz, 8.7, 4.6H, H-3′), 8.25 (dd, $J = 1$ Hz, 8.7, 4.6H, H-3′), 7.66 (d, $J = 1$ Hz, 7.9H, H-4), 7.52 (td, $J = 1$ Hz, 8.4, 2.8H, H-4′), 7.52 (td, $J = 1$ Hz, 8.4, 2.8H, H-4′), 7.38 (d, $J = 1$ Hz, 8.1H, H-7), 7.28 – 7.18 (m, $J = 1$ Hz, -H, H-6), 7.16 – 7.10 (m, $J = 1$ Hz, -H, H-5), 7.09 (d, $J = 1$ Hz, 2.3H, H-2), 3.82 (q, $J = 2$ Hz, 6.7H, H-β), 3.11 (t, $J = 2$ Hz, 7.0H, H-α). ^{13}C NMR (125 MHz, $CDCl_3$) δ 163.5 (Cγ), 162.3 (C5′), 160.2 (C5′), 146.6 (C2′), 146.6 (C2′), 136.8 (C6′), 136.6 (C6′), 136.6 (C7a), 127.5 (C3a), 124.1 (C3′), 124.1 (C3′), 124.0 (C4′), 123.8 (C4′), 122.3 (C6), 122.1 (C2), 119.6 (C5), 119.0 (C4), 113.3 (C3), 111.3 (C7), 39.9 (Cβ), 25.7 (Cα). HRESIMS calculado $[C_{16}H_{15}FN_3O]^+$ 284,1194 obtenido 284,1194

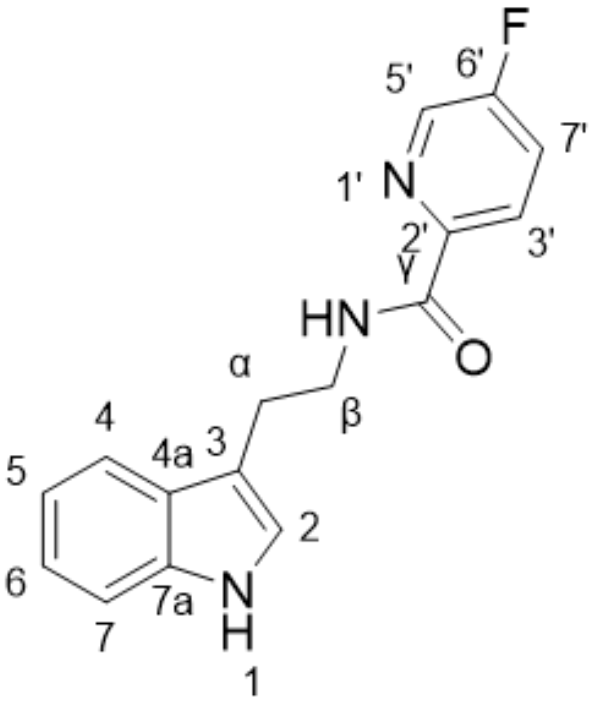


Figura 6.79: Estructura del compuesto 57.

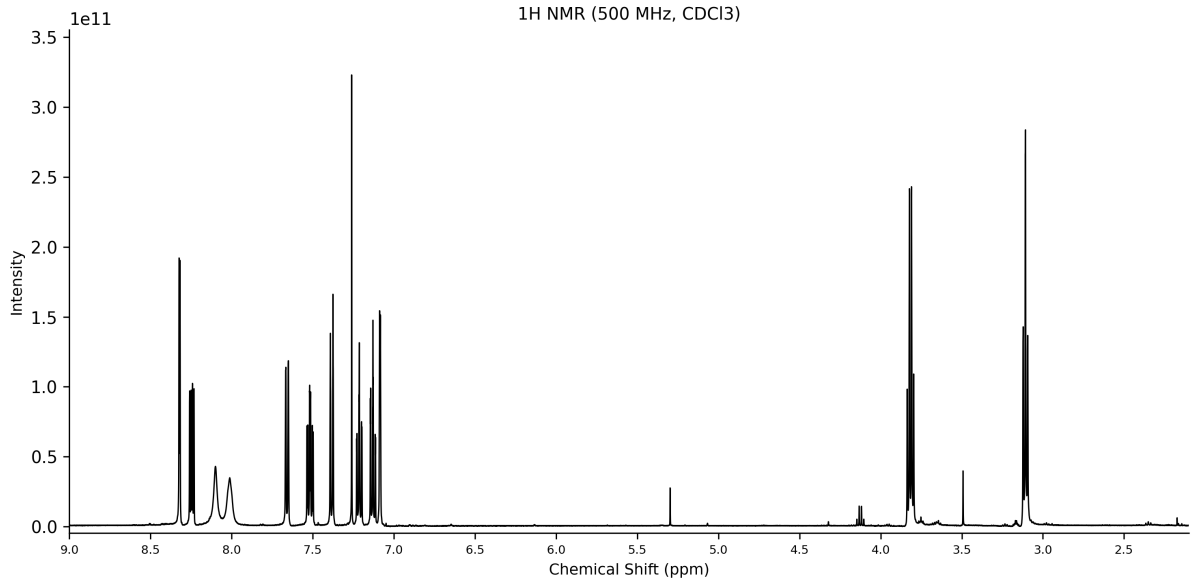


Figura 6.80: ^{1}H NMR del compuesto 57.

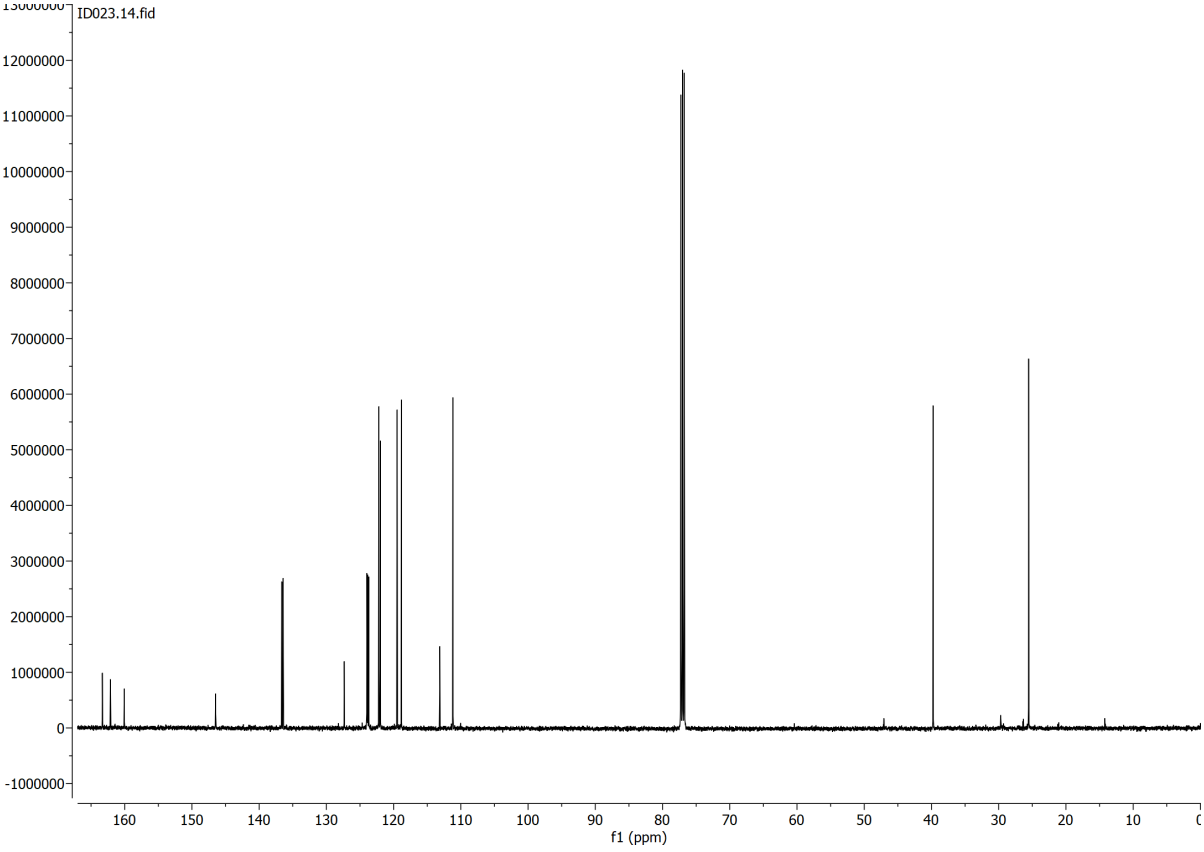


Figura 6.81: ^{13}C NMR del compuesto 57.

6.2.6.22. Compuesto 58

Sólido blanco obtenido mediante la reacción 6.2.5.4, rendimiento: 35 %. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.17 (d, J = 1,4 Hz, 1H, H-5′), 8.96 (d, J = 5,0 Hz, 1H, H-3′), 8.13 (dd, J = 5,0, 1,4 Hz, 1H, H-4′), 7.28 (d, J = 8,8 Hz, 1H, H-7), 7.07 (m, 2H, H-2), 7.07 (m, 2H, H-4), 6.88 (dd, J = 8,8, 2,5 Hz, 1H, H-6), 3.84 (s, 3H, H-ε), 3.82 (q, J = 6,9 Hz, 2H, H-β), 3.09 (t, J = 6,9 Hz, 2H, H-α).^{13}C NMR (125 MHz, $CDCl_3$) δ 162.7 (Cγ), 159.3 (C3′), 157.9 (C5′), 156.5 (C2′), 154.3 (C5), 131.7 (C7a), 127.8 (C3a), 122.9 (C2), 118.6 (C4′), 112.8 (C3), 112.7 (C6), 112.1 (C7), 100.7 (C4), 56.1 (Cε), 40.0 (Cβ), 25.5 (Cα).

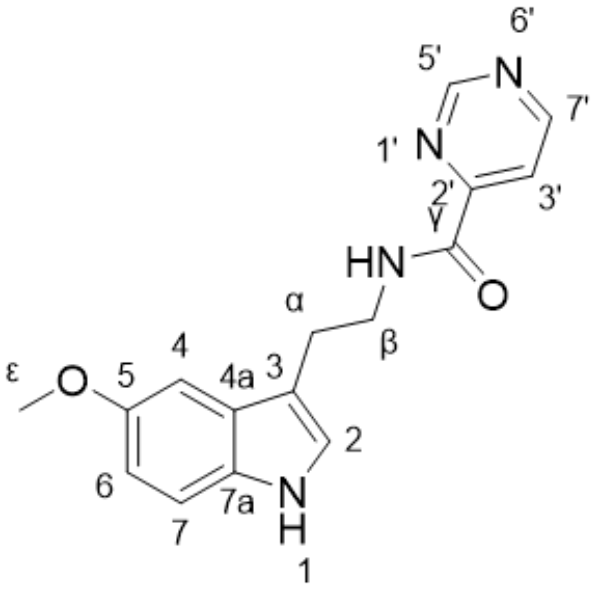


Figura 6.82: Estructura del compuesto 58.

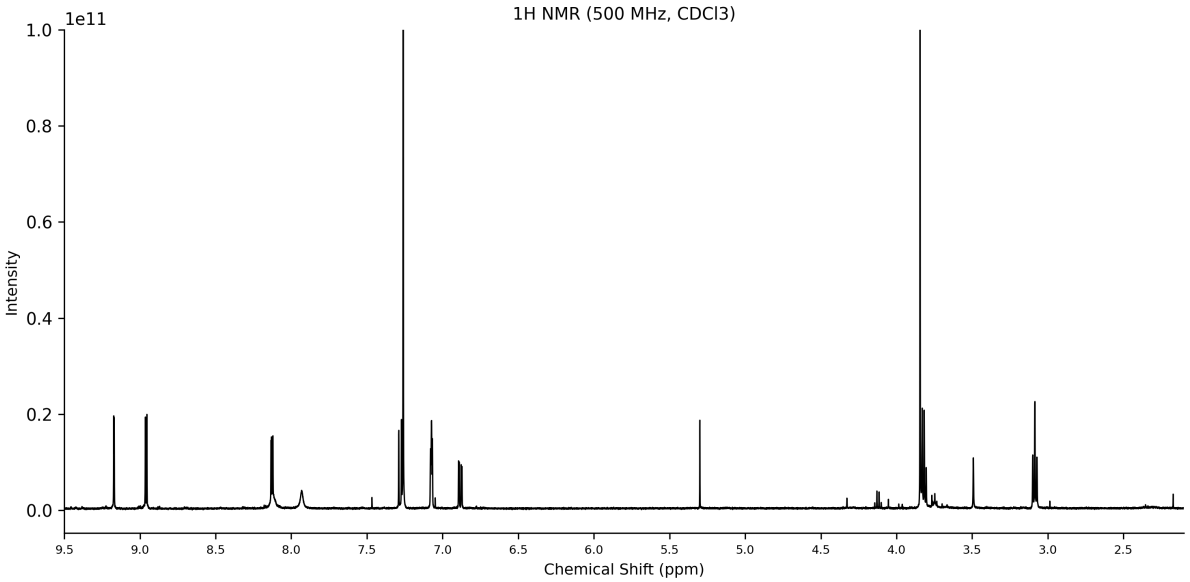


Figura 6.83: ^{1}H NMR del compuesto 58.

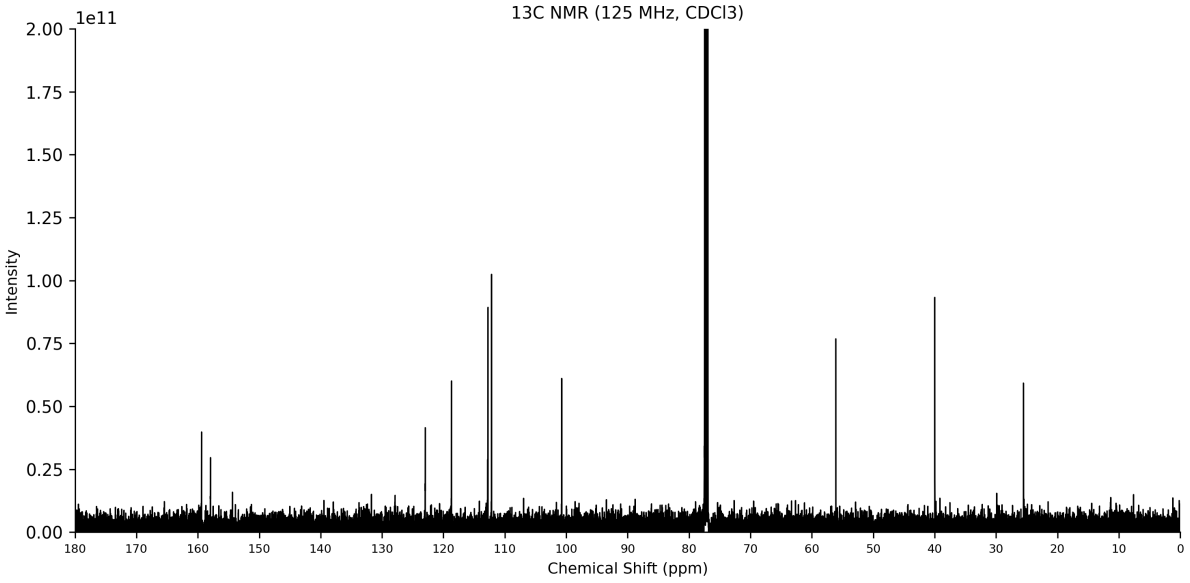


Figura 6.84: ^{13}C NMR del compuesto 58.

6.3 Ensayos de actividad biológica

6.3.1. Ensayos de actividad enzimática

6.3.1.1. Ensayo de inhibición de monoamino oxidasas (MAO-A y MAO-B) humanas

Las mediciones de inhibición de MAO se llevaron a cabo siguiendo el procedimiento general previamente descrito (28). En resumen, los fármacos de prueba y las cantidades adecuadas de hMAO A o hMAO B recombinantes (Sigma Aldrich Chemistry S.A., España), necesarios y ajustados para oxidar 165 pmol de p-tiramina/min en el grupo de control, se incubaron a 37 °C durante 15 minutos en una placa de microtitulación de fondo negro plano de 96 pocillos (BD Biosciences, Franklin Lakes, NJ) colocada en la cámara de fluorímetro oscura. La reacción se inició añadiendo 200 mM de reactivo Amplex Red (Molecular Probes, Inc., Eugene, OR), 1 U/mL de peroxidasa de rábano picante y 1 mM de p-tiramina, y la producción de resorufina se cuantificó a 37 °C en un lector de microplacas multidosis de fluorescencia (FLX800, Bio-Tek Instruments, Inc., Winooski, VT) en función de la fluorescencia generada (excitación, 545 nm; emisión, 590 nm). La emisión específica de fluorescencia se calculó después de la sustracción de la actividad de fondo, que se determinó a partir de los pocillos que contenían todos los componentes excepto las isoformas de hMAO, que fueron reemplazadas por PBS.

6.3.1.2. Inhibición de acetilcolinesterasa y butirilcolinesterasa humanas (hAChE y hBuChE)

Se siguió el método de Ellman, utilizando AChE recombinante humana y BuChE a partir de suero humano (29). La solución de ensayo consistía en un tampón fosfato 0,1 M pH 8,0, 400 µM de 5,5'-ditiobis(2-nitrobenzoico) ácido (DTNB, reactivo de Ellman), 0,05 U/mL de hAChE (Sigma Chemical Co.) o 0,024 U/mL de hBuChE (Sigma Chemical Co.), y 800 µM de ioduro de acetiltiocholina o 500 µM de butiriltiocholina como sustratos de la reacción enzimática, respectivamente. Los compuestos en estudio se añadieron a la solución de ensayo y se preincubaron con la enzima durante 5 minutos a 30 °C. Luego de ese período, se añadió el sustrato. Los cambios de absorbancia a 412 nm se registraron durante 5 minutos con un espectrofotómetro UV/Vis para microplacas, Multiskan Spectrum, Thermo-Electron Co. Se compararon las velocidades de reacción y se calculó el porcentaje de inhibición debido a la presencia del compuesto en estudio. El IC_{50} se definió como la concentración de cada compuesto que reduce al 50 % la actividad enzimática en ausencia de inhibidor.

6.3.1.3. Ensayo de inhibición de GSK-3β

La enzima GSK-3β recombinante humana fue adquirida a *Millipore* (Millipore Ibérica S.A.U.). El sustrato polipeptídico prefosforilado se obtuvo de *Millipore* (Millipore Ibérica S.A.U.). El ensayo luminiscente *Kinase-Glo* fue adquirido de *Promega* (Promega Biotech Ibérica, S.L.). El ATP y el resto de los reactivos fueron comprados a *Sigma-Aldrich* (St. Louis, MO). El buffer de ensayo contenía 50 mM HEPES (pH 7.5), 1 mM EDTA, 1 mM

EGTA y 15 mM de acetato de magnesio.

Se siguió el método descrito por Baki et al.(30) para la inhibición de GSK-3β. Los ensayos *Kinase-Glo* se realizaron en buffer de ensayo utilizando placas negras de 96 pocillos. En un ensayo típico, se añadieron 10 μL (10 μM) del compuesto a evaluar (disuelto en DMSO a una concentración de 1 mM y previamente diluido en buffer de ensayo a la concentración deseada) y 10 μL (20 ng) de enzima en cada pocillo, seguidos de 20 μL de buffer de ensayo que contenía 25 μM de sustrato y 1 μM de ATP. La concentración final de DMSO en la mezcla de reacción no superó el 1 %.

Tras una incubación de 30 minutos a 30 °C, la reacción enzimática fue detenida con 40 μL del reactivo *Kinase-Glo*. La luminiscencia tipo glow fue registrada tras 10 minutos utilizando un lector multimodal *FLUOstar Optima* (BMG Labtechnologies GmbH, Offenburg, Alemania).

La actividad enzimática es proporcional a la diferencia entre el ATP total y el ATP consumido. Las actividades inhibitorias se calcularon con base en las actividades máximas medidas en ausencia de inhibidor. El IC_{50} se definió como la concentración de compuesto que reduce en un 50 % la actividad enzimática respecto a la condición sin inhibidor.

6.3.1.4. Ensayo de inhibición de BACE-1

Este experimento se llevó a cabo de acuerdo con el protocolo descrito por el fabricante (Invitrogen) utilizando un ensayo FRET (31). En resumen, se utilizó un sustrato peptídico basado en APP (rhodamina-EVNLDAEFK-quenching, K_m de 20 μM) que lleva la mutación sueca y contiene una rhodamina como donante de fluorescencia y un quen-

cher como aceptor de fluorescencia en cada extremo. El sustrato intacto es débilmente fluorescente y se vuelve altamente fluorescente tras la escisión enzimática. Los ensayos se realizaron en un tampón de acetato de sodio 50 mM, pH 4.5, con una concentración final de enzima (1 U/mL). La mezcla se incubó durante 60 minutos a 25 °C en condiciones oscuras y luego se detuvo con acetato de sodio 2.5 M. La fluorescencia se midió con un lector de microplacas FLUOstar Optima (BMG Labtechnologies GmbH, Offenburg, Alemania) a una excitación de 545 nm y emisión a 585 nm.

6.3.2. Ensayos de actividad antifúngica

Para evaluar la concentración mínima inhibitoria (MIC) de los compuestos que mostraron actividad antifúngica, se realizaron diluciones seriadas de microdilución del compuesto de interés, como se reportó previamente. El procedimiento consiste en la preparación de diluciones en serie del agente antimicrobiano en una placa de microtitulación de 96 pocillos, que luego se inocula con una suspensión de conidias de concentración conocida (el inóculo se ajusta entre $10^{4\smile}10^{6}$ UFC/mL mediante conteo en cámara de Neubauer). La placa de microtitulación se incuba durante 24/36 h y el punto final, es decir, la concentración a la cual se observa la inhibición total del crecimiento fúngico, es evaluado. Para esta prueba, se utilizaron las siguientes cepas de hongos fitopatógenos: *Fusarium solani* (UFC 3×10^5), *Fusarium tucumaniae* (UFC $2{,}3 \times 10^6$) y *Fusarium graminearum* (UFC 8×10^4). Los compuestos antifúngicos comerciales carbendazim y piraclostrobina (Sigma Aldrich) fueron empleados como controles positivos.

6.3.3. Evaluación de la actividad tripanocida

Todos los compuestos fueron disueltos en dimetilsulfóxido (DMSO). Para el cribado, se siguió el protocolo reportado previamente (32). Brevemente, se sembraron 1×10^4 células HeLa por pocillo en una placa de 96 pocillos con medio DMEM suplementado con 10 % de suero fetal bovino (FBS). Después de 24 horas, las células fueron infectadas con tripomastigotes Y-GFP y se incubaron durante 4 horas. Luego, se lavaron con solución salina tamponada con fosfato (PBS) y el medio fue reemplazado por soluciones de los compuestos a 20 µM, benznidazol a 10 µM (control positivo, C+, correspondiente aproximadamente a la CI_{50}) o DMSO (control negativo, C−) en DMEM (10 % FBS). La incubación se mantuvo durante 72 horas, tras lo cual las células fueron lavadas con PBS y se añadió un tampón de lisis. Después de 30 minutos a 4 °C, se cuantificó la fluorescencia de GFP con un lector de microplacas FLUOStar Optima a 520 nm (excitación a 485/12 nm). La autofluorescencia de lisados de células no infectadas se utilizó como blanco. Luego, se calculó el porcentaje de fluorescencia relativa respecto al control negativo (C−) (n=3) para cada compuesto.

Para estimar la CI_{50} frente a *T. cruzi*, se siguió el mismo procedimiento descrito anteriormente, utilizando diluciones seriadas de aquellos compuestos que mostraron una diferencia significativa respecto al C− en el cribado. La CI_{50} y el intervalo de confianza del 95 % ($n \geq 3$) para cada tratamiento se estimaron aplicando la opción *Dose–Response* del software GraphPad Prism 8.0.1.

Bibliografía

[1] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235–242.

[2] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.

[3] Efremov, R. G.; et al. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **2021**, *61*, 3891–3898.

[4] Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **2009**, *30*, 2785–2791.

[5] Allen, W.; et al. DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **2015**, *36*, 1132–1156.

[6] OpenEye Scientific Software FRED: Fast Rigid Exhaustive Docking. 2021; Version 4.1.0.

[7] OpenEye Scientific Software HYBRID: Ligand-based docking by pose optimization. 2021; Version 4.1.0.

[8] OpenEye Scientific Software OpenEye Toolkits and Applications. 2021; Santa Fe, NM, USA.

[9] Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; others PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **2021**, *49*, D1388–D1395.

[10] Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; others The ChEMBL database in 2017. *Nucleic Acids Research* **2017**, *45*, D945–D954.

[11] Vainio, M. J.; Johnson, M. S. Open source conformer generation with the Balloon toolkit. *Journal of Cheminformatics* **2007**, *1*, 1–10.

[12] Vainio, M. J. Balloon: Conformer generation software. 2007; Available at http://users.abo.fi/mivainio/balloon/.

[13] Hawkins, P.; Skillman, A.; Warren, G.; Ellingson, B.; Stahl, M. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *Journal of Chemical Information and Modeling* **2010**, *50*, 572–584.

[14] Forlilab Meeko: Automatic Ligand Preparation Tool for AutoDock Vina. 2025; https://github.com/forlilab/Meeko, Accessed: 2025-05-09.

[15] GQC MIEF: Molecular Interaction Energy Fingerprints. 2025; https://gitlab.com/gqc/mief, Accessed: 2025-05-09.

[16] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. 2011; https://scikit-learn.org/stable/.

[17] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

[18] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[19] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; others Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

[20] Case, D. A. *et al.* AmberTools20. 2020; University of California, San Francisco.

[21] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

[22] Silverstein, R. M.; Webster, F. X.; Kiemle, D. J.; Bryce, D. L. *Spectrometric Identification of Organic Compounds*; John Wiley & Sons, 2014.

[23] Pretsch, E.; Bühlmann, P.; Badertscher, M. *Tables of spectral data for structure determination of organic compounds*; Springer, 2013.

[24] Dennington, R.; Keith, T. A.; Millam, J. M.; others GaussView, version 6.0. 16. *Semichem Inc Shawnee Mission KS* **2016**, *13*.

[25] Patre, R. E.; Shet, J. B.; Parameswaran, P. S.; Tilve, S. G. Cascade Wittig reaction-double Claisen and Cope rearrangements: one-pot synthesis of diprenylated coumarins gravelliferone, balsamiferone, and 6, 8-diprenylumbelliferone. *Tetrahedron letters* **2009**, *50*, 6488–6490.

[26] Steck, W.; Bailey, B.; Shyluk, J.; Gamborg, O. Coumarins and alkaloids from cell cultures of Ruta graveolens. *Phytochemistry* **1971**, *10*, 191–194.

[27] Macís, F.; Massanet, G.; Rodriguez-Luis, F.; Salvá, J. 13C NMR of coumarins I: 3-(1, 1-Dimethylallyl) derivatives. *Magnetic resonance in chemistry* **1989**, *27*, 705–707.

[28] Herrera-Arozamena, C.; Estrada-Valencia, M.; López-Caballero, P.; Pérez, C.; Morales-García, J. A.; Pérez-Castillo, A.; del Sastre, E.; Fernández-Mendívil, C.; Duarte, P.; Michalska, P.; Lombardía, J.; Senar, S.; León, R.; López, M. G.; Rodríguez-Franco, M. I. Medicinal Chemistry Applications of Natural Products: A Review of 2022. *Journal of Medicinal Chemistry* **2022**, *65*, 4727–4751.

[29] Ellman, G. L.; Courtney, K. D.; Andres Jr, V.; Featherstone, R. M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochemical pharmacology* **1961**, *7*, 88–95.

[30] Baki, A.; Bielik, A.; Molnár, L.; Szendrei, G.; Keserü, G. M. A high throughput luminescent assay for glycogen synthase kinase-3 β inhibitors. *Assay and drug development technologies* **2007**, *5*, 75–84.

[31] Invitrogen BACE1 (b-secretase) FRET assay kit, part # P2985. 2018; http://tools.invitrogen.com/content/sfs/manuals/L0724.pdf, Last accessed: 12 Jun 2018.

[32] García, F. F.; Musikant, D.; Escalona, J. L.; Edreira, M. M.; Liñares, G. G. Lipase-Catalyzed Synthesis and Biological Evaluation of N-Picolineamides as Trypanosoma cruzi Antiproliferative Agents. *ACS Medicinal Chemistry Letters* **2023**, *14*, 59–65.

Capítulo 7

Conclusiones

A lo largo de este trabajo de tesis doctoral se ha desarrollado e implementado una nueva metodología computacional denominada Target Fisher, orientada a la predicción de blancos moleculares para productos naturales y sus derivados sintéticos. A partir de los resultados obtenidos, podemos establecer las siguientes conclusiones:

1. Se ha diseñado e implementado exitosamente una metodología híbrida de predicción de blancos moleculares que integra docking molecular con modelos de aprendizaje automático. Target Fisher emplea huellas de interacción energéticas generadas a partir de complejos proteína-ligando, lo que permite capturar patrones de interacción cruciales no detectables mediante funciones de puntuación convencionales.

2. El desarrollo de la herramienta MIEF (Molecular Interaction Energy Fingerprints), basada en campos de fuerza de mecánica molecular, ha permitido la extracción eficiente de perfiles energéticos de interacción proteína-ligando, superando una limitación importante en el campo del diseño de fármacos asistido por computadora. Dicha herramienta también es de libre disponibilidad.

3. El análisis sistemático de los efectos de diferentes algoritmos de aprendizaje automático (SVM, Random Forest, MLPC) mostró un rendimiento similar entre ellos, validando la elección de SVM como el clasificador principal para la implementación final de Target Fisher.

4. Las pruebas de permutación de etiquetas y de intercambio de conjuntos de ligandos entre blancos moleculares confirmaron que los modelos desarrollados capturan patrones verdaderamente relevantes para la interacción proteína-ligando, y no co-

rrelaciones espurias o sesgos en los datos.

5. Los modelos de aprendizaje automático superaron consistentemente a las aproximaciones basadas únicamente en docking, alcanzando valores de AUC-ROC entre 0.68 y 0.91 según el blanco molecular, con un valor promedio de 0.78 para los 38 blancos incorporados.

6. Los estudios sobre el efecto del filtrado por similitud estructural y el uso de ventanas de separación entre clases revelaron que la inclusión de compuestos estructuralmente similares mejora el poder predictivo de los modelos, lo cual contrasta con la visión tradicional basada en la diversidad química.

7. El uso de compuestos artificiales (decoys) para el entrenamiento de modelos introdujo un sesgo considerable, lo que justifica la decisión metodológica de emplear únicamente compuestos con actividad experimental conocida.

8. El intercambio de conjuntos de compuestos entre blancos evidenció que, aunque existe cierto sesgo debido a las estructuras químicas presentes en cada conjunto, los modelos pueden explotar información estructural para realizar clasificaciones efectivas.

9. Target Fisher ha sido implementado como un software de código abierto y un servidor web accesible, lo que facilita su utilización por parte de la comunidad científica sin requerir conocimientos avanzados en programación o química computacional.

10. La aplicación de Target Fisher en una quimioteca de productos naturales permitió identificar cumarinas como inhibidores potentes de MAO-B, con valores de IC_{50}

submicromolares (0.45 μM para (R)-chalepina y 0.81 μM para (S)-uliscomoncadina A), validando experimentalmente la capacidad predictiva de la herramienta.

11. Los estudios de dinámica molecular realizados sobre los complejos formados entre las cumarinas activas y MAO-B revelaron un modo de unión diferente al observado para inhibidores conocidos, en el que el grupo isoprenil alilo juega un papel fundamental en la selectividad hacia esta enzima.

12. La evaluación de estas cumarinas en otras herramientas de predicción de blancos evidenció que dichos resultados no hubieran sido alcanzados sin el uso de Target Fisher.

13. Comparaciones adicionales con otras metodologías, empleando estructuras novedosas recientemente publicadas, mostraron que Target Fisher presenta una mejor capacidad de generalización.

14. La aplicación de Target Fisher en una quimioteca de compuestos sintéticos permitió identificar una serie de benzotiazepinas como inhibidores de GSK-3β, incluyendo un compuesto con alta potencia (IC_{50} = 0.055 μM), demostrando la versatilidad de la metodología para el cribado virtual tanto inverso como convencional.

15. Se establecieron relaciones estructura-actividad para la serie de benzotiazepinas, identificando la importancia de la presencia de grupos hidrofílicos (carbonilo o oxima) en posiciones específicas de la molécula para mantener la actividad inhibitoria sobre GSK-3β.

16. Se pusieron a punto distintas síntesis para la obtención de análogos de las benzotiacepinas activas a partir de intermediarios disponibles en la quimioteca.

17. Se pusieron a punto diversas rutas sintéticas para la obtención de análogos de Aspidostomida G, incluyendo: (i) formación de amidas mediante acoplamiento con PyBOP, (ii) síntesis de triptaminas sustituidas a través de rutas reductivas, (iii) una estrategia one-pot basada en la síntesis de indoles de Fischer, y (iv) reemplazo isostérico del grupo amida por triazoles mediante química click.

18. Se creó una quimioteca virtual con todos los derivados posibles a partir de las reacciones anteriores, permitiendo el estudio *in silico* de dichos compuestos, abriendo las puertas al diseño racional de derivados.

19. Se puso a punto un modelo de clasificación de compuestos activos frente a inactivos para la enzima CK2, que incluyo docking molecular, modelos de aprendizaje automático y estudios de dinámica molecular y energías de interacción de compuestos con actividad conocida.

20. Se estudió la actividad biológica de algunos de los análogos de aspidostomida G sintetizados durante la puesta a punto de las reacciones, encontrando que los mismos presentan cierta actividad tripanocida.

21. Aprovechando todo lo estudiado a lo largo de la Tesis, se re-enfocó la búsqueda de actividad biológica de los análogos de aspidostomida hacia MAO-B. A partir del modelo disponible en Target Fisher, sumado al análisis de la facilidad sintética de los análogos propuestos, se seleccionaron compuestos a sintetizar con el objetivo de

validar la hipótesis.

7.1 Perspectivas futuras

1. Ampliar el número de blancos moleculares disponibles en Target Fisher, con especial foco en aquellos asociados a enfermedades desatendidas y condiciones médicas con necesidades terapéuticas no cubiertas.

2. Desarrollar una herramienta que automatice el proceso de entrenamiento y validación de modelos, facilitando la incorporación de nuevos blancos.

3. Profundizar en la síntesis de análogos de las benzotiacepinas, a modo tal de mejorar el estudio de las relaciones estructura-actividad identificadas sobre GSK-3β

4. Realizar estudios de toxicidad, tanto sobre las benzotiacepinas como también de las cumarinas estudiadas, a modo tal de incorporar esta como una propiedad a optimizar en el futuro diseño de compuestos de ser necesario.

5. Explorar arquitecturas avanzadas de aprendizaje profundo capaces de realizar predicciones precisas, incluso a partir de conjuntos de datos reducidos, lo que resulta clave para blancos con escasa información disponible.

6. Incorporar enfoques de aprendizaje automático basados en la estructura del ligando como criterio complementario al perfil de interacción proteína-ligando, enriqueciendo así la capacidad predictiva de la metodología.

7.2 Producción científica

Durante el desarrollo de esta tesis doctoral se publicó el trabajo principal que describe en detalle la metodología desarrollada, incluyendo su validación y aplicación experimental:

1. Fernández, J. F., Martinez Heredia, L., Caracciolo, F., Esses, D., Suarez, R., Siless, G., ... Lavecchia, M. (2025). *Target Fisher: A Consensus Structure-Based Target Prediction Tool, and its Application in the Discovery of Selective MAO-B Inhibitors*. *Chemistry–A European Journal*, **31**(1), e202401838.

Además, como resultado de colaboraciones científicas generadas a lo largo del proyecto, se publicaron tres trabajos adicionales en los que se profundizó en temáticas complementarias:

1. Rengifo Carrillo, M., Fernandez, J. F., Macri, G., D'Jonsiles, M. F., Kramer, F., Palermo, J. A., & Duran, F. J. (2025).
Chemical Modifications of 12-Methoxy-N-methylvoachalotine (12-MMV): Enhancing Efficacy against Phytopathogenic Fungi. ACS Agricultural Science & Technology.

2. Martínez Heredia, L., Quispe, P. A., Fernández, J. F., & Lavecchia, M. J. (2025). *NaturAr: A Collaborative, Open-Source Database of Natural Products from Argentinian Biodiversity for Drug Discovery and Bioprospecting. Journal of Chemical Information and Modeling*, **65**(4), 1889–1900.

3. Fernández, J. F., & Lavecchia, M. J. (2022). *Small molecule stabilization of non-native protein-protein interactions of SARS-CoV-2 N protein as a mechanism of action against COVID-19. Journal of Biomolecular Structure and Dynamics*, **40**(10), 4488–4495.

7.3 Conclusión final

Este trabajo de Tesis doctoral busca aportar al campo del descubrimiento de fármacos asistido por computadora mediante el desarrollo de una metodología accesible para la predicción de blancos moleculares de productos naturales y sus derivados. La validación experimental de algunos resultados sugiere que esta aproximación puede resultar útil como herramienta complementaria en etapas tempranas del diseño de compuestos bioactivos. La combinación de técnicas computacionales con química sintética y evaluación biológica ha permitido delinear un posible marco de trabajo que podría ser de utilidad para futuros estudios en este ámbito.